Adaptive Resource Allocation for Wildlife Protection against Illegal Poachers

Rong Yang[+], Benjamin Ford[+], Milind Tambe[+], Andrew Lemieux[*]
[+] University of Southern California, Los Angeles, CA, US
[*] Netherlands Institute for the Study of Crime and Law Enforcement
[+]{yangrong,benjamif,tambe}@usc.edu
[*]ALemieux@nscr.nl

ABSTRACT

Illegal poaching is an international problem that leads to the extinction of species and the destruction of ecosystems. As evidenced by dangerously dwindling populations of endangered species, existing anti-poaching mechanisms are insufficient. This paper introduces the Protection Assistant for Wildlife Security (PAWS) application - a joint deployment effort done with researchers at Uganda's Queen Elizabeth National Park (QENP) with the goal of improving wildlife ranger patrols. While previous works have deployed applications with a game-theoretic approach (specifically Stackelberg Games) for counter-terrorism, wildlife crime is an important domain that promotes a wide range of new deployments. Additionally, this domain presents new research challenges and opportunities related to learning behavioral models from collected poaching data. In addressing these challenges, our first contribution is a behavioral model extension that captures the heterogeneity of poachers' decision making processes. Second, we provide a novel framework, PAWS-Learn, that incrementally improves the behavioral model of the poacher population with more data. Third, we develop a new algorithm, PAWS-Adapt, that adaptively improves the resource allocation strategy against the learned model of poachers. Fourth, we demonstrate PAWS's potential effectiveness when applied to patrols in QENP, where PAWS will be deployed.

Categories and Subject Descriptors

J.m [**Computing Application**]: Game Theory

General Terms

Security, Design

Keywords

Game Theory, Human Behavior, Application, Wildlife Protection

1. INTRODUCTION

This paper introduces a novel multiagent research challenge at the intersection of security games [19] and learning adversarial models [12]: combating wildlife crime [16]. By integrating real-world data into our approach, we can better learn adversarial behavior patterns and adaptively devise optimal strategies for anti-poaching patrols. Tigers, along with many other endangered species,

are in danger of extinction from poaching [8, 16]. The global population of tigers has dropped over 95% from the start of the 1900s, resulting in three out of nine species extinctions [16]. Over the course of 2011, South African rhino poaching reached a rate of approximately one death every 20 hours, and that rate only increased in 2012 [8]. Species extinction can destroy ecosystems and weaken the communities and economies that depend on those ecosystems [16]. In some cases, such as with the tiger trade, poachers are part of well-funded organized crime groups. Many other poachers, however, are not part of organized crime syndicates but are still capable of successfully hunting with less sophisticated means such as snares and poison.

As a result of a joint effort done with researchers at the Queen Elizabeth National Park (QENP), we developed the Protection Assistant for Wildlife Security (PAWS) application to improve the efficiency of the ranger patrols at QENP. Conservation staff at QENP (along with other conservation agencies) suffer from a lack of law enforcement resources to protect a very large, rural area from poachers; compared to urban environments [1], one wildlife crime study in 2007 reported an actual coverage density of one ranger per 167 square kilometers [5]. To assist officials in QENP, we present PAWS- a security game application that demonstrates the success such an approach can have on optimizing wildlife crime patrols and stopping poachers. The important domain properties of wildlife crime are captured and integrated into PAWS, which allows us to generate optimal resource allocation strategies that accurately reflect the many factors that influence the creation of wildlife patrols.

So far, security game research has been primarily motivated by counter-terrorism, and, indeed, this field has been fueled by a range of successfully deployed applications [19, 2]. These games provide a framework for defenders to optimize the use of their limited security resources. In previous security games research, actual adversary data is often missing, and therefore, it is difficult to build accurate models of adversary behavior. In addition, they do not consider the heterogeneity among large populations of adversaries. In domains such as wildlife crime, crime events occur often and generate significant amounts of collectible crime event data. As a result, learning adversary models directly from this data is feasible and thus new opportunities are created for security games towards addressing these new challenges.

In the case of wildlife crime, crime data can be anonymous, or it can be linked to confessed adversaries (i.e., identified). While the latter type of data provides rich information about individual adversaries, that type of data is sparse and hard to collect. The majority

---

[1]Current law enforcement density statistics for New York City show a coverage of approximately 28 officers per square kilometer (34,500 officers over a total land and sea area of 1,213 square kilometers) [11].

of collected data is evidence on crimes committed by anonymous adversaries. Compared to identified data, anonymous data provides no information about the identity of the adversary that committed the crime and therefore cannot be used to build accurate behavioral models on the individual level. The open question here is how to utilize both types of data to build and learn a better model of the large population of criminals. Moreover, how does the learned model help better predict future crime events and thus help law enforcement officials to improve their resource allocation strategies?

This paper makes the following contributions towards answering these open questions. First, we propose a stochastic behavioral model which extends the current state-of-the-art to capture the heterogeneity in the decision making process of a population of poachers. Second, we then demonstrate how to learn the behavioral pattern of the poacher population from both the identified data and the anonymous data. Third, in order to overcome the sparseness of the identified data, we provide a novel algorithm, PAWS-Learn, to improve the accuracy of the estimated behavioral model by combining the two types of data. Fourth, we develop a new algorithm, PAWS-Adapt, which adapts the rangers' patrolling strategy against the poacher population's behavioral model. Fifth, we show the effectiveness of PAWS in a general setting, but our main drive is to deploy PAWS in QENP; we also demonstrate PAW's effectiveness when applied to an area of QENP. Our experiment results and corresponding discussion illustrate the capabilities of PAWS and its potential to improve the efforts of wildlife law enforcement officials in managing and executing their anti-poaching patrols.
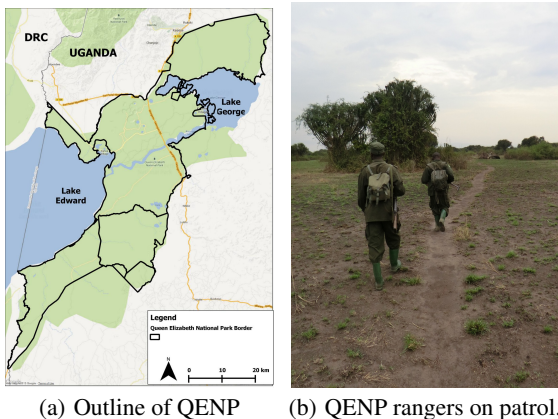

(a) Outline of QENP   (b) QENP rangers on patrol.

