Optimal Cooperative Pursuit on a Manhattan Grid

Krishnamoorthy Kalyanam*
InfoSciTex corporation, Dayton, OH 45431

Swaroop Darbha†
Texas A&M University, College Station, TX 77843

Pramod Khargonekar‡
University of Florida, Gainesville, FL 32611

Meir Pachter§
Air Force Institute of Technology, Wright-Patterson AFB, OH 45433

Phil Chandler¶
Air Force Research Laboratory, Wright-Patterson AFB, OH 45433.

The optimal control of two pursuers searching for a slower moving evader on a Manhattan grid road network is considered. The pursuers do not have on-board capability to detect the evader and rely instead on Unattended Ground Sensors (UGSs) to locate the evader. We assume that all the intersections in the road network have been instrumented with UGSs. When an evader passes by an UGS location, it triggers the UGS and this time-stamped information is stored by the UGS. When a pursuer arrives at an UGS location, the UGS informs the pursuer if and when the evader passed by. When the evader and a pursuer arrive at an UGS location simultaneously, the UGS is triggered and this information is instantly relayed to the pursuer, thereby enabling "capture". We derive exact values for the minimum time guaranteed capture of the evader on the Manhattan grid and the corresponding pursuit policy.

I. Introduction

In this paper, we are concerned with capturing a moving ground target in minimum time. The operational scenario is as follows. The target moves on a Manhattan style road network (2D grid), with an infinite number of rows and columns. All the intersections (grid points) are instrumented with Unattended Ground Sensors (UGSs). When the target, hereafter referred to as the "evader", passes by an UGS, the UGS is triggered. A triggered UGS turns, say, from "green" to "red" and records the evader's time of passage. We consider two pursuers traveling on the network tasked with capturing the evader. The pursuers are twice as fast of the evader, in that, a pursuer takes one time step to travel between two neighboring grid points whereas the evader takes two. When a pursuer arrives at an UGS location, the information stored by the UGS is uploaded to the pursuer, namely, the "green"/"red" status of the UGS and, if the UGS is "red", the time elapsed (delay, D) since the evader's visit to the same location. The evader can be captured in one of two ways: either the evader and a pursuer synchronously arrive at an UGS location, or a pursuer is already loitering/waiting at an UGS location when the evader arrives there. In both cases, the UGS is triggered, informs the pursuer and the evader is captured. The decision problem for the pursuers is to select which neighboring UGS to visit next, including possibly staying at the current UGS location awaiting the arrival of the evader. The decisions are made at discrete time instants immediately after arriving at and interrogating an UGS.

*Research Scientist, InfoSciTex Corporation, Dayton, OH 45431 and AIAA member, *krishnak@ucla.edu*
†Professor, Department of Mechanical Engineering, Texas A&M University, College Station, TX 77843.
‡Professor, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611.
§Professor, Department of Electrical Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433.
¶Tech Advisor, Control & Automation Branch, Air Force Research Laboratory, Wright-Patterson AFB, OH 45433

In this paper, we follow a "min-max" approach, in that we are interested in the minimum time to capture the evader, under worst possible evader actions. Here, the evader actions are assumed to be unknown (as opposed to being random) but belong to a priori known bounded sets.[1] The worst-case approach to solving decision problems dates back to Whitsenhausen[2] and Bertsekas[3,4] and optimal solutions are difficult to obtain, when complete state information is not available to the controller, as is the case here. In this regard, tractable sub-optimal solution strategies for discrete-time systems with imperfect information is provided by Bertsekas.[5] The pursuit problem can also be interpreted as a sequential zero-sum differential game.[6] In this context, the problem of guaranteed capture can be recast as a reachability problem.[7] From the evader's perspective, the desired "target set" would be the subset of the state-space that corresponds to the evader evading capture. We wish to design a pursuit policy that guarantees that the state stays out of the evader's target set under incomplete information. The ensuing state estimation problem, to determine the evader's current position, is related to the work by Schweppe[8] and Bertsekas.[9] A closely related body of work is the guaranteed graph search problem in pursuit evasion games (see references in Fomin and Thilikos[10]). An alternate framework, where the evader's motion is considered to be random and the searcher maximizes the probability of capture has been considered in the past.[11,12] A distinguishing feature of this work is the information structure, which entails delayed and out of sequence evader state information. The simpler scenario wherein the delay i.e., information lag is a known *constant*, has been solved.[13] In the scenario considered herein, the control has a dual effect[14] in that, in addition to aiding capture, it also affects the uncertainty associated with the determination of the evader's state in the future. In a related work, the authors had developed sufficient conditions for guaranteed capture of the evader, using a single pursuer, on an arbitrary road network with a finite number of grid points.[15] We have also considered the case of a single pursuer on a Manhattan (2D) grid with three rows and an infinite number of columns.[16]

In this paper, we investigate the pursuit evasion on the Manhattan road network shown in Fig. 1, where the pursuit policy is based *solely* on the information gathered from the UGSs, by the pursuers. We show that under certain restrictions on the evader's motion, capture is guaranteed and indeed it takes no more than $D + 4$ time steps to capture the evader, for any $D \geq 1$. Here, D is the initial delay observed by (one of) the pursuers.

