

A Markov Chain Game with Dynamic Information¹

G. J. OLSDER² AND G. P. PAPAVALASSILOPOULOS³

Communicated by Y. C. Ho

Abstract. Two players, not knowing each other's position, move in a domain and can flash a searchlight. The game terminates when one player is caught within the area illuminated by the flash of the other. However, if this first player is not in this area, then the other player has disclosed his position to the former one, who may be able to exploit this information. The game is considered on a finite state space and in discrete time.

Key Words. Markov games, zero-sum games, incomplete information, pursuit-evasion, mixed strategies.

1. Introduction

The game to be discussed belongs to the class of two-person zero-sum games. The players move in a certain domain and are unaware of each other's positions unless a player flashes a searchlight that illuminates an area of known shape around this player. By flashing his searchlight, a player discloses his position to the other player, wherever the players are. Termination of the game can occur only if a player flashes his searchlight and the other player finds himself trapped within the area illuminated. Both players

¹ The work of the second author was supported by ZWO, The Netherlands Organization for the Advancement of Pure Research, Contract No. B62-239, by the US Air Force Office of Scientific Research, Grant No. AFOSR-85-0245, and by the National Science Foundation, Grant No. NSF-INT-8504097.

² Professor, Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands.

³ Associate Professor, Department of Electrical Engineering Systems, University of Southern California, Los Angeles, California; Visiting Professor at Delft University of Technology during 1986.

want to catch the other player in their searchlight before they are caught themselves. Therefore, flashing has two aspects; in order for a player to win, he must flash; if, however, during such a flash the other player is not caught, the flashing player is in a more vulnerable position, because he has betrayed his location to the other player. The idea of this game is described in Ref. 1.

The game described above can be interpreted as a duel, such as for instance described in Ref. 2. In a duel, two combatants are each armed with a certain number of bullets (the flashes in our case) and slowly advance toward each other. If a combatant fires a bullet, the probability of success (i.e., hitting the opponent) increases if the distance between the two combatants decreases. If the bullets are noisy (i.e., a combatant hears the firing of a bullet by his opponent) and if this combatant is not hit, he can use this information (fewer bullets left with the opponent) to his advantage. Such information is called dynamic, since it depends on the actions of the players. A feature that makes the game in this paper different is that, in addition to the flashing times, the strategies that decide on their own motion must also be determined.

It is easy to understand that both players will exploit probabilistic strategies, since a pure optimal strategy by one player would certainly lead to a loss for him if the other player would know this optimal strategy. Conceptually, it is very complicated to define mixed and/or behavioral strategies for games that proceed continuously in time (see Ref. 3, Section 5.4). Therefore, the problem has been investigated for a finite state space (i.e., the players move on a network with a finite number of nodes) and for the discrete-time case.

Thus, the realm of Markov chain games has been entered (for a survey, see Ref. 4). For such games, both a normal and an extensive form description can be given; they are equivalent and the former is more amenable for obtaining numerical results. See also Ref. 5, where a specific Markov chain game has been solved and where the players use mixed strategies. A slightly different pursuit evasion game is given in Ref. 6. Both papers deal with a maneuvering evader, while being fired upon by the pursuer.

The problem treated in this paper bears also some relationship to the so-called bomber-versus-battleship game, as originally described in Ref. 7. In Ref. 6, more references with respect to this game are given. In this game, a battleship is floating around on a set of integers, albeit to take one step in either direction in each time unit. The sole aim of the battleship is to avoid a bomber, equipped with only one bomb, which is hovering over it, but suffers from the fact that the bomb takes two time units to drop. Dropping the bomb (when and where) coincides with switching on the searchlight in the current game.

Explicit results have been obtained when the players are confined to the circumference of a circle and, during each time step, each player can either move to one of the two adjacent nodes or stay where he is. Specific results depend on the number of nodes and the shape of the area illuminated (which may be different for the two players). In essence, finding the saddle-point strategies boils down to the solution of a two-level linear programming problem.

In Section 2, the exact problem statement will be given. In the first part of Section 3, a simplified version of the game is considered; the two players know each other's initial positions and only one player can flash (once). Here, it will be shown that mixed and behavioral strategies coincide. In the second part of Section 3, there is again only one player who can flash once; he knows the initial position of the other player, whereas now this latter player does not know the initial position of the player with the flash option. Though in principle the game may last an infinitely long time, it will be shown that, for the solution, a finite-horizon version suffices. In Section 4, the solution will be given for the case that both players have one flash at their disposal and do not know each other's initial positions. Some numerical examples are provided. In Section 5, players with different shapes of their flashing area are considered, which leads to an additional matrix game to be solved. Also, a minor variation of the game described in Section 4 is considered. The variation is that one player has two flash options and the other one has none. Lastly, in Section 6, some possible extensions, conjectures, and limitations will be discussed.

2. Problem Statement

Let us consider n positions arranged on a circle, enumerated by $1, 2, \dots, n$, as in Fig. 1. A pursuer P is at position $i \in S = \{1, \dots, n\}$, and an evader V is at position $j \in S$. At the initial instant of time, $t = 0$, they do not know each other's position, but each one knows his own. Therefore, each player assumes a uniform distribution with respect to the initial position of the other player. At each instant of time, they can move one position to the left or the right or stay where they are. The time is discrete, $t = 1, 2, 3, \dots$

Each one of them has a flashlight that he can flash only once, at some time $t = 1, 2, 3, \dots$. The flashing illuminates the position of the player who flashes, as well as the positions to his left and right. If at the time P (or V) flashes, V (or P) is within the illuminated area, then he is caught and the game terminates; if V is not caught, and has not yet flashed his light, then the game terminates when he also flashes.

