Defending Against Opportunistic Criminals: New Game-Theoretic Frameworks and Algorithms

Chao Zhang[1], Albert Xin Jiang[1] Martin B. Short[2], P. Jeffrey Brantingham[3], and Milind Tambe[1]

[1] University of Southern California, Los Angeles, CA 90089, USA,
zhan661, jiangx, tambe @usc.edu,
[2] Georgia Institute of Technology, Atlanta, GA 30332, USA,
mbshort@math.gatech.edu,
[3] University of California, Los Angeles, CA 90095, USA
pjb@anthro.ucla.edu

Abstract. This paper introduces a new game-theoretic framework and algorithms for addressing opportunistic crime. The Stackelberg Security Game (SSG), which models highly strategic and resourceful adversaries, has become an important computational framework within multiagent systems. Unfortunately, SSG is ill-suited as a framework for handling opportunistic crimes, which are committed by criminals who are less strategic in planning attacks and more flexible in executing them than SSG assumes. Yet, opportunistic crime is what is commonly seen in most urban settings.We therefore introduce the Opportunistic Security Game (OSG), a computational framework to recommend deployment strategies for defenders to control opportunistic crimes. Our first contribution in OSG is a novel model for opportunistic adversaries, who (i) opportunistically and repeatedly seek targets; (ii) react to real-time information at execution time rather than planning attacks in advance; and (iii) have limited observation of defender strategies. Our second contribution to OSG is a new exact algorithm EOSG to optimize defender strategies given our opportunistic adversaries. Our third contribution is the development of a fast heuristic algorithm to solve large-scale OSG problems, exploiting a compact representation.We use urban transportation systems as a critical motivating domain, and provide detailed experimental results based on a real-world system.

1 Introduction

Security is a critical societal challenge. We focus on urban security: the problem of preventing urban crimes. The Stackelberg Security Game (SSG) was proposed to model highly strategic and capable adversaries who conduct careful surveillance and plan attacks [1, 2], and has become an important computational framework for allocating security resources against such adversaries. While there are such highly capable adversaries in the urban security domain, they likely comprise only a small portion of the overall set of adversaries. Instead, the majority of adversaries in urban security are criminals who conduct little planning or surveillance before “attacking” [3]. These adversaries capitalize on local opportunities and react to real-time information. Unfortunately, SSG is ill-suited to model such criminals, as it attributes significant planning and little execution flexibility to adversaries.

Inspired by modern criminological theory [3], this paper introduces the Opportunistic Security Game (OSG), a new computational framework for generating defender strategies to mitigate opportunistic criminals. This paper provides three key contributions. First, we define the OSG model of opportunistic criminals, which has three major novelties compared to SSG adversaries: (i) criminals exhibit Quantal Biased Random Movement, a stochastic pattern of movement to search for crime opportunities that contrasts with SSG adversaries, who are modeled as committed to a single fixed plan or target; (ii) criminals react to real-time information about defenders, flexibly altering plans during execution, a behavior that is supported by findings in criminology literature [4]; (iii) criminals display anchoring bias [5], modeling their limited surveillance of the defender's strategy. Second, we introduce a new exact algorithm, Exact Opportunistic Security Game (EOSG), to optimize the defender's strategy in OSG based on use of a markov chain. The third contribution of this work is a fast algorithm, Compact OPportunistic Security game states (COPS), to solve large scale OSG problems. The number of states in the Markov chain for the OSG grows exponentially with the number of potential targets in the system, as well as with the number of defender resources. COPS compactly represents such states, dramatically reducing computation time with small sacrifice in solution quality; we provided a bound for this error.

Thus, while OSG does share one similarity with SSG — the defender must commit to her strategy first, after which the criminals will choose crime targets — the OSG model of opportunistic adversaries is fundamentally different. This leads us to derive completely new algorithms for OSG. OSG also differs fundamentally from another important class of games, pursuit-evasion games (PEG) [6]; these differences will be discussed in more depth in the related work section.

While OSG is a general framework for handling opportunistic crime, our paper will use as a concrete example crime in urban transportation systems, an important challenge across the world. Transportation systems are at a unique risk of crime because they concentrate large numbers of people in time and space [7]. The challenge in controlling crime can be modeled as an OSG: police conduct patrols within the transportation system to control crime. Criminals travel within the transportation system for such opportunities [8], usually committing crimes such as thefts at stations, where it is easy to escape if necessary [9]. These opportunistic criminals avoid committing crime if they observe police presence at the crime location.

In introducing OSG, this paper proposes to add to the class of important security related game-theoretic frameworks that are widely studied in the literature, including the Stackelberg Security Games and Pursuit Evasion Games frameworks. We use an urban transportation system as an important concrete domain, but OSG's focus is on opportunistic crime in general; the security problems posed by such crime are relevant not only to urban crime, but to other domains including crimes against the environment [10], and potentially to cyber crime [11, 12]. By introducing a new model and new algorithms for this model, we open the door to a new set of research challenges.

2 Related Work

In terms of related work, there are three main areas to consider. First are Pursuit-Evasion Games (PEG), which model a pursuer(s) attempting to capture an evader, often where their movement is based on a graph [6]. However, PEG fail to model criminals who opportunistically and repeatedly strike targets as modeled using QBRM in OSG. Furthermore, in PEG, a pursuer's goal is to capture an evader while in OSG, the defender's goal is to minimize crime; additionally in PEG, the evader's goal is to avoid the pursuer and not seek crime opportunities as in OSG. These critical differences in behaviors of defenders and adversaries lead to new algorithms, i.e., EOGS and COPS, for OSG, that are fundamentally different from algorithms for PEG.

Second are SSG [13–15], which use a model of highly strategic adversaries to generate randomized patrol strategies. The SSG framework has been successfully applied in security domains to generate randomized patrol strategies, e.g., to protect flights [2], for security in the cyber realm [11, 12], and for counter-terrorism and fare evasion checks on trains [16, 17]. Recent work in SSG has begun to consider bounded rationality of adversaries [18] and incorporate some limited flexibility in adversary execution [15]. However, SSG [13–15], again, fails to model criminals who use real-time information to adjust their behavior in consecutive multiple attacks. In SSG, attackers cannot use real-time observation to decide whether to attack at the current time, nor can they use it to update beliefs and plan for their next consecutive attacks. Furthermore, SSG does not investigate efficient algorithms of deriving defender strategies against such opportunistic criminals. The Adversarial Patrolling Game (APG) [19], which is a variant of SSG, does consider the attacker's current observation. However, this game does not consider multiple consecutive attacks. It fails to model attacker's movement during multiple attacks and therefore the influence of current observation on future movement. Recent research has focused on applying game theory in network security [20], especially in communication and computer networks [21, 22]. However, these works again do not consider the flexibility and real-time adjustment of attackers under Stackelberg settings. Besides, the physical constraints (e.g., travel time between targets) in OSG do not exist in communication networks.

A third thread of recent research has made inroads in the modeling of opportunistic criminal behavior, and in how security forces might defend against such adversaries. In [23] burglars are modeled as biased random walkers seeking "attractive" targets, and [24] follows up on this work with a method for generating effective police allocations to combat such criminals. However, these works make the extreme assumption that criminals have no knowledge of the overall strategy of the police, and their behavior is only affected by their observation of the current police allocation in their immediate neighborhood. Also, in [24] police behave in a similarly reactionary way, allocating their resources in an instantaneously optimal way in response to the current crime risk distribution rather than optimizing over an extended time horizon. Furthermore, in [24] there is no notion of the "movement" of police - rather, the distribution of police officers are chosen instantaneously, with no regard for the mechanics of exactly how the allocation may transform from one time step to the next. Our current approach is an attempt to generalize these threads of research.

3 OSG Framework

OSG unfolds on a connected graph that can be seen to model a metro rail system (though many other domains are also possible), where stations are nodes and trains connecting two stations are edges. Fig. 1 shows a simple scenario with three fully connected stations. Stations and trains are collectively referred to as locations. Let the stations be labeled $1, \ldots, N$, with N denoting the number of stations. The train from station i to its neighboring station j is denoted as $i \rightarrow j$. The number of locations is $N_l > N$, e.g., in Fig. 1, $N_l = 9$.

