Exploiting Adversarial Uncertainty in Robotic Patrolling: a Simulation-based Analysis

Pablo J. Villacorta and David A. Pelta

Models of Decision and Optimization Research Group,
Dept. of Computer Science and AI, CITIC,
University of Granada, 18071 Granada, Spain,
{pjvi,dpelta}@decsai.ugr.es

Abstract. Recently, several models for autonomous robotic patrolling have been proposed and analysed on a game-theoretic basis. The common drawback of such models are the assumptions required to apply game theory analysis. Such assumptions do not usually hold in practice, especially perfect knowledge of the adversary's strategy, and the belief that we are facing always a best-responser. However, the agents in the patrolling scenario may take advantage of that fact. In this work, we try to analyse from an empirical perspective a patrolling model with an explicit topology, and take advantage of the adversarial uncertainty caused by the limited, imperfect knowledge an agent can acquire through simple observation. The first results we report are encouraging.

Keywords: Adversarial decision making, Robotic patrolling, Stackelberg games, Empirical study

1 Introduction

The problem of patrolling an area with autonomous mobile robots has been studied more and more in the last years, especially nowadays due to the terrorism threat, and is still open. A patrolling scenario can be described as follows. There is one patroller that must defend an area against one or more robbers, assuming the whole area cannot be protected aginst attacks simultaneously so he must move from one location to another within the area all the time.

This means that the patroller has to move within the area to patrol the locations and get sure nobody is trying to steal them. The locations can represent houses or any other good that is valuable for a possible robber. In principle, not all the locations are equally valuable: both the patroller and the robber have their own preferences. The former knows which locations are more important to keep safe, and the robber knows or can intuit in which locations the most valuable goods are. Robbers may have different preferences over the same set of locations according to the kind of criminal he is and the type of product he is interested in: art works in a museum, cars in a factory, jewellery, money from a house, etc.

This situation admits different models and solution techniques. It is clear that the patroller will visit the locations sequentially, although in a sophisticated model we could think of a technologically-advanced patroller that senses several locations at a time, or even think of a team of coordinated patrollers [1, 2]. The patroller can make a decision over the next location to visit [2] or over a whole route of several locations that will be patrolled sequentially before deciding again a new route [6]. Finally, the model may consider a spatial topology describing the relative positions of the locations [2] or not [6].

Although deterministic algorithms were initially proposed for situations where the patroller cannot defend the whole area at a time, they sometimes show an important drawback: they give to the robber the possibility to rob certain locations when the patroller is far enough, provided that the robber was patient enough to observe and learn the deterministic strategy of the patroller before the attack. To make this learning more difficult, randomized patrolling strategies have been proposed in a lot of models. This idea has been also suggested in other non-patrolling adversarial models as it allows a better equilibrium between the confusion induced in the observer agent and the payoff obtained by the patroller, in terms of the routes chosen [7, 6].

This work is aimed at analysing the so-called BGA model [2], proposed to capture a patrolling situation, from a strictly empirical point of view in order to show deviations from the expected behaviour when applied to a real scenario. Insights are provided on the causes of this, along with a brief discussion on how to take advantage of them. The main reasons are the wrong assumptions made by the model, which hold only partially in real life. Similar attempts were made in [3, 4] but the former does not go in enough detail and the latter still abides by a slightly extended game-theoretical model like [2].

This paper is structured as follows. Section 2 briefly reviews the model we will focus on. A numerical perspective with practical results when applying and simulating the model is given in Section 3, which contains the hypothesis together with the experiments that support them, as they are impossible to separate. Section 3.2 suggests an improvement based on the previous numerical results and shows the effectiveness of the proposal. Finally, section 4 contains a brief discussion and further work.

2 Review of the BGA Model

Maybe the most general model was first proposed in [2] and extended in [4]. It allows for a topological representation of the area being patrolled, which does not need to be a perimeter but admits any arbitrary bi-dimensional disposition of the locations. The environment is divided into a set of $C = \{c_1, ..., c_n\}$ cells, with a directed graph G that specifies how they are connected. G can be represented with its adjacency matrix $A(n \times n)$ so $A(i, j) = 1$ if cells c_i and c_j are connected. In the enviroments of our examples we will take G as non-directed but it could also be directed. A subset T of C contains the cells that have some value for both patroller and intruder. The rest of the cells can be seen merely as a link between the interesting locations.