Figure 1: QENP: The intended site of deployment. Ranger photo taken by Andrew Lemieux.

2. RELATED WORK

In recent years, game theory techniques have been applied and successfully deployed in a variety of security domains, such as airports, airline flights, sea ports, and rapid transit systems [19, 2]. These deployed works all utilize the Stackelberg Game model as the basis of their security strategies, and by doing so, they are able to account for security agencies' common lack of resources without falling into predictable, exploitable patterns. While one example, PROTECT, uses the Quantal Response (QR) model to model the adversary's behavior and bounded rationality [17], we use the Subjective Utility Quantal Response (SUQR) model. As demonstrated in [10], SUQR outperforms the QR model. In addition, our approach features PAWS-Learn, a novel learning technique that continually adapts our adversary behavioral models to collected crime data; this learning affords us additional robustness against adaptable adversaries that has not been realized in the aforementioned deployed works.

Security Games research has also been focusing on uncertainty in Stackelberg Games. Letchford et al. and Marecki et al. study the effects of payoff uncertainty, the uncertainty on the adversary's reward and penalty values, and how they impact the defender's ability to compute an optimal strategy [6, 7]. These works demonstrate that it is possible to reduce payoff uncertainty over time by incorporating data. Instead of payoff uncertainty, however, our approach focuses on behavioral uncertainty. In addition, PAWS is geared for deployment; we are actively working with wildlife crime domain experts to deploy this application to QENP, and thus our assumptions and models must be sufficiently grounded to the realities of the wildlife crime domain in order to be successful.

Learning and predicting individual, heterogeneous behavior has been explored in multi-agent game settings. These works demonstrate that accounting for heterogeneous behavior can lead to richer and more accurate models of human behavior [12, 3]. Like these approaches, PAWS-Learn exploits the value of the natural heterogeneity in human behavior, and our predictions are more accurate as a result. In order to exploit heterogeneity in this domain, however, PAWS-Learn must address the challenge presented by the sparseness of identified data and correctly identify the adversary population's heterogeneous behaviors from this limited data.

Data collection and aggregation software [18, 1] have been enabling conservation managers to more effectively concert their protection efforts. However, these works do not create patrol routes or identify targets to protect; the creation of patrols is still done by an experienced patrol manager. As discussed in previous works, it is extremely difficult for human schedulers to generate feasible schedules that are also unpredictable [21]; game-theory based applications, such as PAWS, are one example of where an automated approach can generate stronger defender strategies (i.e., schedules) than traditional human-scheduler approaches [17].

Poaching is being carefully studied by Criminologists [8, 14], Geographic Information Systems (GIS) experts [13, 4], and Wildlife Management staff [22]. A variety of methods are used to identify critical points in the poaching trade, such as GIS analysis and interviews with apprehended poachers. In spite of all these techniques, one of the primary challenges that conservation agencies still face is a lack of law enforcement resources [14, 4]. This lack of resources significantly impacts a conservation manager's ability to create patrols that adequately defend the entire protected area, even if they focus on key locations. By creating a realistic Stackelberg Game model based on these works, PAWS can solve this resource allocation problem and generate high-quality patrols.

3. DOMAIN

The goal of PAWS is to help conservation agencies improve patrol efficiency such that poachers, from fear of being caught, are deterred from poaching in QENP. Wire snaring is one of the main techniques used by poachers in Africa, including QENP, (as shown in figures 2(a),2(b)); poachers can set and leave snares unattended, and come back when they think an animal has been captured. In addition, poachers can conduct surveillance on rangers' activities and patrol patterns; wildlife rangers are well-aware that some neighboring villagers will inform poachers of when they leave for patrol and where they are patrolling [9]. For any number of reasons, such as changes that impact animal migration habits, rangers may change their patrolling patterns; poachers, in turn, continually conduct surveillance on the rangers' changing patrol strategy and adapt their poaching strategies accordingly. As the law en-

(a) A lioness caught in a snare.

(b) A caught poacher holding up a snare.

Figure 2: Lioness photo courtesy of John Coppinger, Remote Africa Safaris Ltd. Poacher snare photo taken by Andrew Lemieux.

forcement officers of the park, park rangers' primary objective is to stop poaching, and their main method of doing so is to patrol the park. During a patrol, rangers will search for signs of illegal activity inside the park, confiscate any poaching equipment found, and apprehend any persons inside the park illegally (e.g., poachers).

In addition to their normal patrol duties, rangers will collect data on any observed or suspected illegal activity. In most cases, if rangers find wire snares, they will not find the poacher that set them. If the rangers do encounter and apprehend poachers, however, they are sometimes able to make the poachers confess to where they set their snares. After the rangers return to the outpost, collected data is uploaded and analyzed. Eventually, enough data will be collected so that the ranger patrol strategies can be continually updated based on any emerging trends. If snares are found by a ranger patrol, they are recorded as data points. Since it is unknown who placed the snares, we refer to these data points as anonymous data points. Identified data points, when a poacher is captured and divulges where they placed snares, are inherently more useful as they can be used to obtain a more complete behavioral model that can better predict where future poachers will place their traps.

For this deployment, a poacher placing a snare in an area represents an attack. In order to have a tractable space for computing defender strategies, we discretize areas into a grid where each cell represents 1 square kilometer, and every cell in the grid could contain wildlife and is thus a valid target for attackers. Terrain also has an impact; poachers and rangers can travel further if they are traversing grasslands instead of a dense forest of varying elevations. In order to simplify distance calculations in our model, we currently focus on one type of terrain, grasslands. Future work will focus on incorporating different types of terrain into the model. Areas of high animal density, such as areas that contain fresh water (e.g., watering holes, lakes), are known to be high-risk areas for poaching [22, 9, 8]. Distance is also an important factor; a snare density study demonstrated that the density began to decrease significantly once they began travelling more than 4 kilometers away from the international border [22]. This finding is intuitive as poachers need to carry back any poached animals or goods, and longer distances will increase the chances of spoilage and apprehension. Even for Ugandan poachers, distance travelled will still be a factor based on similar concerns. Despite the available information from these studies, there are still too many areas for rangers to patrol, and it is a huge cognitive burden to account for these factors (in addition to physical distance constraints and available rangers) while constantly creating new, unpredictable patrols. Based on all of these factors, PAWS will aid patrol managers and determine an optimal strategy that will enable park rangers to effectively cover these numerous areas with their limited resources.

