Optimal Allocation of Police Patrol Resources Using a Continuous-Time Crime Model

Ayan Mukhopadhyay[1], Chao Zhang[2], Yevgeniy Vorobeychik[1], Milind Tambe[2], Kenneth Pence[1], and Paul Speer[1]

[1] Vanderbilt University
2201 West End Ave, Nashville, TN 37235
[2] University of Southern California,
Los Angeles, California 90089-0894

Abstract. *Police departments worldwide are eager to develop better patrolling methods to manage the complex and evolving crime landscape. Surprisingly, the problem of spatial police patrol allocation to optimize expected crime response time has not been systematically addressed in prior research. We develop a bi-level optimization framework to address this problem. Our framework includes novel linear programming patrol response formulations. Bender's decomposition is then utilized to solve the underlying optimization problem. A key challenge we encounter is that criminals may respond to police patrols, thereby shifting the distribution of crime in space and time. To address this, we develop a novel iterative Bender's decomposition approach. Our validation involves a novel spatio-temporal continuous-time model of crime based on survival analysis, which we learn using real crime and police patrol data for Nashville, TN. We demonstrate that our model is more accurate, and much faster, than state-of-the-art alternatives. Using this model in the bi-level optimization framework, we demonstrate that our decision theoretic approach outperforms alternatives, including actual police patrol policies.*

Keywords: Decision theoretic policing, crime modeling, survival analysis, Bender's decomposition

1 Introduction

Prevention, response and investigation are the three major engagements of police. Ability to forecast and then effectively respond to crime is, therefore, the holy grail of policing. In order to ensure that crime incidents are effectively handled, it is imperative that police be placed in a manner that facilitates quick response. Effective police placement, however, needs crime prediction as a prerequisite. This is one of the reasons why predicting crime accurately is of utmost importance. While a number of techniques have been proposed for characterizing and forecasting crime, optimizing response times has not been addressed so far, to the best of our knowledge.

Our goal is to develop a rigorous optimization-based approach for optimal police placement in space in order to minimize expected time to respond to crime incidents as they occur. For the time being, we assume that a generative model for crime is available; we describe such a model, calibrated on real crime and police patrol data, in Section 4. The key challenge we face is that crime locations and timing are uncertain. Moreover, for a given placement of police resources in space, optimizing crime incident response for a collection of known incidents is itself a non-trivial optimization problem. What makes this problem particularly challenging is that criminals are affected by police, as they avoid committing crimes if the chances of being caught are high; consequently, we expect that police placement will impact spatial and temporal distribution of crime incidents. Our model, therefore, has both decision and game theoretic features, even though we make use of a data-driven generative model of crime that accounts for the impact of police locations, rather than relying on rationality as underpinning criminal behavior.

Formally, we frame the problem of police patrol optimization as a regularized two-stage stochastic program. We show how the second-stage program (computing optimal response to a fixed set of crime incidents) can be formulated as a linear program, and develop a Bender's decomposition method with sample average approximation for the overall stochastic program. To address the fact that the top-level optimization decisions actually influence the probability distribution over scenarios for the second-level crime response optimization problem, we propose a novel iterative stochastic programming algorithm, *IBRO*, to compute approximate solutions to the resulting bi-level problem of finding optimal spatial locations for police patrols that minimize expected response time. We show that our model outperforms alternative policies, including the response policy in actual use by a US metropolitan police department, both in simulation and on actual crime data.

In order to validate our model of police response, we develop a novel crime forecasting model that is calibrated and evaluated using real crime and police patrol data in Nashville, TN. Crime prediction has been extensively studied, and several models for it have been proposed. These include visualization tools, primarily focused on *hotspots*, or areas of high crime incidence [2], spatial cluster analysis tools [17, 15], risk-terrain models [10], leading indicator models [4], and dynamic spatial and temporal models [9, 19, 23]. A major shortcoming of the existing methods is that they do not allow principled data-driven continuous-time spatial-temporal forecasting *that includes arbitrary crime risk factors*. For example, while risk-terrain modeling focuses on spatial covariates of crime, it entirely ignores temporal factors, and does not offer methods to learn a generative model of crime from data. The work by Short et al. [19] on dynamic spatial-temporal crime modeling, on the other hand, does not readily allow inclusion of important covariates of crime, such as locations of pawn shops and liquor stores, weather, or seasonal variations. Including such factors in a spatial-temporal model, however, is critical to successful crime forecasting: for example, these may inform important policy decisions about zoning and hours of operation for liquor stores, and will make the tool more robust to environmental changes that affect such variables. To address these concerns, validate our model, and forecast crimes, we propose a stochastic generative model of crime which is continuous in time and discretized in space, and readily incorporates crime covariates, bridging an important gap in prior art. Our model leverages survival analysis to learn a probability density over time for predicting crime. After creating a model to predict crime, we evaluate its performance by comparing it with a natural adaptation of the Dynamic Spatial Disaggregation Approach (DSDA) algorithm [9] and an Dynamic Bayes Network method [23] using automated abstraction [22].

1.1 Related Work

There has been an extensive literature devoted to understanding and predicting crime incidence, involving both qualitative and quantitative approaches. For example, a number of studies investigate the relationship between liquor outlets and crime [20, 21]. Many of the earlier quantitative models of crime focus on capturing spatial crime correlation (hot spots), and make use of a number of statistical methods towards this end [17, 15]; these are still the most commonly used methods in practice. An alternative approach, risk-terraine modeling, focuses on quantifiable environmental factors as determinants of spatial crime incidence, rather than looking at crime correlation [10]. These two classes of models both have a key limitation: they ignore the temporal dynamics of crime. Moreover, environmental risk factors and spatial crime analysis are likely complementary. Our approach aims to merge these ideas in a principled way.

Recently, a number of sophisticated modeling approaches emerged aiming to tackle the full spatio-temporal complexity of crime dynamics. One of these is based on a spatio-temporal differential equation model that captures both spatial and temporal crime correlation [18, 16]. These models have two disadvantages compared to ours: first, they do not naturally capture crime co-variates, and second, they are non-trivial to learn from data [16], as well as to use in making predictions [18]. Another model in this general paradigm is Dynamic Spatial Disaggregation Approach (DSDA) [9], which combines an autoregressive model to capture temporal crime patterns with spatial clustering techniques to model spatial correlations. The model we propose is significantly more flexible, and combines spatial and temporal predictions in a principled way by using well-understood survival analysis methods. Recently, an approach has been proposed for modeling spatial and temporal crime dynamics using Dynamic Bayes Networks [23, 22]. This approach necessitates discretization of time, as well as space. Moreover, despite significant recent advances, scalability of this framework remains a challenge.