We divide time equally into time steps so that trains arrive at stations at the beginning of each time step. There are two phases in any time step. First is the *decision phase*, the period when trains are at stations for boarding and unboarding. In this phase, each passenger at each location decides where in the system to move next. There are two types of choices available. *Go* $i \rightarrow j$ means that (i) if a passenger is at station i, he gets on the train $i \rightarrow j$; (ii) if he is on a train arriving at station i, he now gets (or stays) on the train $i \rightarrow j$. *Stay* means that the passenger stays at the station, so that if the passenger was on a train, he gets off. After the brief decision phase is the *action phase*, in which trains depart from all stations to all directly connected stations. This model matches the metro systems in Los Angeles, where trains leave stations at regular intervals to all directly connected stations. Without losing generality, we assume that the time it takes to travel between any two adjacent stations is identical; this assumption can be relaxed by including dummy stations. In OSG, the defender ("she") – assisted by our algorithms – is modeled to be perfectly rational. The criminal ("he") is modeled with cognitive biases. Fig. 2 illustrates the OSG flowchart, with relevant equation numbers near variables – these variables and equations are described in the following.

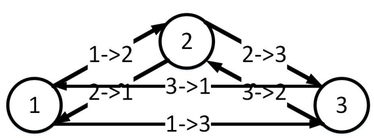


Fig. 1. The metro network

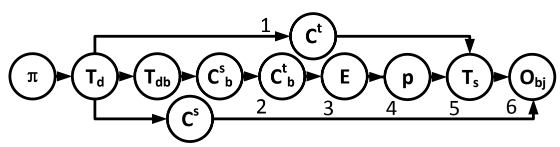


Fig. 2. Flow chart of OSG

3.1 Modeling Defenders

A defender is a team of police officers using trains for patrolling to mitigate crime. We start with a single defender and deal with multiple defenders later. The defender conducts randomized patrols using a *Markov Strategy* π, which specifies for each location a probability distribution over all available actions. At location l, the probabilities of *Go* $i \rightarrow j$ and *Stay* are denoted by $g_l^{i \rightarrow j}$ and s_l respectively.

Example 1: Markov Strategy *In Figure 1, a possible distribution for location* $3 \rightarrow 2$ *in a Markov strategy* π *is,*

$$s_{3\rightarrow 2} = 0.1, \; g_{3\rightarrow 2}^{2\rightarrow 1} = 0.8, \; g_{3\rightarrow 2}^{2\rightarrow 3} = 0.1$$

π	Defender's Markov strategy	$\mathbf{c}_b^s$	Criminal's belief of $\mathbf{c}^s$
T_d	Defender transition matrix	$\mathbf{c}_b^t$	Criminal's belief of $\mathbf{c}^t$
$\mathbf{c}^s$	Defender stationary coverage	T_{db}	Criminal's belief of T_d
$\mathbf{c}^t$	Defender coverage vector at time step t	E	Target expected value for criminals
T_s	Transition matrix for the OSG Markov chain	p	Criminal's next strike probability

Table 1. Notation used throughout this paper.

that is, if the defender is on the train from station 3 to 2, then at the next decision phase: she has probability 0.1 to choose Stay, thereby exiting the train and remaining at station 2; 0.8 to Go $2 \to 1$*, meaning she remains on her current train as it travels to station 1; and 0.1 to Go* $2 \to 3$*, meaning she exits her current train and boards the train heading the opposite direction toward station 3.*

Given π, the defender's movement is a Markov chain over the locations with *defender transition matrix* T_d, whose entry at column k, row l specifies the probability of a defender currently at location k being at location l during the next time step. In T_d, index i ($i \in 1, \dots, N$) represents station i; indexes larger than N represent trains.

Example 2: *For Example 1,* T_d *is as follows:*

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 1\to 2 \\ 1 \to 3 \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & \cdots 2\to 3 & 3 \to 1 & 3 \to 2 \end{array} \\ \left(\begin{array}{cccccc} s_1 & 0 & 0 & s_{3\to 1} & 0 \\ 0 & s_2 & 0 & 0 & s_{3\to 2} \\ 0 & 0 & s_{2\to 3} & 0 & 0 \\ g_1^{1\to 2} & 0 & 0 & g_{3\to 1}^{1\to 2} & 0 \\ g_1^{1\to 3} & 0 & 0 & g_{3\to 1}^{1\to 3} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \right) \end{array}$$

Using T_d and $\mathbf{c}^t = (c_1, c_2, \cdots, c_N, c_{1\to 2}, \cdots)^T$, defined as the probability distribution of a defender's location at time t, we can calculate the coverage vector at time step $t_1 > t$ through the formula

$$\mathbf{c}^{t_1} = (T_d)^{t_1 - t} \cdot \mathbf{c}^t \tag{1}$$

We restrict each element in π to be strictly positive so that T_d is ergodic, meaning it is possible to eventually get from every location to every other location in finite time. For an ergodic T_d, based on Lemma 1, there is a unique *stationary coverage* $\mathbf{c}^s$, such that $T_d \cdot \mathbf{c}^s = \mathbf{c}^s$. The dependence of $\mathbf{c}^s$ on T_d and hence on π is shown in Fig. 2. The defender's initial coverage, $\mathbf{c}^1$, is set to $\mathbf{c}^s$ so that the criminal will face an invariant distribution whenever he enters the system. This invariant initial distribution is analogous to assuming that the defender patrols for a long time and becomes stable, but under our model, criminals can enter the system at any time.

Lemma 1. *(Fundamental Theorem of Markov Chains) For an ergodic Markov chain* P*, there is a unique probability vector* $\mathbf{c}$ *such that* $P \cdot \mathbf{c} = \mathbf{c}$ *and* $\mathbf{c}$ *is strictly positive.*

Proof. This is a very simple restatement of the property of ergodic Markov chain. [25] provides detailed proof. □

3.2 Modeling Opportunistic Criminals

Our model of the criminal consists of three components.

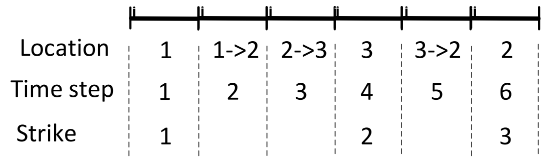


Fig. 3. Example of strikes

Criminal's probability to commit a crime at the current time step: We assume the criminal will only commit crimes at stations, as discussed earlier [9], and only during action phases, since decision phases are considered instantaneous. The probability of such a crime is determined by two factors. The first is the *attractivenes* of each target station [23], which measures the availability of crime opportunities at a station. Attractiveness measures how likely a criminal located at that station during an action phase is to commit a crime *in the absence of defenders*; $\mathbf{Att} = (Att_1, Att_2, \cdots, Att_N)$ is the N vector composed of station attractiveness. The second factor is the defender's presence; i.e., if a criminal is at the same station as a defender, he will not commit a crime. Thus, his probability of committing a crime at station i will be influenced by $c^t(i)$. Using this strategy, the criminal will never be caught red handed by the defender, but may be forced toward a less attractive target. Thus, the probability of the criminal committing a crime if located at station i during the action phase of time step t, is denoted as $q_c(i,t) = (1 - c^t(i))Att(i)$.

Criminal's belief state of the defender: During the decision phase, the criminal decides the next target station; he then moves directly to that station at the next action phase(s). Hence, the criminal's motion within the metro system can be distilled down to a sequence of stations where he chooses to locate; we refer to these instances of attempted crime as *Strikes*. Figure 3 is a toy example showing the relationship between the time steps and strikes for a criminal. As shown in the figure, only the time steps when the criminal is at stations are counted as strikes.

When making these target decisions, the criminal tends to choose stations with high *expected utilities*. He uses his knowledge of π and his real-time observations to make such decisions. Let T_{db}, $\mathbf{c}_b^t$, and $\mathbf{c}_b^s$ be his *belief* of T_d, $\mathbf{c}^t$, and $\mathbf{c}^s$, respectively. As the criminals have limited surveillance capability, these beliefs may not be the same as T_d, $\mathbf{c}^t$, and $\mathbf{c}^s$. To model the criminal's surveillance imperfection we use *anchoring bias* – a cognitive bias, with extensive experimental backing, which reveals the human bias toward choosing a uniform distribution when assigning probabilities to events under imperfect information [18, 5]. We denote the level of the criminal's anchoring bias with the parameter b, where $b = 0$ indicates no anchoring bias, and $b = 1$ indicates complete reliance on such bias. We set $T_{db} = (1 - b) \cdot T_d + b \cdot T_u$, with corresponding stationary coverage $\mathbf{c}_b^s$, where T_u corresponds to the uniform distribution.