At one turn, the patroller can move from one cell to any of the adjacents and patrol the destination cell. Every cell c_i has associated an integer d_i indicating the number of turns needed to rob it. To rob that cell, the intruder directly enters the cell at once (i.e. the model does not initially consider doors or paths through the environment, although an extension does [4]) and must stay in it for d_i turns from the entering turn. If the patroller reaches the cell during that period, then the intruder will have been caught.

It is assumed that the intruder first observes the patroller without being detected and perfectly learns the patrolling strategy. It is a so-called leader-follower situation, because the patroller can impose the strategy he wants and the intruder will try his best to respond to that strategy. In game-theoretic terms, this can be modeled as a strategic-form game, i.e. a game in which both players, the patroller and the intruder, only play once and act simultaneously. To achieve this, the temporal component has to be eliminated. This can be done by defining the possible actions available to the intruder as *enter-when*(c_i, c_j), meaning that the intruder enters cell c_i the turn after the patroller was in c_j. The randomized strategy of the patroller are the set of probabilities $\{\alpha_{ij}\}$ that indicate the probability that the patroller moves to cell c_j when it is in cell c_i. The intruder is assumed to perfectly know these probabilites because it has been observing the patroller for long enough before choosing a location to rob.

The possible outcomes of this strategic-form game are *penetration-*c_i, *intruder-capture* and *no-attack* if the best choice for the intruder is not to attack any cell ever. The payoffs attained respectively by the patroller and the intruder for each of these outcomes are (X_i, Y_i), $(X_0, 0)$ and (X_0, Y_0), with the restrictions: $X_i < X_0$ (capturing the intruder or persuading it not to attack always reports a higher payoff for the patroller) and $Y_0 \leq 0 < Y_i$ for all c_i (being captured is worse for the intruder than successfully robbing any cell).

According to game theory, the solution of this game is the *leader-follower equilibrium*, which gives the leader (patroller) the maximum expected utility when the follower acts as a best responser to the strategy imposed by the leader, i.e. the follower tries to maximize its own payoff according to the strategy (probabilities) imposed by the leader [5, 2]. This constitutes a bi-level optimization problem whose solution is computed using mathematical programming techniques. It can be proved that, when the leader imposes a randomized strategy, the best response of the follower is always a deterministic action and not another randomization over its actions. In other words, there always exists an action *enter-when*(c_i, c_j) that reports the follower a higher payoff than any randomization over all the *enter-when*$(\cdot,\cdot)$ actions available.

3 An Experimental Approach

The authors of the model state in [3] that some of the assumptions are not realistic and prevent the model to be applied in real situations. A 3D simulator is used to evaluate the impact of the violation of some assumptions. Here we try to provide insights on the causes of the deviations from the expected behaviour.

3.1 Discrete Perception of Probabilities

The hypothesis of perfect knowledge of the patrolling strategy by the intruder will be relaxed and modeled in a more realistic manner. We will now consider an intruder which does not know precisely the strategy of the patroller, but only knows what it observes turn after turn, in a discrete way. Suppose the intruder has an *observation matrix* O with dimensions $n \times n$. O_{ij} is the number of times that, in the past, the patroller has moved to c_j from c_i. This number can be expressed as a probability: $\hat{\alpha}_{ij} = O_{ij} / \sum_{j=1}^{n} O_{ij}$, where $\hat{}$ indicates that it is a discrete estimation of the true probability α_{ij}. Notice that $\sum_{j=1}^{n} O_{ij}$ is the number of times that the patroller has passed over c_i. This means that the probabilites are perceived as a relative frequency over the number of observations made, so the observed probabilites will approach the true ones only after a long time. Probably, in a real scenario the intruder does not have so much time to observe and learn but it will attack much earlier. As a result, the most interesting part is not the asymptotically stable behaviour predicted by the leader-follower equilibrium, but a *transient*, difficult-to-predict behaviour in early stages of the patrolling situation. This is why empirical simulations are required.