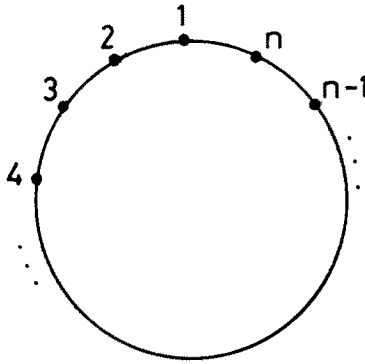


Fig. 1. State space.

Let us point out that, if P flashes at some instant of time, before V does, and V is not caught, then P has revealed his position to V , who saw where P was when he flashed, whereas P knows only that V is in some of the remaining $n - 3$ positions. At this instant of time, the information pattern of both P and V , concerning each other's position, changes; thus, we are dealing with a problem with dynamic information. Or, to put it in another way, each player receives information about the position of his opponent only when he or his opponent flashes his light; thus, there is a learning aspect with respect to the flashing, but not with respect to the motion of the players.

3. Solution of Two Simplified Versions of the Problem

In this section, two simple versions of the general problem are solved. The combination of these two versions will provide the solution to the general problem, and that solution is given in Section 4.

Case 1. Both P and V know each other's initial position; V has no flash.

Here, the situation is exactly as before, but V has no flash and both P and V know each other's initial positions. If the motion of P is completely deterministic, then V can always avoid capture; correspondingly, P can capture V if V 's motion is deterministic. Thus, an equilibrium solution in pure strategies does not exist, and we will consider that P and V move according to some randomized decision schemes.

P 's decision of when and where to flash is also considered to be the result of a randomized decision scheme.

Let Λ denote an $n \times n$ Markovian matrix of the following structure:

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_{1,n} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_{32} & \lambda_{33} & \lambda_{34} & 0 & 0 & & 0 & 0 \\ 0 & 0 & \lambda_{43} & \lambda_{44} & \lambda_{45} & 0 & & 0 & 0 \\ \vdots & & & & & & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & \lambda_{n-1,n-2} & \lambda_{n-1,n-1} & \lambda_{n-1,n} \\ \lambda_{n,1} & 0 & \cdots & \cdots & \cdots & 0 & 0 & \lambda_{n,n-1} & \lambda_{nn} \end{bmatrix}, \quad (1a)$$

$$\lambda_{ij} \geq 0, \quad \sum_j \lambda_{ij} = 1, \quad (1b)$$

$$\lambda_{ij} = 0, \quad \text{if } i = 2, 3, \dots, n-1 \text{ and } |j-i| > 1, \quad (1c)$$

$$\lambda_{13} = \cdots = \lambda_{1,n-1} = \lambda_{n,2} = \cdots = \lambda_{n,n-2} = 0. \quad (1d)$$

The motion of V is described as follows. Let V be at position i at $t = 0$, and let

$$e_i = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{R}^n.$$

\uparrow
 ith position

At times $t = 0, 1, 2, \dots$, the evader V chooses matrices $\Lambda_1, \Lambda_2, \Lambda_3, \dots$ of the type (1), so that the probability of being at any position in S , at time t , is given by the row vector

$$e_i^T \Lambda_1 \Lambda_2 \cdots \Lambda_t. \quad (2)$$

Actually, only the row i of Λ_1 and rows $i-1, i$, and $i+1$ of Λ_2 , and so on, need to be considered.

The motion of P as well as his decision of whether to flash or not and where and when are described as follows. At time $t = 0$, P is at position j . Consider matrices P_1, P_2, \dots of the type (1). The motion of P is described by the analogue of (2):

$$e_j^T P_1 P_2 \cdots P_t. \quad (3)$$

At time t , P may decide to flash at any of the positions where he may be, so that, at time t , he also has a vector (P_{t1}, \dots, P_{tn}) , where P_{ik} is chosen as a function of the previous decisions of P ; P_{ik} is zero if position k is not accessible for him at time t , given that he was at position j at $t = 0$. It holds that

$$\sum_{i,k} P_{ik} = 1, \quad P_{ik} \geq 0.$$

A little reflection will persuade the reader that P can dispense with deciding on both P_1, P_2, \dots and P_{ik} and consider only a probability vector $\{q_{it}\}$:

$$q_{it} \geq 0, \quad \sum_{i,t} q_{it} = 1, \quad q_t = \begin{bmatrix} q_{1t} \\ \vdots \\ q_{nt} \end{bmatrix}, \quad (4a)$$

$$q_{it} = 0, \quad \begin{array}{l} \text{if position } i \text{ is not accessible at } t, \\ \text{if } P \text{ is at } j \text{ at } t = 0; \end{array} \quad (4b)$$

i.e., q_{it} = probability with which P will be at time t at position i and flash there. This is essentially due to the open-loop character of P 's information, since his knowledge concerning V 's position is not altered by any information about V 's actual motion.

As an example of the q_{it} 's in (4a), let $e_j = (1, 0, \dots, 0)$; i.e., P starts at time $t = 0$ at position 1. Then,

$$q_1 = \begin{bmatrix} q_{11} \\ q_{21} \\ 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \\ 0 \\ q_{n,1} \end{bmatrix}, \quad q_2 = \begin{bmatrix} q_{12} \\ q_{22} \\ q_{32} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \\ q_{n-1,2} \\ q_{n,2} \end{bmatrix}, \quad q_3 = \begin{bmatrix} q_{13} \\ q_{23} \\ q_{33} \\ q_{43} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ q_{n-2,3} \\ q_{n-1,3} \\ q_{n,3} \end{bmatrix}, \dots, 0. \quad (4c)$$

Let Π be the $n \times n$ matrix

$$\Pi = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 1 & 1 & \cdots & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & & & 1 & 1 & 1 & \\ 1 & 0 & & & 0 & 1 & 1 & \end{bmatrix}; \quad (5)$$

i.e., $\Pi_{ij} = 1$ if position j is illuminated when P is at position i and flashes.