At any given time step t when the criminal is at a station, i.e., a strike, he may be modeled as using his belief and observations to estimate $\mathbf{c}_b^t$. We assume the opportunistic criminal only uses his current observation, $\mathbf{c}_b^s$ and T_{db} to estimate $\mathbf{c}_b^t$ (criminal's belief of defender's location distribution). Specifically, if the criminal is at station i and the defender is also there, then $\mathbf{c}_b^t$ is $(0, 0, ..., 1, 0, ..., 0)^T$, where row i is 1 and all others are 0. Otherwise the defender is not at i, and

$$\mathbf{c}_b^t = \frac{(c_b^s(1), c_b^s(2), ..., 0, c_b^s(i+1), ..., c_b^s(N_l))^T}{[1 - c_b^s(i)]}, \tag{2}$$

where row i is 0 and other rows are proportional to the corresponding rows in $\mathbf{c}_b^s$. Our approach to compute $\mathbf{c}_b^t$ is justified on two grounds. First, it is computationally cheap. Second, as we show in experimental results, even perfect knowledge provides very limited improvement in the criminal's performance given our modeling of the criminal's bounded rationality and anchoring bias; thus a more complex procedure is unnecessary. Given $\mathbf{c}_b^t$ and T_{db}, the belief coverage vector at time step t_1 ($t_1 > t$), $\mathbf{c}_b^{t_1}$, is calculated via Eq. 1.

Input: i: the criminal's station; π: defender's Markov strategy; m: the defender's location;
b: parameter of criminal's anchoring bias
Output: $p(\cdot|i, \mathbf{c}_b^{t_0})$: The criminal's probability distribution for next target

Initial N with the number of stations ;
Initial T_d by π;
Initial $\mathbf{c}^s$ with stationary coverage of T_d;
Initial $\mathbf{c}_b^{t_0}$ with a $1 \times (3N-2)$ zero vector ;
$T_{db} = (1-b) \cdot T_d + b \cdot T_u$;
$\mathbf{c}_b^s = (1-b) \cdot \mathbf{c}^s + b \cdot \mathbf{c}_u^s$;
if $i == m$ **then**
    $\mathbf{c}_b^{t_0}(i) = 1;$
end
if $i \neq m$ **then**
    for $j \in Location$ **do**
        $\mathbf{c}_b^{t_0}(j) = \frac{\mathbf{c}_b^s(j)}{1 - \mathbf{c}_b^s(i)}$;
    end
    $\mathbf{c}_b^{t_0}(i) = 0$;
end
for $j \in Station$ **do**
    $t = |i - j| + 1$;
    $\mathbf{c}_b^{t_0+t} = (T_{db})^t \cdot \mathbf{c}_b^{t_0}$;
    $E(j|i, \mathbf{c}_b^{t_0}) = \frac{\left(1 - \mathbf{c}_b^{t_0+t}(j)\right) Att(j)}{t}$;
end
for $j \in Station$ **do**
    $p(j|i, \mathbf{c}_b^{t_0}) = \frac{E(j|i, \mathbf{c}_b^{t_0})^\lambda}{\sum_{h=1}^{N} E(h|i, \mathbf{c}_b^{t_0})^\lambda}$;
end
return $p(\cdot|i, \mathbf{c}_b^{t_0})$;

Algorithm 1: BIASED RANDOM WALK ALGORITHM

We set the actual payoff for a crime to 1, but this can be generalized. The expected payoff for the criminal when choosing station j as the next strike, given that the current strike is at station i at time step t, is $q_{cb}(j, t+\delta_{ij})$, where $\delta_{ij} \geq 1$ is the minimum time needed to arrive at j from i. But, criminals are known to discount more distant locations when choosing targets. Therefore, the *utility* that the criminal places on a given payoff is discounted over time. We implement this by dividing the payoff by the time taken. Finally, the criminal must rely on his belief of the defender's coverage when evaluating $q_{cb}(j, t+\delta_{ij})$. Altogether, station j has the expected utility $E(j|i, \mathbf{c}_b^t) = \frac{q_{cb}(j,t+\delta_{ij})}{\delta_{ij}}$, which is

$$E(j|i, \mathbf{c}_b^t) = \frac{\left(1 - \left[(T_{db})^{\delta_{ij}} \cdot \mathbf{c}_b^t\right](j)\right) Att(j)}{\delta_{ij}}. \tag{3}$$

The criminal's Quantal Biased Random Movement (QBRM): Finally, we propose QBRM to model the criminal's bounded rationality based on other such models of criminal movements in urban domains [23]. Instead of always picking the station with highest expected utility, his movement is modeled as a random process biased toward stations of high expected utility. Given the expected value for each station $\mathbf{E}(\cdot|i, \mathbf{c}_b^t)$, the probability distribution for each being chosen as the next strike, $\mathbf{p}(\cdot|i, \mathbf{c}_b^t)$ is:

$$p(j|i, \mathbf{c}_b^t) = \frac{E(j|i, \mathbf{c}_b^t)^\lambda}{\sum_{h=1}^{N} E(h|i, \mathbf{c}_b^t)^\lambda}, \tag{4}$$

where $\lambda \geq 0$ is a parameter that describes the criminal's level of rationality. This is an instance of the *quantal response* model of boundedly rational behavior [26]. The criminal may, as an alternative to choosing a further strike, leave the system at *exit rate* α. Therefore, the criminal eventually leaves the system with probability 1, and in expectation receives a finite utility; he cannot indefinitely increase his utility.

Given the criminal's QBRM, the Opportunistic Security Game can be simplified to a Stackelberg game for specific value of the parameters describing criminal's behaviour (Theorem 1).

Theorem 1. *When the criminal's rantionality level parameter $\lambda = 0$, the defender's optimal strategy is a stationary strategy, meaning that the defender picks a station and does not move in the patrol.*

Proof. According to Eqn. 4, when $\lambda = 0$, $p(j|i, \mathbf{c}_b^t) = \frac{1}{N}$ for all targets, which is independent of defender's Markov strategy π. Therefore, the OSG is equivalent to a Stackelberg Game where the leader (the criminal) makes his choice first, which is independent of the follower's (defender's) choice. Then the follower can decide her action given the leader's action. Therefore, as in a Stackelberg game, the follower's (defender's) optimal strategy is a pure strategy. Furthermore, we know that in this Stackelberg game, the leader (the criminal) is making a uniform random choice, meaning that he chooses each target with the same probability. Therefore, the defender's optimal strategy is staying at the station with highest attractiveness. □

To summarize, as shown in Figure 2, the opportunistic criminal is modeled as follows: First, he decides whether to commit a crime or not based on the defender's presence at his station at each strike. Next, he uses his imperfect belief T_{db} of the defender's strategy, which is affected by anchoring bias, and his real-time observation to update his belief $\mathbf{c}_b^t$ using a simple scheme (Eq. 2). Finally, we use QBRM to model his next attack (Eq. 4) based on the expected utility of different targets (Eq. 3). Algorithm 1 is a full mathematical description of the criminal's movement. In Algorithm 1, steps 1-4 are initialization; steps 5-6 model how the criminal generates his imperfect belief; steps 7-15 model how the criminal updates his belief given his real-time observation; steps 16-20 model how the criminal evaluates each station based on his updated belief; and steps 21-24 use QBRM to model his probability distribution of visiting each station in his next strike.

4 Exact OSG (EOSG) algorithm

Given the defender and criminal models, the EOSG algorithm computes the optimal defender strategy by modeling the OSG as a finite state Markov chain. As all the criminals behave identically, we can focus on the interaction between the defender and one criminal without loss of generality.

