

Relaxation of the Optimal Search Path Problem with the Cop and Robber Game

Frédéric Simard *, Michael Morin **,
Claude-Guy Quimper, François Laviolette, and Josée Desharnais

Department of Computer Science and Software Engineering
Université Laval, Québec Qc G1V 0A6, Canada,
{frederic.simard.10,Michael.Morin.3}@ulaval.ca,
{Claude-Guy.Quimper,Francois.Laviolette,Josee.Desharnais}@ift.ulaval.ca

Abstract. In the Optimal Search Path problem from search theory, the objective is to find a finite length searcher's path that maximizes the probability of detecting a lost wanderer on a graph. We introduce a novel bound on the probability of finding the wanderer in the remaining search time and discuss how this bound is derived from a relaxation of the problem into a game of cop and robber from graph theory. We demonstrate the efficiency of this bound on a constraint programming model of the problem.

Keywords: Optimal Search Path, Cop and Robber, constraint relaxation, pursuit games

1 Introduction

The Optimal Search Path (OSP) problem [1–3] is a classical problem from search theory [4]. The goal is to find the best finite length path a searcher has to take in order to maximize his probability of finding a lost wanderer moving randomly on a graph. SAROPS [5] and the SARPlan [6] systems are two examples of search theory applications. We introduce a novel bound on the objective function of the problem, described in Section 2. This bound is obtained by relaxing the OSP into a game of cop and robber [7–10], which pertains to a well-studied game from graph theory (see Section 3). In Section 4, we show the benefits of using this bound by providing experimental results obtained on an existing constraint programming (CP) model of the problem [11]. We conclude in Section 5.

2 Optimal Search Path

The searcher and the wanderer are moving on the vertices $V(G)$ of a graph G , each of their moves, that we can assume simultaneous, being performed along one of the edges $E(G)$ of the graph. There is a maximal number T of steps

* MSc student, main author. Co-supervised by F. Laviolette and J. Desharnais.

** PhD candidate, co-author. Co-supervised by C.G. Quimper and I. Abi-Zeid.

taken. The wanderer \mathcal{W} is invisible, but each time he shares a vertex v with the searcher \mathcal{S} , the latter has a probability ($\text{pod}(v) \leq 1$) of detecting and removing him from the graph. Let $\pi_t = v_1, v_2, \dots, v_t$ be a path of length $t \leq T$ representing the t first moves of the searcher ($v_i \in V(G)$, $i = 1, \dots, t$). Following Morin et al. [11], we have:

- $\text{pod}(v)$:= probability of detecting \mathcal{W} when both \mathcal{W} and \mathcal{S} are on vertex v ,
- $\text{poc}_{\pi_t}(v)$:= probability¹ of \mathcal{W} to be on vertex v at time t and not have been detected yet given that \mathcal{S} 's first t steps were along π_t ,
- COS_{π_t} := Cumulative probability Of Success of detecting \mathcal{W} up to time t when \mathcal{S} follows π_t . I.e., $\sum_{i=1}^t \text{poc}_{v_1, \dots, v_i}(v_i) \text{pod}(v_i)$,
- $\text{COS}_{\pi_t}^*$:= maximal cumulative probability of success up to T if \mathcal{S} 's t first steps are along π_t , that is, $\max\{\text{COS}_{\pi_T} \mid \pi_T \text{ has prefix } \pi_t\}$.

The goal of the searcher is to find π_T with maximal COS_{π_T} , or, equivalently, $\text{COS}_{\pi_0}^*$, where π_0 is the empty path. Note that our definition of poc_{π_t} implies that if the wanderer is detected at time $i < t$, then $\text{poc}_{\pi_t}(v) = 0$ for all v , meaning that the wanderer has been removed from the graph.

3 A Bound on the OSP

The cop and robber game on a graph consists in finding a winning strategy for a cop to catch a robber, considering perfect information for both players. We focus on a recent variant where the robber is random (or drunk) [12–14]. We show how a valid upper bound on the objective function of the OSP can be derived by allowing the searcher (cop) to see the wanderer (robber).