Using (2), (4a), (4b), (5), we find that the probability of capture is

$$e_i^T \Lambda_1 \Pi q_1 + e_i^T \Lambda_1 \Lambda_2 \Pi q_2 + \cdots + e_i^T \Lambda_1 \Lambda_2 \cdots \Lambda_t \Pi q_t + \cdots. \quad (6)$$

The summation stops at $t = T$ for a finite time case, and extends to infinity for the infinite time one.

Clearly, (q_1, q_2, \dots) lies in a multidimensional simplex, and so does $(\Lambda_1, \Lambda_1 \Lambda_2, \dots)$, as the following lemma demonstrates, enabling us to reformulate (6) as a classical matrix game.

Lemma 3.1. The set of vectors

$$\{(e_i^T \Lambda_1, e_i^T \Lambda_1 \Lambda_2, \dots, e_i^T \Lambda_1 \Lambda_2 \cdots \Lambda_t)^T \in R^n | \Lambda_1, \dots, \Lambda_t \text{ of type (1)}\} \quad (7)$$

is a convex polyhedron in R^n .

Proof. Let

$$e_i^T \bar{\Lambda}_1 = \bar{y}_1, \quad e_i^T \bar{\Lambda}_1 \bar{\Lambda}_2 = \bar{y}_2, \dots, e_i^T \bar{\Lambda}_1 \bar{\Lambda}_2 \dots \bar{\Lambda}_t = \bar{y}_t,$$

or equivalently,

$$e_i^T \bar{\Lambda}_1 = \bar{y}_1, \quad \bar{y}_1 \bar{\Lambda}_2 = \bar{y}_2, \dots, \bar{y}_{t-1} \bar{\Lambda}_t = \bar{y}_t.$$

Let also

$$e_i^T \bar{\bar{\Lambda}}_1 = \bar{\bar{y}}_1, \quad \bar{\bar{y}}_1 \bar{\bar{\Lambda}}_2 = \bar{\bar{y}}_2, \quad \dots, \quad \bar{\bar{y}}_{t-1} \bar{\bar{\Lambda}}_t = \bar{\bar{y}}_t.$$

Let $0 < \lambda < 1$. We will show that there exists $\Lambda_1, \Lambda_2, \dots, \Lambda_t$, so that

$$e_i^T \Lambda_1 = \lambda \bar{y}_1 + (1 - \lambda) \bar{\bar{y}}_1, \tag{8a}$$

$$(\lambda \bar{y}_1 + (1 - \lambda) \bar{\bar{y}}_1) \Lambda_2 = \lambda \bar{y}_2 + (1 - \lambda) \bar{\bar{y}}_2, \tag{8b}$$

⋮

$$(\lambda \bar{y}_{t-1} + (1 - \lambda) \bar{\bar{y}}_{t-1}) \Lambda_t = \lambda \bar{y}_t + (1 - \lambda) \bar{\bar{y}}_t. \tag{8c}$$

Obviously,

$$\Lambda_1 = \lambda \bar{\Lambda}_1 + (1 - \lambda) \bar{\bar{\Lambda}}_1$$

satisfies (8a). Assume that $\Lambda_1, \dots, \Lambda_{k-1}$, have been determined so as to satisfy the first $k - 1$ of Eqs. (8). We have to choose Λ_k so as to satisfy the k th of Eqs. (8), i.e.,

$$(\lambda \bar{y}_{k-1} + (1 - \lambda) \bar{\bar{y}}_{k-1}) \Lambda_k = \lambda \bar{y}_{k-1} \bar{\Lambda}_k + (1 - \lambda) \bar{\bar{y}}_{k-1} \bar{\bar{\Lambda}}_k. \tag{9}$$

Let

$$\bar{y}_{k-1} = (a_1, \dots, a_n), \quad \bar{\bar{y}}_{k-1} = (b_1, \dots, b_n),$$

and

$$\Lambda_k = \begin{bmatrix} \lambda a_1 / [\lambda a_1 + (1 - \lambda) b_1] & & 0 & & \\ & 0 & \lambda a_2 / [\lambda a_2 + (1 - \lambda) b_2] & & 0 \\ & & & \dots & \\ & 0 & & & \lambda a_n / [\lambda a_n + (1 - \lambda) b_n] \end{bmatrix} \bar{\Lambda}_k + \begin{bmatrix} (1 - \lambda) b_1 / [\lambda a_1 + (1 - \lambda) b_1] & & & & 0 \\ & & & & \\ & & & & \\ & 0 & & & (1 - \lambda) b_n / [\lambda a_n + (1 - \lambda) b_n] \end{bmatrix} \bar{\bar{\Lambda}}_k. \tag{10}$$

If $a_i = b_i = 0$, then use $\mu, 1 - \mu$ in place of $\lambda a_i / [\lambda a_i + (1 - \lambda) b_i], (1 - \lambda) b_i / [\lambda a_i + (1 - \lambda) b_i]$ in (10), where $0 \leq \mu \leq 1$.

Obviously, Λ_k is of type (1) and satisfies the k th of Eqs. (8). The process can now be continued. The elements of each $\Lambda_i, i = 1, \dots, t$, satisfy

certain linear inequality constraints as indicated after formula (1). Since each Λ_i appears linearly in the vectors forming the set indicated by (7), it is clear not only that this set of vectors spans a convex area, but also that this area is bounded by a finite number of linear equality constraints. Thus, the set is a convex polyhedron. \square

It is clear now that (6), at least for the finite summation case (i.e., finite game duration, $t=0, 1, \dots, T$) can be reduced to a classical matrix game. Namely, let $\delta_1, \dots, \delta_k$ be the extreme points of the polyhedron (7). Then, any element of (7) can be expressed as Δy , where

$$\Delta = [\delta_1, \dots, \delta_k], \quad y = (y_1, \dots, y_k)^T, \quad y_i \geq 0, \quad \sum y_i = 1,$$

and thus (6) assumes the form

$$y^T A_T \tilde{q}, \quad \sum y_i = \sum \tilde{q}_i = 1, \quad (11)$$

where A_T is the matrix resulting from the multiplication of rows of Δ with columns of Π and disregarding the q_{ii} 's that are zero. As a matter of fact, the same game (11) would result if one were to mix directly the pure strategies of P and V , as a little reflection will convince the reader. In (11), V chooses y and minimizes; P chooses \tilde{q} and maximizes. The classical minimax theorem yields existence of solutions and uniqueness in value for this game.