4. MODEL IN PAWS

4.1 Stackelberg Game Formulation

Based on our discussion of the wildlife crime domain and its various parameters of interest, we apply a game theoretic framework, more specifically Stackelberg Security Games (SSGs) [19], to the problem and first model the interaction between the rangers and the poachers. In a SSG, there are two types of players: the defender (leader) commits to a strategy first; the follower then responds after observing the leader's strategy. The defender's goal is to protect a set of targets, with limited security resources, from being attacked by the adversary. The adversary will first conduct surveillance to learn about the defender's strategy, and then he (he by convention) will select a target to attack.

In the wildlife crime problem, the ranger plays as the leader and the poachers are the followers. While the rangers are trying to protect animals by patrolling locations where they frequently appear, the poachers are trying to poach the animals at these areas. As discussed earlier, we discretize the area into a grid where each cell represents 1 square kilometer. We use $\mathcal{T}$ to denote the set of locations that can be targeted by the poacher, where $i \in \mathcal{T}$ represents the ith target. If the poacher selects target i and it is covered by the rangers, he receives a utility of $U_{p,i}^c$. If the selected target is not covered by rangers, he receives a utility of $U_{p,i}^u$. The ranger's utility is denoted similarly by $U_{r,i}^c$ and $U_{r,i}^u$. As a key property of SSG, we assume $U_{p,i}^c \leq U_{p,i}^u$ and $U_{r,i}^c \geq U_{r,i}^u$. Simply put, adding resources to cover a target hurts poachers and helps the rangers.

As discussed in the previous section 3, animal density is a key factor in determining poaching risk, and we thus model it as the primary determinant of reward for poachers (i.e., $U_{p,i}^u$) and penalty for rangers (i.e., $U_{r,i}^u$). Areas with a high density of animals are attractive to poachers since they are more likely to have a successful hunt. Similarly, rangers will view these areas as costly if left unprotected. Distance is also a determining factor in poaching reward. Although a poacher may view an area with a large density of animals as attractive, it may be too far away to be rewarding. We also need to model the penalty for poachers (i.e., $U_{p,i}^c$) and reward for rangers (i.e., $U_{r,i}^c$). If the poachers attack a defended area, they will incur a fixed penalty that represents a fine. The poachers will also incur an additional penalty that increases with the distance that they travel from their starting point. Rangers will receive a flat (i.e., uniform) reward based on the poacher's fixed penalty but not on the distance travelled. This uniform reward represents the ranger's lack of preference on where or how poachers are found; as long as poachers are apprehended, the patrol is considered a success.

In our SSG model for this wildlife crime problem, we assume a single leader (i.e., a single group of rangers who are executing the same patrolling strategy) and a population of poachers. We also assume that poachers respond to the rangers' patrolling strategy independently, and we defer to future work to consider potential collaboration between poachers. We adopt a compact representation of the rangers' patrolling strategy: $\boldsymbol{x} = \langle x_i \rangle$ where x_i denotes the probability that i will be covered by the rangers. The actual patrol can be derived from this compact representation using sampling techniques similar to those in previous SSG applications [17, 20]. Given a defender strategy $\boldsymbol{x}$, we denote the response of a poacher as $\langle q_i(\omega|\boldsymbol{x}) \rangle$, where $q_i(\omega|\boldsymbol{x})$ represents the probability that the poacher will select target i. The parameter ω is associated with the poacher's behavioral model, which we will discuss in more details in Section 4.2. Table 1 lists key notations used in this paper.

We model the repeated crime activities of the poachers as the following: in each round of the interaction between the rangers and the poachers, the ranger executes the same mixed strategy over

Table 1: Notations used in this paper

$\mathcal{T}$	Set of targets; $i \in \mathcal{T}$ denotes target i
x_i	Probability that target i is covered by a resource
$U_{r,i}^c$	Ranger utility for covering i if it's selected by the poacher
$U_{r,i}^u$	Ranger utility for not covering i if it's selected
$U_{p,i}^c$	Poacher utility for selecting i if it's covered
$U_{p,i}^u$	Poacher utility for selecting i if it's not covered
ω	Parameter of the SUQR model
$f(\omega)$	Probability density function of ω
$U_r(\boldsymbol{x}\|\omega)$	Ranger expected utility by playing strategy $\boldsymbol{x}$ against a poach with the model parameter ω
$U_r(\boldsymbol{x})$	Ranger expected utility by playing strategy $\boldsymbol{x}$ against the whole population of the poachers
$q_i(\omega\|\mathcal{G})$	Probability that poacher with parameter ω selects target i in game $\mathcal{G}$

Table 2: Log-likelihood

	Single Parameter Setting	Parameter Learned for each subject
Training Set	-1.62	-1.09
Testing Set	-1.68	-1.15

dimensions. Furthermore, we report in Table 2 the average log-likelihood of the SUQR model with the parameter value learned for each subject. We also include in Table 2 the log-likelihood of the SUQR model with the assumption that the parameter value is the same for all the subjects. The results are evaluated using cross-validation. Table 2 shows that the predictive power of the model improves by tuning the parameter for each subject, since the log-likelihood of the prediction by the model is increased. On average, the log-likelihood of the SUQR model with the parameter learned for each subject is 0.53 higher than that with a uniform parameter across all subjects. In other words, the prediction of the former model is 1.70 (i.e. $e^{0.53}$) times more likely than that of the latter.

a period of time (e.g., a month); the poachers will first conduct surveillance on the rangers' patrolling strategy and then respond. If the ranger switches the patrolling strategy, a new round starts. We assume that the poachers are myopic (i.e., they make their decision based on their knowledge of the ranger's strategy in the current round). In this paper, we also assume the poachers' surveillance grants them perfect knowledge about the rangers' strategy; we defer to future work to consider the noise in poachers' understanding of the rangers' strategy due to limited observations.