2 Optimizing Police Placement

Our goal is to address a fundamental decision theoretic question faced by police: how to allocate limited police patrols so as to minimize expected response time to occurring crime. In reality, this is a high-dimensional dynamic optimization problem under uncertainty. In order to make this tractable in support of practical decision making, we consider a simplified two-stage model: in the first stage, police determines spatial location of a set of patrol vehicles, P, and in the second stage, vehicles respond to crime incidents which occur. The decisions in the first stage are made under uncertainty about actual crime incidents, whereas for second-stage response decisions, we assume that this uncertainty is resolved. *A key strategic consideration in police placement is its impact on crime incidence.* In particular, it is well known that police presence has some deterrence effect on crime, which in spatio-temporal domains takes two forms: reduced overall crime frequency, and spatial crime shift [12, 19]. We assume below that the effect of police presence on crime distribution is captured in a stochastic crime model. Later, we describe and develop the stochastic crime model where we use real crime and police patrol data.

We present the problem formulation of allocating police given a stochastic generative model of crime. We divide the available area under police patrol into discrete grids. Formally, we define q as the vector of police patrol decisions, where q_i is the number of police vehicles place in grid i. Let s be a random variable corresponding to a batch of crime incidents occurring prior to the second stage. The two-stage optimization problem for police placement then has the following form:

$$\min_{q} \mathbb{E}_{s \sim f}[D(q; s)], \tag{1}$$

where $D(q; s)$ is the minimal total response time of police located according to q to crime incidents in realization s, which is distributed according to our crime distribution model f described in Section 4, associated with each grid (and the corresponding spatial variables). The model implicitly assumes that crime occurrence is distributed i.i.d. for each grid cell, conditional on the feature vector, where the said feature vector captures the inter-dependence among grids. While the crime prediction model is continuous in time, we can fix a second-stage horizon to represent a single *time zone* (4-hour interval), and simply consider the distribution of the crime incidents in this interval.

The optimization problem in Equation (1) involves three major challenges. First, even for a given s, one needs to solve a non-trivial optimization problem of choosing which subset of vehicles to send in response to a collection of spatially dispersed crime incidents. Second, partly as a consequence of the first, computing the expectation exactly is intractable. Third, the probability distribution of future crime incidents, f, depends on police patrol locations q through the features that capture deterrence effects as well as spatial crime shift to avoid police. We address these problems in the following subsections.

2.1 Minimizing Response Time for a Fixed Set of Crime Incidents

While our goal is to minimize total response time (where the total is over the crime incidents), the information we have is only about spatial locations of crime and police in discretized space. As a result, we propose using distance traveled as a proxy. Specifically, if a police vehicle located at grid i is chosen to respond to an incident at grid j, the distance traveled is d_{ij}, distance between grids i and

j. Assume that these distances d_{ij} are given for all pairs of grids i, j. Next, we assume that a single police vehicle is sufficient to respond to all crime incidents in a particular grid j. This is a reasonable assumption, since the number of crime incidents in a given cell over a 4-hour interval tends to be relatively small, and this interval is typically sufficient time to respond to all of them.

Given this set up, we now show how to formulate this response distance minimization problem as a linear integer program by mapping it to two classical optimization problems: the transportation [1] and *k-server* problems [3].

In the transportation problem, there are m suppliers, each with supply s_i, n consumers, each with demand r_j, and transportation cost c_{ij} between supplier i and consumer j. The goal is to transport goods between suppliers and consumers to minimize total costs. To map crime response to transportation, let police vehicles be suppliers, crime incidents be consumers, and let transportation costs correspond to distances d_{ij} between police vehicle and crime incident grids, with each grid being treated as a node in the network. While the transportation problem offers an effective means to compute police response, it requires that the problem is balanced: supply must equal demand. If supply exceeds demand, a simple modification is to add a dummy sink node. However, if demand exceeds supply, the problem amounts to the multiple traveling salesman problem, and needs a different approach.

To address excess-demand settings, we convert the police response to a more general k-server problem. The k-server problem setting involves k servers in space and a sequence of m requests. In order to serve a request, a server must move from its location to the location of the request. The k-server problem can be reduced to the problem of finding minimum cost flow of maximum quantity in an acyclic network [3]. Let the servers be $s_1, ..., s_k$ and the requests be $r_1, ..., r_m$. A network containing $(2 + k + 2m)$ nodes is constructed. In the formulation described in [3], each arc in the network has capacity one. The arc capacities are modified in our setting, as described later in the problem formulation. The total vertex set is $\{a, s_1, ..., s_k, r_1, ..., r_m, r_1^{'}, ..., r_m^{'}, t\}$. a and t are source and sink respectively. There is an arc of cost 0 from a to each of s_i. From each s_i, there is an arc of cost d_{ij} to each r_j, where d_{ij} is the actual distance between locations i and j. Also, there is an arc of cost 0 from each s_i to t. From each r_i, there is an arc of cost $-K$ to each $r_i^{'}$, where K is an extremely large real number. Furthermore, from each $r_i^{'}$, there is an arc of cost d_{ij} to each r_j where $i < j$ in the given sequence. In our setting, servers and requests correspond to grids with police and crime respectively. In the problem setting we describe, G is the set of all the nodes in the network. We term the set $\{s_i \ \forall i \in G\}$ as G^1, the set $\{r_i \ \forall i \in G\}$ as G^2 and the set $\{r_i^{'} \ \forall i \in G\}$ as G^3. The structure of the network is shown in Fig. 1, which shows how the problem can be framed for a setting with 6 discrete locations. Shaded nodes represent the presence of police and crime in their respective layers.

The problem of finding placement of k-servers in space to serve an unordered set of requests is the same as the multiple traveling salesperson problem (mTSP), a generalization of the TSP problem, which is NP-hard. The offline k-server problem gets around this by having a pre-defined sequence of requests. By sampling crimes from the spatio-temporal model, although we can create a sequence of crimes by ordering them according to their times of occurrence, this sequence need not necessarily provide the least time to respond to all the crimes. In order to deal with this problem, we leverage the fact that crimes are relatively rare events. In order to find the ordering of crimes that provides the least response time, we solve the problem for each possible ordering of crimes. Despite this, the k-server solution approach is significantly less scalable than the transportation formulation. Consequently, we make use of it only in the (rare) instances when crime incidents exceed the number of available police.