Let us consider the size of the matrix A_T in (11). After one time step, P can reach three positions; after two time steps, five positions; etc. Since he can flash at any of these time instants, his total number of strategies, if there are T time steps to go, is $3+5+\dots+(2T+1) = T^2+2T$. A tacit assumption here is that $T \leq [n/2]$. For $T \geq [n/2]$, all points on the circle can be reached. Player V can choose from three directions during each time step, and his total number of different strategies is therefore 3^T . The size of A_T is $3^T * (T^2+2T)$. The elements of A_T are either 0 (no capture) or 1 (capture).

Example 3.1. Take $n=7$ and $T=2$. Then, A is a 9×8 matrix. Assume that initially P is two positions to the right of V . An optimal mixed strategy for V consists of the following maneuvers: two times left (probability $1/3$); first left and then stay there (probability $1/3$); and lastly two times right (probability $1/3$). The directions left and right are defined for V (and P) with the nose toward the center of the circle. Only three pure strategies form part of the optimal mixed strategy for V . Another way of describing V 's behavior is by means of behavioral strategies; during the first time step, V moves to the right (probability $1/3$) or to the left (probability $2/3$). If, during the first time step, V moved to the right, then during the second

time step he will again move to the right (probability 1). If, during the first time step, V had moved to the left, then during the second time step he stays there (probability 1/2) or moves again to the left (probability 1/2). Whatever P does, the value of the game will not be more than 1/3. If we change the initial positions somewhat, such that P is initially three instead of two positions to the right of V , then it turns out somewhat surprisingly that the value is 1/2 (P can now catch V by either moving along the left bound or by moving along the right bound, which apparently increases his changes). \square

The solution of (11) can be obtained by solving the following linear programming (LP) problem (see Ref. 3):

$$\max_{\bar{y}}(\bar{y}^T l_K), \quad \text{subject to } A_T^T \bar{y} \leq l_{\bar{T}}, \quad \bar{y}_i \geq 0,$$

where l_m stands for a vector of m elements that are all equal to one, \bar{T} is the number of elements of \bar{q} , and the superscript T denotes transpose. The solution to this LP problem yields the mixed strategy of player P , normalized with respect to the saddle-point value of the game. The saddle-point value equals the maximum value v of the criterion of the LP problem.

The dual problem,

$$\min_{\bar{q}}(\bar{q}^T l_{\bar{T}}), \quad \text{subject to } A_T \bar{q} \geq l_K, \quad \bar{q}_i \geq 0,$$

yields the optimal mixed strategy for player V , also normalized with respect to the saddle-point value v .

For different T 's, (11) has different solutions. It is our intention now to show that no time larger than $\lceil n/2 \rceil$ needs to be considered, since the solution for $T = \lceil n/2 \rceil$ provides the solutions for any arbitrarily large T .

Lemma 3.2. The value J_T of (11) is an increasing function of T .

Proof. Intuitively, one sees that, the larger the time interval, the better are the chances for P to capture V . More rigorously, consider the summation (6) up to time T . If P imposes on himself the restriction $q_T = (0, \dots, 0)^T$, then V can exploit only $\Lambda_1, \dots, \Lambda_{T-1}$ and the choice of Λ_T is irrelevant. On the other hand, V cannot terminate effectively the game at $T-1$ for arbitrary T , since no choice of Λ_T nullifies the last term. \square

Lemma 3.3. $J_1 \leq J_2 \leq \dots \leq J_{\lceil n/2 \rceil} = J_{\lceil n/2 \rceil + 1} = J_{\lceil n/2 \rceil + 2} = \dots$

Proof. Continuing the rationale of the proof of Lemma 3.2, one sees that a choice by V of $\Lambda_T = I$ will result in

$$\begin{aligned} & e_i^T \Lambda_1 \Pi q_1 + \dots + e_i^T \Lambda_1 \dots \Lambda_{T-1} \Pi q_{T-1} + e_i^T \Lambda_1 \dots \Lambda_{T-1} \Pi q_T \\ & = e_i^T \Lambda_1 \Pi q_1 + \dots + e_i^T \Lambda_1 \dots \Lambda_{T-1} [q_{T-1} + q_T]. \end{aligned} \tag{12}$$

q_T can be added to q_{T-1} , and (12) can be considered as the payoff of a $T-1$ period game, if the addition of q_T to q_{T-1} does not alter the character of q_{T-1} , i.e., does not create nonzero components in $\tilde{q}_{T-1} = q_T + q_{T-1}$ in places where q_{T-1} has to have zeros [see (4a), (4c)]. This can happen only if $T \geq [n/2]$, i.e., the minimum time needed for P to go to any position on the circle, when he starts moving one step at a time at a time at $t=0$. Thus, if $T \geq [n/2]$, V can terminate effectively the game at time T , by choosing

$$\Lambda_{[n/2]+1} = \Lambda_{[n/2]+2} = \dots = I,$$

which is obviously to his interest given the increasing character of J_T . \square

Remark 3.1. If T is larger than $[n/2]$, it holds that $J_T \geq 3/n$, since P can flash randomly at some position, after time $t \geq [n/2]$ has elapsed and V has spread himself uniformly on the circle. It is natural to expect that $J_{[n/2]}$ will be greater than or equal to K/n , if K positions can be flashed instead of 3. For $K = 1$ [the Π matrix of (5) becomes the unit matrix], we have $J_{[n/2]} = 1/n$, as the following reasoning shows.