3.1 A Game of Cop and Drunk Robber

In this game, a cop and a robber move in turn on a graph G , the cop moving first. In contrast with the OSP, the cop has probability 1 of catching the robber when sharing the same vertex, and he *sees* the robber, who, as does the wanderer in the previous game, walks randomly according to a stochastic transition matrix M . The value $M(r, r')$ is the probability that the robber moves to r' provided he is in r at the beginning of the turn. It is positive only if $[r, r'] \in E(G)$. The game ends whenever the cop and the robber share the same vertex. The following definition generalizes Nowakowski and Winkler's [7] relational characterization for the classic cop and robber game to the stochastic one.

Definition 1. *Given r, c the respective positions of the robber and the cop in G , M the robber's random walk matrix and provided the cop moves first, we define:*

$$w_0^M(r, c) := 1 \text{ if } r = c; \text{ otherwise, it is } 0;$$

$$w_n^M(r, c) := \begin{cases} 1 & \text{if } c \in N[r], n \geq 1; \\ \max_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c') & \text{if } c \notin N[r], n \geq 1. \end{cases}$$

where $N[z] := \{z\} \cup \{v \in V(G) \mid [v, z] \in E(G)\}$ is the closed neighbourhood of z .

¹ $\text{poc}_t(v)$ also stands for the probability of containment [11].

The following shows the correctness of the previous definition.

Proposition 1. *If M governs the robber's random walk on G , then $w_n^M(r, c)$ is the probability that a cop in c captures the drunk robber in r in n steps or less.*

Proof. By induction. The base case, with $n = 0$, is clear. Suppose the proposition holds for $n - 1 \geq 0$ and let us now prove it for n . If $c \in N[r]$, then $w_n^M(r, c) = 1$ and the result follows because the cop indeed catches the robber. If $c \notin N[r]$, then let the cop move to some vertex c' . The position of the robber at the end of the round is r' with probability $M(r, r')$. The probability that the cop catches the robber depends on this last one's next move and on w_{n-1} following the expression $\sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c')$. Hence, the best possible move for the cop is $\operatorname{argmax}_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c')$. The wanted probability is thus $w_n^M(r, c)$. \square

Proposition 1 could provide a trivial upper bound on the probability of finding the wanderer, but a tighter one is presented in the following section.

3.2 Markov Decision Process (MDP) and the OSP Bound

In Section 3.1 we presented a game where the robber is visible, a setting that we use as a bound on the OSP's objective function. A MDP is a stochastic process in which an agent has to take the best possible decisions in an environment evolving stochastically [15–17]. At each time step, the agent receives a local reward; he tries to maximize his expected total reward within a time bound T . We first formulate the game of cop and robber defined above as an MDP and then deduce a valid OSP bound. The following definition presents the said MDP; its solution encodes the optimal strategy for the cop. Pralat and Kehagias [13] also formulated the game of cop and visible drunk robber as an MDP but in a different way.

Definition 2 (MDP Cop and Drunk Defending Robber). *Let G be a graph, M a stochastic matrix on $V(G)$ and T the maximal number of time steps. We define an MDP $\mathcal{M} = (S, A, P, R)$ as follows.*

$$\begin{aligned} S &:= (V(G) \cup \{\text{jail}\}) \times V(G) \times \{1, 2, \dots, T + 1\}. \\ A &:= V(G). \end{aligned}$$

$$P\left[(r', c', t') | (r, c, t), a\right] := 0 \text{ whenever } a \neq c' \text{ or } c' \notin N[c] \text{ or } t' \neq t + 1$$

otherwise, P is defined as:

$$P\left[(r', c', t + 1) | (r, c, t), c'\right] := \begin{cases} 1 & \text{if } r = r' = \text{jail}; \\ \operatorname{pod}(r) & \text{if } r = c', r' = \text{jail}; \\ (1 - \operatorname{pod}(r))M(r, r') & \text{if } r = c', r' \neq \text{jail}; \\ M(r, r') & \text{if } r \notin \{c, c', \text{jail}\}. \end{cases} \quad (1)$$

