CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems
School of Information Systems

7-2014

STREETS: Game-Theoretic Traffic Patrolling with Exploration and Exploitation

Matthew BROWN
University of Southern California

Sandhya SAISUBRAMANIAN
Singapore Management University, sandhyas@smu.edu.sg

Pradeep VARAKANTHAM
Singapore Management University, pradeepv@smu.edu.sg

Milind TAMBE
University of Southern California

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Artificial Intelligence and Robotics Commons, and the Operations Research, Systems Engineering and Industrial Engineering Commons

Citation

BROWN, Matthew; SAISUBRAMANIAN, Sandhya; VARAKANTHAM, Pradeep; and TAMBE, Milind. STREETS: Game-Theoretic Traffic Patrolling with Exploration and Exploitation. (2014). *Proceedings of 26th IAAAI Conference on Innovative Applications in Artificial Intelligence: 29-31 July 2014, Quebec City*. 2966-2971. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/2222

This Conference Proceeding Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

STREETS: Game-Theoretic Traffic Patrolling with Exploration and Exploitation

Matthew Brown*, **Sandhya Saisubramanian**$^+$, **Pradeep Varakantham**$^+$, **Milind Tambe***
*University of Southern California, Los Angeles, CA 90089, USA
$^+$Singapore Management University, Singapore 178901
*{mattheab, tambe}@usc.edu, $^+${sandhyas, pradeepv}@smu.edu.sg

Abstract

To dissuade reckless driving and mitigate accidents, cities deploy resources to patrol roads. In this paper, we present STREETS, an application developed for the city of Singapore, which models the problem of computing randomized traffic patrol strategies as a defender-attacker Stackelberg game. Previous work on Stackelberg security games has focused extensively on counter-terrorism settings. STREETS moves beyond counter-terrorism and represents the first use of Stackelberg games for traffic patrolling, in the process providing a novel algorithm for solving such games that addresses three major challenges in modeling and scale-up. First, there exists a high degree of unpredictability in travel times through road networks, which we capture using a Markov Decision Process for planning the patrols of the defender (the police) in the game. Second, modeling all possible police patrols and their interactions with a large number of adversaries (drivers) introduces a significant scalability challenge. To address this challenge we apply a compact game representation in a novel fashion combined with adversary and state sampling. Third, patrol strategies must balance exploitation (minimizing violations) with exploration (maximizing omnipresence), a tradeoff we model by solving a bi-objective optimization problem. We present experimental results using real-world traffic data from Singapore. This work is done in collaboration with the Singapore Ministry of Home Affairs and is currently being evaluated by the Singapore Police Force.

Introduction

Traffic safety is a significant concern in cities throughout the world. Of the large number of people injured or killed in traffic accidents, a vast majority of these casualties are a direct result of reckless driving. It is for this reason, that the Singapore Police Force and their counterparts in other cities use traffic patrols to persuade drivers to comply with traffic laws through the threat of citations and fines. Such patrols must be randomized to avoid predictability and provide adequate coverage of different areas of a city. Yet, lack of randomization is a well-known problem in human patrol scheduling (Tambe 2011) and when such randomization must also take into account speed-distance calculations, potential traffic delays, and historical data on traffic violations to ensure appropriate coverage of different areas in a city like Singapore, it presents a very difficult challenge for human schedulers.

Stackelberg security games (SSG) have become an increasingly popular paradigm for modeling security patrolling problems. In SSGs, the defender (i.e., the security agency) commits to a mixed strategy that the adversary (i.e., criminal, terrorist, or in our domain, reckless driver) is able to first *observe* and then best respond (Korzhyk, Conitzer, and Parr 2010; Basilico, Gatti, and Amigoni 2009). This mixed strategy represents a probability distribution over the possible patrol schedules. Research on SSGs has resulted in several real-world systems deployed to protect transportation infrastructure such as airports, ports, and train stations (Tambe 2011). These systems have focused predominantly on counter-terrorism domains. Of the few applications that have branched out from counter-terrorism, e.g., TRUSTS (Yin et al. 2012), none have focused on traffic patrolling.