Each state of the EOSG Markov chain is a combination of the criminal's station and the defender's location. Here we only consider situations where the criminal is at a station as states because he only makes decisions at stations. Since there are N stations and N_l locations, the number of states is $N \cdot N_l$ in the EOSG markov chain. State transitions in this EOSG markov chain are based on *strikes* rather than *time steps*. The transition matrix for this Markov chain, denoted as T_s, can be calculated by combining the defender and criminal models. For further analysis, we pick the element $p_{S1\to S2}$ in T_s that represents the transition probability from state $S1$ to $S2$. Suppose in $S1$ the criminal is at station i while the defender is at location m at time step t, and in $S2$, the criminal is at station j while the defender is at location n at time step $t+\delta_{ij}$. We need two steps to calculate the transition probability $p_{S1\to S2}$. First, we find the transition probability of the criminal from i to j, $p(j|i, \mathbf{c}_b^t)$. Then, we find the defender's transition probability from m to n, which is $c^{t+\delta_{ij}}(n) = \left((T_d)^{\delta_{ij}} \cdot \mathbf{e}_m\right)(n)$, where $\mathbf{e}_m$ is a basis vector for the current location m. The transition probability $p_{S1\to S2}$ is therefore given by

$$p_{S1\to S2} = p(j|i, \mathbf{c}_b^t) \cdot c^{t+\delta_{ij}}(n). \tag{5}$$

Since $p(j|i, \mathbf{c}_b^t)$ and $c^{t+\delta_{ij}}(n)$ are determined by π, $p_{S1\to S2}$ is also in terms of π (see Fig. 2), and hence so is T_s.

Given this EOSG model, we can calculate the defender's expected utility at each strike. For each successful crime, the defender receives utility $u_d < 0$, while if there is no crime, she receives utility 0. We do not consider the time discount factor in the defender's expected utility, as the goal of the defender shall be to simply minimize the total expected number of crimes that any criminal will commit. Formally, we define a vector $\mathbf{r}_d \in \mathbf{R}^{N\cdot N_l}$ such that entries representing states with both criminal and defender at the same station are 0 while those representing states with criminal at station i and defender not present are $Att(i) \cdot u_d$. Then, the defender's expected utility $V_d(t)$ during strike number t is $V_d(t) = \mathbf{r}_d \cdot \mathbf{x}_t$, where $\mathbf{x}_t$ is the state distribution at strike number t. $\mathbf{x}_t$ can be calculated from the initial state distribution $\mathbf{x}_1$, via $\mathbf{x}_t = ((1-\alpha)\cdot T_s)^{t-1}\mathbf{x}_1$.

The initial state distribution $\mathbf{x}_1$ can be calculated from the initial criminal distribution and $\mathbf{c}^s$. The defender's total expected utility over all strikes is thus

$$\begin{aligned} Obj &= \lim_{\ell \to \infty} \sum\nolimits_{t=1}^{\ell} V_d(t) \\ &= \mathbf{r}_d \cdot (I - (1-\alpha)T_s)^{-1} \mathbf{x}_1 \ , \end{aligned} \tag{6}$$

where I is an identity matrix. In this equation we use the geometric sum formula and the fact that the largest eigenvalue of T_s is 1, so that $I - (1-\alpha)T_s$ is nonsingular for $0 < \alpha < 1$.

The objective is a function of the transition matrix T_s and $\mathbf{x}_1$, which can be expressed in terms of π via Eqs. (1), (3), (4), and (5). We have thus formulated the defender's problem of finding the optimal Markov strategy to commit to as a nonlinear optimization problem, specifically to choose π to maximize Obj (that is, minimize the total amount of crime).

5 OSG for multiple defenders

If K multiple defenders all patrol the entire metro, using the same π, which is denoted as *full length patrolling*, then they will often be at the same station simultaneously, which carries no benefit. On the other hand if we allow arbitrary defenders' strategies that are correlated, we will need to reason about complex *real-time* communication and coordination among defenders. Instead, we divide the metro into K contiguous segments, and designate one defender per segment, as in typical real-world patrolling of a metro system. Each defender will have a strategy specialized to her segment.

Defenders: In the k-th segment, the number of locations is n_l^k. Defender k patrols with the Markov strategy π_k. Her transition matrix is $T_{dk} \in \mathbf{R}^{n_l^k \times n_l^k}$. Her coverage vector at time t is $\mathbf{c}_k^t$, and $\mathbf{c}_k^s$ is her stationary coverage. Hence, defender k's behavior is the same as that in a single-defender OSG, while the collective defender behavior is described by the Markov strategy $\pi = (\pi_1, \pi_2, ..., \pi_K)$. The transition matrix T_d is as follows, where we have dropped the trains between segments from the basis for T_d and ensured that station numbering is continuous within segments:

$$T_d = \begin{pmatrix} T_{d1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & T_{dK} \end{pmatrix} . \tag{7}$$

The coverage of all units at time step t is $\mathbf{c}^t$, and is defined as the concatenation of coverage vectors $(\mathbf{c}_1^t; \mathbf{c}_2^t; ...; \mathbf{c}_K^t)$. $\mathbf{c}^t$ sums to K since each $\mathbf{c}_k^t$ sums to 1. The vector $\mathbf{c}^t$ evolves to future time steps t_1 in the same way as before, via Eq. 1. The overall stationary coverage is $\mathbf{c}^s = (\mathbf{c}_1^s; \mathbf{c}_2^s; ...; \mathbf{c}_K^s)$.

Opportunistic criminals: The previous model for criminals still applies. However, any variables related to defenders (T_d, $\mathbf{c}^t$, $\mathbf{c}^s$) are replaced by their counterparts for the multiple defenders. Furthermore, the criminal in segment k at time t cannot observe defenders other than k. As a result, his belief of defender coverage is updated only

Input: i: the criminal's station; π: vector of defender Markov strategies; $\mathbf{m}$: vector of defender locations; b: parameter of criminal's anchoring bias

Output: $p(\cdot|i, \mathbf{c}_b^{t_0})$: The criminal's probability distribution for next target

Initial N with the number of stations ;
Initial K with the number of defenders ;
Initial k_i with the segment that station i is in ;
for $k \leq K$ **do**
    Initial T_{dk} by π_k;
    Initial $\mathbf{c}_k^s$ by stationary coverage of T_{dk};
    $T_{dbk} = (1-b) \cdot T_{dk} + b \cdot T_{uk}$;
    $\mathbf{c}_{bk}^s = (1-b) \cdot \mathbf{c}_k^s + b \cdot \mathbf{c}_{uk}^s$;
    $\mathbf{c}_{bk}^{t_0} = \mathbf{c}_{bk}^s$
    if $k == k_i$ **then**
        Initial $\mathbf{c}_{bk}^{t_0}$ with a $1 \times n_l^k$ zero vector ;
        if $i == \mathbf{m}(k)$ **then**
            $\mathbf{c}_{bk}^{t_0}(i) = 1$;
        end
        if $i \neq \mathbf{m}(k)$ **then**
            for $j \in Location\ in\ segment\ k$ **do**
                $\mathbf{c}_{bk}^{t_0}(j) = \dfrac{\mathbf{c}_{bk}^s(j)}{1 - \mathbf{c}_{bk}^s(i)}$;
            end
            $\mathbf{c}_{bk}^{t_0}(i) = 0$;
        end
    end
end
$T_{db} = \begin{pmatrix} T_{db1} & 0 & \dots & 0 \\ 0 & T_{db2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T_{dbK} \end{pmatrix}$
$\mathbf{c}_b^{t_0} = (\mathbf{c}_{b1}^{t_0}; \mathbf{c}_{b2}^{t_0}; ...; \mathbf{c}_{bK}^{t_0})$.
for $j \in Station$ **do**
    $t = |i - j| + 1$;
    $\mathbf{c}_b^{t_0+t} = (T_{db})^t \cdot \mathbf{c}_b^{t_0}$;
    $E(j|i, \mathbf{c}_b^{t_0}) = \frac{\left(1 - \mathbf{c}_b^{t_0+t}(j)\right) Att(j)}{t}$;
end
for $j \in Station$ **do**
    $p(j|i, \mathbf{c}_b^{t_0}) = \frac{E(j|i, \mathbf{c}_b^{t_0})^\lambda}{\sum_{h=1}^N E(h|i, \mathbf{c}_b^{t_0})^\lambda}$;
end
return $p(\cdot|i, \mathbf{c}_b^{t_0})$;

Algorithm 2: BIASED RANDOM WALK ALGORITHM WITH MULITPLE DEFENDERS for segment k, i.e., $\mathbf{c}_b^t = (\mathbf{c}_{b1}^s; \mathbf{c}_{b2}^s; ...; \mathbf{c}_{b(k-1)}^s; \mathbf{c}_{bk}^t; \mathbf{c}_{b(k+1)}^s; ...; \mathbf{c}_{bK}^s)$. Algorithm 2 describes a criminal's behavior in the multiple defenders settings. Similar to Algorithm 1, in Algorithm 2, steps 1-3 are initialization; steps 4-22 model how the criminal generates and updates his imperfect belief for each defender, such that for defender $k(k \leq K)$, the process of calculating the criminal's belief is exactly the same as the single defender scenario; steps 23-24 combine the criminal's belief for each defender as his belief for all the defenders; steps 25-29 model how the criminal evaluates each station based on his belief; and steps 30-34 use QBRM to model his probability distribution of visiting each station in his next strike.