$$R((r', c', t') | (r, c, t), a) := \begin{cases} 1 & \text{if } r' = \text{jail} \neq r, t \leq T; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The game is initialized as follows: the cop chooses his initial position c on the graph, and then the initial position r of the robber is picked at random according to the probability distribution poc_0 , which results in an initial state $(r, c, 1)$ for the MDP. A turn starts with a cop move. If the cop transits to the robber state ($r = c'$), then there is probability $\text{pod}(r)$ that he catches the robber, which results in the robber going to jail ($r' = \text{jail}$) and staying there for the rest of the game. The cop then receives a reward of 1. If the catch fails ($r' \neq \text{jail}$, with probability $1 - \text{pod}(r)$) or if the cop did not transit to the robber state ($r \neq c'$), the robber is still free to roam, following M . Note that the state transition probabilities (1) are non-null only when valid moves are considered (time goes up by one and $a = c' \in N[c]$). Note also that, when the robber is in jail, no more reward is given to the cop. In the MDP \mathcal{M} , the cop's goal is to find a strategy, also called policy, to maximize his expected reward, that is, the probability to capture the robber before a total of T steps is reached. A strategy in \mathcal{M} consists in first choosing an initial position (for the cop) and then in following a function $f : S \rightarrow A$ that, given the current state, tells the cop which action to take, that is, which state to transit to.

Because of \mathcal{M} 's Markov property, whenever a strategy is fixed, one can compute the *value* $u(r, c, t)$ of each state (r, c, t) of the MDP, that is, the expected reward the cop can obtain from that state when following the strategy. The optimal strategy is therefore the one that gives the highest value $u^*(r, c, t)$ on all states. The cop's optimal strategy consists in moving to the robber's position if possible, and then trying to capture him. If the robber is not positioned on one of the cop's neighbours, the cop moves to the position that maximizes his probability of capture in the remaining time allowed. According to Proposition 1, the value of this optimal strategy is:

$$u^*(r, c, t) = \begin{cases} \max_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') u^*(r', c', t + 1) & \text{if } r \notin N[c]; \\ 1 - (1 - \text{pod}(r))^{T+1-t} & \text{if } r \in N[c]. \end{cases} \quad (3)$$

If $r \notin N[c]$ the cop, who moves first, must choose a next state that will result in the best probability of eventual capture, given the robber's present position, and knowing that the robber's next move is governed by M . If $r \in N[c]$, the cop tries to catch the robber with probability of success $\text{pod}(r)$; if he fails, the robber will transit to one of his neighbours, and hence the cop can keep on trying to catch the robber until success or until the maximal time has been reached. The formula follows from the fact that at the beginning of time step t , the cop has $T + 1 - t$ remaining attempts.

Since the optimal probability of capture in the OSP problem is always lower than the optimal probability of capture in the cop and robber game, we have the following proposition:

Proposition 2. *The probability $\text{COS}_{\pi_0}^*$ of finding the wanderer is always at most that of catching the robber:*

$$\text{COS}_{\pi_0}^* \leq \max_{c \in V(G)} \sum_{r \in V(G)} \text{poc}_{\pi_0}(r) u^*(r, c, 1).$$

Proof. Clearly, $\text{COS}_{\pi_0}^*$ is bounded by the optimal probability of capture of the cop and robber game. In the MDP, the optimal probability of capture is obtained if the cop’s first choice maximizes his probability of capture considering that at that moment the robber is not yet positioned on the graph but will be according to the probability distribution poc_0 . \square

Unfortunately, Proposition 2’s bound is of no use in a branch-and-bound attempt for solving the OSP problem. The next proposition generalizes it appropriately.