The purpose of this paper is to introduce a new game-theoretic application, STREETS (**ST**rategic **R**andomization with **E**xploration and **E**xploitation in **T**raffic patrol **S**chedules), which we developed to assist the Singapore Ministry of Home Affairs (MHA) in scheduling randomized traffic patrols on the Singapore road network. We model this problem as a Stackelberg game with one defender (the police) and multiple adversaries (drivers). STREETS represents a novel application of Stackelberg games and required addressing several research challenges. First, road networks are complex and dynamic systems, with unpredictable delays associated with congestion, traffic signals, etc. The presence of this type of uncertainty complicates the process of planning traffic patrols. Second, the game being played at the heart of STREETS is massive in scale in terms of both the number of possible patrol strategies as well as the number of adversaries representing the thousands of drivers who use the Singapore road network. Third, the repeated nature of the traffic patrolling domain results in an abundance of data on traffic, accidents, citations, etc. However, this data is collected when the defender issues citations and thus is inherently available only for patrolled locations. Therefore, it is important to avoid confirmation bias (Nickerson 1998) from over relying on the data, which can lead to self-reinforcing behavior and undesired consequences.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

No previous work on SSGs has addressed these challenges in combination, and in fact none has addressed the challenge of avoiding confirmation bias – leading us to introduce a new concept of exploration versus exploitation in SSGs. Therefore, STREETS required us to develop a new SSG game model and an entirely new algorithm combining three key features. First, to capture the inherent stochasticity of a road network, we use a Markov Decision Process (MDP) to model the defender's patrol scheduling problem. Second, to formulate a game with an exponential number of patrol strategies and a large number of adversaries, we adopt a compact game representation which converts the defender's strategy space to a network flow through a transition graph. Additionally, we use two sampling approaches that improve efficiency by considering only a subset of either adversary types or game states when solving the game. Third, while we exploit all available data to improve patrol effectiveness, to prevent overfitting this data, we introduce an entropy-based approach. The idea being that the defender should patrol all areas of the road network with at least some probability to avoid confirmation bias and to give the perception of omnipresence to drivers. This creates a tradeoff between exploitation (minimizing reckless driving by focusing on high violation areas) and exploration (maximizing omnipresence by dispersing patrols). We explicitly formulate this tradeoff as a bi-objective optimization problem. Rather than having one optimal patrol strategy, the patrolling agency can now choose from the space of optimal tradeoff strategies located on the Pareto frontier.

STREETS was developed in collaboration with the Singapore Ministry of Home Affairs. STREETS is currently being evaluated by Singapore Police Force.

Domain

Traffic safety is a significant concern in cities throughout the world. Of the large number of people injured or killed in traffic accidents, a vast majority of these casualties are a direct result of reckless driving. For example, Singapore experienced 7,188 injury accidents in 2012, resulting in 168 fatalities. Perhaps just as alarming is the 330,909 traffic violations recorded during that same period for a vehicle population of only 965,192 (SPF 2013). It is sobering statistics like these that compel the Singapore Traffic Police (TP) and their counterparts in other cities to use traffic patrols to enforce traffic laws through the threat of citations and fines.

Since the number of roads and highways is typically very large, it is not possible to have enough resources to patrol every road and highway at every time. Therefore, a major challenge for TP is to compute patrol strategies on when and where different groups have to patrol so as to reduce the number of violations and accidents.

Due to our collaboration with the Future Urban Mobility (FM) [1] center in Singapore, we are able to obtain both the traffic volumes, violations, and accidents occurring on all the major roads and highways across Singapore. By using this data, we construct models of traffic behavior on various roads and then using the techniques developed in the next section, we generate randomized patrol strategies.

[1]FM is part of the Singapore MIT Alliance for Research and Technology (SMART) initiative.

