

## Security via Strategic Randomization

Dr. Milind Tambe and Dr. Fernando Ordonez, University Of Southern California  
tambe@usc.edu, fordon@usc.edu

### 1. Overview

The ARMOR (Assistant for Randomized Monitoring Over Routes) project has been developing methods for creating randomized plans and processes, so that even if an attacker observes the plans over time, he/she cannot predict its progression or how it will unfold in the future, thus providing risk reduction while guaranteeing a certain level of protection quality. Our key research accomplishments have been developing efficient algorithms to address the problem of providing decision support to a patrolling or security service in an adversarial domain. The idea is to create patrols that can achieve a high level of coverage or reward while taking into account the presence of an adversary. We assume that the adversary can learn or observe the patrolling strategy and use this to its advantage.

We follow two different approaches depending on what is known about the adversary. If there is no information about the adversary we use a Markov Decision Process (MDP) to represent patrols and identify randomized solutions that minimize the information available to the adversary. This leads to algorithms BRLP, for policy randomization of MDPs, and RDR, for Decentralized-Partially Observable MDPs. Second, when there is partial information about the adversary we decide on efficient patrols by solving a Bayesian Stackelberg game. Here, the leader decides first on a patrolling strategy and then an adversary, of possibly many adversary types, selects its best response for the given patrol. We provide an efficient MIP formulation to solve this NP-hard problem. Our experimental results show the efficiency of these algorithms and illustrate how these techniques provide optimal and secure patrolling policies.

The result of this research is embedded in the ARMOR system currently being applied at the Los Angeles International Airport.

The ARMOR system, the result of our project, has been applied on a trial basis for randomizing checkpoints at the Los Angeles International Airport (LAX). This trial has been on-going since August 2007. Please see below for a checkpoint as recommended by ARMOR. We will soon be applying our system for randomizing deployment of K-9 units at LAX.

In accomplishing this goal, we collaborated with the Los Angeles World Airport (LAWA) police.



"This research was supported by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under grant number 2007-ST-061-000001. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security."

## 2. Research Accomplishments

Effective patrolling strategies are a key ingredient in being able to manage and deter risk posed by terrorists or adversaries. Our work has developed a series of models and algorithms to decide best patrolling strategies that can achieve a high level of coverage or reward while taking into account the presence of an adversary. We assume the adversary can observe and learn the patrolling strategy and use it to its advantage. This makes ineffective classical optimal patrolling strategies which can be predictable and new models must be developed. We develop two types of models, depending on what is known of the adversary:

- When there is no adversary model we use MDPs and POMDPs to model patrols of an agent/team of agents respectively. On these decision processes we obtain solutions that trade-off the reward and the randomness of the patrolling policy.
- When we assume partial information of the adversary we model the interaction as a Bayesian Stackelberg game, where the agent is the leader and decides a patrol with the adversary, of possibly different types, following.

For both types of models we introduce a series of efficient algorithms to compute the solutions to real sized examples.

Security, commonly defined as the ability to deal with intentional threats from other agents is a major challenge for agents or agent-teams deployed in these adversarial domains (note that the word “agents” here broadly refers to robots, software agents or people). Such adversarial scenarios arise in a wide variety of situations that are becoming increasingly important such as agents patrolling to provide perimeter security around critical infrastructure or performing routine security checks. These domains have the following characteristics: (a) The agent or agent-team needs to commit to a security policy while the adversaries may observe and exploit the policy committed to. (b) The agent/agent-team potentially faces different types of adversaries and has varying information available about the adversaries (thus limiting the agents’ ability to model its adversaries).

To address security in such domains, we developed two types of algorithms. First, when the agent has no model of its adversaries, our key idea is to randomize agent’s policies to minimize the information gained by adversaries. To that end, we developed algorithms for generating randomized policies (i.e. plans) for both the Markov Decision Processes (MDPs) and the Decentralized-Partially Observable MDPs (Dec POMDPs) that maintain quality guarantees. Second, when the agent has partial model of the adversaries, we model the security domain as a Bayesian Stackelberg game where the agent’s model of the adversary includes a probability distribution over possible adversary types. While the optimal policy selection for a Bayesian Stackelberg game is known to be NP-hard, our solution approach based on an efficient Mixed Integer Linear Program (MILP) provides significant speedups over existing approaches while obtaining the optimal solution.

Minimum Information Markov Decision Process - We represent the agent’s decision process to select a patrolling policy as MDPs and Dec-POMDPs, since they are able to address uncertainties in real-world domains. It is known that optimal solutions to these models will be deterministic and hence vulnerable to adversaries. Therefore in addition to a reward or coverage objective, randomness is desirable in the selected policy. Since arbitrary randomization can violate quality constraints (for example, the resource usage should be below a certain threshold or key areas must be patrolled with a certain frequency), our algorithms guarantee quality constraints on the randomized policies generated. We quantify the randomness of a solution to an MDP or Dec-POMDP as a (weighted) sum of the entropy of the decision policy at each stage and we consider the problem of solving for the maximum entropy policy subject to

reward constraints. The measures of randomness, however, are non-linear non-convex functions of the decision variables in the patrolling policy and thus solving these problems is difficult for large problems.

For efficiency, we provide a novel linear program based algorithm for randomized policy generation in MDPs, and then build on this program for a heuristic solution for Dec-POMDPs. The basis of the algorithm is to solve a linear program that obtains the maximum reward solution with constraints on the randomness of the policy and use this problem in a Binary Search procedure to find a solution with a required reward threshold.