Note that limiting the knowledge of the intruder about the patroller strictly to *what it has observed* has important implications. The most important is that it allows strategic manipulation. At the same time, it causes more complicated strategies for the patroller to be very difficult to perceive. The observation matrix we have introduced above is the simplest way to model the observations, but an agent with this kind of memory is unable to detect complex behaviours, such as non-Markovian, or partially deterministic routes. Such behaviours will be wrongly perceived just as probabilities, which can induce the intruder not to act as a *true* best-responser but just as a best-responser regarding its own perceptions about the patroller. This represents a subtle form of manipulation and has been analysed on a more abstract model in [9, 8].

3.2 Experimental Settings and Results

We have used the map proposed in [2], with its corresponding optimal strategy as given by the authors. Both are reproduced in Fig. 1. We implemented the BGA model in a general-purpose language and simulated a patrolling situation in that map. Three different experiments were conducted. The first one is aimed at determining how important it is for the intruder to follow the prescribed best-response action. The second one shows the empirical perception that the intruder has of the patroller's randomized movement, and how the deviation from the true probabilities affect the best-response the intruder would choose at every moment. Finally, the third one is aimed at evaluating the impact of including occasional deterministic movements in the patroller's strategy to anticipate the intruder's attacks and increase the capture chance.

01 (1,0) $d_{01} = 1$	**02** (1,0) $d_{02} = 1$	**03** (1,0) $d_{03} = 1$	**04** (.8,.4) $d_{04} = 6$	**05** (1,0) $d_{05} = 1$
06 (.7,.5) $d_{06} = 4$		**07** (1,0) $d_{07} = 1$		**08** (1,0) $d_{08} = 1$
09 (1,0) $d_{09} = 1$	**10** (1,0) $d_{10} = 1$	**11** (1,0) $d_{11} = 1$	**12** (.8,.4) $d_{12} = 5$	**13** (1,0) $d_{13} = 1$

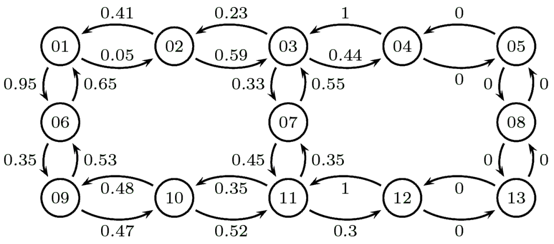


Fig. 1. A simple map used in the experiments and the optimal strategy for the patroller prescribed by the leader-follower equilibrium. Figure reproduced from [2]

Expected Payoff for a Best Responser In the first experiment, we did not use the implementation of the model but only tried to gain a clearer idea on how important it is to follow the strategy prescribed by the leader-follower equilibrium. According to the optimal probabilities of Fig. 1, the expected payoffs attained by the intruder for every *enter-when*(·,·) action are shown in Table 1. Only the actions whose expected payoff is greater than 0 are listed. The *success probability*, or probability of not being detected, is given for informational purposes. For a given action *enter-when*(q, s), such probability is the value $\alpha = \sum_{i \in C \setminus q} \gamma_{s,i}^{d_q,i}$ mentioned in [2], which stands for the sum of the probabilities of all the paths that the patroller can follow in d_q turns, starting in cell s and not passing through cell q. We can see the payoff for that action as a random variable that takes value Y_0 with probability (1 - α) and Y_q with probability α. The variance of such variable can be very informative so it has been added to the table. An enumeration like Table 1 is exactly what a best-responser is expected to do to make a decision. With those results, the intruder would choose *enter-when*(6, 12). It gives him the highest expected payoff, and it is preferable to *enter-when*(6,11) because, despite having the same expected payoff, cell 11 is nearer target cell 6 than cell 12. However, notice both actions have a high variance in relation to the payoff. On the contrary, actions *enter-when*(4,1), *enter-when*(4,6) and *enter-when*(4,9) yield a slightly smaller payoff but their probability of success is 6 % higher and their variance is smaller. In simple words, what we see is that a cell with a high payoff is worth running a risk to rob it because of the expected payoff, although such expectation has a high variance (risk) associated. However, if two cells are very similar in payoff, maybe

Table 1. Intruder's expected payoff together with the corresponding variance and the probability of not being detected