Model

We formally model the interaction between the police and drivers as a defender-attacker Stackelberg game. This game played by the defender and the adversaries takes place on a graph which models a road network where vertices represent intersections and edges represent road segments. The graph features a temporal dimension, where traversing a road segment takes some (non-deterministic) amount of time. The defender has a maximum patrol duration of h hours. The defender (the police) commits to a randomized patrol strategy, which is used to generate daily patrol schedules for each of the r resources. A daily patrol schedule consists of a trajectory through the graph, i.e., a sequence of road segments to patrol and the times they are to be patrolled.

The adversaries (drivers) also follow a schedule but we assume this trajectory through the graph is fixed on a daily basis (travelling to work, school, etc.). Adversaries are able to observe the presence (or lack thereof) of police patrols over a period of time, in the process obtaining an accurate estimation of the probability of encountering the police on any given day. To construct the graph for the road network in the Singapore Central Business District (CBD), shown in Figure 1, we used data from OpenStreetMap (OSM)[2].

A normal form representation of this game, as used in the original work on Stackelberg security games (Paruchuri et al. 2008), would require us to explicitly enumerate pure strategies for the defender (patrol schedules) as well as for all of the adversaries (obey or violate decisions). This would be an extremely large number of player actions, even for small instances of our traffic patrolling domain. Therefore, we need a technique that allows us to scale up.

Achieving Scaleup

We adopt a compact representation in the form of a transition graph, which converts the game, from the defender's perspective, into a spatio-temporal flow problem. Rather than computing a probability distribution over full patrol schedules, the defender now has to compute the optimal flow through the transition graph. Such a flow can be interpreted as a marginal coverage vector. These marginals can then be used to reconstruct daily patrol schedules for the defender.

This transition graph formulation is similar to the approach used in TRUSTS which modeled patrolling a train line. However, the traffic patrolling domain features a number of complexities that make our use of a transition graph within a Stackelberg game novel. One of the biggest complexities is the continuous nature of traffic patrolling. Not tied to following predetermined transportation schedules (e.g. train schedules in TRUSTS), a traffic patroller, generally speaking, can be almost anywhere within the road network at any given time. To avoid having to adopt a continuous-time model, and the associated computational overhead, we discretize time to a granularity of m minutes. Therefore, a vertex is added to the transition graph for every

[2]http://www.openstreetmap.org/

Figure 1: Singapore OSM Graph

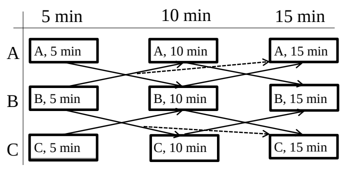


Figure 2: Toy MDP Example

intersection in the road network every m minutes until the patrol duration of h hours is reached.

Defender Model In reality, there may be unexpected delays that disrupt the defender's daily patrol schedules. In a road network, a patroller can be delayed from its schedule due to a variety of factors including congestion or traffic signals. The defender must account for stochasticity in traffic delays when planning patrols. Therefore, we now define an MDP $\langle S, A, T, R \rangle$ to represent the defender's patrol scheduling problem:

- S is a finite set of states. Each state $s \in S$ is a tuple (l, τ), where l is the current location (i.e., intersection in the road network) of the defender and τ is the current time.
- A is a finite set of actions. The set of actions available from a given state $s = (l, \tau)$, $A(s)$, is the set of road segments which originate from location l.
- $T(s, a, s')$ is the probability of ending up in the state s' after performing action a in state s.
- $R(s, a, s')$ is the immediate reward for the defender from ending up in state s' after performing action a in state s. However, our main focus is on the game-theoretic reward (i.e., expected number of violations) as a result of the defender patrolling strategy. Thus, for the remainder of this paper, we assume, without loss of generality, that $R(s, a, s') = 0$, $\forall s, a, s'$.