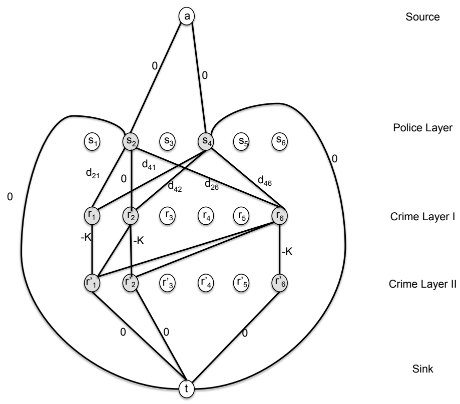


Fig. 1: Network Structure

2.2 Stochastic Programming and Sample Average Approximation for Police Placement

Now that we have two ways of addressing the problem of minimizing response time given a *known* set of crime incidents, we consider the original problem of optimizing allocation of police patrols. As a first step, we point out that the resulting stochastic program is intractable in our setting because of the large space of possible crime incident realizations. We therefore make use of sample average approximation, whereby we estimate the expectation using a collection of i.i.d. crime incident realization samples (henceforth, scenarios) generated according to f. For each scenario, we represent the presence of crimes in the grids by a binary vector z and total available police by k. The decision variable, x_{ij}^s refers to the number of police vehicles traveling from grid i to grid j in scenario s. Under such a setting, the optimization program with transportation problem in the second level can be formulated as:

$$\min_{q} \quad \sum_{s \in S} \left[\min_{x^s \geq 0} \sum_{ij} d_{ij} x^s_{ij} \right] \tag{2a}$$

$$\text{s.t. :} \quad q_i \in \mathbb{Z}_+ \ \forall i \in G$$

$$\sum_{i \in G} q_i = k \tag{2b}$$

$$\sum_{j \in G} x^s_{ij} = q_i, \ \ \forall i \in G, \forall s \in S \tag{2c}$$

$$\sum_{i \in G} x^s_{ij} = z^s_j, \ \ \forall j \in G, \forall s \in S, \tag{2d}$$

$$x^s_{ij} \geq 0 \ \forall i, j \in G \tag{2e}$$

The optimization program leveraging the k-server problem, on the other hand, can be formulated as:

$$\min_{q} \quad \sum_{s \in S} \left[\min_{x^s \geq 0} \sum_{ij} d_{ij} x^s_{ij} \right] \tag{3a}$$

$$\tag{3b}$$

$$\text{s.t. :} \quad q_i \in \mathbb{Z}_+ \ \forall i \in G$$

$$\sum_{j \in \{G^2, t\}} x^s_{ij} = q_i \quad \forall i \in G^1, \forall s \in S \tag{3c}$$

$$\sum_{i \in G} x^s_{ij} = z^s_j \quad \forall j \in G^2, \forall s \in S \tag{3d}$$

$$\sum_{j \in G} x^s_{ij} - \sum_{l \in G} x^s_{li} \ = \ s_i \quad \forall i \in G, \forall s \in S \text{ where } s_i = \begin{cases} k \text{ if } i = a \\ -k \text{ if } i = t \\ 0 \text{ otherwise} \end{cases} \tag{3e}$$

$$x^s_{ij} \leq 1 \quad \forall i, j \in \{i, j \in G\} \setminus \{\{i, j \in G \text{ and } i = a \text{ and } j \in G^1\} \cup \{i, j \in G \text{ and } i \in G^1 \text{ and } j = t\}\}, \forall s \in S \tag{3f}$$

$$x^s_{ij} \geq 0 \ \forall i, j \in G, \forall s \in S \tag{3g}$$

The overall optimization problem then becomes

$$\min_{q \geq 0} \mathbb{E}_{s \sim f} \left[\mathbb{1}(k \geq m_s) \min_{x^s \in C_1^{s(q)}} \sum_{ij} d_{ij} x^s_{ij} + \mathbb{1}(k < m_s) \min_{x^s \in C_2^{s(q)}} \sum_{ij} d_{ij} x^s_{ij} \right] \tag{4}$$

where $C_1^{s(q)}$ includes the Constraints 2c, 2d, and $C_2^{s(q)}$ includes Constraints 3c 3d and 3e, as well as the capacity constraints, for all realizations of crime incidents s, that are drawn from the distribution f.

We propose to solve this stochastic program using Bender's decomposition [1]. The first step is to represent the inner (lower-level) optimization problems using their duals, which for the transportation problem, is represented as:

$$\max_{\alpha,\beta} \sum_{i\in G} q_i\alpha_i^s + \sum_{j\in G} z_j^s\beta_j^s \tag{5a}$$

$$s.t.: \quad d_{ij} - \alpha_i^s - \beta_j^s \geq 0 \quad \forall i,j \in G, \tag{5b}$$

where $\{\alpha_1^s, ..., \alpha_g^s\}$ are the dual variables for Constraints 2c and $\beta_1^s, ..., \beta_g^1$ are dual variables for Constraints 2d. The dual for the k-server problem is represented as:

$$\max_{\lambda,\delta,f,c} - \sum_{i\in G^1} \lambda_i^s q_i - \sum_{j\in G^2} \delta_j^s z_j - \sum_{i,j\in C_c} c_{ij}^s - \sum_{i\in G} f_i^s s_i \tag{6a}$$

$$s.t. \tag{6b}$$

$$\mathbb{1}(i,j\in C_\lambda)\lambda_i^s + \mathbb{1}(i,j\in C_\delta)\delta_j^s + f_i^s - f_j^s + \mathbb{1}(i,j\in C_c)c_{ij}^s + d_{ij} \geq 0 \quad \forall i,j\in G \tag{6c}$$

where

$$i,j\in C_\lambda \quad \text{if} \quad i,j\in G \quad \text{and} \quad i\in G^1, j\in \{G^2,t\}$$
$$i,j\in C_\delta \quad \text{if} \quad i,j\in G \quad \text{and} \quad i\in G^2$$
$$i,j\in C_c \quad \text{if} \quad i,j\in \{i,j\in G\}\setminus\{\{i,j\in G \text{ and } i=a \text{ and } j\in G^1\}$$
$$\cup\{i,j\in G \text{ and } i\in G^1 \text{ and } j=t\}\}$$

We introduce dual variables $\lambda_i^s, ..., \lambda_k^s$ for constraints 3c, $\delta_i^s, ..., \delta_m^s$ for constraints 3d, $f_i^s, ..., f_n^s$ for constraints 3e and $c_{11}^s, c_{12}^s..., c_{nn}^s$ for constraints 3f.