Intruder's action	Expected payoff	Sucess probability	Payoff variance
enter-when(12,10)	0.0582	0.7558	0.3617
enter-when(12,7)	0.1125	0.7947	0.3198
enter-when(4,11)	0.1680	0.8343	0.2710
enter-when(4,12)	0.1680	0.8343	0.2710
enter-when(12,9)	0.2203	0.8716	0.2193
enter-when(4,10)	0.2624	0.9017	0.1737
enter-when(12,3)	0.2692	0.9066	0.1660
enter-when(12,4)	0.3376	0.9555	0.0834
enter-when(12,2)	0.3632	0.9737	0.0502
enter-when(12,1)	0.3640	0.9743	0.0491
enter-when(12,6)	0.3641	0.9743	0.0490
enter-when(4,1)	0.3645	0.9746	0.0485
enter-when(4,6)	0.3646	0.9747	0.0483
enter-when(4,9)	0.3648	0.9748	0.0481
enter-when(6,3)	0.3656	0.9104	0.1835
enter-when(6,4)	0.3656	0.9104	0.1835
enter-when(6,7)	0.3660	0.9107	0.1831
enter-when(6,11)	0.3664	0.9110	0.1825
enter-when(6,12)	0.3664	0.9110	0.1825

the one with a slightly smaller payoff but smaller variance and higher success probability is a better choice because the risk of being detected is smaller. This is another practical issue an intruder could consider but is not taken into account by strict game-theoretic models.

Intruder's Deviation from the Predicted Action The second experiment required running simulations of the model. The intruder was provided with an observation matrix as explained in the previous section. The patroller moved all the time along the map following the strategy described in Fig. 1, starting at cell 1[1]. The intruder was observing the patroller all the time and recording the observations in the observation matrix. We wanted to check how his observations matched the true probabilities of the patroller, and if the differences changed his decision regarding what we could predict about the intruder as a best-responser. Every 100 turns, the intruder evaluated the expected payoff it could attain for each *enter-when*(q, s) combination, based on the empirical probabilities he had observed up to that turn. The results after two independent runs (in the same conditions and with the same map) are shown in Fig. 2.

Each plot shows two lines. Every change in the action chosen by a best-responser is shown so that every action is at a different height in the graph. The height of an *enter-when* action does not have a meaning nor is it related to its

[1] Since the experiments are long enough, the starting point is not important.

payoff; its purpose is just to clearly show changes in the action preferred by the intruder along the time. The continuous, thick line measures the distance between the observed probability distribution, built with the discrete observations recorded in the intruder's observation matrix, and the true probability distribution used by the patroller, namely the leader-follower equilibrium strategy. There are many ways to measure the distance between two discrete probability distributions; a very simple one is the *maximum norm* between two vectors. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors of R^n. The distance between them, according to the maximum norm, is

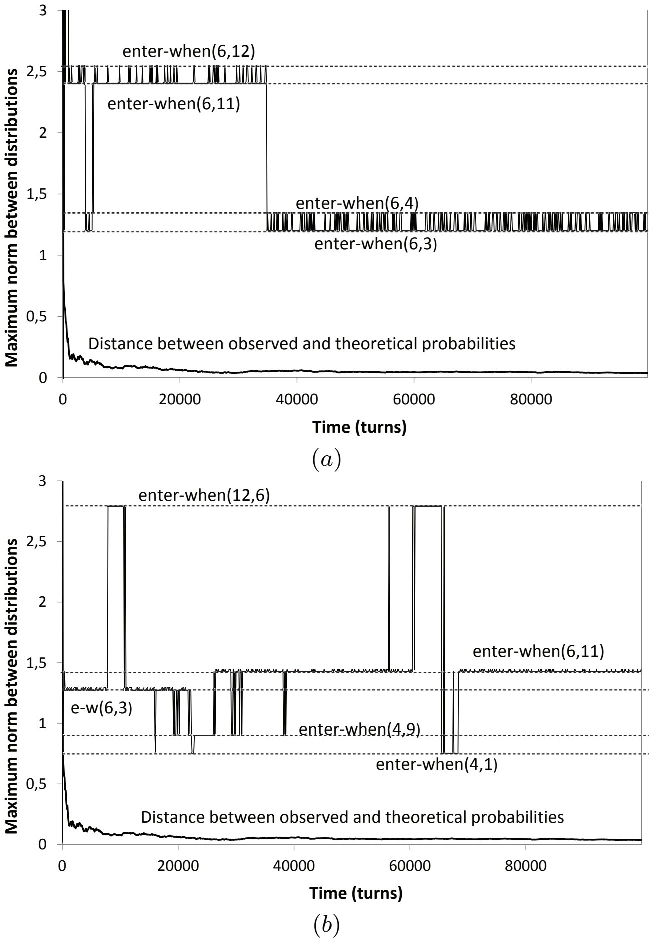
$$||\mathbf{x} - \mathbf{y}||_\infty = \max(|x_1 - y_1|, ..., |x_n - y_n|) \tag{1}$$