Figure 2 shows a toy example of the MDP with three locations (A,B,C) and three time periods (5,10,15). The solid black arrows indicate the transitions available from each vertex. The dashed arrows represent uncertainity in the domain, e.g., anticipating going from $(B, 5) \rightarrow (A, 10)$ but being delayed and ending up in $(A, 15)$. The defender strategy is represented by the probability placed on each edge in the MDP rather than over whole patrols.

Adversary Model The set of adversaries consists of the drivers using the road network, who are assumed to always violate the law in the absence of police presence. A driver type is defined for each state-action pair s, a in the MDP and we refer to this type as $<s, a>$. This formulation represents the driver entering the transition graph (road network) at a specified vertex (intersection) and time, traversing an edge (road segment), and then exiting at the destination vertex at a later time. Thus, the trajectory of each driver type in the game is modeled as a single road segment. The reasoning being that a driver may change their behavior for different roads, choosing to violate the law on some road segments and comply with the law on others. Thus, if the decision to violate or not is made on a road-by-road basis and the decision for one road segment does not affect the decision at another, then there is no need to model driver types with trajectories with multiple road segments.

Given a fixed trajectory consisting of a single road segment, the only decision made by each individual driver type is the frequency with which they will obey the law as opposed to violate the law. This decision is influenced by the defender's patrol strategy, which we assume to be known to the drivers. If the perceived likelihood of encountering a police officer is high, then the driver will choose to obey the law more frequently (Koper 1995). More precisely, we define a coverage threshold $t(s, a)$ for driver type $<s, a>$ that represents the probability of encountering a patroller above which the driver will always obey the law. Starting from always violating in the absence of police patrols, we model that the probability of violating the law decreases linearly as the frequency patrols increases until the threshold $t(s, a)$ is reached and driver type $< s, a >$ no longer violates.

We use $v(s, a)$ to denote the average daily traffic volume along the road segment during the time range $[\tau, \tau + m)$. Similarly, we use $c(s, a)$ to denote the yearly violation / citation count along the road segment during the time range $[\tau, \tau + m)$. We combine the traffic volume and violation count data to define the prior associated with type $<s, a>$ as $p(s, a) = \frac{c(s,a)}{365 \times v(s,a)}$. This provides the defender with a distribution over all the adversary types in the game. Through the Future Urban Mobility (FM) research centre, we were able to obtain traffic volume and violation count data for the Singapore CBD. We processed this data and utilized it to populate the values of $v(s, a)$, $c(s, a)$, and $p(s, a)$ which serve as input in our game.

Generating Randomized Patrols

Remember that the defender is trying to achieve two objectives simultaneously: (1) minimize violations; and (2) maximize omnipresence. These objectives are conflicting as they drive the defender towards different patterns of behavior. The desire to minimize violations incentivizes the defender to exploit the traffic data and patrol only in areas where violations have occurred before. Meanwhile, the desire to maximize omnipresence incentivizes exploration so that all areas of the road network are patrolled at least occasionally. Given two conflicting objectives, some tradeoff between exploration and exploitation must be made.

We borrow from work on randomized MDPs (Paruchuri et al. 2006) to formalize the tradeoff between exploration and exploitation. For discrete probability distributions, we know that entropy provides a useful measure of randomness. The MDP policy which maximizes entropy is a uniform random policy $\hat{\pi}$. Given this, one way to evaluate the level of exploration achieved is to determine the ratio of randomness compared to $\hat{\pi}$. To do this, we introduce a parameter $\beta = [0, 1]$. An MDP policy π is said to be β-random if the following condition holds $\pi(s, a) \geq \beta \hat{\pi}(s, a) \ \forall s, a$. Thus, fixing a β value can be thought of as placing constraints on π, forcing it to perform a certain amount of exploration.

Variable	Definition
$c(s,a)$	yearly violation count for type $<s,a>$
$v(s,a)$	daily traffic volume for type $<s,a>$
$p(s,a)$	prior for type $<s,a>$ set to $\frac{c(s,a)}{365\times v(s,a)}$
$t(s,a)$	coverage threshold for type $<s,a>$
$o(s,a)$	probability of type $<s,a>$ obeying the law
$\hat{\pi}$	uniform Markov policy (maximizes entropy)
β	tradeoff parameter between violations / entropy

Figure 3: LP formulation definitions.