4.2 Behavioral Heterogeneity

The model we use to predict the behavior of the poachers is based on the SUQR model [10] and replaces the assumption of a single parameter setting with a probabilistic distribution of the model parameter in order to incorporate the heterogeneity among a large population of adversaries. SUQR extends the classic quantal response model by replacing the expected utility function with a subjective utility function:

$$SU_i(\omega) = \omega_1 x_i + \omega_2 U_{p,i}^u + \omega_3 U_{p,i}^c \qquad (1)$$

where the parameter $\omega = \langle \omega_1, \omega_2, \omega_3 \rangle$ measures the weight of each factor in the adversary's decision making process. In previous work, ω was learned using data collected with human subjects from Amazon Mechanical Turk and assumed that there was a single parameter ω. We will show that the parameters learned for individuals in the data set differ from each other. We then show that the model's predictive power significantly improves if the parameter is changed from a single value to a probabilistic distribution.

In the data set collected in [15, 10], each subject played 25-30 games. In total, data was collected on about 760 subjects. We learn the SUQR parameter for each individual by maximizing the log-likelihood defined in Equation (2)

$$\log L(\omega) = \sum_k \log(q_{c_k}(\omega\|\mathcal{G}_k)) \qquad (2)$$

where, $\mathcal{G}_k$ denotes the kth game played by the subject. c_k is the index of the target selected by the subject in this game. q_{c_k} represents the probability that target c_k will be selected by the subject predicted by the SUQR model, which is computed as the following:
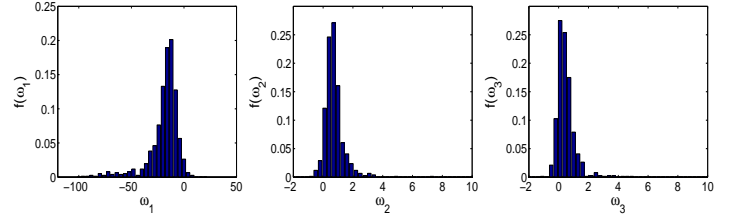
$$q_{c_k}(\omega\|\mathcal{G}_k) = \frac{e^{SU_{c_k}(\omega\|\mathcal{G}_k)}}{\sum_i e^{SU_i(\omega\|\mathcal{G}_k)}} \qquad (3)$$

where, $SU_i(\omega\|\mathcal{G}_k)$ is the subjective utility function as described in Equation (1) given a game instance $\mathcal{G}_k$. Figure 3 displays the empirical PDF of ω. It shows a shape of normal distribution in all three

Figure 3: Empirical Marginal PDF of the SUQR parameter among all the 760 subjects

Given the results shown in Figure 3, we assume a probabilistic, normal distribution of the SUQR parameter ω in order to incorporate the heterogeneity of the decision-making process of the whole population of poachers. The SUQR model with a specific value of ω essentially represents one type of poacher. With the continuous distribution of ω, we are indeed facing a Bayesian Stackelberg game with infinite types. We denote the probability density function of ω as $f(\omega)$.

4.3 Adapting patrolling strategy using historical crime data

In the domain of wildlife crime, if the behavioral model of the whole adversary population is given, the optimal patrolling strategy $\boldsymbol{x}^*$ is the one that maximizes the expected utility of the rangers.

$$\boldsymbol{x}^* = arg \max_{\boldsymbol{x}} \oint U_r(\boldsymbol{x}\|\omega)f(\omega)d\omega \qquad (4)$$

where, $U_r(\boldsymbol{x}\|\omega)$ is the rangers' expected utility by executing strategy x against a poacher that has a model parameter of ω. $U_r(\boldsymbol{x}\|\omega)$ is computed as the following,

$$U_r(\boldsymbol{x}\|\omega) = \sum_i U_{r,i}(\boldsymbol{x}\|\omega)q_i(\omega\|\mathcal{G}) \qquad (5)$$

where, $U_{r,i}(\boldsymbol{x}\|\omega)$ is the rangers' expected utility if target i is selected by the poachers. $U_r(\boldsymbol{x}\|\omega)$ is a nonlinear fractional function given that $q_i(\omega\|\mathcal{G})$ follows the prediction of the SUQR model.

In reality, the behavioral model of the adversary population is unknown to the rangers. Thus, a key challenge for obtaining an optimal patrolling strategy is to learn the poachers' behavioral models. More specifically, we want to learn the distribution of the SUQR

model parameter. In the wildlife crime problem, data is often available about historical crime activities. Recall that these data points record the snares found by the rangers, which can be either anonymous or identified. The identified crime data that is linked to an individual poacher can be used to learn his behavioral model (i.e., estimate the SUQR model parameter for that poacher). In contrast, it is impossible to directly use anonymous crime data to build a behavioral model for any individuals. In theory, with enough identified crime data, we could estimate the underlying population distribution of ω directly. In reality, however, identified crime data is rare compared to anonymous crime data.

5. RESEARCH ADVANCES IN PAWS

Recall that existing techniques in SSG cannot be applied directly to PAWS due to the new challenges coming from this new domain. In this section, we describe the novel research advances developed for solving the SSG in PAWS.

5.1 Learn the Behavioral Model

At the beginning of the game, rangers only know that the distribution of the poacher population's model parameter follows a normal distribution. The goal is to learn the multi-variable normal distribution (i.e., the mean μ and the covariance matrix Σ) of the 3-dimensional SUQR model parameter ω as data becomes available.

As previously discussed, identified data, although sparse, can be used to directly learn poachers' individual behavioral models. Since it is sparse, it takes a much longer time to collect enough data to learn a reasonable distribution. In contrast, there is much more anonymous crime data collected. As we will show, we can learn the behavioral model of the poacher population using these two types of data. Furthermore, we combine the use of the sparse identified data to boost the convergence of the learning.