EOSG: In optimizing defender strategies via a Markov chain, each state records the station of the criminal and the location of *each* defender. As a result, each state is denoted as $S = (i, m_1, ..., m_K)$, where the criminal is at station i and defender k is at location m_k. Since defender k can be at n_l^k different locations, the total number of states is $N \cdot n_l^1 \cdots n_l^K$. To apply EOSG for multiple defenders, T_s is still calculated using the defender and criminal models. The transition probability $p_{S_1 \to S_2}$ from $S_1 = (i, m_1, ..., m_K)$ at time t to $S_2 = (j, n_1, ..., n_K)$ at time $t + \delta_{ij}$ is

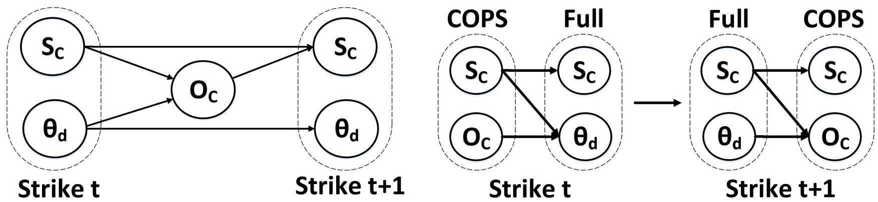
$$p_{S_1 \to S_2} = p(j|i, \mathbf{c}_b^t) \prod_k c^{t+\delta_{ij}}(n_k),$$

where $c^{t+\delta_{ij}}(n_k) = ((T_d)^{\delta_{ij}} \cdot \mathbf{e}_{m_1,m_2,...,m_K})(n_k)$ and $\mathbf{e}_{m_1,m_2,...,m_K}$ is an indicator vector with 1 at entries representing locations $m_1, m_2, ..., m_K$ and 0 at all others. The state distribution $\mathbf{x}$ and revenue $\mathbf{r}_d$ are both $N \cdot n_l^1 \cdots n_l^K$ vectors. The defenders' total expected utility is given by Eq. (6); our goal remains to find a π to maximize Obj.

6 The COPS Algorithm

The objective of EOSG can be formulated as a non-linear optimization. Unfortunately, as we will show in our experiments, the EOSG algorithm fails to scale-up to real-world sized problem instances due to the size of T_s in Eq. (6), which is exponential ($N \cdot n_l^1 \cdots n_l^K$ by $N \cdot n_l^1 \cdots n_l^K$) for K defenders. We propose the Compact OPportunistic Security game state (COPS) algorithm to accelerate the computation. COPS simplifies the model by compactly representing the states. The size of the transition matrix in COPS is $2N \times 2N$, *regardless of the number of defenders*, which is dramatically smaller than in the exact algorithm. The COPS algorithm is inspired by the Boyen-Koller(BK) algorithm for approximate inference on Dynamic Bayesian Networks [27]. COPS improves upon a direct application of BK's factored representation by maintaining strong correlations between locations of players in OSG.

In OSG with a single defender, there are two components in a Markov chain state for strike t: the station of the criminal S_c^t and the location of the defender θ_d^t. These two components are correlated when they evolve. We introduce an intermediate component, the criminal's observation O_c^t, which is determined by both S_c^t and θ_d^t. Given the criminal's current station and his observation, we can compute his distribution over the next strike station. At the same time, the evolution of θ_d^t is independent of S_c^t. Such evolution is shown in Figure 4(a). This is an instance of a Dynamic Bayesian Network: S_c^t, O_c^t, and θ_d^t are the random variables, while edges represent probabilistic dependence.



(a) Full state evolution (b) COPS state evolution

Fig. 4. COPS algorithm

A direct application of the Boyen-Koller algorithm compactly represents the states by using the marginal distribution of these two components, S_c^t and θ_d^t, as approximate states. The marginal distributions of S_c^t and θ_d^t are denoted as $\Pr(S_c^t)$ and $\Pr(\theta_d^t)$ respectively, and it is assumed that these two components are independent, meaning we can restore the Markov Chain states by multiplying these marginal distributions. Note that in Section 4.2, we set $\Pr(\theta_d^t) = \mathbf{c}^s$ for all strikes. Thus, we do not need to store θ_d^t in the state representation. Therefore, the total number of the approximate states in this case is just N. However, such an approximation throws away the strong correlation between the criminal's station and defender unit's location through the criminal's real-time observation. Our preliminary experiments showed that this approximate algorithm leads to low defender expected utility.

To design a better algorithm, we should add more information about the correlation between the criminal and defenders. To that end, our COPS algorithm compactly represents our Markov Chain states with less information lost. Instead of just considering the marginal distributions of each component $\Pr(\theta_d^t)$ and $\Pr(S_c^t)$, we also include the observation of the criminal O_c^t while constructing the approximate states. The criminal's observation is binary: 1 if the defender is at the same station with him, 0 otherwise. The new approximate states, named *COPS states*, only keep the marginal probability distribution of $\Pr(S_c^t, O_c^t)$. So, the new state space is the Cartesian product of the sets of S_c^t and O_c^t, which has size $2N$.

One subtask of COPS is to recover the distributions over the full state space (S_c^t, θ_d^t), given our state representation $\Pr(S_c^t, O_c^t)$. We cannot restore such distribution by multiplying $\Pr(\theta_d^t)$ and $\Pr(S_c^t)$ in COPS. This is because S_c^t, O_c^t, and θ_d^t are not independent. For example, in COPS state $S_c^t = 1$, $O_c^t = 1$, θ_d^t cannot be any value except 1. In other words, the defender's location distribution $\Pr(\theta_d^t | S_c^t, O_c^t)$ is no longer $\mathbf{c^s}$. Instead, we approximate $\Pr(\theta_d^t | S_c^t, O_c^t)$ as follows. In each COPS state (S_c^t, O_c^t), the *estimated marginal distribution* for the defender, $\widehat{\Pr}(\theta_d^t | S_c^t, O_c^t)$, is found in a manner similar to that used to find the criminal's belief distribution $\mathbf{c}_b^t$. Specifically, if $O_c^t = 1$, $\widehat{\Pr}(\theta_d^t | S_c^t, O_c^t) = (0, 0, ..., 1, 0, ..., 0)^T$, where the row representing station S_c^t is 1 and all others are 0; if $O_c^t = 0$, then $\widehat{\Pr}(\theta_d^t | S_c^t, O_c^t)$ is found through Equation 2, but with the $c_b^s(j)$ replaced by the true stationary coverage value $c^s(j)$. We can then recover the estimated distribution over full states $\widehat{\Pr}(S_c^t = i, \theta_d^t | S_c^t = i, O_c^t) = \widehat{\Pr}(\theta_d^t | S_c^t = i, O_c^t)$ for all i and $\widehat{\Pr}(S_c^t = j, \theta_d^t | S_c^t = i, O_c^t) = 0$ for all $j \neq i$. Estimated full distributions evolve the same way as exact distributions do, as described in Section 4. At the future strike, we can then project the evolved estimated full distribution to distributions over COPS states. Figure 4(b) shows the whole process of the evolution of COPS states. However, such a process would appear to involve representing a full T_s, negating the benefit of the factored representation; we avoid that by using T_{COPS}, discussed below.