However, it is difficult to know *a priori* how to balance the objectives. Therefore, our approach is to generate a set of optimal compromise solutions which form the Pareto frontier using the tradeoff parameter β, where $\beta = 0$ represents full exploitation and $\beta = 1$ represents full exploration. We present a bi-objective linear program which takes β as input and can be solved to generate a point on the Pareto frontier. Different points on the Pareto frontier can be generated by varying the value of β. The Pareto frontier can then be presented to the end user, who selects their desired solution based any qualitative or quantitative measures they choose.

LP Formulation

We can construct a linear program (LP) to solve the MDP formulation of the defender's problem. Let $x(s,a,s')$ denote the marginal probability of the defender reaching state s, executing action a, and ending up in state s'. Similarly, let $w(s,a)$ be the marginal probability of the defender reaching state s and performing action a. The probability of adversary type $<s,a>$ obeying the law which is denoted by $o(s,a)$.

We define the bi-objective linear program as follows:

$$\min_{w,\,x} \sum_{s,a} p(s,a)\,[1-o(s,a)] \tag{1}$$

$$\text{s.t. } x\,(s,\,a,\,s') = w\,(s,\,a)\,T\,(s,\,a,\,s')\,,\forall s,\,a,\,s' \tag{2}$$

$$\sum_{s',\,a'} x\,(s',\,a',\,s) = \sum_{a} w\,(s,\,a)\,,\forall s \tag{3}$$

$$\sum_{a} w(s^{+},\,a) = r \tag{4}$$

$$\sum_{s,\,a} x(s,\,a,\,s^{-}) = r \tag{5}$$

$$w(s,a) \geq 0, \forall s, a \tag{6}$$

$$o(s,a) \leq \frac{w(s,a)}{t(s,a)}, \forall s,\,a \tag{7}$$

$$0 \leq o(s,a) \leq 1, \forall s,\,a \tag{8}$$

$$w(s,a) \geq \beta\hat{\pi}(s,a) \sum_{a'} w(s,a'), \forall s,\,a \tag{9}$$

Equation 1 is the objective function which minimizes the total expected number of violations in the system. This is a zero-sum game where each violation has the same utility and thus our goal of minimizing the total expected violations means that the minimax defender strategy is also the Strong Stackelberg Equilibrium (SSE) strategy. Constraints 2-6 are flow constraints, which combine to enforce that $\mathbf{x}$ and $\mathbf{w}$ represent feasible patrolling strategies with respect to the transition function T. Constraints 2 and 3 define the relationship between $\mathbf{x}$ and $\mathbf{w}$, while Constraints 4 and 5 ensure the flow out of the dummy source state s^{+} as well as into the dummy sink state s^{-} are equal to r. Constraint 7 computes $o(s,a)$ as the ratio between the coverage placed on the road segment $w(s,a)$ and the coverage threshold of adversary type $<s,a>$, $t(s,a)$. For $0 \leq w(s,a) \leq t(s,a)$, adversary type $<s,a>$ will obey the law a fraction of the time, specifically $w(s,a)/t(s,a)$. Constraint 8 is used to ensure that $o(s,a)$ represents a valid probability, i.e., $o(s,a) \in [0,1]$, when $w(s,a) > t(s,a)$. (This places no restrictions on $w(s,a)$, as Constraint 7 is an inequality constraint.)