This distance gives an idea of how well the relative frequency of the observations matches the true probabilities used by the patroller during the simulation. As could be expected, it becomes smaller as the simulation goes on, because a lot of samples can approximate a probability distribution better than just a few.

The figures confirm that it is not possible to predict exactly the action of the intruder in a real scenario, because its perceptions do not exactly match the probabilities of the patroller, and the deviation can change the decision. This happens even more if the payoff of two actions is very similar or if the map contains zones with symmetric access paths like in our case. The fact that after a long time the best action for the intruder continues unstabilized like in Fig.2 can be explained by those reasons, and actually the target cell is the same; only the entering moment (the cell where the patroller should be) varies. What is more surprising is that we can still find some changes in the target cell in the long run. The expected payoff between target cells 4 and 6 (see Table 1) is so similar and may lead the intruder to prefer any of them. However, recall that they do have very different success probabilities as can be seen in the table, but the intruder did not take this into account.

Occasional Deterministic Paths As stated in Section 3.1, the limited perception of the patroller's behaviour as a set of probabilities opens the door to the possibility of manipulating the observer with a behaviour that cannot be described in terms of probabilities. This form of manipulation may improve the results of the patroller.

Assume the patroller can take into account his own actions just as the intruder is doing, by recording his own movements in an observation matrix identical to that used by the intruder. That way the patroller is able to know exactly what is being perceived about his strategy for an external observer, and which seems to be the best-response cell at each moment, disregarding the true best-response. It is a way to anticipate the most feasible enter cell for the intruder. But in order for this to be effective, it is necessary to have a mechanism to increase vigilance over the cell that is a best-response at every moment, no matter if this violates the movement probabilities used by the patroller. To achieve that, occasional deterministic paths arriving to such probable target cell have been added to the patroller's normal behaviour. Since they are still perceived as



(a)

(b)

Fig. 2. Intruder's best choice every 100 turns according to the observed probabilities in two independent runs (a) and (b).

samples of the probabilistic movement, they should not be too frequent because otherwise they would influence and change the observed relative frequencies very quickly and as a result, that cell would not be the best response anymore for the intruder.

In order to test this, the following experiment was done. The patroller and the intruder engage in a simulation so that every 100 turns, both update their observation matrices (which are identical) and consequently, both update the current best-response for the intruder, according to the observed probabilities, i.e. the relative frequencies of the patroller's past actions. Now, as the patroller also has this information, it decides to occasionally move straightforward to that best-response cell in order to increase the probability of capturing it. This movement is a deterministic path from cell s to cell q, since the intruder will only

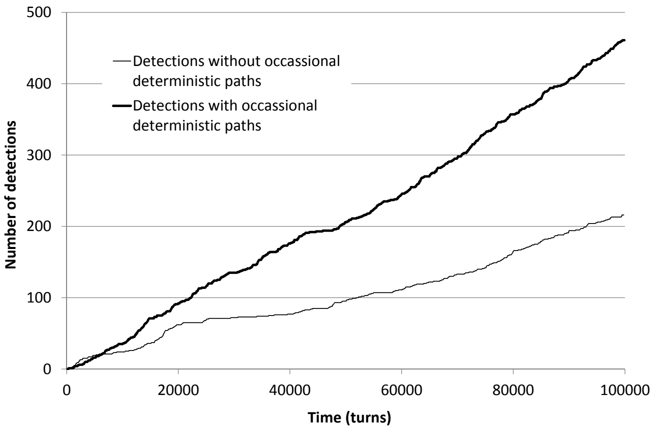


Fig. 3. Accumulated number of times the intruder was detected when trying to rob the cell prescribed by the optimal *enter-when* action according to observed probabilities.