Consider $n = 4$. For $t = 0$ and $t = 1$, the probabilities with which V moves are denoted by arrows on the outside of the circle, whereas in the inside of the circle the total probability of V being at the corresponding positions are given. The possible positions of P at each time are marked by a star. Obviously, for the strategies of V denoted in Fig. 2, P has no more than probability $1/4$ of capturing V , and one can easily construct other strategies with the same property. That the strategy of V delineated in Fig. 2 corresponds to an equilibrium for J_2, J_3, \dots is easy to show. One can generalize this example for $n > 3$, by considering strategies of V that essentially assign total probability $1/n$ of V being in any position accessible by P at any instant of time. It should be pointed out, that the case $\Pi = \text{unit}$, i.e., that P can illuminate only his position and no adjacent ones, results in making P indifferent to knowing V 's initial position, since he can get the same payoff $1/n$ by flashing randomly at some position after time

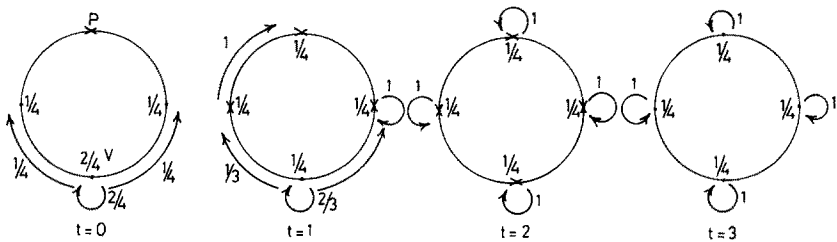


Fig. 2. Mixed strategies of V .

$T \geq [n/2]$ has elapsed. On the other hand, if $T \not\geq [n/2]$, V can escape obviously with probability 1, by going with probability 1 to the position(s) not accessible by P within this time.

Case 2. P knows V 's position; V does not know P 's position; V has no flash.

Here, we examine the case where P knows V 's initial position, whereas V does not know P 's position, but only that P is not at the position of V or the adjacent ones at $t = 0$. P is the only one who can flash. This situation arises by considering the general problem of the first part of this section, in which V flashes and fails before P flashes. Thus, P can use his flash knowing V 's position (betrayed by V 's own flashing). The time interval T is finite (the infinite time case can be reduced to the finite one by arguments essentially the same as those employed in Lemmas 3.2 and 3.3).

Consider Fig. 3. The evader V is at position $n - 1$, knows that P is at position 1, or 2, or \dots , or $n - 3$. The pursuer P knows that V is in position $n - 1$ at time V . V solves the following problem:

$$\min_y \max_{\tilde{q}_1, \dots, \tilde{q}_{n-3}} \left[\frac{1}{n-3} y^T A_{1,T} \tilde{q}_1 + \frac{1}{n-3} y^T A_{2,T} \tilde{q}_2 + \dots + \frac{1}{n-3} y^T A_{n-3,T} \tilde{q}_{n-3} \right], \tag{13}$$

where each $y^T A_{iT} \tilde{q}_i$ is exactly of the type (11), with A_{iT} corresponding to time interval T , with initial positions i for P and $n - 1$ for V . The rationale of V employing (13) is that, since P can be anywhere in $1, 2, \dots, n - 3$, with equal probability $1/(n - 3)$, a weighted sum of problems of type (11) needs to be solved. In (13), y is a probability vector and so is each \tilde{q}_i . Clearly, $(\tilde{q}_1^T, \dots, \tilde{q}_{n-3}^T)^T$ is not a probability vector, but obviously spans a convex polyhedron; thus, it can be expressed as $E\lambda$, where E is the matrix whose columns are the extreme points of this polyhedron and λ is a

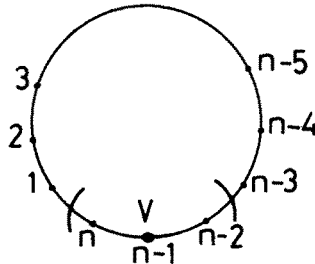


Fig. 3. Location of P and V .

probability vector of appropriate dimension. Thus, (13) is transformed easily to a classical matrix game.

The transformation of (13) to an LP problem is as follows:

$$\begin{aligned} & \min \sum_{i=1}^{T^2+T} (\bar{q}_1)_i, \\ & \text{s.t. } [A_{1,T} \quad A_{2,T} \quad \cdots \quad A_{n-3,T}] \begin{bmatrix} \bar{q}_2 \\ \bar{q}_2 \\ \vdots \\ \bar{q}_{n-3} \end{bmatrix} \geq l_{(n-3)(T^2+2T)}, \\ & \sum_{i=1}^{T^2+2T} (\bar{q}_1)_i = \sum_{i=1}^{T^2+2T} (\bar{q}_j)_i, \quad j = 2, \dots, n-3, \\ & (\bar{q}_j)_i \geq 0. \end{aligned}$$

The solution to this LP problem, with both equality and inequality constraints, equals the inverse of the saddle-point value of the game. Because of symmetry arguments on the circle, the number of elements in the vector

$$\bar{q} = (q_1^T, \dots, q_{n-3}^T)^T$$

can be reduced by a factor of about 2.

Another way of solving (13) is to solve first for $\tilde{q}_i, i = 1, \dots, n-3$, as functions of y (these functions are continuous and piecewise linear), then substitute these $\tilde{q}_i(y)$ functions and minimize the resulting function with respect to y . This latter function is continuous, piecewise linear, and convex; hence the (or a) minimum exists.