By construction, the primal transportation problem always has a feasible solution as it is balanced, and the primal k-server problem always has a feasible solution provided $\sum_i q_i > 0$, which is ensured by always having a budget greater than 0. Consequently, there always exists an optimal dual solution which is one of the (finite number of) extreme points of the polyhedron comprised from Constraints 5b and 6c for the corresponding problems. Since these constraints do not depend on the police patrol allocation decisions q, the set of extreme points of the constraint polyhedra $E^s = \{(\lambda^s, \delta^s, f^s, c^s)\}$ and $E^s = \{\alpha^s, \beta^s\}$ for both the problems are independent of q. Thus, we can then rewrite the stochastic program as

$$\min_q \quad \sum_{s\in S}\Bigg[\mathbb{1}(k<m_s)\{\max_{(\lambda^s,\delta^s,f^s,c^s)\in E^s} - \sum_{i\in G^1}\lambda_i^s q_i - \sum_{j\in G^2}\delta_j^s z_j - \sum_{i,j\in C_c} c_{ij}^s - \sum_{i\in G} f_i^s s_i\} + \mathbb{1}(k\geq m_s)\{\max_{\alpha,\beta}\sum_{i\in G} q_i\alpha_i^s + \sum_{j\in G} z_j^s\beta_j^s\}\Bigg] \tag{7}$$

Since E^s is finite, we can rewrite it as

$$\min_{q,u^s} \sum_s u^s \tag{8a}$$

$$\text{s.t.}: \quad q_i \in \mathbb{Z}_+ \quad \forall\, i \in G$$

$$u^s \geq -\sum_{i \in G^1} \lambda_i^s q_i - \sum_{j \in G^2} \delta_j^s z_j - \sum_{i,j \in C_c} c_{ij}^s - \sum_{i \in G} f_i^s s_i \quad \forall\, s, (\lambda^s, \delta^s, f^s, c^s) \in \tilde{E}^s \tag{8b}$$

$$u^s \geq \sum_{i \in G} q_i \alpha_i^s + \sum_{j \in G} z_j^s \beta_j^s \quad \forall\, s, (\alpha^s, \beta^s) \in \tilde{E}^s \tag{8c}$$

where $\tilde{E}^s$ is a subset of the extreme points which includes the optimal dual solution and Constraints 8b and 8c are applicable based on whether the particular scenario is mapped to the transportation problem or the k-server problem. Since this subset is initially unknown, Bender's decomposition involves an iterative algorithm starting with empty $\tilde{E}^s$, and iterating solutions to the problem with this subset of constraints (called the *master* problem), while generating and adding constraints to the master using the dual program for each s, until convergence (which is guaranteed since E^s is finite).

A problem remains with the above formulation: if police vehicles significantly outnumber crime events, we only need a few of the available resources to attain a global minimum, and the remaining vehicles are allocated arbitrarily. In practice, this is unsatisfactory, as there are numerous secondary objectives, such as overall crime deterrence, which motivate allocations of police which are geographically diverse. We incorporate these considerations informally into the following heuristic objectives:

- There should be more police coverage in areas that observe more crime, on average, and
- Police should be diversely distributed over the entire coverage area.

We incorporate these secondary objectives by modifying the objective function in (3) to be

$$\min_q \; -\gamma h_i q_i + \kappa q_i + \min_{x^s \geq 0} \sum_{s \in S} \sum_{ij} d_{ij} x_{ij}^s \tag{9}$$

where h_i is the observed frequency of crimes in grid i and γ and κ are parameters of our model. The first term $\gamma h_i q_i$ forces the model to place police in high crime grids. The second term κq_i penalizes the placement of too many police vehicles in a grid and thus forces the model to distribute police among grids.

2.3 Iterative Stochastic Programming

Bender's decomposition enables us to solve the stochastic program under the assumption that f is stationary. A key challenge identified above however, is that the distribution of future crime actually depends on the police placement policy q. Consequently, a solution to the stochastic program for a fixed set of samples s from a distribution f is only optimal if this distribution reflects the distribution of crime conditional on q, turning stochastic program into a fixed point problem. We propose to use an iterative algorithm, *IBRO (Iterative Bender's Response Optimization)* (Algorithm 1), to address this issue. Intuitively, the algorithm provides police repeated chances to react to crimes, while updating the distribution of crimes given current police positions. In the algorithm, MAX_ITER is an upper limit on the number of iterations, e is the set of all evidence (features) except police presence and $\tau|z$ is the response time to crime z. q and z, as before, refer to vectors of police placements and crime locations and $q_i|z_i$ refers to police placement given a particular set of crimes.

Algorithm 1 IBRO

INPUT: q_0: Initial Police Placement
OUTPUT: q^*: Optimal Police Placement
for $i = 1..MAX\_ITER$ **do**
    Sample Crime z_i from $f(t|e, q_{i-1})$
    Find Optimal Police Placement $q_i|z_i$ by Stochastic Programming.
    Calculate $\mathbf{E_i}(\tau|z_i)$
    if $\mathbf{E_i}(\tau|z_i) > \mathbf{E_{i-1}}(\tau|z_{i-1})$ **then**
        Return q_{i-1}
    end if
    if $|\mathbf{E_i}(\tau|z_i) - \mathbf{E_{i-1}}(\tau|z_{i-1})| \leq \epsilon$ **then**
        Return q_i
    end if
end for
Return q_i

3 Crime and Police Data

In order to validate the decision theoretic model above, we used the following data to learn the parametric model of crime described in Section 4. We use burglary data from 2009 for Davidson County, TN, a total of 4,627 incidents, which includes coordinates and reported occurrence times. Observations that lacked coordinates were geo-coded from their addresses. In addition, we used police vehicle patrol data for the same county, consisting of GPS dispatches sent by county police vehicles, for a total of 31,481,268 data points, where each point consists of a unique vehicle ID, time, and spatial coordinates. A total of 624 retail shops that sell liquor, 2494 liquor outlets, 41 homeless shelters, and 52 pawn shops were taken into account. We considered weather data collected at the county level. Additional risk-terrain features, included population density, housing density, and mean household income at a census tracts level.