Let's first define the format of the data collected in each round of the game. Let $N_a^{(t)}$ be the number of anonymous crimes observed by the ranger in round t and $N_c^{(t)}$ be the number of captured poachers in round t. Furthermore, let $\mathcal{A}^{(t)} = \{a_j^{(t)}|j=1,...,N_a^{(t)}\}$ denote the set of targets chosen by the anonymous poachers in round t and $\Omega^{(t)} = \{\omega_k^{(t)}|k=1,...,N_c^{(t)}\}$ denote the set of parameter values associated with the captured poachers in round t. $\omega_k^{(t)}$ is the SUQR parameter of the kth captured poacher in round t. We assume that a captured poacher will confess his entire crime history in all previous rounds. For the kth captured poacher in round t, we denote $\mathcal{C}_k^{(t)} = \{c_{k,l}^{(t)}\}$ as the set of crimes committed by him, where the index l in $c_{k,l}^{(t)}$ represents the lth crime committed by him. $c_{k,l}^{(t)} = (\alpha_{k,l}^{(t)}, \boldsymbol{x}_{k,l}^{(t)})$ includes the target chosen by the poacher when the crime was committed (denoted as $\alpha_{k,l}^{(t)}$) and the resource allocation strategy of the rangers at the time (denoted as $\boldsymbol{x}_{k,l}^{(t)}$). To simplify the notation, we denote $\alpha_{k,l}^{(t)}$ as α_l and $\boldsymbol{x}_{k,l}^{(t)}$ as χ_l in the following part of the paper.

5.1.1 Learning with the identified data

For each captured poacher, the associated SUQR model parameter can be estimated with Maximum Likelihood Estimation (MLE).

$$\omega_k^{(t)} = arg\max_{\omega} \log L(\omega|\mathcal{C}_k^{(t)})$$
$$= arg\max_{\omega} \sum_l \log(q_{\alpha_l}(\omega|\chi_l)) \qquad (6)$$

where, $q_{\alpha_l}(\omega|\chi_l)$ is the predicted probability that the kth captured poacher chooses target α_l when he committed the crime after ob-

serving χ_l as the resource allocation strategy of the rangers. It can be shown that $\log L(\omega|\mathcal{C}_k^{(t)})$ is a concave function, since the Hessian matrix is negative semi-definite.

At round t, there are in total $\sum_{\tau=1}^{t} N_c^{(\tau)}$ poachers captured. After learning the model parameter ω for each of these poachers, there are $\sum_{\tau=1}^{t} N_c^{(\tau)}$ data samples collected from the distribution of the poacher population. By applying MLE, the distribution of ω can be learned from these data samples. Given that ω follows a 3-dimensional normal distribution, the mean and the covariance matrix learned with MLE is calculated as the following:

$$\mu^{(t)} = \frac{1}{\sum_{\tau=1}^{t} N_c^{(\tau)}} \sum_{\tau=1}^{t} \sum_{\omega \in \Omega^{(\tau)}} \omega \qquad (7)$$

$$\Sigma^{(t)} = \frac{1}{\sum_{\tau=1}^{t} N_c^{(\tau)}} \sum_{\tau=1}^{t} \sum_{\omega \in \Omega^{(\tau)}} (\omega - \mu^{(t)})(\omega - \mu^{(t)})^T \qquad (8)$$

5.1.2 Learning with the anonymous data

Each anonymous data item records the target selected by the poacher. Since no information is recorded about the individual poacher who committed the crime, it is impossible to estimate the model parameter like is done with identified data. One potential approach is to treat each anonymous data point as committed by different independent poachers.

$$\omega_j^{(t)} = arg\max_{\omega} \log L(\omega|a_j^{(t)}, \boldsymbol{x}^{(t)}) \qquad (9)$$

where, $\boldsymbol{x}^{(t)}$ is the strategy of the rangers in round t. $\omega_j^{(t)}$ denotes the estimated model parameter of the anonymous poacher who committed the jth crime in round t. Note that in each round, the log-likelihood of any given value of ω only depends on the target that was selected by the poacher. Different poachers with different model parameters will be treated the same if they choose the same target in the same round. Let $\tilde{\Omega}^{(t)} = \{\omega_j^{(t)}|j=1,...,N_a^{(t)}\}$ represent the set of estimated model parameters associated with the $N_a^{(t)}$ anonymous crimes recorded in round t. Similar to how the identified data was used, the maximum likelihood estimation of the mean and covariance matrix of the distribution of the model parameter can be computed as:

$$\tilde{\mu}^{(t)} = \frac{1}{\sum_{\tau=1}^{t} N_a^{(\tau)}} \sum_{\tau=1}^{t} \sum_{\omega \in \tilde{\Omega}^{(\tau)}} \omega \qquad (10)$$

$$\tilde{\Sigma}^{(t)} = \frac{1}{\sum_{\tau=1}^{t} N_a^{(\tau)}} \sum_{\tau=1}^{t} \sum_{\omega \in \tilde{\Omega}^{(\tau)}} (\omega - \tilde{\mu}^{(t)})(\omega - \tilde{\mu}^{(t)})^T \qquad (11)$$

5.1.3 Combining the two kinds of data

Identified data provides an accurate measurement of an individual poacher's behavior. However, it leads to slow learning convergence for the population's behavioral model due to its sparseness. While anonymous data provides a noisy estimation of an individual poacher's behavioral model, it gives a sufficiently accurate measurement of the crime distribution of the poacher population due to the large amount of data points. We propose PAWS-Learn, an algorithm to improve the estimation of the model parameter by combining both the identified data and the anonymous data. Algorithm 1 shows the outline of PAWS-Learn.

At round t, PAWS-Learn first uses the identified data to learn the mean and the covariance, as shown in Line (2). It then measures the accuracy of this estimation using the mean square error (MSE) of

---

Algorithm 1: PAWS-Learn

1 Input: $t, \mathcal{C}^{(\tau)}, \mathcal{A}^{(\tau)}, \boldsymbol{x}^{(\tau)}, \forall \tau = 1 \ldots t - 1$;
2 $(\mu^{(t)}, \Sigma^{(t)}) \leftarrow \texttt{Learn}(\{\mathcal{C}^{(\tau)}, \tau = 1, ..., t - 1\})$;
3 $(\{\omega_n\}) \leftarrow \texttt{Sample}(\mu_{(t)}, \Sigma_{(t)}, N_s)$; ;
4 $\pi_n^o \leftarrow \frac{f(\omega_n|\mu^{(t)}, \Sigma^{(t)})}{\sum_{n'} f(\omega_{n'}|\mu^{(t)}, \Sigma^{(t)})}, \forall n$;
5 $\langle \pi_n \rangle \leftarrow \texttt{Refine}(\{\mathcal{C}^{(\tau)}, \tau = 1, ..., t - 1\}, \langle \pi_n^o \rangle)$; ;
6 Return $(\{\omega_n\}, \langle \pi_n \rangle)$;