Let us call \tilde{J}_T the value of the game formulated in (13). Obviously, it does not depend on the initial position of V , and the coefficient $1/(n-3)$ in the weighted sum of (13) can be disregarded, since each position of P at $t = 0$ is assigned equal weight. A rationale similar to those used in Lemmas 2 and 3 shows that

$$\tilde{J}_1 \leq \tilde{J}_2 \leq \dots \leq \tilde{J}_{[n/2]} = \tilde{J}_{[n/2]+1} = \tilde{J}_{[n/2]+2} = \dots;$$

thus, only a finite number of problems of type (13) need to be considered. Before leaving this section, let us point out that, if $y^*, \tilde{q}_1^*, \dots, \tilde{q}_{n-3}^*$ solves (13), then the payoff of V is \tilde{J}_T , whereas the payoff of P is $y^{*T} A_{jT} \tilde{q}_j^*$, where j is the true position of P known to him. Essentially, in this subsection we dealt with a nonzero sum game, because P and V have different information, but one of a very special type, since the information of V is included in that of P .

4. Solution of the Problem of Section 1

We are now ready to address the main problem of this paper, which was described in Sections 1 and 2. If both P and V flash at time $t = 1$, then, because of the assumed uniform distribution, the probability for V to capture P is $3/n$ and so is the probability of P to capture V . Thus, the payoff is $3/n - 3/n = 0$. The reason why the probability of P being captured is $3/n$ is that V has no information about P 's initial and current position; thus, from V 's point of view, P can be anywhere with equal probability. Let us consider now that V flashes at time $t = 1$ and P does not. Then, the payoff for V is

$$\begin{aligned} & (3/n) - (1 - (3/n))\tilde{J}_{T-1} \\ & = (\text{probability that } V \text{ captures } P) \\ & - (\text{probability that } V \text{ does not capture } P) \\ & \times (\text{probability that } P \text{ captures } V \\ & \quad \text{during the remaining } T-1 \text{ period game}). \end{aligned}$$

Similarly, if V flashes at time $t = 2$ and P does not flash at $t = 1$ or 2 , the payoff for V is $(3/n) - (1 - (3/n))\tilde{J}_{T-2}$.

In general, let

$$c_K = (3/n) - (1 - (3/n))\tilde{J}_{T-K}, \quad K = 1, 2, 3, \dots, T-1, \quad (14)$$

and consider the matrix

$$M = \begin{bmatrix} 0 & c_1 & c_1 & c_1 & c_1 & \dots & \dots & c_1 \\ -c_1 & 0 & c_2 & c_2 & c_2 & & & c_2 \\ -c_1 & -c_2 & 0 & c_3 & c_3 & & & c_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & & \vdots \\ -c_1 & -c_2 & -c_3 & & & -c_{T-2} & 0 & c_{T-1} \\ -c_1 & -c_2 & -c_3 & \dots & \dots & -c_{T-2} & -c_{T-1} & 0 \end{bmatrix}. \quad (15)$$

Let

$$s = (s_1, s_2, \dots, s_T)^T, \quad r = (r_1, r_2, \dots, r_T)^T$$

represent probability vectors, according to which P and V choose to flash at times $t = 1, 2, \dots, T$. Then, V is interested in maximizing, whereas P in minimizing $r^T Ms$; i.e.,

$$\min_s \max_r (r^T Ms). \quad (16)$$

The solution of the matrix game (16) provides the solution to our problem. The rationale for introducing (16) is the following. Let r_t be the probability

that V flashes at time t . Then, his incurred payoff, given that P flashes after him, is

$$(3/n) - (1 - (3/n))\tilde{J}_{T-t}.$$

This is incurred with probability equal to the product of probabilities that V flashes at time t and P does not flash before time t , i.e.,

$$r_t(1 - s_1 - s_2 - \dots - s_t) = r_t(s_{t+1} + \dots + s_T).$$

Using formula (14) and a symmetrical rationale with P flashing before V , we arrive at (15), (16).

Lemma 4.1. It holds that:

- (i) $c_1 \leq c_2 \leq \dots \leq c_{T-1} = 1/n$;
- (ii) if $T \geq \lceil n/2 \rceil$, then

$$c_1 = c_2 = \dots = c_{T-\lceil n/2 \rceil} \leq c_{T-\lceil n/2 \rceil+1} \leq c_{T-\lceil n/2 \rceil+2} \leq \dots \leq c_{T-1} = \begin{cases} 1/n, & n \geq 7, \\ 1/4, & n = 6, \\ 2/5, & n = 5, \\ 5/8, & n = 4; \end{cases}$$

- (iii) if $T \geq \lceil n/2 \rceil$ and $n \geq 11$, then $c_1 \leq 0$.

Proof. (i) This is an immediate consequence of (14) and the fact that $\tilde{J}_1 \leq \tilde{J}_2 \leq \dots$.

- (ii) A rationale similar to the one used in Lemma 3.2 yields

$$\tilde{J}_{\lceil n/2 \rceil} = \tilde{J}_{\lceil n/2 \rceil+1} = \tilde{J}_{\lceil n/2 \rceil+2} = \dots$$

Thus,

$$c_1 = c_2 = \dots = c_{T-\lceil n/2 \rceil}.$$

Consider now Fig. 3, and assume that V is at position $n-1$, whereas P is at one of $1, 2, \dots, n-3$. We will calculate \tilde{J}_1 of Section 3. Let $n \geq 7$. We have only one step to go. Let V go to position n with probability a , to position $n-2$ with probability b , and stay at position $n-1$ with probability $1-a-b$. If P is at position 1, he goes to position n and flashes, thus capturing V with probability $a + (1-a-b) = 1-b$. If he is at position 2, he goes to position 1 and flashes, thus capturing V with probability a . If P is at position $n-4$, he goes to $n-3$, flashes, and captures V with probability b . If P is at $n-3$, he goes to $n-2$, flashes, and captures V with probability

$b + (1 - a - b)$. Any other position of P in 3, 4, . . . , $n - 5$ yields zero probability of capture. Thus, P captures V with probability

$$\frac{1}{n-3} (1 - a - b + a) + \frac{1}{n-3} a + \frac{1}{n-3} b + \frac{1}{n-3} (1 - a - b + b) = \frac{2}{n-3}.$$

Thus,

$$c_{T-1} = (3/n) - (1 - (3/n))[2/(n-3)] = 1/n, \quad \text{if } n \geq 7.$$

If $n \leq 7$, say $n = 6$, then the motion of P is not necessarily as described above. Let us consider the cases $n = 4, 5, 6$ individually. First, let $n = 4$ and consider Fig. 4.