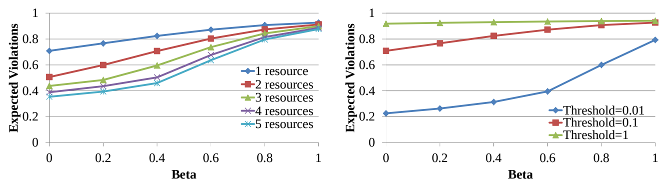
Given β and $\hat{\pi}$ as input, Constraint 9 ensures that the patrolling strategy achieves at least a fraction (i.e., β) of the randomness of the maximal entropy policy, $\hat{\pi}$, which is a uniform random policy. For example, if two actions a_1 and a_2 are available from state s, then $\hat{\pi}(s,a_1)$ and $\hat{\pi}(s,a_2)$ would both be 0.5. For $\beta = 0.2$, Constraint 9 specifies that at least 10% (0.5×0.2) of the flow coming out of state s, i.e., $\sum_a w(s,a)$, must be directed to each action available from s, in this case a_1 and a_2. This constraint allows for a tradeoff between two objectives: (1) minimizing violations ($\beta=0$), and (2) maximizing entropy ($\beta=1$). The Pareto frontier can be generated by solving the LP for different values of β.

Additional Scaleup

For longer patrol lengths, the resulting linear program can grow quite large. To address this challenge we used constraint and state sampling (De Farias and Van Roy 2004).

Driver Type Sampling One approach for using constraint sampling in our problem is driver type sampling. Sampling a subset of the driver types reduces the size of the LP, as only the constraints (i.e., Constraints 7 and 8) and variables (i.e., $o(s,a)$) associated with the sampled driver types are considered. Evaluation becomes more complicated after introducing constraint sampling as we can no longer just look at the objective value obtained by solving the sampled LP, as it only accounts for violations committed by sampled driver types. However, the defender may still implicitly influence the behavior of unsampled driver type $<s,a>$ by placing coverage on the road segment associated with $<s,a>$ in order to position themselves to interact with the sampled driver types. Thus, we use Monte Carlo simulation to sample patrol schedules from the Markov strategy computed for the sampled LP and evaluate the schedules against all driver types.

State Sampling We can also improve efficiency by only considering a sampled subset of states obtained in a principled manner by using a coarser time granularity. For example, doubling the time granularity m cuts the size of the state space in half. However, some extra steps are required when generating patrol schedules or evaluating the patrol strategy generated from the state-sampled LP on the original MDP using Monte Carlo sampling. In either case, if a state $s = (l,\tau)$ is reached which does not exist in the set of sampled states, then a look up is performed for the pol-



(a) Defender Resources (b) Driver Threshold

Figure 4: Effect of Parameters on Expected Violations.

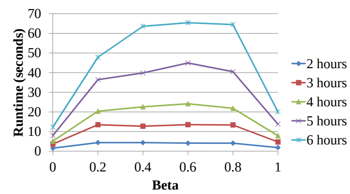


Figure 5: Effect of Patrol Duration on Runtime.

icy from state $s' = (l, \tau')$, where s' is the state in the set of sampled states closest in time to s with the same location l.

Evaluation

To evaluate STREETS, we conducted a set of simulations using actual traffic volume and violation count data from the Central Business District of Singapore provided to us by the Singapore LTA. For each simulation, we compute the Pareto frontier with an granularity of 0.2 on the β parameter which controls the tradeoff between minimizing violations and maximizing omnipresence. The Pareto frontier allows us to compare a fully game-theoretic approach with $\beta = 0$ (all exploitation) against a uniform random approach with $\beta = 1$ (all exploration), as well as everything in between. Unless otherwise specified, the default experimental setup features a patrol length of 240 minutes, a 5 minute time granularity, 1 defender resource, and a coverage threshold $t(s, a)$ of 0.1 for all drivers. All results are averaged over 30 simulations.

Analysis of Tradeoffs

Defender Resources In Figure 4a, we evaluate the effect on the number of expected violations as we vary the number of defender resources r. The x-axis is the value of β used when solving the LP formulation, while the y-axis is the total expected number of violations in the game, i.e., $\sum_{s,a} p(s, a) [1 - o(s, a)]$, achieved by the defender's (Pareto) optimal patrol strategy. As a baseline, we can use these experiments to compare a game-theoretic approach ($\beta = 0$) against a uniform random approach ($\beta = 1$).