To streamline the process of evolving COPS states, in practice we use a transition matrix $T_{COPS} \in \mathbb{R}^{2N\times 2N}$. Each element of T_{COPS}, i.e., transition probability $\Pr(S_c^{t'}, O_c^{t'}|S_c^t, O_c^t)$, can be calculated as follows:

$$
\begin{aligned}
&\Pr(S_c^{t'}, O_c^{t'}|S_c^t, O_c^t)\\
&= \sum\nolimits_{\theta_d^{t'}} \sum\nolimits_{\theta_d^t} \Pr(S_c^{t'}, O_c^{t'}|S_c^{t'}, \theta_d^{t'}) \cdot \Pr(S_c^{t'}, \theta_d^{t'}|S_c^t, \theta_d^t) \cdot \widehat{\Pr}(S_c^t, \theta_d^t|S_c^t, O_c^t)\\
&= \Pr(S_c^{t'}|S_c^t, O_c^t) \sum\nolimits_{\theta_d^{t'}} \Pr(O_c^{t'}|S_c^{t'}, \theta_d^{t'}) \cdot \sum\nolimits_{\theta_d^t} \Pr(\theta_d^{t'}|S_c^{t'}, S_c^t, \theta_d^t) \cdot \widehat{\Pr}(\theta_d^t|S_c^t, O_c^t),
\end{aligned} \tag{8}
$$

where $\Pr(S_c^{t'}|S_c^t, O_c^t)$ and $\Pr(\theta_d^{t'}|S_c^{t'}, S_c^t, \theta_d^t)$ correspond to $p(j|i, \mathbf{c}_b^{t_0})$ and $c^{t_0+|i-j|+1}(n)$, respectively, in Section 4.

The defenders' total expected utility in COPS is calculated in a similar way as the exact algorithm, which is

$$
Obj_{COPS} = \mathbf{r}_{d,COPS} \cdot (I - (1-\alpha)T_{COPS})^{-1}\mathbf{x}_{1,COPS}, \tag{9}
$$

where $\mathbf{r}_{d,COPS}$, $\mathbf{x}_{1,COPS}$ are the expected utility vector and the initial distribution for COPS states. Similar to $\mathbf{r}_d$, $\mathbf{r}_{d,COPS}(S)$ is 0 if in state S the defender is at the same station with the criminal, else $\mathbf{r}_{d,COPS}(S) = u_d$. COPS is faster than the exact algorithm because the number of states is reduced dramatically. Meanwhile, the approximation error of COPS algorithm is bounded according to Theorem 2.

Definition 1. *Let m_i be the location corresponding to station i. For a distribution over OSG full states $\mathbf{x}$, the corresponding distribution over COPS states $\mathbf{x}_{COPS}$ is:*

$$
\mathbf{x}_{COPS}(i,o) = \begin{cases} \mathbf{x}(i, m_i) & \textit{if } o = 1 \\ \sum_{m\neq m_i} \mathbf{x}(i,m) & \textit{if } o = 0 \end{cases}
$$

For a distribution over COPS states $\mathbf{x}_{COPS}$, the corresponding approximate distribution over OSG full states $\mathbf{x}'$ is:

$$
\mathbf{x}'(i,m) = \begin{cases} \mathbf{x}_{COPS}(i,1) & \textit{if } m = m_i \\ \mathbf{x}_{COPS}(i,0) \cdot \frac{c^s(m)}{1-c^s(i)} & \textit{otherwise} \end{cases}
$$

This conversion can be summarized through a single matrix multiplication, such that $\mathbf{x}' = A\mathbf{x}$.

Lemma 2. *Let μ_2 be the magnitude of the second largest eigenvalue of transition matrix T_s. Let δ be the largest possible L_2 approximation error introduced when full state distribution $\mathbf{x}$ is transformed into the COPS representation vector $\mathbf{x}_{COPS}$ and back into the approximate distribution $\mathbf{x}'$ over full states: $||\mathbf{x} - A\mathbf{x}|| \leq \delta$. At strike number t, the L_2 norm between the EOSG distribution $\mathbf{y}_t$ and the distribution found through COPS algorithm $\mathbf{x}_t$ is bounded, such that $||\mathbf{y}_t - \mathbf{x}_t||_2 \leq (1-\alpha)^{t-1}\frac{\delta(1-\mu_2^t)}{1-\mu_2}$.*

Proof. Let $\mathbf{x}_t$ be the state vector as found through the COPS algorithm at time t. The time evolution for $\mathbf{x}$ proceeds then as follows: $\mathbf{x}_t = (1-\alpha)^{t-1}(AT_s)^{t-1}\mathbf{x}_1$, where $\mathbf{x}_1 = A\mathbf{y}_1$, and $\mathbf{y}_1$ is the initial state vector for the EOSG algorithm. So, consider the L_2 error introduced at iteration t by the COPS approximation alone

$$||T_s\mathbf{x}_t - AT_s\mathbf{x}_t||_2 = (1-\alpha)^{t-1}||T_s(AT_s)^{t-1}\mathbf{x}_1 - AT_s(AT_s)^{t-1}\mathbf{x}_1||_2.$$

Since the vector $T_s(AT_s)^{t-1}\mathbf{x}_1$ is a full state vector, the error bound here is simply

$$||T_s\mathbf{x}_t - AT_s\mathbf{x}_t||_2 \leq \delta(1-\alpha)^{t-1}. \tag{10}$$

Now, assume that the error between the state vectors $\mathbf{x}_t$ and $\mathbf{y}_t$ at some time t is bound by ϵ: $||\mathbf{y}_t - \mathbf{x}_t||_2 \leq \epsilon$. Since in the EOSG Markov chain it is possible to travel from any state to any other state in a finite amount of time, this Markov chain is ergodic. Let the stationary distribution of T_s be $\mathbf{x^s}$, which is normalized such that $\overrightarrow{1} \cdot \mathbf{x}^s = 1$. $\mu_1 = 1 > \mu_2 \geq ... \geq \mu_{N \cdot N_l}$ are the magnitudes of the eigenvalues of T_s corresponding to eigenvectors $v_1(= \mathbf{x^s}), v_2, ..., v_{N \cdot N_l}$. Since T_s is the transition matrix of an ergodic Markov chain, $\mu_k < 1$ for $k \geq 2$. For eigenvectors $v_k, k \geq 2$, we have $|T_s \cdot v_k| = |\mu_k \cdot v_k|$. Multiplying by $\overrightarrow{1}$ and noting that $\overrightarrow{1} \cdot T_s = \overrightarrow{1}$, we get $|\overrightarrow{1} \cdot v_k| = |\mu_k \cdot \overrightarrow{1} \cdot v_k|$. Since $\mu_k \neq 1$, $\overrightarrow{1} \cdot v_k = 0$.

Write $\mathbf{x}_t$ and $\mathbf{y}_t$ in terms of $v_1, v_2, ..., v_{N \cdot N_l}$ as:

$$\mathbf{y}_t = \beta_1\mathbf{x^s} + \sum_{i=2}^{N \cdot N_l} \beta_i v_i$$
$$\mathbf{x}_t = \beta_1'\mathbf{x^s} + \sum_{i=2}^{N \cdot N_l} \beta_i' v_i$$

Since $\mathbf{y}_t = (1-\alpha)^{t-1}T_s^{t-1}\mathbf{y}_1$, then $\overrightarrow{1} \cdot \mathbf{y}_t = (1-\alpha)^{t-1}$; similarly, $\overrightarrow{1} \cdot \mathbf{x}_t = (1-\alpha)^{t-1}$. Multiplying both equations above by $\overrightarrow{1}$, we get $\beta_1 = \beta_1' = (1-\alpha)^{t-1}$. Therefore,

$$\begin{aligned}
||T_s \cdot \mathbf{y}_t - T_s \cdot \mathbf{x}_t||_2 &\leq ||\sum_{i=2}^{N \cdot N_l}(\beta_i - \beta_i')\mu_i v_i||_2 \\
&\leq |\mu_2|\sqrt{(\beta_2 - \beta_2')^2 + (\beta_3 - \beta_3')^2 + \cdots + (\beta_{N \cdot N_i} - \beta_{N \cdot N_i}')^2} \\
&\leq \mu_2||\mathbf{x}_t - \mathbf{y}_t||_2 \\
&\leq \mu_2\epsilon
\end{aligned}$$

Accordingly, at $t = 1$, we have

$$||\mathbf{y}_1 - \mathbf{x}_1||_2 = ||\mathbf{y}_1 - A\mathbf{y}_1||_2 \leq \delta\ .$$

At $t = 2$, we have

$$\begin{aligned}
||\mathbf{y}_2 - \mathbf{x}_2||_2 = (1-\alpha)||T_s\mathbf{y}_1 - AT_s\mathbf{x}_1|| = (1-\alpha)||T_s\mathbf{y}_1 - AT_s\mathbf{x}_1 + T_s\mathbf{x}_1 - T_s\mathbf{x}_1|| = \\
(1-\alpha)||T_s\mathbf{y}_1 - T_s\mathbf{x}_1 + T_s\mathbf{x}_1 - AT_s\mathbf{x}_1||_2 \leq \\
(1-\alpha)||T_s\mathbf{y}_1 - T_s\mathbf{x}_1||_2 + (1-\alpha)||T_s\mathbf{x}_1 - AT_s\mathbf{x}_1||_2\ .
\end{aligned}$$