Bayesian Stackelberg Games - In a Stackelberg game, a leader commits to a strategy first, and then a follower (or group of followers) selfishly optimizes their own rewards, considering the action chosen by the leader. Stackelberg games are commonly used to model attacker-defender scenarios in security domains as well as in patrolling. To illustrate this game, consider for example a single security agent that is responsible for patrolling a region, searching for robbers. Since the security agent (the leader) cannot be in all areas of the region at once, it must instead choose some strategy of patrolling various areas within the region, one at a time. This strategy could be a mixed strategy in order to be unpredictable to robbers (followers). The robbers, after observing the pattern of patrols over time, can then decide their own strategy by choosing a location to rob. Because we are unsure of which follower we face, and assume only an a-priori distribution over multiple follower types, the patrolling domain is a Bayesian Stackelberg game. The resulting policy of the Bayesian Stackelberg game randomizes the agent's possible strategies, while taking into account the probability distribution over adversary types.

The problem of choosing an optimal strategy for the leader to commit to in a Stackelberg game is known to be NP-hard for Bayesian games with multiple types of followers. Existing methods for finding optimal strategies for non-Bayesian games can be applied to the Bayesian Stackelberg game by converting the Bayesian game into a normal-form game by the Harsanyi transformation. However, by transforming the game, the compact structure of the Bayesian game is lost leading to an exponential increase in the size of the game. In addition, this prior method to solve Stackelberg games requires running a large number of linear programs, some of which may be infeasible. If, on the other hand, we wish to compute the highest-reward Nash equilibrium, new methods such as MIP-Nash, using mixed-integer linear programs (MILPs) may be used, since the highest-reward Bayes-Nash equilibrium is equivalent to the corresponding Nash equilibrium in the transformed game. However, as stated above, the compactness in structure of the Bayesian game is lost. In addition, since the Nash equilibrium assumes a simultaneous choice of strategies, the advantages of being the leader are not considered.

We have developed an exact and an approximate technique to solve Bayesian Stackelberg games for the patrolling domain. These algorithms exploit the compact Bayesian representation and solve a single mixed integer programming problem to decide the optimal solution for the leader with constraints that enforce that every follower selects an optimal pure strategy given a strategy for the leader. The approximate algorithm further limits the choice of optimal strategy for the leader by only considering policies that have a multiset support. Both algorithms require the solution to a single mixed integer programming problem that is not exponential in size of the input and can be solved efficiently for reasonably large problems. This method of solving Bayesian Stackelberg problems has three key advantages. First, the method allows for a Bayesian game to be expressed compactly without requiring conversion to a normal-form game via the Harsanyi transformation. Second, the method requires only one mixed-integer linear program to be solved, rather than a set of such programs, thus leading to a further performance improvement. Third, it directly searches for an optimal strategy, rather than a Nash (or Bayes-Nash) equilibrium, thus allowing it to find high-reward non-equilibrium strategies. Our experimental results illustrate that the proposed algorithms outperform existing algorithms and have enabled us to find optimal secure policies efficiently for an increasingly important class of security domains.

### 3. Applied Relevance

The problem of developing more effective/secure patrolling policies is relevant for many security applications, in particular to border security and container inspections/port security. To date we have begun exploring the use of these enhanced patrolling strategies for police patrols and checkpoints at the Los Angeles International Airport and with USC's Department of Public Safety. The idea of using randomness to thwart the objectives of adversaries could also be applied more broadly to any decision that a leader can make as a probability distribution over actions.

For our collaboration with Los Angeles World Airports (LAWA) police and USC's DPS we have developed a computer system interface (ARMOR) to these algorithms to determine best vehicle checkpoint locations and patrol routes. The system is being used in a pilot run with the LAWA police at LAX.

We have developed four different models to obtain effective patrolling policies. These are:

- BRLP (Binary Search for Randomization with Linear Programming) - A linear programming based algorithm that uses binary search to adjust the randomness level of a solution to obtain a target reward value for a single agent MDP.
- RDR (Rolling Down Randomization) - An efficient heuristic that sequentially uses BRLP to distribute the allowable reduction in reward among a team of agents. This heuristic aims for a solution with a high randomness to a Dec-POMDP with a quality guarantee.
- ASAP (Agent Security via Approximate Policies) - Solves approximately a Bayesian Stackelberg Game by constraining the solution of the leader to solutions with a support in a multiset of the possible actions. Requires the solution to a single mixed integer programming problem.
- DOBSS (Decomposed Optimal Bayesian Stackelberg Solver) - Solves a Bayesian Stackelberg Game as a single mixed integer programming problem.

We have developed an interface for the models above called ARMOR (Assistant for Randomized Monitoring Over Routes). This computer system has two flavors, for checkpoint scheduling (ARMOR-Checkpoints) and for scheduling of K-9 routes (ARMOR-K9). In addition to allow for the input of the specific problem constraints in each domain, and interface with the DOBSS algorithm to obtain the optimal policy, the computer system provides a number of output options to analyze and change the solution by adding additional constraints, and to generate reports using the standard LAWA police format.

We have begun a pilot run of the ARMOR-Checkpoints and ARMOR-K9 in collaboration with the Los Angeles World Airport (LAWA) police. This collaboration has involved extensive meetings with LAWA police personnel to calibrate the data and include specific constraints of each problem to the DOBSS model. We have provided LAWA police with alternative schedules of vehicle checkpoints and K-9 patrols obtained from the ARMOR software and have received feedback on the ease of use and effectiveness of the recommendations.

We also had meetings with the chief of police at USC's Department of Public Safety (DPS) to study how to adapt the ARMOR system to help design DPS patrols at USC and around campus.