Algorithm 2: PAWS-Adapt

1 Input: $N_c^{(t)}, N_a^{(t)}$;
2 $\boldsymbol{x}^{(1)} \leftarrow \textsc{Maximin}$;
3 **for** $\tau = 1, ...$ **do**
4 $(\mathcal{C}^{(\tau)}, \mathcal{A}^{(\tau)}) \leftarrow \texttt{CollectData}(\boldsymbol{x}^{(\tau)})$;
5 $(\{\omega_n\}, \{\pi_n\}) \leftarrow \text{PAWS} - Learn(\mathcal{C}^{(\tau)}, \mathcal{A}^{(\tau)}, \boldsymbol{x}^{(\tau)})$;
6 $\boldsymbol{x}^{(\tau+1)} \leftarrow \texttt{ComputeStrategy}(\{\omega_n\}, \{\pi_n\})$

---

the predicted crime distribution recorded by the anonymous data.

$$MSE^{(t)}(\mu^{(t)}, \Sigma^{(t)}) = \sum_{i \in \mathcal{T}} (\bar{q}_i(\boldsymbol{x}^{(t)}|\mu^{(t)}, \Sigma^{(t)}) - y_i^{(t)})^2 \quad (12)$$

where $y_i^{(t)}$ is the proportion of crimes found at target i as recorded by the anonymous data[2]. $\bar{q}_i(\boldsymbol{x}|\mu^{(t)}, \Sigma^{(t)})$ is the predicted probability that target i will be selected by the poacher population, given $\mathcal{N}(\mu^{(t)}, \Sigma^{(t)})$. Ideally, $\bar{q}_i(\boldsymbol{x}|\mu^{(t)}, \Sigma^{(t)})$ is calculated as

$$\bar{q}_i(\boldsymbol{x}|\mu^{(t)}, \Sigma^{(t)}) = \oint_{\Omega} q_i(\omega, \boldsymbol{x}^{(t)}) f(\omega|\mu^{(t)}, \Sigma^{(t)}) d\omega$$

Let $\pi = \langle \pi_n \rangle$ denote the vector of probabilities associated with the sampled parameter values, where $\sum_n \pi_n = 1$ due to normalization. The predicted probability that target i will be selected by the poacher population in round t is approximated as

$$\bar{q}_i(\boldsymbol{x}^{(t)}) = \sum_n \pi_n q_i(\omega_n, \boldsymbol{x}^{(t)})$$

The quadratic program formulation for minimizing the MSE of the observed crime distribution is shown in Equations (13)-(15).

$$\min_{\pi} \sum_{i \in \mathcal{T}} (\sum_n \pi_n q_i(\omega_n, \boldsymbol{x}^{(t)}) - y_i^{(t)})^2 \quad (13)$$

$$s.t. \sum_n \pi_n = 1, \quad \pi_n \in [0, 1], \forall n \quad (14)$$

$$|\pi_n - \pi_n^o| \leq \beta \pi_n^o, \forall n \quad (15)$$

Equation (15) is to ensure the smoothness of $\langle \pi_n \rangle$ since the values are essentially samples from the probability density function of a normal distribution. More specifically, it constrains π_n to be within a certain distance of the initial value π_n^o. The parameter β is set to decide the range of π_n proportion to π_n^o. As shown in Line (4),π_n^o is set to the pdf of the current estimated distribution $\mathcal{N}(\mu^{(t)}, \Sigma^{(t)})$: $\pi_n^o = C \cdot f(\omega_n|\mu^{(t)}, \Sigma^{(t)})$, where $C = \frac{1}{\sum_n f(\omega_n|\mu^{(t)}, \Sigma^{(t)})}$ is the constant to make sure that $\sum_n \pi_n = 1$. As shown in Line (5), PAWS-Learn refines the probabilities of the sampled parameter values by solving the above quadratic programming problem.

5.2 Adapting patrolling strategy

We propose PAWS-Adapt, a framework to adaptively design the patrolling strategy for the rangers. Let $(\mathcal{C}^o, \mathcal{A}^o, \boldsymbol{x}^o)$ be the initial data set. At round t, PAWS-Adapt first estimates the behavioral model with all the historical data by calling PAWS-Learn. Let $(\omega_n^{(t)}, \langle \pi_n^{(t)} \rangle)$ be the learning results of the poacher population's behaviorial model by PAWS-Learn. PAWS-Adapt then computes the optimal patrolling strategy, based on the current learning result, to execute in the next round.

---

[2]PAWS-Learn currently assumes that in the anonymous data collected by rangers in each round, the observed crime distribution is close to the true distribution.

In computing the optimal patrolling strategy under the given behavioral model, we need to solve the optimization problem in Equation (4), which is equivalent to a Bayesian Stackelberg Game with infinite types. With the representation of discretized samples, we approximate the infinite types with a set of sampled model parameters $\{\omega_n\}$. Given that the objective function in Equation (4) is non-convex, we solve it by finding multiple local optima with random restarts. Let $\boldsymbol{x}^{(t)}$ be the patrolling strategy computed by PAWS-Adapt for round t. The rangers then update their strategy in the new round. As shown in Algorithm 2, the rangers will update the poachers' behavioral models each round after more data is collected. They then switch to a new strategy that was computed with the new model.