From above, the bound for the first term is $\mu_2\delta$, given the error bound at $t = 1$. The bound for the second term is directly given by (10), and is simply δ. Hence

$$||\mathbf{y}_2 - \mathbf{x}_2||_2 \leq (1-\alpha)\delta(\mu_2 + 1) \; .$$

At $t = 3$, we have

$$||\mathbf{y}_3 - \mathbf{x}_3||_2 = (1-\alpha)||T_s\mathbf{y}_2 - AT_s\mathbf{x}_2|| = (1-\alpha)||T_s\mathbf{y}_2 - AT_s\mathbf{x}_2 + T_s\mathbf{x}_2 - T_s\mathbf{x}_2|| =$$
$$(1-\alpha)||T_s\mathbf{y}_2 - T_s\mathbf{x}_2 + T_s\mathbf{x}_2 - AT_s\mathbf{x}_2||_2 \leq$$
$$(1-\alpha)||T_s\mathbf{y}_2 - T_s\mathbf{x}_2||_2 + (1-\alpha)||T_s\mathbf{x}_2 - AT_s\mathbf{x}_2||_2 \; .$$

From above, the bound for the first term is $\mu_2(1-\alpha)\delta(\mu_2+1)$, given the error bound at $t = 2$. The bound for the second term is taken from (10), and is $\delta(1-\alpha)$. Hence

$$||\mathbf{y}_3 - \mathbf{x}_3||_2 \leq (1-\alpha)^2\delta(\mu_2^2 + \mu_2 + 1) \; .$$

By extension, then, the error bound at time step t between EOSG and COPS states is:

$$||\mathbf{y}_t - \mathbf{x}_t||_2 \leq (1-\alpha)^{t-1}\delta \sum_{i=0}^{t-1} \mu_2^i = (1-\alpha)^{t-1}\delta \, \frac{1-\mu_2^t}{1-\mu_2} \; .$$

□

Lemma 3. *For any two vectors v_1, v_2, the relationship between the L_1 distance and L_2 distance is: $||v_1 - v_2||_2 \leq ||v_1 - v_2||_1 \leq \sqrt{n}||v_1 - v_2||_2$, where n is the dimension of the vectors.*

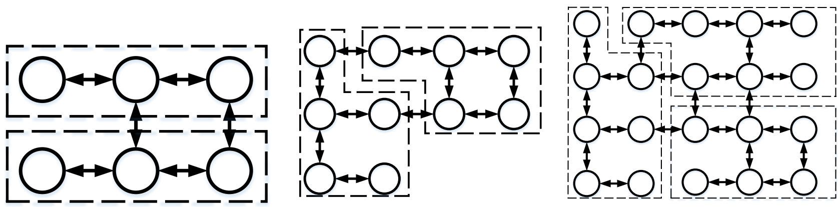
Theorem 2. *The difference between the EOSG objective and the COPS approximate objective $|Obj - Obj_{COPS}|$ is bounded by $\frac{\sqrt{N \cdot N_l}\delta|u_d|}{[1-(1-\alpha)\mu_2] \; \alpha}$*

Proof. Given Lemma 2 and 3, $||\mathbf{y}_t - \mathbf{x}_t||_1 \leq \frac{\sqrt{N \cdot N_l}(1-\alpha)^{t-1}(1-\mu_2^t)\delta}{1-\mu_2}$. Hence we have:

$$\begin{aligned}
|Obj - Obj_{COPS}| &= \sum_{t=1}^{\infty} |\mathbf{r}_d \cdot \mathbf{y}_t - \mathbf{r}_d \cdot \mathbf{x}_t| \\
&= \sum_{t=1}^{\infty} |\mathbf{r}_d \cdot (\mathbf{y}_t - \mathbf{x}_t)| \\
&\leq \sum_{t=1}^{\infty} |r_{max}|||\mathbf{y}_t - \mathbf{x}_t||_1 \\
&\leq |r_{max}| \sum_{t=1}^{\infty} \frac{\sqrt{N \cdot N_l}(1-\alpha)^{t-1}(1-\mu_2^t)\delta}{1-\mu_2} \\
&= |r_{max}| \frac{\sqrt{N \cdot N_l} \; \delta}{[1-(1-\alpha)\mu_2] \; \alpha}
\end{aligned}$$

where r_{max} is the element in $\mathbf{r}_d$ with largest magnitude, which is $min(Att(i) \cdot u_d)$ because $\mathbf{r}_d$ is a non-positive vector by definition. Given $Att(i) \leq 1$, we have $|Obj - Obj_{COPS}| \leq \frac{\sqrt{N \cdot N_l}\delta|u_d|}{[1-(1-\alpha)\mu_2] \; \alpha}$ □

7 Experimental Results



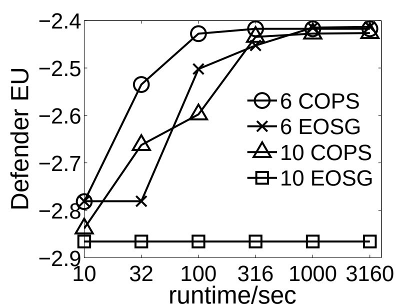
(a) 6 stations (b) 10 stations (c) 20 stations

Fig. 5. Part of metro systems in mega cities

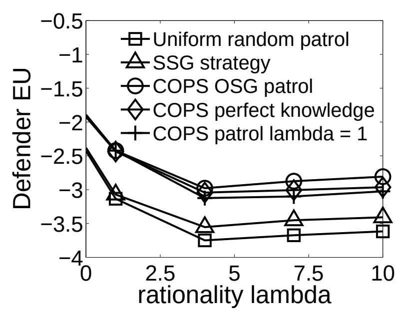
Settings: We use the graphs in Figure 5 – metro structures commonly observed in the world's mega cities – in our experiments. We also tested our algorithm on line structure systems, and the results are similar (online appendix: http://osgcops.webs.com/). We solve the non-linear optimization in OSG using the `FindMaximum` function in Mathematica, which computes a locally optimal solution using an Interior Point algorithm. Each data point we report is *an average of 30 different instances*, each based on a different attractiveness setting; these instances were generated through a uniform random distribution from 0 to 1 for the attractiveness of each station. For multiple patrol unit scenarios, we use segment patrolling (except for Fig. 6(d)), and divide the graph so that the longest distances in each segments are minimized; the dashed boxes in Fig. 5 show the segments used. Results for other segmentations are similar (online appendix). The defender's utility of a successful crime is $u_d = -1$. The criminal's initial distribution is set to a uniform distribution over stations. The criminal exit rate is $\alpha = 0.1$. Strategies generated by all algorithms are evaluated using Equation 6. All key results are *statistically significant* ($p < 0.01$).

Results: Fig. 6(a) shows the performance of the COPS algorithm and the EOSG algorithm using the settings from Fig. 5(a) and Fig. 5(b). In both, we set $\lambda = 1$. The Interior Point algorithm used by Mathematica is a locally optimal solver and there is always a current best feasible solution available, although the quality of the solution keeps improving through iterations. Therefore, one practical way to compare solutions is to check the solution quality after a fixed run-time. The x-axis in this figure shows runtime in seconds on a log scale, while the y-axis maps the defenders' average expected utility against one criminal, achieved by the currently-best solution at a given run time. Focusing first on results of 6 stations, where we have one defender, COPS outperforms EOSG for any runtime within 100 s, even though COPS is an approximate algorithm. This is because COPS reaches a local optimum faster than EOSG. Further, even for runtime long enough for EOSG to reach its local optimum (3160 s), where it outperforms COPS, the difference in solution quality is less than 1%. Focusing next on results of 10 stations with 2 defenders (using segment patrolling), the conclusions are similar to 6 stations, but the advantage of COPS is more obvious in this larger scale problem. In most instances, COPS reaches a local optimum in 1000 s while the output of EOSG are the same as initial values in 3160 s.