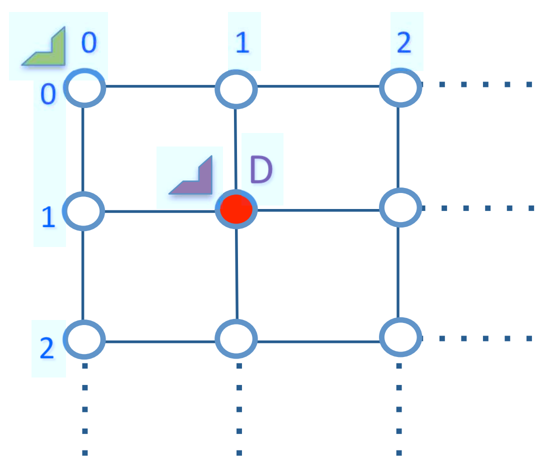

Figure 1. Manhattan Grid Layout with 2 pursuers

II. Pursuit-Evasion on a Manhattan Grid

Consider a Manhattan grid network (see figure 1) consisting of an infinite number of rows and columns. The rows/columns are indexed by the set of non-negative integers, $\mathcal{Z} = \{0, 1, 2, \ldots\}$. We denote the coordinates of a grid point by (i, j), where i indicates the row and j denotes the column index. The grid points are the nodes of the network and the horizontal and vertical connecting roads are the edges. We assume that a pursuer travels from one node to a neighboring node in exactly one time step whereas the slower moving evader takes two time steps to travel between neighboring nodes. At time k, the allowed actions for the evader at a node are indicated by $w(k) \in \{N, E, S\}$, where N, E and S indicate going North, East and South respectively, provided a path exists in that direction. The evader is constrained to be always on the move and also prohibited from visiting any given node, more than once. This assumption precludes cycles in the graph and the ensuing intractability in solving the search problem. The pursuer, in addition to being able to move North (N), East (E) and South (S), can also wait at a node location. So, the allowed pursuer actions at time k are given by $u(k) \in \{N, E, S, L\}$, where L indicates loitering/waiting at the same node location until the next time step. Note that neither the pursuers nor the evader can move west.

II.A. Pursuer-Evader Dynamics

Let the m^{th} pursuer's state i.e., location on grid, at time k, denoted by $p_m(k) = (i_m(k), j_m(k))$, follow the discrete-time dynamics:

$$p_m(k+1) = f_p(p_m(k), u_m(k)),\ u_m(k) \in \mathcal{U},\ m = 1, 2, \tag{1}$$

where the allowed set of pursuer actions, $\mathcal{U} = \{N, S, E, L\}$. Let the evader's state at time k be denoted by $e(k) = (p_e(k), o_e(k))$, where $p_e(k) = (i_e(k), j_e(k))$ is its position on the grid and $o_e(k) \in \{N, E, S\}$ is its current orientation/heading. Let the evader's state follow the discrete-time dynamics:

$$e(k+2) = f_e(e(k), w(k)),\ w(k) \in \mathcal{W}(o_e(k)), \tag{2}$$

where $\mathcal{W}(o_e(k))$ is the set of allowed evader actions from the state $e(k)$. We retain the evader's orientation, in addition to its location, so as to enforce additional motion constraints, such as preventing the evader from making a U-turn, and so, we have:

$$\mathcal{W}(N) = \{N, E\},\ \mathcal{W}(S) = \{S, E\},\ \text{and}\ \mathcal{W}(E) = \{N, S, E\}.$$

Also, we increment time by 2 in Eq. (2), so as to indirectly impose the $2\times$ speed advantage of the pursuer.

Recall that the pursuer, upon arriving at a grid point, interrogates the UGS at that location and makes one of two observations: status "red" with delay $D \geq 0$ indicating that the evader was at that location D time steps ago or status "green" indicating that the location has not been visited by the evader. So, we denote the observation made by the m^{th} pursuer at time k by,

$$z_m(k) = \begin{cases} \ell, \ \text{if}\ p_e(k-\ell) = p_m(k),\ \text{for some}\ \ell \in [0, k], \\ G, \ \text{otherwise}. \end{cases} \tag{3}$$

Here G indicates "green" status of the UGS at $p_m(k)$.