enter cell q the turn after the patroller is in cell s according to the best-response *enter-when*(q,s) action. In our simulation, every time the patroller reaches cell s, it takes a probabilistic decision: it will go straight to cell q with a probability u, following the shortest path through the environment[2]. Once the patroller arrives at cell q, it returns to normal probabilistic movement. In the experiments, u was set to 0.1. Independently, at every turn, the intruder decides probabilistically whether to execute the best-response action *enter-when*(q,s) with probability r, or not to attack. In the experiments, r was set to 0.1. If it decides to attack, it waits until the patroller reaches cell s and then enters the cell. The model prescribes that the game is one-shot so the simulation should end after an attack, be it successful or not. In our experiments, we annotated the result after every attack but did not inform any of the agents of the result of the attack, so for them the simulation continues as if nothing had happened.

Results are depicted in Fig.3 and confirm the hypothesis. Two independent runs were made, one using with occasional deterministic paths with probability $u = 0.1$ and the other without such paths.The attack probability r remained the same in both cases, $r = 0.1$ so approximately the same number of attacks are expected to arise in both simulations. The graph shows the accumulated number of times the patroller detected the intruder during the simulation. As can be seen, the patroller performed better with occasional deterministic paths, and they were not perceived very clearly by the intruder because otherwise the prescribed best-response would have changed the target cell to avoid being captured.

[2] We assume the patroller can compute itself the shortest path or it is provided in the control software

4 Conclusions and Further Work

A patrolling model for topologically-represented environments has been analysed following a strictly empirical approach. Practical issues concerning the application of the model have been addressed, specially the deviation from the expected optimal behaviour. We have studied how to take advantage of such deviations, and have concluded that the limitation on the perception of a movement strategy through discrete observations can be exploited by the patroller with more sophisticated strategies that cannot be described only in terms of probabilities. This has been demonstrated by adding a deterministic component to a randomized strategy, which improves the patroller's performance. These early results are encouraging and pave the way for further research on more complex movement patterns for the patroller, and also for other kinds of manipulation exploiting the limited perception abilities of the intruder.

5 Acknowledgments

This work has been partially funded by projects TIN2008-01948, TIN2008-06872-C04-04, TIN2011-27696-C02-01 from the Spanish Ministry of Science and Innovation and P07-TIC-02970, TIC 08001 from the Andalusian Government. First author acknowledges a FPU scholarship from Spanish Ministry of Education.

References

1. Agmon, N., Kraus, S., Kaminka, G.: Multi-robot perimeter patrol in adversarial settings. In: Proc. of the IEEE Conf. on Robotics and Automation. pp. 2339–2345 (2008)
2. Amigoni, F., Basilico, N., Gatti, N.: Finding the optimal strategies for robotic patrolling with adversaries in topologically-represented environments. In: Proc. of the IEEE Conf. on Robotics and Automation. pp. 819–824 (2009)
3. Amigoni, F., Basilico, N., Gatti, N., Saporiti, A., Troiani, S.: Moving Game Theoretical Patrolling Strategies from Theory to Practice: an USARSim Simulation. In: Proc. of the IEEE Conf. on Robotics and Automation. pp. 426–431 (2010)
4. Basilico, N., Gatti, N., Rossi, T.: Capturing augmented sensing capabilities and intrusion delay in patrolling-intrusion games. In: Proc. of the 5th Int. Conf. on Computational Intelligence and Games. pp. 186–193 (2009)
5. Osborne, M., Rubinstein, A.: A Course in Game Theory. MIT Press, Cambridge, MA (1994)
6. Paruchuri, P., Pearce, J., Tambe, M., Ordonez, F., Kraus, S.: An efficient heuristic approach for security against multiple adversaries. In: Proc. of the 6th Int. Conf. on Autonomous Agents and Multiagent Systems. pp. 311–318 (2007)
7. Pelta, D., Yager, R.: On the conflict between inducing confusion and attaining payoff in adversarial decision making. Information Sciences 179, 33–40 (2009)
8. Villacorta, P.J., Pelta, D.A.: Theoretical analysis of expected payoff in an adversarial domain. Information Sciences 186(1), 93–104 (2012)
9. Villacorta, P., Pelta, D.: Expected payoff analysis of dynamic mixed strategies in an adversarial domain. In: Proc. of the IEEE Symposium on Intelligent Agents (IA 2011). IEEE Symposium Series on Computational Intelligence. pp. 116 – 122 (2011)