From these results, we observe three general trends. First, increasing r leads to a reduction in the expected number of traffic violations in the road network. Second, the benefit of each additional defender resource diminishes as r increases. Third, as β increases, so does the number of expected violations. This makes sense, as the defender is moving closer to a uniform random strategy and farther away from optimizing based on the violations data. It is interesting to see that $\beta = 1$ yields almost the same number of expected violations for all values of r because a uniform random strategy does not allow for coordination (even implicitly) between resources.

Coverage Threshold In Figure 4b, we evaluate the effect on the number of expected violations as we test three different values for driver coverage threshold, $t(s, a)$. For $t(s, a) = 1$, we observe the highest level of violations as well as minimal difference between the performance of the full game-theoretic strategy ($\beta = 0$) and the full uniform random strategy ($\beta = 1$). This seems reasonable given that for $t(s, a) = 1$ it is difficult to dissuade drivers, who are fully deterred from violating only if their road segment is patrolled with probability 1. Decreasing $t(s, a)$ to 0.1, yields a similar level of violations for $\beta = 1$, but with $\beta = 0$, the game-theoretic approach, which is very deliberate in how it allocates it patrols, results in a reasonable decrease in violations. Finally, at $t(s, a) = 0.01$, essentially any amount of patrolling on a road segment will convince the driver types to obey. As a result, the game-theoretic strategy leads to an even greater reduction in the expected number of violations.

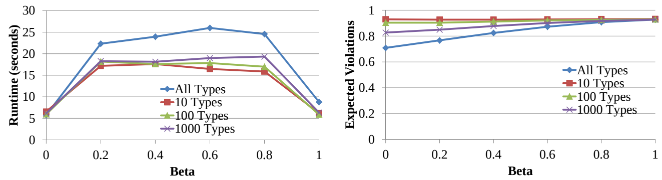
Patrol Duration In Figure 5, we evaluate the effect on runtime as we vary the patrol duration between 2 and 6 hours. Once again the x-axis is β, but now the y-axis is the runtime needed to solve the LP formulation. Intuitively, the results show that the runtime increases as the patrol duration is increased. Additionally, as β is varied, we observe significantly reduced runtimes at the two extremes ($\beta = 0$ and $\beta = 1$), as in both cases, the LP is a single objective optimization problem where the other objective is ignored.

Scalability

STREETS is currently focused on generating randomized traffic patrols for the Singapore CBD. However, the eventual goal for STREETS is to scale to the entire city. Therefore, we evaluate two scaleup approaches to project how they would perform on larger problem sizes.

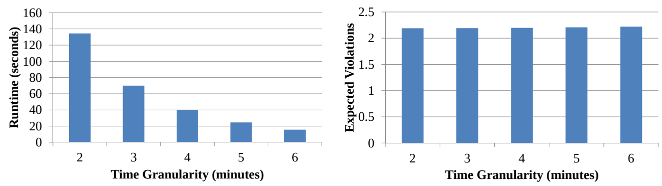
Driver Type Sampling In Figure 6a, we evaluate the effect on runtime for different orders of magnitude of sampled driver types. The original game contains 10346 driver types. Reducing the number of driver types to 1000 via uniform random sampling results in a reasonable decrease in runtime. Further decreasing the number of sampled driver types to 100 and 10 only marginally improves the runtime. Meanwhile, in Figure 6b, we evaluate the effect on solution quality as we vary the number of sampled driver types. For the smallest number of sampled types, the game-theoretic strategy performs only as well as the uniform random strategy which ignores information about the driver types. Furthermore, the number of violations goes down as the number of sampled types goes up. However, the modest runtime improvements combined with the non-negligible loss in solution quality suggests there are limitations on driver type sampling as a technique for improving scalability.

State Sampling In Figure 7a, we evaluate the effect on runtime as we vary the time granularity m between 2 and 6 minutes. The x-axis is the time granularity and the y-axis is



(a) Runtime (b) Expected Violations

Figure 6: Effect of Driver Type Sampling.