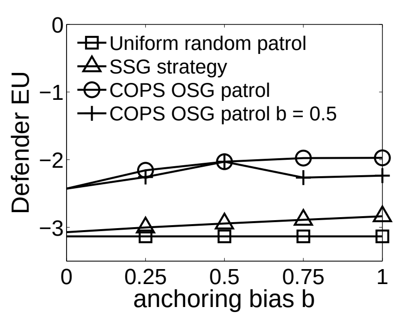
Figure 6(b) employs criminals with varying levels of rationality to compare the performance of three different strategies: the uniform random strategy, which is a Markov strategy with equal probability for all available actions at each location; an SSG strategy, which is the optimal strategy against a strategic attacker that attacks a single target; and a COPS OSG strategy (given 1800 s so it reached a local optimum). In Fig. 6(b), we set $b = 0$; results with other b are similar, which are shown in online appendix. The system consists of 10 stations and 2 defenders. The COPS OSG strategy outperforms the random and SSG strategies significantly for any λ. Next, two more settings are tested: the first is the OSG strategy against criminals who have perfect knowledge of defenders' current location. This is a purely hypothetical setting, and created only to check if a more complex criminal belief model than the one in Eq. 2 would have led to significantly different defender performance. The degradation in performance against perfect criminals is less than 6%, indicating that a more complex belief update for defenders' current location would have insignificant impact on the results. The second is also an OSG strategy, but the defenders set a fixed λ during computation to test performance when the defender has an inaccurate estimate of λ. We picked $\lambda = 1$ from a set of sampled λ, since the OSG strategy with $\lambda = 1$ performs best against criminals with various levels of rationality. Even though the OSG strategy assuming $\lambda = 1$ performs slightly worse than that using the correct λ, it is still better than SSG and uniform strategies. We conclude that OSG is a better model against opportunistic criminals even with an inaccurate estimation of λ.



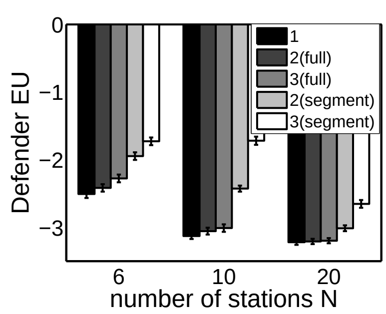
(a) varying runtime



(b) varying λ



(c) varying b



(d) varying N

Fig. 6. Experimental Results

The COPS strategy, the SSG, and the uniform random strategy are compared again in Fig. 6(c), this time against criminals with different levels of anchoring bias b. In order to evaluate the performance of COPS when the defender has an inaccurate estimate of the anchoring bias b, we plotted both the expected utility of COPS where the defender has an accurate estimate of the criminal's anchoring bias and that using a fixed anchoring bias $b = 0.5$. $b = 0.5$ was picked from a set of sampled b since the OSG strategy with this b performs best. In Fig. 6(c), λ is fixed to 1, but experiments with other λ generate similar results, which are shown in the online appendix. Again, COPS outperforms uniform random and SSG strategies with accurate and an inaccurate fixed b. Thus, OSG generates a better strategy even with an imprecise assumption of the criminal's anchoring bias.

To show COPS's scalability, we compare its performance with different numbers of defenders in metro systems with a varying number of stations; 20 stations is very comparable to the number of stations patrolled by police units in major metros, such as the Los Angeles Metro system. Five different settings are compared in Fig. 6(d): one defender, two defenders with full length patrolling, three defenders with full length patrolling, two defenders with segment patrolling, and three defenders with segment patrolling. The max runtime is 1800 s. With the same patrol techniques, more defenders provide higher expected utility. But, with the same amount of resources, segment patrolling outperforms full length patrolling.

8 Summary

This paper introduces OSG, a new computational framework to address opportunistic crime, opening the door for further research on this topic. Furthermore, we propose a new exact algorithm, EOSG, to compute defender resource allocation strategies, and an approximate algorithm, COPS, to speed up defender allocation to real-world scale scenarios. Our experimental results show that the OSG strategy outperforms baseline strategies with different types of criminals. We also show that COPS is more efficient than EOSG in solving real-world scale problems. Given our experimental results, COPS is being evaluated in the Los Angeles Metro system. Finally, in introducing OSG, this paper has added to the class of important security-focused game-theoretic frameworks in the literature, opening the door to a new set of research challenges for the community of researchers focused on game theory for security.

References

1. Letchford, J., MacDermed, L., Conitzer, V., Parr, R., Isbell, C.L.: Computing optimal strategies to commit to in stochastic games. In: AAAI. (2012)
2. Tambe, M.: Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press (2011)
3. Brantingham, P.J., Tita, G.: Offender mobility and crime pattern formation from first principles. Artificial crime analysis systems (2008)
4. Ratcliffe, J.H.: A temporal constraint theory to explain opportunity-based spatial offending patterns. Journal of Research in Crime and Delinquency (2006)

5. See, K.E., Fox, C.R., Rottenstreich, Y.S.: Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. Journal of Experimental Psychology (2006)
6. Hespanha, J.P., Prandini, M., Sastry, S.: Probabilistic pursuit-evasion games: A one-step nash approach. (2000)
7. Brantingham, P., Brantingham, P.: Criminality of place. European Journal on Criminal Policy and Research **3**(3) (1995) 5–26
8. Cornish, D.B., Clarke, R.V.: Understanding crime displacement: An application of rational choice theory. Criminology (1987)
9. Loukaitou-Sideris, A., Liggett, R., Iseki, H.: The geography of transit crime documentation and evaluation of crime incidence on and around the green line stations in los angeles. JPER (2002)
10. Qian, Y., Jiang, X., Haskell, W., Tambe, M.: Online planning for optimal protector strategies in resource conservation games, AAMAS (2014)
11. Zhu, Q., Clark, A., Poovendran, R., Basar, T.: Deceptive routing games. In: Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, IEEE (2012) 2704–2711
12. Clark, A., Zhu, Q., Poovendran, R., Başar, T.: Deceptive routing in relay networks. In: Decision and Game Theory for Security. Springer (2012) 171–185
13. Basilico, N., Gatti, N., Amigoni, F.: Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In: AAMAS. (2009)
14. Basilico, N., Gatti, N.: Automated abstractions for patrolling security games. In: AAAI. (2011)
15. Basilico, N., Gatti, N., Rossi, T., Ceppi, S., Amigoni, F.: Extending algorithms for mobile robot patrolling in the presence of adversaries to more realistic settings. In: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 02. (2009)
16. Jiang, A.X., Yin, Z., Zhang, C., Tambe, M., Kraus, S.: Game-theoretic randomization for security patrolling with dynamic execution uncertainty. In: AAMAS. (2013)
17. Varakantham, P., Lau, H.C., Yuan, Z.: Scalable randomized patrolling for securing rapid transit networks. In: IAAI. (2013)
18. Pita, J., Jain, M., Ordóñez, F., Tambe, M., Kraus, S., Magori-Cohen, R.: Effective solutions for real-world stackelberg games: When agents must deal with human uncertainties. In: AAMAS. (2009)
19. Vorobeychik, Y., An, B., Tambe, M.: Adversarial patrolling games. In: In AAAI Spring Symposium on Security, Sustainability, and Health. (2012)
20. Manshaei, M.H., Zhu, Q., Alpcan, T., Bacşar, T., Hubaux, J.P.: Game theory meets network security and privacy. ACM Computing Surveys (CSUR) (2013)
21. Zhu, Q., Basar, T.: Dynamic policy-based ids configuration. In: CDC/CCC 2009. Proceedings of the 48th IEEE Conference on, IEEE (2009)
22. Nguyen, K.C., Alpcan, T., Basar, T.: Stochastic games for security in networks with interdependent nodes. In: GameNets' 09, IEEE (2009)
23. Short, M.B., D'Orsogna, M.R., Pasour, V.B., Tita, G.E., Brantingham, P.J., Bertozzi, A.L., Chayes, L.B.: A statistical model of criminal behavior. Mathematical Models and Methods in Applied Sciences **18**(supp01) (2008) 1249–1267
24. Zipkin, J.R., Short, M.B., Bertozzi, A.L.: Cops on the dots in a mathematical model of urban crime and police response. (2013)
25. Plavnick, A.: The fundamental theorem of markov chains, University of Chicago VIGRE REU (2008)
26. McKelvey, R.D., Palfrey, T.R.: Quantal Response Equilibria for Normal Form Games. Games and Economic Behavior **10**(1) (1995) 6–38
27. Boyen, X., Koller, D.: Tractable inference for complex stochastic processes. In: UAI, Morgan Kaufmann Publishers Inc. (1998) 33–42