4 Continuous-Time Crime Forecasting

4.1 Model

Crime models commonly fall into three categories: purely spatial models, which identify spatial features of previously observed crime, such as hot spots (or crime clusters), spatial-temporal models which attempt to capture dynamics of attractiveness of a discrete set of locations on a map, and risk-terrain models, which identify key environmental determinants (risk factors) of crime, and create an associated time-independent risk map. A key gap in this prior work is the lack of a spatial-temporal generative model that can capture both spatial and temporal correlates of crime incidents, such as time of day, season, locations of liquor outlets and pawn shops, and numerous others. We propose to learn a density $f(t|w)$ over time to arrival of crimes for a set of discrete spatial locations G, allowing for spatial interdependence, where w is a set of crime co-variates.

A natural choice for this problem is survival analysis [6] which allows us to represent distribution of time to events as a function of arbitrary features. Formally, the survival model is $f_t(t|\gamma(w))$, where f_t is a probability distribution for a continuous random variable T representing the inter-arrival time, which typically depends on covariates w as $\log(\gamma(w)) = \rho_0 + \sum_i \rho_i w_i$. A key component in a survival model is the survival function, which is defined as $S(t) = 1 - F_t(t)$, where $F_t(t)$ is the cumulative distribution function of T. Survival models can be parametric or non-parametric in nature, with parametric models assuming that *survival time* follows a known distribution. In order to model and learn $f(t)$ and consequently $S(t)$, we chose the exponential distribution, which has been widely used to model inter-arrival time to events and has the important property of being memoryless. We use Accelerated Failure Model (AFT) for the survival function over the semi-parametric Cox's proportional hazard model (PHM) and estimate the model coefficients using maximum likelihood estimation (MLE), such that in our setting, $S(t|\gamma(w)) = S(\gamma(w)\,t)$. While both the AFT and PHM models measure the effects of the given covariates, the former measures it with respect to survival time and the latter does so with respect to the hazard. The AFT model thus allows us to offer natural interpretations regarding how covariates affect crime rate.

A potential concern in using survival analysis in this setting is that grids can experience multiple events. We deal with this by learning and interpreting the model in a way that the multiple events in a particular grid are treated as single events from multiple grids and prior events are taken into consideration by updating the temporal and spatial covariates.

In learning the survival model above, there is a range of choices about its spatial granularity, from a single homogeneous model which captures spatial heterogeneity entirely through the model parameters w, to a collection of distinct models f_i for each spatial grid $i \in G$. For a homogeneous model it is crucial to capture most of the spatial variation as model features. Allowing for a collection of distinct models f_i, on the other hand, significantly increases the risk of overfitting, and reduces the ability to capture generalizable spatially-invariant

Table 1: Variables for Crime Prediction

Type of Feature	Sub-Type	Variable	Description
Temporal	Temporal Cycles	Time of Day	Each day was divided into 6 equal time zones with binary features for each.
		Weekend	Binary features to consider whether crime took place on a weekend or not.
		Season	Binary features for winter, spring, summer and fall seasons.
	Weather	Mean Temperature	Mean Temperature in a day
		Rainfall	Rainfall in a day
		Snowfall	Snowfall in a day
	Effect of Police	Police Presence	Number of police vehicles passing through a grid and neighboring grids over past 2 hours
Spatial	Risk-Terrain	Population Density	Population density (Census Tract Level)
		Household Income	Mean Household Income (Census Tract Level)
		Housing Density	Housing Density (Census Tract Level)
Spatial-Temporal	Spatial Correlation	Past Crime	Separate variables considered for each discrete crime grid representing the number of crimes in the last two days, past week and past month. We also looked at same crime measures for neighbors of a grid.
	Effect of Police	Crime Spillover	Number of police vehicles passing in the past two hours through grids that are not immediately adjacent, but farther away.

knowledge about crime co-variates. To balance these considerations, we split the discrete spatial areas into two coarse categories: high-crime and low-crime, and learned two distinct homogeneous models for these. We do this by treating the count of crimes for each grid as a data point and then splitting the data into two clusters using *k-means* clustering.

The next step in the modeling process is to identify a collection of features that impact crime incidence, which will comprise the co-variate vector w. In doing this, we divide the features into temporal (those that only change with time), spatial (those capturing spatial heterogeneity), and spatio-temporal (features changing with both time and space).

4.2 Temporal Features

Temporal Crime Cycles: Preliminary analysis and prior work [7, 13] were used to identify the set of covariates, such as daily, weekly and seasonal cycles, that affect crime rate. Crime rates have also been shown to depend on seasons (with more crime generally occurring in the summer) [14]. Thus, we consider seasons as binary features. In order to incorporate crime variation throughout the day, each day was divided into six zones of four hours each, captured as binary features. Similarly, another binary feature was used to encode weekdays and weekends.

Temporal Crime Correlation: It has previously been observed that crime exhibits inter-temporal correlation (that is, more recent crime incidents increase thelikelihood of subsequent crime). To capture this aspect, we used recent crime counts in the week and month preceding time under consideration.

Weather: It is known that weather patterns can have a significant effect on crime incidence [5]. Consequently, we included a collection of weather-related features, such as rainfall, snowfall, and mean temperature.
Police Presence: The final class of features that are particularly pertinent to our optimization problem involves the effect of police presence on crimes. Specifically, it is often hypothesized that police presence at or near a location will affect future crime at that location [12]. We try to capture this relationship, by including a feature in the model corresponding to the number of police vehicles passing within the grid, as well as its immediate neighboring grid cells, over the previous two hours.