Proposition 3. *Let $\pi_t = v_1, v_2, \dots, v_t$. Then*

$$\text{COS}_{\pi_t}^* \leq \text{COS}_{\pi_t} + \max_{c' \in N[v_t]} \sum_{r' \in V(G)} \text{poc}_{\pi_t}(r') u^*(r', c', t + 1). \quad (4)$$

Proof. Similarly as in the preceding proof, $\text{COS}_{\pi_t}^*$ is bounded by the optimal reward obtained when first playing the OSP game along π_t and then (if the wanderer is not yet detected) switching to the cop and robber game: this is done by making the wanderer (robber) visible to the searcher (cop). When starting this second phase (at the $t + 1$ step), the cop must choose his next position to maximize the probability of capture; at this moment, the robber is not yet visible but his next position is governed by poc_{π_t} . If the cop chooses c' , his probability of capture will be $\sum_{r' \in V(G)} \text{poc}_{\pi_t}(r') u^*(r', c', t + 1)$ and the result follows. \square

In the next section, we use this bound jointly with a CP model of the OSP. The solver branches in a static order of the searcher’s positions which are the decision variables (ordered by time steps). The CP variables representing the probabilities, including the variables representing $\text{COS}_{\pi_t}^*$ (i.e., the objective to maximize), are entirely determined by the searcher’s path. When opening a node for a searcher’s position at a time t , we are able to compute, using our novel bound (4), a tighter upper bound on the domain of the objective value variables. We proved that our bound is admissible, that is, it never underestimates the real objective value. Whenever the bound computed on the opened node of the searcher’s position at a time t is lower than the best known lower bound on the objective value (e.g., the objective value of the current incumbent solution) that searcher’s move is proven unpromising. The bound is not limited to CP usage as it may be applied to the branch and bound OSP framework.

4 Experimentation and Results

All the experiments were run on an Intel(R) Core(TM) i7 2.6 GHz with 8 GB of RAM using the Choco 2.1.5 solver [18] along with the Java Universal Network/Graph (JUNG) 2.0.1 framework [19]. We implemented the bound on two

known OSP CP models: a basic model (*max*), and the *max* model with the total detection heuristic implemented as a value selection heuristic (TD) [11]. The TD heuristic is on many account similar to the bound derived in this paper which gives its theoretical justification. We generated six OSP instances² using the graph G^L of the Université Laval’s tunnels [11] with $T = 17$. For each instance, we chose $\text{pod}(r) \in \{0.3, 0.9\}$ (for all vertex r) and generated a transition matrix as follows: (1) the wanderer has a probability ρ of staying in place and (2) the remaining probability mass $1 - \rho$ was uniformly distributed on the neighbouring vertices. We chose $\rho \in \{0.3, 0.6, 0.9\}$. The total allowed CPU time is 5 minutes (or 5 million backtracks).³ From Table 1, we deduce that the bound consistently

Table 1. Results on OSP problem instances without and with a bound (+B)

$\text{pod}(r)$ ρ		Max model						TD model					
		Objective		Time [†]		Backtracks		Objective		Time [†]		Backtracks	
		+B		+B		+B		+B		+B		+B	
G^L with $T = 17$													
0.3	0.3	0.318	0.346	291 164	21649 14548	0.519	0.519	4	3	282	181		
	0.6	0.359	0.392	299 215	21706 18639	0.585	0.586	130 188	10391	16055			
	0.9	0.480	0.520	294 209	21707 17955	0.749	0.749	147 78	12160	6715			
0.9	0.3	0.651	0.683	300 247	22855 21410	0.831	0.838	188 128	15000	11076			
	0.6	0.704	0.740	299 267	22796 23074	0.870	0.881	177 204	14628	17604			
	0.9	0.829	0.858	276 271	20263 23360	0.939	0.945	179 262	15212	22999			

[†] The time, in seconds, to the last incumbent solution.

[‡] Bold values are better.

finds a better (or equal) objective value than when not used. When the objective value stays the same, the number of backtracks used by the bound is lower or equal than with the bound. Equality of objective values occur only when using the TD heuristic, which already leads the solver towards good solutions, explaining why the bound seems less efficient in this case. Therefore, either with the bound a better solution is found, or the same is found faster than without.