III. Optimization Problem Statement

Suppose the evader is at the location $(1, 1)$ at time 0. After some delay $D > 0$, the 1^{st} pursuer (colored purple) arrives (for the first time) at the same location $(1, 1)$, knowing that the evader entered $(1, 1)$ from outside the grid- see figure 1. Further suppose that the second pursuer (colored light green) is at location $(0, 0)$, and the two are tasked with cooperatively capturing the evader. We shall use $\mathcal{C}_D$ to denote this initial condition. There is no loss of generality here, since the pursuer who discovers the evader for the first time (say pursuer 1) can always wait at the same location, until the other pursuer (say pursuer 2) arrives at the location immediately above and to the left of pursuer 1. Since the pursuers have no direct knowledge of the evader's position, except the information they gathered from the UGSs, we define the information state at time $k \geq D$ by: $\mathcal{I}(k) = (x_p(D), \ldots, x_p(k))$, where $x_p(k) = (p_1(k), z_1(k), p_2(k), z_2(k))$ i.e., the locations visited thus far, and the observations made therein. Let $\mathcal{A}$ denote the set of all admissible control polices, where any $A \in \mathcal{A}$ is a (vector) sequence of maps from the information state $\mathcal{I}(k)$ to the available control actions, $\mathcal{U} \times \mathcal{U}$. Let $J(D, A, w(2k); k \geq 0)$ be the number of steps to capture, when the initial configuration is $\mathcal{C}_D$ and the pursuers' execute the control policy, $A : [u_1(k),\ u_2(k)] = A(\mathcal{I}(k)); k \geq D$, thereafter, while the evader implements the actions, $w(2k); k \geq 0$. If capture never occurs, we set J to ∞.

III.A. Definition:

For any $A \in \mathcal{A}$:

$$T_A(D) = \sup_{\substack{w(2k)\in\mathcal{W}(o_e(2k))\\k=0,1,\ldots}} J(D, A, w(2k); k \geq 0).$$

III.B. Definition:

The worst-case minimum time to capture is then given by:

$$\begin{aligned} T(D) &= \min_{A\in\mathcal{A}} \sup_{\substack{w(2k)\in\mathcal{W}(o_e(2k))\\k=0,1,\ldots}} J(D, A, w(2k); k \geq 0) \\ &= \min_{A\in\mathcal{A}} T_A(D). \end{aligned}$$

For any policy $A \in \mathcal{A}$, we also define A_D to be its restriction to the initial delay: D. To be clear, the restriction A_D specifies (completely) what the pursuers should do starting from the configuration $\mathcal{C}_D$ and thereafter.

IV. Dynamic Programming solution

Suppose at time $k > D$, the evader is yet to be captured and we are interested in obtaining the min-max steps to capture thereafter. Let $V^*(\mathcal{I}(k))$ denote the minimum time (or number of time steps) for guaranteed interception of the evader, with the information $\mathcal{I}(k)$. We note that the information state evolves according to:

$$\mathcal{I}(k+1) = (\mathcal{I}(k), f_p(p_1(k), u_1(k)), z_1(k+1), f_p(p_2(k), u_2(k)), z_2(k+1)), \tag{4}$$

which shows the explicit dependence on the future course of action of, and the likely observations made by, the pursuers. Clearly, it will take at least one more time step to capture the evader. So, one can write the following Dynamic Programming (DP) recursion to obtain $V^*(\mathcal{I}(k))$:

$$V^*(\mathcal{I}(k)) = 1 + \min_{\mathbf{u}(k)} \max_{\mathbf{z}(k+1)} V^*(\mathcal{I}(k+1)), \tag{5}$$

where $\mathbf{z}(k+1) = (z_1(k+1), z_2(k+1)) \in \mathbf{Z}(k+1)$. Here, $\mathbf{Z}(k+1)$ denotes the set of all possible observations at $k+1$ that are consistent with a given $\mathcal{I}(k)$ and control actions, $\mathbf{u}(k) = (u_1(k), u_2(k))$. The recursion terminates when the evader is captured, i.e., $V^*(\mathcal{I}(t)) = 0$, if $z_1(t) = 0$ or $z_2(t) = 0$. By definition (III.B), we have $T(D) = V^*(\mathcal{I}(D))$ and so, one can obtain $T(D)$ and the corresponding optimal pursuit policy $A^* \in \mathcal{A}$ by solving the recursion Eq. (5). It becomes readily apparent that the DP method quickly becomes intractable for large delays D, unless some structure in the problem is exploited. This is primarily due to the combinatorial explosion in the number of pursuer and evader actions and the partial and delayed information structure, i.e., the information state grows unbounded with time. Luckily, the problem exhibits structure characterized by the translational and reflection symmetries in the Manhattan grid and motion constraints imposed on the pursuer and evader. The structure enables us to *indirectly* solve the DP recursion in an efficient manner. We show that $T(D) < \infty$ and derive an induction based solution method that computes $T(D)$ and the corresponding pursuit policy for any $D \geq 1$.

V. Optimal Solution to Worst Case Capture Problem

In this section, we shall first show that on the Manhattan grid depicted in figure 1, the minimum time to capture under worst-case evader action, is bounded. We shall do so by demonstrating that there exists a policy $\mu \in \mathcal{A}$ such that:

$$T_\mu(D) = D + 4, \forall\ D \geq 1. \tag{6}$$

Since the optimal policy can do no worse than the policy μ, we have $T(D) \leq T_\mu(D), D > 0$.