4.3 Spatial and Spatio-Temporal Features

Risk-Terrain Features: We leveraged the risk-terrain modeling framework [10], as well as domain experts, to develop a collection of spatial features such as population density, mean household income, and housing density at the census tract level. We used the location of pawn shops, homeless shelters, liquor stores, and retail outlets that sell liquor as the observed spatial-temporal variables (note that temporal variation is introduced, for example, as new shops open or close down).
Spatial Crime Correlation: One of the most widely cited features of crime is its spatial correlation (also referred to as *repeat victimization* [11]), a phenomenon commonly captured in hot-spotting or spatial crime clustering techniques. We capture spatial correlation as follows. For each discrete grid cell in the space we first consider the number of crime incidents over the past two days, past week, and past month, as model features, capturing repeat victimization within the same area. In addition, we capture the same features of past crime incidents for neighboring grid cells, capturing spatial correlation.
Spatial Effects of Police Presence: Aside from the temporal effect of police on crime (reducing its frequency at a particular grid cell), there is also a spatial effect. Specifically, in many cases criminals may simply commit crime elsewhere [8]. To capture this effect, we assume that the spillover of crime will occur between relatively nearby grid cells. Consequently, we add features which measure the number of police patrol units over the previous two hours in grid cells that are not immediately adjacent, but are several grid cells apart. In effect, for a grid cell, we hypothesize that cells that are very close push crime away or reduce it, whereas farther away grids spatially shift crime to the concerned grid, causing spillover effects. The list of all the variables is summarized in Table 1.

5 Results

5.1 Experiment Setup

We used python and R to learn the model parameters, with *rpy2* acting as the interface between the two. We make direct comparison of our model to the discrete-time non-parametric Dynamic Bayes Network model [23, 22] and the DSDA continuous-time model [9]. We used CPLEX version 12.51 to solve the optimization problem described in Section 2. The experiments were run on a 2.4GHz hyperthreaded 8-core Ubuntu Linux machine with 16 GB RAM.

5.2 Evaluation of Crime Prediction

Our first step is to evaluate the ability of our proposed continuous-time model based on survival analysis to forecast crime. Our parametric model is simpler (in most cases, significantly) than state-of-the-art alternatives, and can be learned using standard maximum likelihood methods for learning survival models. Moreover, it is nearly homogeneous: only two distinct such models are learned, one for low-crime regions, and another for high-crime regions. This offers a significant advantage both in interpretability of the model itself, as well as ease of use. Moreover, because our model incorporates environmental factors, such as locations of pawn shops and liquor stores, it can be naturally adapted to situations in which these change (for example, pawn shops closing down), enabling use in policy decisions besides police patrolling. On the other hand, one may expect that such a model would result in significant degradation in prediction efficacy compared to models which allow low-resolution spatial heterogeneity. As we show below, remarkably, our model actually outperforms alternatives both in terms of prediction efficacy, and, rather dramatically, in terms of running time.

For this evaluation, we divided our data into 3 overlapping datasets, each of 7 months. For each dataset, we used 6 months of data as our training set and 1 month's data as the test set. For spatial discretization, we use square grids of sides 1 mile throughout, creating a total of 900 grids for the entire area under consideration. While our model is continuous-time, we draw a comparison to both a continuous-time and a discrete-time models in prior art. However, since these are not directly comparable, we deal with each separately, starting with the continuous-time DSDA model. We refer to the DSDA model simply as *DSDA*, the model based on a Dynamic Bayes Network is termed *DBN*, and our model is referred to as *PSM* (parametric survival model).

Prediction Effectiveness Comparison with DSDA Our first experiments involve a direct performance comparison to a state-of-the-art DSDA model due to Ihava et al. [9]. We chose this model for two reasons. First, DSDA provides a platform to make a direct comparison to a continuous time model. Second, it uses time series modeling and CrimeStat, both widely used tools in temporal and spatial crime analysis.

We introduce the underlying concept of the model before comparing our results. DSDA segregates temporal and spatial aspects of crime prediction and learns them separately. In the temporal model, days like Christmas, Halloween, and football match days that are expected to show deviation from the usual crime trend are modeled using hierarchical profiling (HPA) by using the complement of the gamma function:

$$y = a_p - b_p t^{c_p - 1} e^{-d_p t}$$

where y is observed count, t is time and a_p, b_p, c_p and d_p are the parameters to be estimated using ordinary least squares (OLS).

All other days are initially assumed to be part of a usual average weekly crime rate, which is modeled using the following harmonic function

$$y = a_a - b_a t + c_a t^2 + \sum_{i=1}^{26} \left[d_a \cos\left(\frac{i\pi t}{26}\right) + e_a \sin\left(\frac{i\pi t}{26}\right) \right]$$

where y_a is the weekly crime average, t is time and a_a, b_a, c_a, d_a and e_a are the parameters that are estimated using OLS. Then, the deviations are calculated from the observed data and these are again modeled using the harmonic function. This forms the deterministic part of the model $f(t)$. The error Z from the observed data is modeled using seasonal ARIMA, and the final model is $y = f(t) + Z$. The spatial component of DSDA was evaluated using STAC [9], which is now a part of CrimeStat [15].

In order to make a comparative analysis, we considered a natural adaptation of the HPA-STAC model, which enables us to compare likelihoods. We use the outputs (counts of crime) from the HPA model as a mean of a Poisson random variable, and sample the number of crimes from this distribution for each day. For the spatial model, HPA-STAC outputs weighted clusters in the form of standard deviation ellipses, a technique used commonly in crime prediction. Here, we consider that:

$$P(x_i) = P(c(x_i))P(x_i^{c(x_i)})$$

where $P(x_i)$ is the likelihood of a crime happening at a spatial point x_i which belongs to cluster c_i, $P(c(x_i))$ is the probability of choosing the cluster to which point x_i belongs from the set of all clusters and $P(x_i^{c(x_i)})$ is the probability of choosing point x_i from its cluster c_i. We assume that $P(x_i^{c_i}) \propto \frac{1}{Area_{c(x_i)}}$. Finally, we assume that the total likelihood is proportional to the product of the spatial and temporal likelihoods.

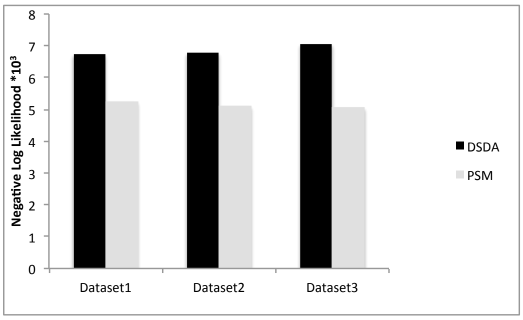


Fig. 2: Likelihood comparison of PSM vs DSDA.