6. EVALUATION

6.1 General Game Settings

In the first set of our experiments, we generate a random payoff matrix similar to that in [10]. The crime data points are simulated as the following: given the true distribution of ω, we first draw a set of random parameters for ω to represent the whole poacher population. Let N_p be the total number of poachers. In each round, we first draw a subset of ω from these N_p values to represent the subset of poachers who are going to commit crimes in the current round. Given the patrolling strategy, we then simulated the target choices made by this subset of poachers. These choices are recorded as the anonymous data. Meanwhile, we randomly select a given number of poachers from this subset to represent the poachers that are captured by the rangers in the current round. Once a poacher is captured, the choices he made in the previous round will be linked and recorded as the identified data points.
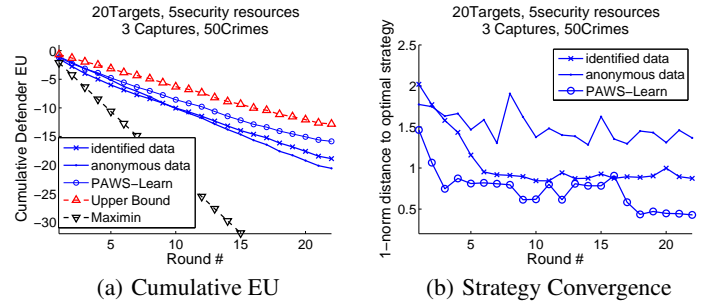

(a) Cumulative EU   (b) Strategy Convergence

Figure 4: Simulation results over round

In Figure 4(a), we show the cumulative expected utility (EU) of the rangers over the round. We compare three different approaches: PAWS-Learn, learning from only the identified data, and learning from only the anonymous data. We also included the maximin strategy as the baseline. The upper bound is computed assuming the rangers know the true distribution of ω. In Figure 4(a), we show the
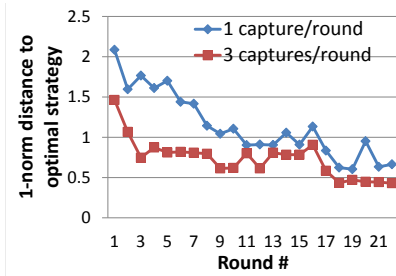
Figure 5: Slow Capture v.s. Fast Capture

average result over 20 random game instances. We set the number of targets to 20 and the number of security resources to 5. The true distribution of ω is the same as that learned in Section 4.2. In each round, 50 anonymous data points are generated, and 3 poachers are captured. As can be seen in the figure, PAWS-Learn outperforms the other two learning approaches that use one type of data. Furthermore, learning indeed helps improve the patrolling strategy since the three solid lines are much closer to the upper bound compared to the baseline solution maximin strategy.

In Figure 4(b), we show the convergence of the patrolling strategy from the three different learning methods. The figure shows that PAWS-Learn converges faster than the other two methods. Thus, combining the two types of data indeed boosts the learning of the poacher population's behavioral model.

In order to show how the speed of capturing poachers impacts the performance of PAWS, we fix the number of anonymous data points to 50 and simulate the captured poachers in each round at two different paces: 1 poacher vs. 3 poachers. Figure 5 shows the convergence of PAWS-Learn in these two cases. It is clear that the strategy converges faster if more poachers are captured.

We compare the cumulative EU achieved by the three different methods under varying number of targets and varying amount of resources. In both Figure 6(a) and 6(b), the y-axis displays the cumulative EU of the rangers at the end of round 20. In both figures, we simulate 50 crimes and randomly generate 3 captured poachers each round. In Figure 6(a), we vary the number of resources on the x-axis while fixing the number of targets to 20. It shows that the cumulative EU increases as more resources are added. In addition, PAWS-Learn outperforms the other two approaches regardless of resource quantity. Similarly, we vary the number of targets on the x-axis in Figure 6(b) while fixing the amount of resources to 5. The better performance of PAWS-Learn over the other two learning methods can be seen from the figure regardless of the number of targets.
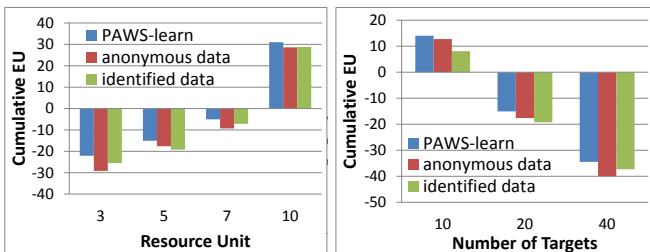
(a) 20 Targets, 3 Captures, 50 Crimes (b) 5 security resources, 3 Captures, 50 Crimes

Figure 6: Comparing cumulative EU at round 20

6.2 Results for the Deployment Area
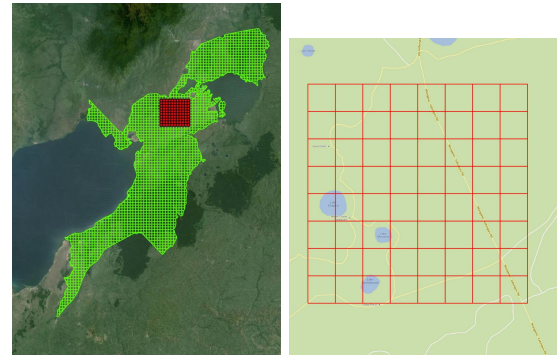
(a) A zoomed out view of the simulation area. (b) The 64 sq. km grid overlaid on the simulation area.

Figure 7: The QENP area of interest for our simulation

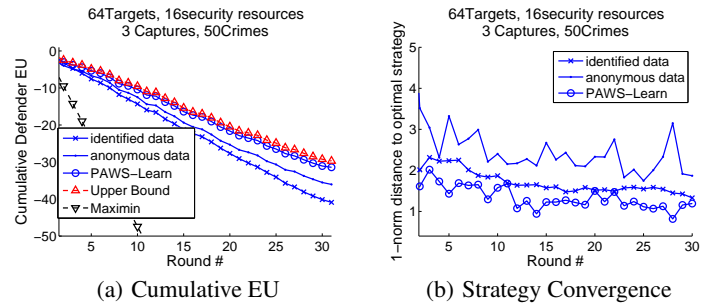
(a) Cumulative EU (b) Strategy Convergence

Figure 8: Simulation results over round for the 64 sq. km grid area in QENP

We now show the experiment results of applying PAWS to QENP. We focus on a 64 square kilometer area in QENP that features flat grasslands, an international trade route that connects nearby Democratic Republic of the Congo, smaller roads, and fresh water. In our simulation area 7(b), the series of lakes are modeled as areas of high animal density. Since the roads in this area provide multiple access points for poachers and rangers, they can leave the closest road at the closest point to their targeted cells. We calculate travel distances according to that rationale. These representations of animal density and distance form the primary basis for the payoffs for both the rangers and the poachers. The poachers' reward (i.e., if they choose a cell not covered by rangers) depends on the relative animal density of the cell and the travelling cost to that cell. The travelling cost depends on the distance from the closest entry point (e.g., a road). Therefore, a lake close to a road is at high risk for poaching and is thus modelled as an area of high reward to the poachers. In turn, the poachers' penalty (i.e., the chosen cell is covered by rangers) is decided by the travelling cost to a cell and the loss of being captured by the rangers. The rangers' reward is considered to be uniform since their goal is to search for snares and capture poachers regardless of the location. The penalty for the rangers (i.e., fail to find snares at a place) is decided by the animal density of the cell. Further discussion of the rationale can be found in the domain section 3.