V.A. Definition:

For any given evader position, $p_e = (i_e, j_e)$ and pursuers' position, $p_m = (i_m, j_m),\ m = 1, 2$ on the grid, we define the "Manhattan distance" between the m^{th} pursuer and the evader to be:

$$\mathcal{M}_m(p_e, p_m) = |i_e - i_m| + |j_e - j_m|.$$

Suppose the evader's position is known to the pursuers at all time and also that the pursuers can move in all four directions. Given the $2\times$ speed advantage of the pursuers, it is clear that the minimum steps to capture, starting from the evader position, p_e and pursuer positions, $p_m,\ m = 1, 2$, is given by $2 \times \min_{m=1,2} \mathcal{M}_m(p_e, p_m)$. This is so because whichever pursuer is closer to the evader can capture it in the aforementioned number of steps under full information. Let $\mathcal{O}(D)$ be the set of all evader positions' at time D that are consistent with the initial condition, $\mathcal{C}_D$ (see figure 1). For example, $\mathcal{O}(2) = \{(0, 1), (1, 2), (2, 1)\}$. With the above distance metric, we have the following (non-trivial) lower bound on the optimal steps to capture.

Lemma 1

$$T(D) \geq 2 \min_{m=1,2} \mathcal{M}_m(p_e, p_m),\ \forall p_e \in \mathcal{O}(D), \tag{7}$$

where $p_1 = (1, 1)$ *and* $p_2 = (0, 0)$.

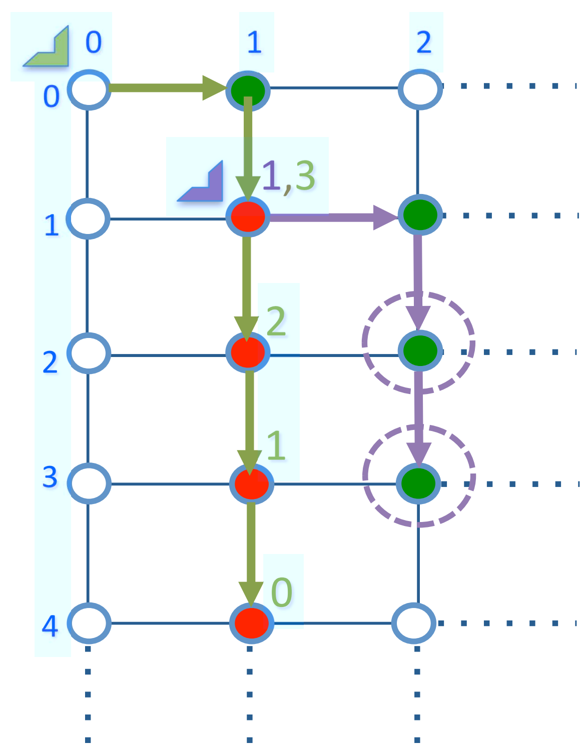

Figure 2. Initial condition, $\mathcal{C}_1$: Pursuer 1 at $(1, 1)$ observes delay 1 and pursuer 2 is at $(0, 0)$, Evader strategy: S^3

Before establishing the policy, μ that satisfies Eq. (6), we introduce the following notation for pursuer/intruder trajectories. We shall represent a finite sequence of pursuer/ evader actions by a string, e.g., SN^2E represents the sequence of moves: South, followed by going North twice, followed by going East.

Lemma 2 *There exists a pursuit policy μ, such that $T_\mu(1) = 5$.*

Proof. This is the scenario $\mathcal{C}_1$, where the 1^{st} pursuer arrives at a red UGS at location $(1, 1)$ with delay $D = 1$ (and 2^{nd} pursuer is at location $(0, 0)$ - see figure 2). The restriction of the policy μ to this case, referred to as μ_1, is specified as follows: the 1^{st} pursuer executes the sequence of actions $ESLSL$, in that order, and the 2^{nd} pursuer executes the sequence of actions ES^4, in that order, until the evader is captured. Table 1 lists the time to capture, the capture location (and by whom), for all possible evader strategies, when μ_1 is executed. A worst-case sequence of evader moves is given by $W_1 = S^3$ and the corresponding

Table 1. $\mathcal{C}_1$: Steps to capture for different evader strategies

Evader Strategy	Steps to capture	$p_e(k)$	Pursuer ID
N	1	$(0,1)$	2
E	1	$(1,2)$	1
SE	3	$(2,2)$	1
S^2E	5	$(3,2)$	1
S^3	5	$(4,1)$	2

steps to capture as per table 1 is 5. So, we have $T_\mu(1) = 5$. The corresponding trajectory for both pursuers, along with the UGS statuses (green or red with delay) encountered by them, are shown in figure 2. Here, the "red" observations made by the pursuers are color coded as per the pursuer 's color.