(a) Runtime (b) Expected Violations

Figure 7: Effect of State Sampling.

the runtime need to solve the LP formulation for $\beta = 0.5$. We observe an exponential decay in runtime as m is increased. This results in an almost order-of-magnitude runtime decrease by going from $m=2$ to $m=6$. Meanwhile, in Figure 7b, we evaluate the effect on solution quality as we vary m. The x-axis is still the time granularity m, but the y-axis is now the expected violations of the state-sampled strategy when evaluated on the MDP for $m = 2$. We chose to evaluate on $m = 2$ as it was the smallest value of m that we could solve exactly without any sampling. Despite increasing m, the number of expected violations is virtually unchanged. The combination of these runtime and solution quality results are a clear sign that state sampling via adjusting the time granularity can provide the type of scalability needed to handle patrolling over entire cities.

Related Work and Conclusion

In this paper we presented STREETS, a application which we developed to assist the Singapore MHA in scheduling randomized traffic patrols in the Singapore CBD. STREETS is currently in the process of being evaluated by the Singapore Police Force. We have already discussed how this work introduces novelties (MDP formulation, compact game representation, exploration / exploitation) over previous game-theoretic approaches for patrolling domains (Tambe 2011). There is a body of literature examining how to allocate traffic patrols (Adler et al. 2013; Lee, Franz, and Wynne 1979; Koper 1995) as well as how to influence driver behavior (Ritchey and Nicholson-Crotty 2011). That work has established the relation between traffic patrols and their impact on improving traffic safety, which is the basis off which STREETS is built. Much of the related research is prescriptive in nature, offering guidelines and suggestions, but stopping short of providing an implementable approach for patrolling. Our work presents a new perspective on the problem by modeling the interaction between the police and drivers as a game. Importantly, we provide a principled approach for generating randomized schedules.

Acknowledgements

This research is supported in part by the National Research Foundation (NRF) Singapore through the Singapore MIT Alliance for Research and Technology (SMART) and its Future Urban Mobility (FM) Interdisciplinary Research Group.

References

Adler, N.; Hakkert, A. S.; Kornbluth, J.; Raviv, T.; and Sher, M. 2013. Location-allocation models for traffic police patrol vehicles on an interurban network. *Annals of Operations Research* 1–23.

Basilico, N.; Gatti, N.; and Amigoni, F. 2009. Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

De Farias, D. P., and Van Roy, B. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of operations research* 29(3):462–478.

Koper, C. S. 1995. Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. *Justice Quarterly* 12(4):649–672.

Korzhyk, D.; Conitzer, V.; and Parr, R. 2010. Complexity of computing optimal stackelberg strategies in security resource allocation games. *In AAAI*.

Lee, S. M.; Franz, L. S.; and Wynne, A. J. 1979. Optimizing state patrol manpower allocation. *Journal of the Operational Research Society* 885–896.

Nickerson, R. S. 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology* 2(2):175.

Paruchuri, P.; Tambe, M.; Ordonez, F.; and Kraus, S. 2006. Security in multiagent systems by policy randomization. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Paruchuri, P.; Pearce, J. P.; Marecki, J.; Tambe, M.; Ordonez, F.; and Kraus, S. 2008. Playing games with security: An efficient exact algorithm for bayesian stackelberg games. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Ritchey, M., and Nicholson-Crotty, S. 2011. Deterrence theory and the implementation of speed limits in the american states. *Policy Studies Journal* 39(2):329–346.

SPF. 2013. Annual road traffic situation 2012. Technical report, Singapore Police Force.

Tambe, M. 2011. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press.

Yin, Z.; Jiang, A.; Johnson, M.; Tambe, M.; Kiekintveld, C.; Leyton-Brown, K.; Sandholm, T.; and Sullivan, J. 2012. Trusts: Scheduling randomized patrols for fare inspection in transit systems. In *Conference on Innovative Applications of Artificial Intelligence (IAAI)*.