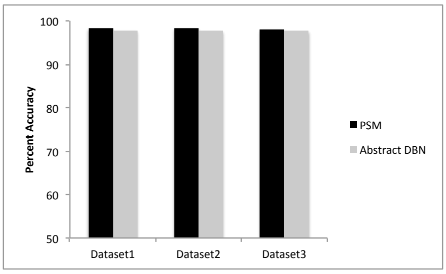
Figure 2 shows the comparison of DSDA log-likelihood (on test data) for the three datasets described above. Indeed, our model outperforms DSDA in both the temporal and the spatial predictions by a large margin (overall, the improvement in log-likelihood is 25-30%).

Prediction Effectiveness Comparison with the Dynamic Bayes-Network Model Next, we compare our model to the framework proposed by Zhang et al. [23], which looks at crime prediction by learning a non-parametric Dynamic Bayes Network (DBN) representation, and applying abstraction techniques to improve scalability [22]. The DBN includes three sets of state variables: numbers of police vehicles in each grid i at time t, denoted by D_{it}, the number of criminals in grid i at time t, X_{it}, and the number of crimes Y_{it} in each grid i at time t. The main assumptions of this DBN are that a) police vehicle dynamics are known (so they are not random variables), b) locations of criminals at time $t+1$ only depends on patrol and criminal (but not crime) locations at time t, and c) crime incidents at time t only depend on locations of criminals and police at time t. Consequently, the problem involves learning two sets of transition models: $P(X_{i,t+1}|D_{1,t},...,D_{N,t},X_{1,t},...,X_{N,t})$ and $P(Y_{i,t}|D_{1,t},...,D_{N,t},X_{1,t},...,X_{N,t})$ for all grid cells i, which are assumed to be independent of time t. Since the model involves hidden variables X, Zhang et al. learn it using the Expectation-Maximization framework. While the model is quite general, Zhang et al. treat X, Y, and D as binary.

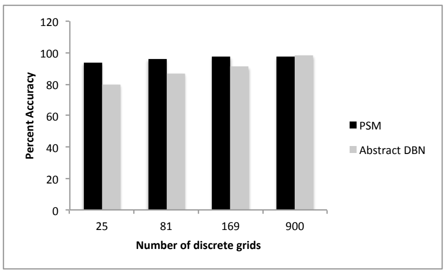
Since our proposed model is continuous-time, whereas Zhang et al. model is in discrete-time, we transform our model forecasts into a single probability of at least one crime event occurring in the corresponding interval. Specifically, we break time into 8-hour intervals (same temporal discretization as used by Zhang et al.), and derive the conditional likelihood of observed crime as follows. Given our distribution $f(t|w)$ over inter-arrival times of crimes, and a given time interval $[t_1,t_2]$, we calculate the probability of observing a crime in the interval as $F(t \leq t_2|w) - F(t \leq t_1|w)$, where F represents the corresponding cumulative distribution function (cdf).

To draw the most fair comparison to DBN, we use an evaluation metric proposed by Zhang et al. [22] which is referred to as *accuracy*. Accuracy is calculated as a measure of correct predictions made for each grid and each time-step. For example, if the model predicts a probability of crime as 60% for a target, and the target experiences a crime, then the accuracy is incremented by 0.6. Formally, let p_i be the predicted likelihood of observing a crime count for data point i. Then accuracy is defined as $\frac{1}{m}\sum_i p_i$, where i ranges over the discrete-time sequence of crime counts across time and grids and m the total number of such time-grid items.

Figure 3(a) shows the results of accuracy comparison (with the accuracy measure defined above) between the DBN model and our model (PSM). We can observe that both models perform extremely well on the accuracy measure, with our model very slightly outperforming DBN. We also make comparisons by varying the number of grids, shown in Figure 3 (b), starting around downtown Nashville and gradually moving outwards. Our model outperforms DBN in all but one case, in which the accuracies are almost identical.



(a)



(b)

Fig. 3: Accuracy comparison between PSM and Abstract DBN. (a) Varying data subsets. (b) Varying the number of grids.

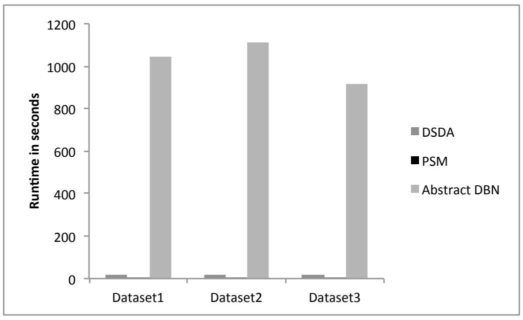


Fig. 4: Runtime comparison (seconds) between DSDA, Abstract DBN, and PSM.

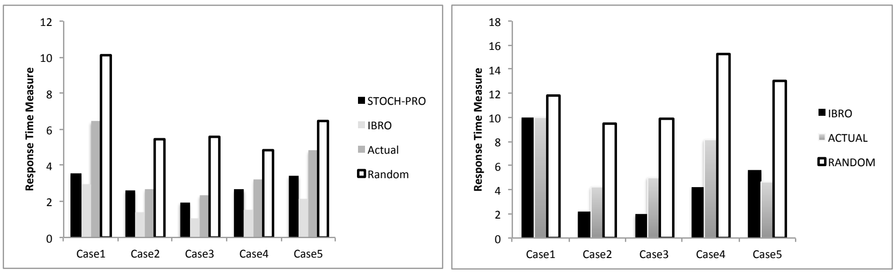
Runtime Comparison with DSDA and DBN We already saw that our PSM model, despite its marked simplicity, outperforms two state-of-the-art forecasting models, representing continuous-time and discrete-time prediction methods, in terms of prediction efficacy. An arguably more important technical advantage of PSM over these is running time. Figure 4 shows running times (of training) for PSM, DSDA, and DBN (using the abstraction scheme proposed by Zhang et al. [22]). The DBN framework is significantly slower than both DSDA and PSM. Indeed, PSM running time is so small by comparison to both DSDA and DBN that it is nearly invisible on this plot.

5.3 Effectiveness of the Response Time Optimization Method

Next, we evaluate the performance of our proposed framework combining iterative stochastic programming with sample average approximation. To do this, we randomly select timezones of 4 hours each from our dataset and sample 100 sets of crimes for each. In practice, although the number of police vehicles is significantly higher than the number of crimes in a 4-hour zone, all police vehicles are not available for responding to a specific type of crimes, due to assigned tasks.