To illustrate the policy μ, we enumerated the restriction μ_1 for the initial condition $\mathcal{C}_1$. We now provide the policy specification that satisfies Eq. (6) for the initial condition $\mathcal{C}_D, \forall D > 0$. We specify the actions to be taken by both pursuers' per policy μ in a time-iterative fashion. At the initial time, $k = D$, both pursuers' go East, i.e., $u_m(D) = E,\ m = 1, 2$. For all future time, $k > D$, the pursuit policy is given by:
1) If both pursuers had moved East in the last time step, i.e., $u_m(k-1) = E,\ m = 1, 2,$

$$u_m(k) = \begin{cases} E,\ z_1(k) > 0, \\ N,\ z_1(k) = G \text{ and } z_2(k) = z_1(k-1) - 1, \\ S,\ \text{otherwise},\ m = 1, 2. \end{cases} \tag{8}$$

2) If both pursuers had moved North in the last time step, i.e., $u_m(k-1) = N,\ m = 1, 2,$

$$u_m(k) = \begin{cases} E,\ z_1(k) > 0, \\ N,\ \text{otherwise},\ m = 1, 2. \end{cases} \tag{9}$$

3-a) If both pursuers had moved South in the last time step, i.e., $u_m(k-1) = S,\ m = 1, 2,$

$$u_m(k) = \begin{cases} E,\ z_1(k) > 0, \\ S,\ z_1(k) = G \text{ and } z_2(k) > 3,\ m = 1, 2. \end{cases} \tag{10}$$

3-b) If both pursuers had moved South in the last time step, $u_m(k-1) = S,\ m = 1, 2$, and the observations, $z_1(k) = G$ and $z_2(k) = 3$, the pursuers' execute the end-game strategy:

$$u_2(k+j) = S,\ j = 0, \ldots, 2, \text{ and } u_1(k) = L, u_1(k+1) = S \text{ and } u_1(k+2) = L. \tag{11}$$

The pursuers' follow the above policy prescription until the evader is captured. This completes the specification of the policy μ. Note that the above specification, when applied to $\mathcal{C}_1$, collapses to the restriction, μ_1, specified earlier. Indeed, the end-game (**3-b**) is the same play employed by the pursuers, if the evader is still at large 2 time steps after the initial condition, $\mathcal{C}_1$. At that time, pursuer 2 observes a red UGS with delay 3 and pursuer 1 observes a green UGS after (both) having moved south in the earlier time step. Thereafter they follow the reminder of the policy restriction μ_1 i.e., pursuer 1 executes the sequence LSL and pursuer 2 executes the sequence S^3, until the evader is captured. Notice that under the policy μ, the pursuers maintain their relative positions, since they both make the same move, until the end-game maneuver is employed. We shall now demonstrate that if the pursuit policy μ is employed, the evader will be captured in no more than $D + 4$ steps for any the initial condition: $\mathcal{C}_D,\ D > 0$.

Theorem 1 $T_\mu(D) = D + 4,\ D \geq 1.$

Proof. We shall prove the result via mathematical induction.
Basic step: The result holds for $D = 1$, since $T_\mu(1) = 5$ from Lemma 2.

Induction step: Assume that for some $D \geq 2$,

$$T_\mu(M) = M + 4,\ M = 1, \ldots, (D-1). \tag{12}$$

We proceed to show that $T_\mu(D) = D + 4$. Recall that $\mathcal{C}_D$ is the initial condition, where the 1^{st} pursuer is at location $(1,1)$ and observes delay, $D > 1$ and the 2^{nd} pursuer is at location $(0,0)$ - see figure 1. To illustrate the result, we shall consider a few example evader strategies and show that capture occurs in no more than $D + 4$ steps, when pursuit policy μ is employed.

1. Suppose the evader follows the sequence of actions: $E * W_{D-1}$, i.e, it goes east first, followed by any allowed sequence of actions indicated by W_{D-1}. In this case, under policy μ, both pursuers move east first and so, the 1^{st} pursuer will observe delay $D-1$ and the 2^{nd} pursuer will observer a green UGS. But this situation is identical to $\mathcal{C}_{D-1}$. So, using the induction hypothesis, we have the worst-case steps to capture given by,
$$1 + T_\mu(D-1) = 1 + (D-1) + 4 = D + 4.$$
2. Suppose the evader follows the sequence of actions: $N^{D-i}E * W_i$, for some $i = 2, \ldots, D-1$. As per policy μ, both the pursuers will execute EN^{D-i} and the 1^{st} pursuer will observe a red UGS with delay $i-1$ at time $2D+1-i$. Again, using the induction hypothesis, we have the worst-case steps to capture given by,
$$D + 1 - i + T_\mu(i-1) = D + 1 - i + (i-1) + 4 = D + 4.$$
3. Suppose the evader follows the sequence of actions: $N^{D-1}E$. As per policy μ, both the pursuers will execute EN^{D-1} and the 1^{st} pursuer will capture the evader at time $2D$. So, the number of steps to capture equals D.
4. Suppose the evader follows the sequence of actions: $S^{D-j}E * W_j$, for some $j = 2, \ldots, D-1$. As per policy μ, both the pursuers will execute ES^{D-j} and the 1^{st} pursuer will observe a red UGS with delay $j-1$ at time $2D+1-j$. Again, using the induction hypothesis, we have the worst-case steps to capture given by,
$$D + 1 - j + T_\mu(j-1) = D + 1 - j + (j-1) + 4 = D + 4.$$