Let the motions of P and V be described probabilistically as indicated. Then,

$$\begin{aligned} \tilde{J}_1 &= \min_{\alpha_i} \max_{\mu_i} [\mu_1(\alpha_1 + \alpha_2) + \mu_2(\alpha_2 + \alpha_3)] \\ &= \min_{\alpha_i} \max_{\mu_i} [\mu_1, \mu_2] \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \end{aligned}$$

$$\mu_1 + \mu_2 = 1, \quad \alpha_1 + \alpha_2 + \alpha_3 = 1,$$

$$\mu_i \geq 0, \quad \alpha_i \geq 0.$$

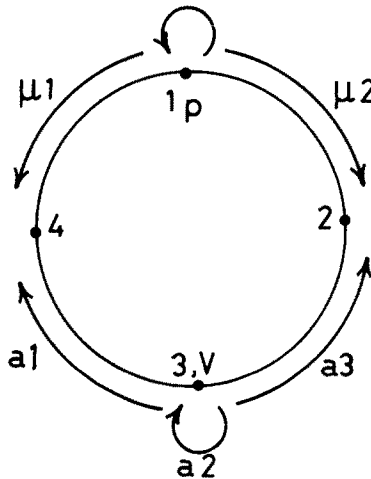


Fig. 4. Mixed strategies of P and V .

It is easy to see that

$$\mu_1^* = \mu_2^* = \alpha_1^* = \alpha_3^* = 1/2, \quad \alpha_2^* = 0, \quad \tilde{J}_1 = 1/2.$$

Thus, for $n = 4$,

$$c_{T-1} = (3/4)(1 - (3/4))(1/2) = 3/4 - (1/4)(1/2) = 5/8.$$

Similarly, one can examine the cases $n = 5$ and $n = 6$. For $n = 5$, $c_{T-1} = 2/5$; and, for $n = 6$, $c_{T-1} = 1/4$.

(iii) Consider Fig. 3. We consider that, if P is at any of the positions 2, 3, $n - 4$, $n - 5$, the pursuer P acts by flashing at the next instant of time, yielding a probability of capture $2/(n - 3)$, whereas for the other positions 4, 5, 6, ..., $n - 6$, P waits for a sufficient time and then flashes randomly, so that, for these positions, the probability of capture is $3/n$. Considering that there are $n - 7$ such positions, each one weighted by $1/(n - 3)$, we have that

$$\tilde{J}_T \geq 2/(n - 3) + (3/n)[(n - 7)/(n - 3)],$$

since such a policy is obviously suboptimal for P . It is easy to see now that

$$c_1 \geq 0 \Leftrightarrow \tilde{J}_T \leq 3/(n - 3).$$

But, for $n \geq 11$, it holds that

$$3/(n - 3) < 2/(n - 3) + (3/n)[(n - 7)/(n - 3)].$$

Thus, if T is sufficiently large and $n \geq 11$, c_1 has to be negative. □

Use of Lemma 4.1 enables one to see that the matrix game (16) has pure strategies solutions

$$r^* = s^* = (0, \dots, 0, 1, 0, \dots, 0),$$

with the position of 1 determined by the first c_i (starting from $i = T - 1$ and going backward in time) which becomes negative, or equivalently by the first \tilde{J}_i , going forward in time, which becomes smaller than $3/(n - 3)$. For example, consider M to have the form

$$M = \begin{bmatrix} 0 & - & - & - & - \\ + & 0 & - & - & - \\ + & + & 0 & + & + \\ + & + & - & 0 & + \\ + & + & - & - & 0 \end{bmatrix}, \quad T = 5.$$

Then,

$$r^* = s^* = (0, 0, 1, 0, 0).$$

Notice that, for a given n and different T 's, the distance of the $r_i^* = 1$ from the last equation remains fixed, so that, independently of the duration of the game, the players will flash at a fixed time distance from the final time. If the game is of infinite duration, then they never flash, a fact that is in agreement with the fact that an infinite matrix M of the form

$$M = \begin{bmatrix} 0 & -1 & -1 & -1 & \dots \\ +1 & 0 & -1 & -1 & \dots \\ +1 & +1 & 0 & & \\ \vdots & \vdots & \vdots & & \end{bmatrix}$$

does not have a saddle point (see Ref. 3, page 164). Of course, for particular cases where $n \leq 10$, one does not have the guarantee of Lemma 4.1(iii) that c_1 will be negative, and one has to calculate the c_i 's in order to find whether c_i changes sign. Clearly, if for some n and T , $c_1 > 0$, then both players flash in the beginning.

5. Nonidentical Players

We now make two remarks with respect to games in which the players enter the game nonsymmetrically.

Envisage the situation that the players have different flashing shapes. This is, for instance, the case if P can flash three positions and V can flash five positions. Then, the matrix M in (15) will become

$$M = \begin{bmatrix} 0 & c_1 & \dots & \dots & \dots & c_1 \\ -d_1 & 0 & & c_2 & \dots & c_2 \\ \vdots & -d_2 & & & & \\ \vdots & \vdots & & 0 & & c_{T-1} \\ -d_1 & -d_2 & \dots & -d_{T-1} & & 0 \end{bmatrix}.$$

and it is no longer skew-symmetric. The elements $-d_i$ are defined similarly to the c_i elements, the only difference being that a different flashing area has been used. Define i_c to be the lowest index of c such that c_{i_c} is positive, and define i_d to be the lowest index of f such that d_{i_d} is positive (or, equivalently, $-d_{i_d}$ is negative). It follows easily that the matrix game corresponding to the current M does not have pure solutions if $i_c \neq i_d$. It is easy to show that the optimal mixed strategies are a mixture of the pure strategies corresponding to row or column numbers

$$\min(i_c, i_d), \min(i_c, i_d) + 1, \dots, \max(i_c, i_d).$$

These strategies can again be found by solving an LP problem.