5 Conclusion

We developed, in this paper, a bound (4) based on MDPs for the OSP problem. This bound has been proven efficient on an existing CP model for the OSP using a state-of-the-art CP solver. A more general method from MDP theory [20] could be applied to compute tighter bounds than the one presented here. It consists in subtracting a function z_{π_t} , with $\mathbb{E}[z_{\pi_t} | \text{OSP}] \leq 0$, to the right-hand side of (4). This leads to tighter bounds, but determining z_{π_t} is an optimization problem in itself [21]. Such an observation leads to a whole family of OSP bounds.

² The detailed results, including all instances of [11], are available at: <http://graal.ift.ulaval.ca/downloads/osp-and-cr-games/>

³ In all cases, the solver failed to prove the optimality within 5 minutes.

References

1. Trummel, K., Weisinger, J.: The complexity of the optimal searcher path problem. *Operations Research* **34**(2) (1986) 324–327
2. Stewart, T.: Search for a moving target when the searcher motion is restricted. *Computers and Operations Research* **6**(3) (1979) 129–140
3. Washburn, A.R.: Branch and bound methods for a search problem. *Naval Research Logistics* **45**(3) (1998) 243–257
4. Stone, L.: *Theory of Optimal Search*. Academic Press, New York (2004)
5. Netsch, R.: The USCG search and rescue optimal planning system (SAROPS) via the commercial/joint mapping tool kit (c/jmtk). In: *Proceedings of the 24th Annual ESRI User Conference, August. Volume 9.* (2004)
6. Abi-Zeid, I., Frost, J.: A decision support system for canadian search and rescue operations. *European Journal of Operational Research* **162**(3) (2005) 636–653
7. Nowakowski, R., Winkler, P.: Vertex-to-vertex pursuit in a graph. *Discrete Mathematics* **43**(2) (1983) 235–239
8. Quilliot, A.: *Problème de jeux, de point fixe, de connectivité et de représentation sur des graphes, des ensembles ordonnés et des hypergraphes*. PhD thesis, Université de Paris VI (1983)
9. Bonato, A., Nowakowski, R.: The game of cops and robbers on graphs. *American Mathematical Soc.* (2011)
10. Fomin, F.V., Thilikos, D.M.: An annotated bibliography on guaranteed graph searching. *Theoretical Computer Science* **399**(3) (2008) 236–245
11. Morin, M., Papillon, A.P., Abi-Zeid, I., Laviolette, F., Quimper, C.G.: Constraint programming for path planning with uncertainty. In: *Principles and Practice of Constraint Programming.* (2012) 988–1003
12. Kehagias, A., Mitsche, D., Prałat, P.: Cops and invisible robbers: The cost of drunkenness. *Theoretical Computer Science* **481** (2013) 100–120
13. Kehagias, A., Prałat, P.: Some remarks on cops and drunk robbers. *Theoretical Computer Science* **463** (2012) 133–147
14. Komarov, N., Winkler, P.: Cops vs. gambler. *arXiv preprint* (2013)
15. Bäuerle, N., Rieder, U.: *Markov Decision Processes with Applications to Finance*. Springer (2011)
16. Puterman, M.L.: *Markov decision processes: discrete stochastic dynamic programming*. Volume 414. John Wiley & Sons (2009)
17. Barto, A.G., Sutton, R.S.: *Reinforcement learning : An introduction*. MIT press (1998)
18. Laburthe, F., Jussien, N.: *Choco solver documentation*. Ecole de Mines de Nantes (2012)
19. O’Madadhain, J., Fisher, D., Nelson, T., White, S., Boey, Y.B.: The jung (java universal network/graph) framework. (2010)
20. Brown, D.B., Smith, J.E., Sun, P.: Information relaxations and duality in stochastic dynamic programs. *Operations research* **58**(4-part-1) (2010) 785–801
21. Ye, F., Zhou, E.: Parameterized penalties in the dual representation of Markov decision processes. In: *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, IEEE* (2012) 870–876