In a similar fashion, one can show that regardless of the evader's strategy, the pursuers while employing μ will capture the evader in no more than $D+4$ steps. Indeed, table 2 enumerates all possible evader strategies and the corresponding steps to capture (and by whom) under policy μ. It is clear that the evader can do no

Evader Strategy	Steps to capture	Pursuer ID
$E * W_{D-1}$	$D+4$	1 or 2
$N^{D-1} * W_1$	D	1 or 2
$N^{D-i}E * W_{i-1},\ i = 2, \ldots, D-1$	$D+4$	1 or 2
$S^{D-j}E * W_{j-1},\ j = 2, \ldots, D-1$	$D+4$	1 or 2
$S^{D-1}E$	D	1
$S^{D}E$	$D+2$	1
$S^{D+1} * W_1$	$D+4$	1 or 2

Table 2. Number of steps/moves to capture the evader for different evader strategies.

better than $D + 4$ steps and so, $T_\mu(D) = D + 4$. Since the optimal policy can do no worse than the policy μ, we have from Theorem 1:

$$T(D) \leq D + 4, \forall\, D > 0. \tag{13}$$

We shall now proceed to show that μ *is* a optimal pursuit policy. In order to prove the optimality of μ, it suffices to show that $T(D) \geq D + 4,\ \forall D > 0$. The basic idea behind the construction of the lower bound is the following:

- If one were to restrict the set of trajectories of the evader in (III.B) by requiring $w(2k) \in \hat{\mathcal{W}}(e(2k)) \subset \mathcal{W}(e(2k))$, one gets a lower bound for $T(D)$. Essentially, this is tantamount to optimizing the pursuer's policy over a restricted set of evader trajectories.

Before establishing the lower bound for delays $D > 1$, we first show that the policy restriction μ_1 is optimal.

Lemma 3 $T(1) = 5$.

Proof. We have:

$$\begin{aligned} T(1) &= \min_{A \in \mathcal{A}} \max_{\substack{w(2k) \in \mathcal{W}(o_e(2k)) \\ k=0,1,\ldots}} J(1, A, w(2k); k \geq 0) \\ &\geq \max_{\substack{w(2k) \in \mathcal{W}(o_e(2k)) \\ k=0,1,\ldots}} \min_{A \in \mathcal{A}} J(1, A, w(2k); k \geq 0), \end{aligned} \tag{14}$$

with the last inequality resulting from weak duality.

Any pursuit policy A will specify, in particular, the first move for the two pursuers. Since the last inequality (14) involves maximization over evader's trajectories for a given policy A, it is always possible for the evader to pick a initial move, $w(0)$, such that at time 2, the Manhattan distance (V.A) between the evader and either pursuer is at least two. For example, if both pursuers were to move east, i.e., $u_m(1) = E,\ m = 1, 2$, the evader can choose $w(0) = S$. In this case, $p_1(2) = (1, 2)$, $p_1(2) = (0, 1)$, $p_e(2) = (2, 1)$ and so, $\mathcal{M}_m(p_e(2), p_m(2)) = 2,\ m = 1, 2$. Instead, if $u_m(1) = S,\ m = 1, 2$, then the evader can again go east and the resulting distance, $\mathcal{M}_m(p_e(2), p_m(2)) = 2,\ m = 1, 2$. As long as the pursuers move, the evader can always pick $w(0)$ such that its Manhattan distance (V.A) from either pursuer is at least 2. If the pursuers choose instead to loiter (stay) at their initial locations, the evader can still pick an initial move $w(0)$, such that its Manhattan distance from either pursuer is at least 2, immediately after they make their initial move(s). Note that for capture to occur, the pursuers will have to necessarily make a move. Hence, the minimum number of moves required to capture the evader after the first move (the pursuers make), as per Lemma 1 is 4. So, we get:

$$\max_{\substack{w(2k) \in \mathcal{W}(o_e(2k)) \\ k=0,1,\ldots}} \min_{A \in \mathcal{A}} J(1, A, w(2k); k \geq 0) \geq 1 + 4 \Rightarrow T(1) \geq 5. \tag{15}$$

From Theorem 1, we have $T(1) \leq T_\mu(1) = 5$. Hence, $T(1) = 5$ and the policy restriction μ_1, as defined earlier, is optimal.

We now define the restriction on the evader's trajectory that enables us to derive the lower bound. Also, we provide an intermediate result on the *first* move made by the pursuers. This result is necessary to establish the lower bound.

V.B. Restriction:

For any $D > 1$, let $T_{res}(D)$ denote the worst-case, minimum time to capture, when the evader's motion is further restricted as follows: the evader's trajectory must contain the segment E^{D-1} starting at the red UGS location at $(1, 1)$, i.e., $w(2k) = E,\ k = 0, 1, \ldots, D-2$.

Lemma 4 *There exists an optimal pursuit policy such that for any initial delay $D > 1$, under the restriction (V.B) and the evader entering the network at location $(1, 1)$ going east, we have, $u_m(D) = E,\ m = 1, 2$.*

Suppose ν is any *optimal* policy (under the restrictions stated above) such that the first move prescribed by the policy is not E for either pursuer. It suffices to construct another policy $\bar{\nu}$ which does no worse, in terms of the time taken to capture the evader, but prescribes the pursuers to go east first.