Example 5.1. Consider $n=7$, P 's flashing area is three positions ($k=3$), one to the left and right and his own position, E 's flashing area is only his own position ($k=1$). By solving problem (13) for this problem, it was found that

$$c_1 = c_2 = c_3 = 0.1428,$$

$$d_1 = -0.1428, \quad d_2 = -0.144, \quad d_3 = -0.0476.$$

Therefore, the M matrix becomes

$$M = \begin{bmatrix} 0.0000 & 0.1428 & 0.1428 & 0.1428 \\ 0.1428 & 0.0000 & 0.1428 & 0.1428 \\ 0.1428 & 0.114 & 0.0000 & 0.1428 \\ 0.1428 & 0.114 & 0.0476 & 0.0000 \end{bmatrix},$$

and $i_c = 1$, $i_d = 4$. The value of this game related to the M matrix (the row chooser maximizes and the column-chooser minimizes) turns out to be 0.0962. The optimal strategies are totally mixed and are given by

$$s = \begin{bmatrix} 0.326 \\ 0.326 \\ 0.261 \\ 0.087 \end{bmatrix}, \quad r = \begin{bmatrix} 0.326 \\ 0.239 \\ 0.109 \\ 0.326 \end{bmatrix}.$$

The next game to be discussed, in which the players enter the game nonsymmetrically, is again along the circumference of a circle. During each time step, both P and V can go one position to the left, stay where they are, or go one position to the right. In this game, P has two flash options and V has none. Suppose that initially the players do not know each other's position (they assume a random distribution for each other's position). If at the first flash of P the other player has not been caught, then all that P knows is that V is situated at one of the remaining $n-3$ points (if $k=3$) according to a uniform distribution. On the other hand, V then knows P 's position. In order to decide on when to flash the searchlight for the second time, a minimax problem of the kind (13) must be solved again.

6. Conclusions

We have discussed a discrete-time game with dynamic information in a finite state space. We have considered in detail the state space consisting of n elements that are positioned around the circumference of a circle. Each state has thus two adjacent states. The two players move in the state space and can terminate the game by flashing the searchlight. A surprising result

of the analysis is that, provided that both players enter the game with the same capacities (i.e., the flashing areas of the players have identical shape), the time of flashing can be determined by means of a pure strategy. This time is determined by the size of the flashing area and the number of states. If the players have different flashing areas, then the best flashing time can be determined by means of a mixed strategy.

Extensions of the state space are possible, and it seems that the techniques developed in the paper extend directly to connected networks. An obvious adjustment must then be made with respect to the weights $1/(n-3)$ in (13). These weights came from a uniform distribution. For a general network, one obtains these weights by solving a static game in which the players can position themselves arbitrarily on the network (hence, no restrictive dynamics). Also, different shapes of illumination do not seem to cause any difficulty; they can be attacked by the same methods.

The techniques developed in the paper can be used directly for solving the dynamic game in which one player has two flashes at his disposal and the other has none. Extensions to more than two flashes are nontrivial, since, by having flashed all but the last time, a game of the kind described by (13) still has to be solved (to determine the time of the last flash), but now the weights are functions of the previous flash times.

In order to solve the problem numerically, one must solve a series of linear programming (LP) problems that have been given explicitly in the text. The size of these LP problems increases drastically with the number of states n . We have not studied possible reductions in the size of these LP problems, due to redundancy and/or symmetries in the problem statement.

The continuous-time, continuous-state-space version of the problem treated seems hard to tackle. One needs conceptual extensions of mixed and behavioral strategies as functions of a continuously evolving time. It seems, however, that the ratio k/n (flashing area/total length or surface of the state space) and the maximum speeds are the crucial parameters. The ratio k/n is also a crucial parameter in the current setup. If for instance k/n decreases (and k is the same for both players), then the time of flashing will be closer to the end of the game (if one deals with a finite horizon).

References

1. HO, Y. C., and OLSDER, G. J., *Differential Games: Concepts and Applications*, Mathematics of Conflict, Edited by M. Shubik, North-Holland, Amsterdam, The Netherlands, pp. 127-186, 1983.
2. KIMELDORF, G., *Duels: An Overview*, Mathematics of Conflict, Edited by M. Shubik, North-Holland, Amsterdam, The Netherlands, pp. 55-72, 1983.

3. BASAR, T., and OLSDER, G. J., *Dynamic Noncooperative Game Theory*, Academic Press, New York, New York, 1982.
4. PARTHASARATHY, T., and STERN, M., *Markov Games: a Survey*, Differential Games and Control Theory II, Edited by E. O. Roxin *et al.*, Marcel Dekker, New York, New York, pp. 1-46, 1977.
5. BERNHARD, P., COLOMB, A. L., and PAPA VASSILOPOULOS, G. P., *Rabbit and Hunter Game: Two Discrete Stochastic Formulations*, Computers and Mathematics with Applications, Vol. 13, pp. 205-225, 1987.
6. KUMAR, P. R., *Optimal Mixed Strategies in Dynamic Games*, IEEE Transactions on Automatic Control, Vol. AC-25, pp. 743-749, 1980.
7. ISAACS, R., *A Game of Aiming and Evasion: General Discussion and the Marksman's Strategies*, Rand Corporation, Report No. 1385, 1954.