We consider a maximum of a single police vehicle per grid and we consider that only a fraction ($1/6th$) of the them are available to respond to burglaries. In order to simulate the actual crime response by the police department (in order to evaluate actual spatial allocation policy of police vehicles within our data), we greedily assign the closest police vehicle to a crime in consideration.

Our first evaluation uses our crime prediction model f to simulate crime incidents in simulation, which we use to both within the IBRO algorithm, as well as to evaluate (by using a distinct set of samples) the policy produced by our algorithm in comparison with three alternatives: a baseline stochastic programming method (using Bender's decomposition) which ignores the fact that distribution of crimes depends on the police allocation (*STOCH-PRO*), b) actual police location in the data (*Actual*), and c) randomly assigning police vehicles to grids (*Random*). Figure 5(a) demonstrates that IBRO systematically



(a) (b)

Fig. 5: Response Times (lower is better): (a) using simulated crimes, (b) observed crimes.

outperforms these alternatives, usually by a significant margin.

Our next experiment evaluates performance of IBRO in comparison to others with respect to *actual crime incident data*. Note that this is inherently disadvantageous to IBRO in the sense that actual data is not adaptive to the police location as accounted for by IBRO. Nevertheless, Figure 5(b) shows that IBRO typically yields better police patrol location policies than either actual (in the data) or random.

6 Conclusion

We develop a novel bi-level optimization method for allocating police patrols in order to minimize expected crime incident response time. Our approach makes use of stochastic programming, with a Bender's decomposition and constraint generation framework offering a scalable solution approach. Moreover, we introduce a novel iterative stochastic programming algorithm which allows us to account for the dependence of the spatio-temporal crime incidence distribution on police location. To evaluate this optimization framework, we presented a novel discrete-space continuous-time model for forecasting crime as a function of a collection of co-variates which include vehicular police deployment. Our model, which makes use of survival analysis, allows for spatial as well as temporal crime correlation, and effectively captures the effect of police presence both temporally and spatially. This model is learned from burglary incident data in a major US metropolitan area. Our experiments demonstrate that this model outperforms state of the art continuous- and discrete-time crime prediction models both in terms of prediction effectiveness and running time.

7 Acknowledgments

This research was partially supported by the NSF (IIS-1526860), ONR (N00014-15-1-2621), ARO (W911NF-16-1-0069), ARO MURI (W911NF-111-0332), and Vanderbilt University.

References

1. Bertsimas, D., Tsitsiklis, J.N.: Linear Optimization. Athena Scientific, 3rd edn. (1997)
2. Brantingham, P.J., Brantingham, P.L.: Patterns in crime. Macmillan New York (1984)
3. Chrobak, M., Karloof, H., Payne, T., Vishwnathan, S.: New ressults on server problems. SIAM Journal on Discrete Mathematics 4(2), 172–181 (1991)
4. Cohen, J., Gorr, W.L., Olligschlaeger, A.M.: Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. Geographical Analysis 39(1), 105–127 (2007)
5. Cohn, E.G.: Weather and crime. British journal of criminology 30(1), 51–64 (1990)
6. Cox, D.R., Oakes, D.: Analysis of survival data, vol. 21. CRC Press (1984)
7. Felson, M., Poulsen, E.: Simple indicators of crime by time of day. International Journal of Forecasting 19(4), 595–601 (2003)
8. Hope, T.: Problem-oriented policing and drug market locations: Three case studies. Crime prevention studies 2(1), 5–32 (1994)
9. Ivaha, C., Al-Madfai, H., Higgs, G., Ware, J.A.: The dynamic spatial disaggregation approach: A spatio-temporal modelling of crime. In: World Congress on Engineering. pp. 961–966 (2007)
10. Kennedy, L.W., Caplan, J.M., Piza, E.: Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. Journal of Quantitative Criminology 27(3), 339–362 (2011)
11. Kleemans, E.R.: Repeat burglary victimisation: results of empirical research in the netherlands. Crime prevention studies 12, 53–68 (2001)
12. Koper, C.S.: Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. Justice Quarterly 12(4), 649–672 (1995)
13. Landau, S.F., Fridman, D.: The seasonality of violent crime: the case of robbery and homicide in israel. Journal of research in crime and delinquency 30(2), 163–191 (1993)

14. Lauritsen, J.L., White, N.: Seasonal Patterns in Criminal Victimization Trends, vol. 245959. US DOJ, Office of Justice Program, Bureau of Justice Statistics (2014)
15. Levine, N., et al.: Crimestat III: a spatial statistics program for the analysis of crime incident locations (version 3.0). Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice (2004)
16. Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E.: Self-exciting point process modeling of crime. Journal of the American Statistical Association 106(493), 100–108 (2011)
17. Murray, A.T., McGuffog, I., Western, J.S., Mullins, P.: Exploratory spatial data analysis techniques for examining urban crime implications for evaluating treatment. British Journal of criminology 41(2), 309–329 (2001)
18. Short, M.B., D'Orsogna, M.R., Pasour, V.B., Tita, G., Brantingham, P.J., Bertozzi, A.L.., Chayes, L.B.: A statistical model of criminal behavior. Mathematical Models and Methods in Applied Sciences pp. 1249–1267 (2008)
19. 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), 1249–1267 (2008)
20. Speer, P.W., Gorman, D.M., Labouvie, E.W., Ontkush, M.J.: Violent crime and alcohol availability: Relationships in an urban community. Journal of Public Health Policy 19(3), 303–318 (1998)
21. Toomey, T.L., Erickson, D.J., Carlin, B.P., Quick, H.S., Harwood, E.M., Lenk, K.M., Ecklund, A.M.: Is the density of alcohol establishments related to nonviolent crime? J Stud Alcohol Drugs 73(1), 21–25 (2012)
22. Zhang, C., Bucarey, V., Mukhopadhyay, A., Sinha, A., Qian, Y., Vorobeychik, Y., Tambe, M.: Using abstractions to solve opportunistic crime security games at scale. In: International Conference on Autonomous Agents and Multiagent Systems (2016), to appear
23. Zhang, C., Sinha, A., Tambe, M.: Keeping pace with criminals: Designing patrol allocation against adaptive opportunistic criminals. In: International Conference on Autonomous Agents and Multiagent Systems. pp. 1351–1359 (2015)