One may emulate the two policies with two sets of pursuers, with the first set of pursuers following the policy ν and the second set following policy $\bar{\nu}$. One may also assume that the policy ν is known to the second set of pursuers and this knowledge be used in the construction of the policy $\bar{\nu}$. Let $\mathcal{I}^\nu(k)$ and $\mathcal{I}^{\bar{\nu}}(k)$ denote the information states of the first and second set of pursuers at time $k,\ k \geq D$. At time D, they are both initialized to: $\mathcal{I}^\nu(D) = \mathcal{I}^{\bar{\nu}}(D) = \mathcal{I}(D)$. While the information state associated with the two sets of pursuers is the same initially, the evolution of the information state can be different. Construction of the policy $\bar{\nu}$ is equivalent to specifying the actions for every information state, $\mathcal{I}^{\bar{\nu}}(k)$, that the 2^{nd} set of pursuers will likely encounter. Let $\nu_m(\mathcal{I})$ denote the action of the m^{th} pursuer corresponding to information state $\mathcal{I}$ under the policy ν. So, we have the initial actions: $\nu_m(\mathcal{I}(D)) \neq E,\ m = 1, 2$, and $\bar{\nu}_m(\mathcal{I}(D)) = E,\ m = 1, 2,$.

Under the restriction (V.B), the evader makes $D - 1$ moves to the east from its initial position. Clearly, if the evader were to be ever captured, it can only happen in columns numbered D or higher. Hence, at least one of the pursuers following ν must go east. Our strategy for construction of $\bar{\nu}$ is to ensure that each pursuer following $\bar{\nu}$ will be able to meet up with the corresponding pursuer following ν and thereafter mimic its moves.

Under the restriction (V.B), there can only be one location in columns 1 through D (all of them in row 1), where a red UGS will be encountered. Since the policy ν prescribes that $\nu_m(\mathcal{I}(D)) \neq E,\ m = 1, 2$, the pursuers either go north or south or wait at time D. Let the first and second pursuer following ν make $k_1(\geq 1)$ and $k_2(\geq 1)$ moves respectively before going east for the first time, i.e., $\nu_m(I^\nu(D+k)) \neq E,\ k = 0, \ldots, (k_m - 1)$ and $\nu_m(I^\nu(D + k_m)) = E,\ m = 1, 2$. So, both pursuers following ν will move in their respective columns (1 and 0) until time $D + k_1$ and $D + k_2$ respectively. Furthermore, pursuer 1 alone will encounter a red UGS if and when it re-visits row 1 (all other observations by both pursuers being green). Hence, with the knowledge of ν, one can infer the information states $\mathcal{I}^\nu(D+k)$, for all $k = 1, \ldots, \min(k_1, k_2)$ i.e., until one of the pursuers go east. Note that k_1 and k_2 are also specified, once we know ν. Using this information, we construct the policy $\bar{\nu}$ as follows:

$$
\begin{aligned}
\bar{\nu}_m(\mathcal{I}(D)) &:= E, \quad m = 1, 2, \\
\bar{\nu}_m(\mathcal{I}^{\bar{\nu}}(D+k)) &:= \begin{cases} \nu_m(\mathcal{I}^\nu(D+k-1)),\ 1 \leq k \leq k_m, \\ \nu_m(\mathcal{I}^\nu(D+k)),\ k > k_m, \end{cases} \quad m = 1, 2,
\end{aligned}
$$

Following the policy $\bar{\nu}$, after $\max(k_1, k_2)$ moves, both pursuers will be co-located with the corresponding pursuer following the policy ν. The number of steps to capture, following policy ν, is greater than $\min(k_1, k_2)$, since at least one pursuer has to move east, to enable capture. Since ν is optimal, the policy $\bar{\nu}$ also cannot result in capture, in $\min(k_1, k_2)$ steps or less. If following policy $\bar{\nu}$, capture occurs in j steps; $\min(k_1, k_2) < j < \max(k_1, k_2)$, and say $k_1 < k_2$, then the evader must necessarily have been captured by pursuer 1 (regardless of the policy) and since pursuer 1 following $\bar{\nu}$ is co-located with the corresponding pursuer following ν, the time to capture is the same for both policies. If following policy $\bar{\nu}$, capture occurs in j steps; $j \geq \max(k_1, k_2)$, then the time to capture is again identical, since the policies are identical (and the pursuers are in the same locations regardless of the policy) after $\max(k_1, k_2)$ steps. This implies that policy $\bar{\nu}$ results in the exact same number of steps to capture, as does policy ν, and so is optimal.

We now have all the tools required to establish the lower bound on the min-max steps to capture and we do so in the following result.