We run simulations with a sample game, similar to that in the general setting as explained in Section 6.1. Figure 8 displays the simulation results, where the number of resources is set to 16, indicating that a single patrol covers 16 grid areas in the map. The ranger's cumulative EU is shown in Figure 8(a). It can be seen
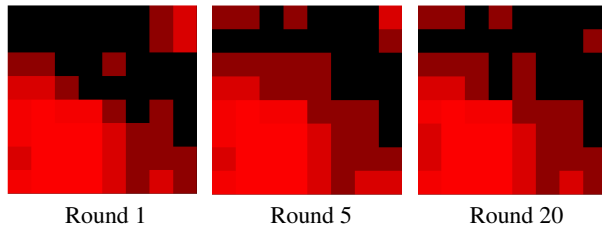
| Round 1 | Round 5 | Round 20 |

Figure 9: Patrolling coverage density in the park

that PAWS-Learn achieves very close performance to the optimal strategy. The convergence of the patrolling strategy to the optimal strategy is shown in Figure 8(b).

In order to help visualize the change of ranger's patrolling strategy, we show the coverage density in the 8-by-8 area at three different rounds in Figure 9. Darker colors indicate less coverage in the area. Note that there are three lakes located in the lower-left area, where the density of animals is higher. It is clear that these areas are covered more by the rangers. The figure also shows a clear shift of the patrolling coverage over the rounds.

These results are enthusiastically received by our collaborator at QENP. While the existing framework requires manual analysis of the snare data, PAWS provides a systematic way of generating patrolling strategies based on automatic analysis of the data. PAWS will start to be tested in the field in March 2014 with actual deployment planned for the latter portion of 2014.

7. SUMMARY

This paper introduced the innovative application PAWS, the result of the joint effort with researchers at QENP *where* PAWS *will be deployed*. Wildlife crime patrols, while essential to combating wildlife poaching, are difficult to create and introduce a large cognitive burden on outpost commanders due to the large number of factors involved. As demonstrated in our experimental results, PAWS successfully models the wildlife crime domain and optimizes wildlife crime patrols while remaining flexible enough to operate generally and in a specific deployed area. Due to the unique challenges introduced by wildlife crime, we have also made a series of necessary technical contributions[3]. Specifically, the success of PAWS depend on the following novel contributions: 1. a stochastic behavioral model extension that captures the population's heterogeneity; 2. PAWS-Learn, which combines both anonymous and identified data to improve the accuracy of the estimated behavioral model; 3. PAWS-Adapt, which adapts the rangers' patrolling strategy against the behavioral model generated by PAWS-Learn.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] S. Collaboration. http://www.smartconservationsoftware.org/.

[2] V. Conitzer. Computing game-theoretic solutions and applications to security. In *AAAI*, pages 2106–2112, 2012.

[3] Y. Gal, A. Pfeffer, F. Marzo, and B. J. Grosz. Learning social preferences in games. In *AAAI*, 2004.

[4] M. Hamisi. *Identification and mapping risk areas for zebra poaching: A case of Tarangire National Park, Tanzania*. Thesis, ITC, 2008.

[5] T. Holmern, M. J., and R. E. Local law enforcement and illegal bushmeat hunting outside the serengeti national park, tanzania. *Environmental Conservation*, 34:55–63, 2007.

[6] J. Letchford, V. Conitzer, and K. Munagala. Learning and approximating the optimal strategy to commit to. In *Proceedings of the 2nd International Symposium on Algorithmic Game Theory*, pages 250–262, 2009.

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

[8] M. Montesh. Rhino poaching: A new form of organised crime. Technical report, College of Law Research and Innovation Committee of the University of South Africa, 2013.

[9] W. Moreto. *To Conserve and Protect: Examining Law Enforcement Ranger Culture and Operations in Queen Elizabeth National Park, Uganda*. Thesis, Rutgers, 2013.

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

[11] C. of New York. Nypd website- frequently asked questions, 2013.

[12] Y. Oshrat, R. Lin, and S. Kraus. Facing the challenge of human-agent negotiations via effective general opponent modeling. In *AAMAS*, pages 377–384, 2009.

[13] E. Ouko. *Where, when, and why are there elephant poaching hotspots in Kenya?* Thesis, University of Twente, 2013.

[14] S. Pires. *The Illegal Parrot Trade in the Neo-Tropics: The Relationship between Poaching and Illicit Pet Markets*. Thesis, Rutgers, 2012.

[15] J. Pita, R. John, R. Maheswaran, M. Tambe, and S. Kraus. A robust approach to addressing human adversaries in security games. *In ECAI*, 2012.

[16] G. T. I. Secretariat. Global tiger recovery program implementation plan: 2013-14. Report, The World Bank, Washington, D.C., 2013.

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

[18] E. J. Stokes. Improving effectiveness of protection efforts in tiger source sites: Developing a framework for law enforcement monitoring using mist. *Integrative Zoology*, (5):363–377, 2010.

[19] M. Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, New York, NY, 2011.

[20] J. Tsai, S. Rathi, C. Kiekintveld, F. Ordonez, and M. Tambe. Iris - a tool for strategic security allocation in transportation networks. *In AAMAS*, 2009.

[21] W. A. Wagenaar. Generation of random sequences by human subjects: A critical survey of the literature. 1972.

[22] Y. A. Wato, G. M. Wahungu, and M. M. Okello. Correlates of wildlife snaring patterns in tsavo west national park, kenya. *Biological Conservation*, 132(4):500–509, 2006.

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

---

[3] A key difference of this paper from [23] is the focus on learning the behavioral model of the poachers, while Qian et al. focus on learning the payoff matrix of the game.