Lemma 5 $T(D) \geq D - 1 + T(1),\ \forall D \geq 2.$

Proof. By definition (III.B), the worst-case minimum time to capture, $T(D) \geq T_{res}(D)$. From Lemma 4, there exists an optimal pursuit policy, $\bar{\nu}$, such that $\bar{\nu}_m(\mathcal{I}(D)) = E,\ m = 1, 2$. By following the policy $\bar{\nu}$, under the restriction (V.B), the 1^{st} pursuer will encounter another red UGS with delay $D-1$ at time $D+1$. So, by repeated application of Lemma 4, the optimal actions for the pursuers are to go east at every one of the subsequent $D-1$ time steps; until at time $2D-1$, pursuer 1 encounters a red UGS with delay 1. Hence, $T_{res}(D) = D - 1 + T(1)$. Since $T(D) \geq T_{res}(D)$, the result follows.

Theorem 2 *For any $D \geq 1$, $T(D) = D + 4$, and μ is an optimal policy.*

Proof. By combining Lemmas 3 and 5 , we readily obtain,

$$T(D) \geq D - 1 + T(1) = D + 4,\ D \geq 2.$$

From Theorem 1, we have $T(D) \leq T_\mu(D) = D + 4,\ D \geq 1$ and so, $T(D) = D + 4,\ D \geq 1$. It follows that the policy, μ, is optimal.

Remarks:

1. As far as implementing the policy μ is concerned, the pursuers only need to remember their last actions and the current observations, to determine the current action. In addition, they also need to remember the sequence of actions associated with the end-game.

2. It is noteworthy to mention that μ is an optimal policy, despite not being dependent on the entire history of observations.

VI. Conclusions

The solution to a deterministic cooperative search and capture problem on a Manhattan grid, under partial and delayed information, is obtained. We have shown that capture is guaranteed for any initial delay $D > 0$ and furthermore, it takes exactly $D + 4$ steps to capture the evader, in the worst-case. We have also derived the optimal pursuit policy that enables capture. For any delay $D > 0$, the pursuers maintain their relative staggered positions, until the evader is captured or the end-game is initiated. From here on, it takes no more than 3 steps to capture the evader.

References

[1] Witsenhausen, H., *Mimimax control of uncertain systems*, Ph.D. thesis, Massachusetts Institute of Technology, 1966.

[2] Witsenhausen, H., "A minimax control problem for sampled linear systems," *IEEE Transactions of Automatic Control*, Vol. 13, Feb. 1968, pp. 5–21.

[3] Bertsekas, D. P., *Control of uncertain systems with a set-membership description of the uncertainty*, Ph.D. thesis, Massachusetts Institute of Technology, 1971.

[4] Bertsekas, D. P. and Rhodes, I. B., "Sufficiently Informative Functions and the Minimax Feedback Control of Uncertain Dynamic Systems," *IEEE Transactions of Automatic Control*, Vol. 18, No. 2, April 1973, pp. 117–124.

[5] Bertsekas, D. P., "On the solution of some minimax problems," *Conference on Decision and Control*, Vol. 11, IEEE, Dec 1972, pp. 328–332.

[6] Issacs, R., *The theory of Differential Games*, Wiley, New York, 1965.

[7] Bertsekas, D. P. and Rhodes, I. B., "On the Minimax Reachability of Target Sets and Target Tubes," *Automatica*, Vol. 7, 1971, pp. 233–247.

[8] Schweppe, F., "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Transactions of Automatic Control*, Vol. 13, 1971, pp. 22–28.

[9] Bertsekas, D. P. and Rhodes, I. B., "Recursive state estimation for a set-membership description of uncertainty," *IEEE Transactions of Automatic Control*, Vol. 16, April 1971, pp. 117–128.

[10] Fomin, F. V. and Thilikos, D. M., "An annotated bibliography on guaranteed graph searching," *Theoretical Computer Science*, Vol. 399, 2008, pp. 236–245.

[11] Eagle, J. K., "The optimal search for a moving target when the search path is constrained," *Operations Research*, Vol. 32, No. 5, 1984, pp. 1107–1115.

[12] Thomas, L. C. and Washburn, A. R., "Dynamic Search Games," *Operations Research*, Vol. 39, No. 3, 1991, pp. 415–422.

[13] Dzyubenko, G. T. and Pshenichnyi, B. N., "Discrete Differential Games with Information Lag," *Cybernetics and Systems Analysis*, Vol. 8, No. 6, 1972, pp. 947–952.

[14] Bar-Shalom, Y. and Tse, E., "Dual effect, certainty equivalence, and seperation in stochastic control," *IEEE Transactions of Automatic Control*, Vol. 19, No. 5, 1974, pp. 494–500.

[15] Krishnamoorthy, K., Casbeer, D., Chandler, P., Pachter, M., and Darbha, S., "UAV Search and Capture of a Moving Ground Target under Delayed Information," *Conference on Decision and Control*, IEEE, Maui, HI, Dec 2012, pp. 3092–3097.

[16] Krishnamoorthy, K., Darbha, S., Khargonekar, P., Casbeer, D., Chandler, P., and Pachter, M., "Optimal Minimax Pursuit Evasion on a Manhattan Grid," *American Control Conference*, IEEE, Washington D.C., June 2013, pp. 3427–3434.