

Asymmetric Information Acquisition Games

Vartika Singh and Veeraruna Kavitha,
IEOR, Indian Institute of Technology Bombay, India

Abstract—We consider a stochastic game with partial, asymmetric and non-classical information, where the agents are trying to acquire as many available opportunities/locks as possible. Agents have access only to local information, the information updates are asynchronous and our aim is to obtain relevant equilibrium policies. Our approach is to consider optimal open-loop control until the information update, which allows managing the belief updates in a structured manner. The agents continuously control the rates of their Poisson search clocks to acquire the locks, and they get rewards at every successful acquisition; an acquisition is successful if all the previous stages are successful and if the agent is the first one to complete. However, none of them have access to the acquisition status of the other agents, leading to an asymmetric information game. Using standard tools of optimal control theory and Markov decision process (MDP) we solved a bi-level control problem; every stage of the dynamic programming equation of the MDP is solved using optimal control tools. We finally reduced the game with an infinite number of states and infinite-dimensional actions to a finite state game with one-dimensional actions. We provided closed-form expressions for Nash Equilibrium in some special cases and derived asymptotic expressions for some more.

I. INTRODUCTION

We consider non-classical information games (as described in [1]) inspired by the full information games considered in [2]. In [2], agents attempt to acquire M available destinations; each agent controls its rate of acquisition (advertisement through a Social network) to increase its chances of winning the destinations, while trading off the cost for acquisition. They considered full information games, wherein the agents at any time know the number of destinations available (not acquired by some agent); they also considered no-information games, where the agents would not even know the number of destinations/locks acquired by themselves.

It is more realistic to assume that the agents have access only to local information; they might know the number of locks acquired by themselves, but not the numbers acquired by others. This leads to partial, asymmetric and non-classical information games; Basar et. al in [1] describe a game to be of non-classical information type, and we describe the same in our words: if the state of agent i depends upon the actions of agent j , and if agent j has some information which is not available to agent i we have a non-classical information game. These kind of games are typically hard to solve ([1]); when one attempts to find best response against a strategy profile of others, one would require belief of others states, belief about the belief of others, and so on.

We have some initial results related to this asymmetric information game in [4], where we solved the problem for the case with two destinations and two agents. We also

considered that the locks (represent destinations) have to be acquired in a given order and only the agent that first wins all the locks would receive a (single) prize. We now consider a significant generalization of the game: a) we consider general number of locks and agents; b) the agent that wins all the first k locks before the others would receive prize c_k and this is true for any lock.

Our approach to this problem can be summarized as “open loop control till information update” (as in [4]). With no-information, one has to resort to open loop policies (action changes with time, but, oblivious to the state). This is the best when one has no access to information updates. With full information one can have closed loop policies (actions can depend on state). Further, in full information controlled Markov jump processes, every agent is informed immediately of the jump in the state and can change its action based on the change. In our case we have access to partial information, the agents can observe only some jumps and not all; thus we need policies that are open loop type till an information update. At every information update, one can choose a new open loop control depending upon the new information.

The agents have no access to the information of others, however upon contacting a lock they would know if they are the first to contact. Any strategy profile in our game is described by one open loop policy for each state, each state is primarily described by the time of acquisition of the previous lock; thus we have an infinite dimensional (strategy space is polish space) game. We used the tools of optimal control theory (Hamiltonian Jacobi equations) and Markov decision process to reduce this infinite dimensional game to a one dimensional game; we showed that the best response against any strategy profile of others includes a time threshold policy (acquisition attempts to be made with full intensity till a threshold of time, after which there would be no attempts); more importantly we showed that these thresholds can be specified by a single constant (for each lock), irrespective of the time of acquisition of the previous lock. We finally showed that the *reduced game is a strict concave game, proved the existence of unique Nash equilibrium* and provided a simple algorithm to compute the same. Our results matched with those in [4], for all the cases considered there. We have closed form expressions (some are asymptotic) for the Nash equilibrium for some cases.

Our models can capture a variety of applications, e.g., social network problem as in [2] with one destination or a block chain problem as in [4]. For general M , we solve problem of winning a project (unaware of others success) with multiple completion phases.

II. PROBLEM STATEMENT

There are n agents competing to win a project. The project is to acquire M locks before the deadline T . The locks are ordered, i.e., all the agents will compete for the first lock in the beginning, after which they will compete for the second lock and so on. The agent that contacts all the locks successfully wins the project and gets a terminal reward. Further we say k -th contact (of k -th lock) is successful if this contact happens before time T and if the agent was the first one to contact all the previous $(k-1)$ locks and the k -th lock; the agent gets some reward at every successful contact. The acquisition/contact process of each agent is modelled by independent (possibly non-homogeneous) Poisson Processes; they can choose the rate of their (Poisson) contact process as a function of time $t \in [0, T]$. A higher rate of contact will increase the chances of success but will also incur higher cost. The aim of the agent is to maximize its expected reward.

Information structure: The agents have partial/asymmetric information about the number of locks acquired by various agents and would use the available information to design their rate functions optimally. Any agent would know at all the times information related to its contact attempts, however has limited access to that of the others. When it contacts a lock, it would know if it is successful; if not, it gets to know that it is not the first one to contact. So, agent gets some partial information about the state of others.

Decision epochs: As already mentioned, an agent has access only to partial information. There is a (major) change in the available information at the lock-contact epochs¹; it would know successful/unsuccessful status of the lock immediately after contact, and based on this information the agent can choose an optimal action. Hence contact epochs form the natural decision epochs. Further, these epochs will be exponentially distributed random variables with the parameters chosen by the agents, so it is clear that the decision epochs of different agents will not be synchronous.

State: The state of any agent at any time $t \in [0, T]$ is the information available to it at that time. The information available to the agent i after contacting $(k-1)$ -th lock, z_k^i , has two components: i) a flag l_k^i represents whether the contact was successful ($l_k^i = 1$) or unsuccessful ($l_k^i = 0$); and ii) the time of $(k-1)$ -th contact denoted by τ_{k-1}^i . Thus $z_k^i = (l_k^i, \tau_{k-1}^i)$ is the state of agent i at decision epoch k . The state remains the same in the time interval $[\tau_{k-1}^i, \tau_k^i)$. The initial state $z_1^i = (l_1^i, \tau_0^i)$ is simply set to $(1, 0)$; the process starts at 0 and l_1^i is set as 1 for convenience.

Actions: The agents choose the rate functions (defined till T) at their respective decision epochs based on their states; these functions are *open loop policies*, wherein the dynamic action is independent of the state of the system; the agents change their rate function only at next decision epoch. Such an approach is called ‘‘Open loop policy till information update’’ as in [4]. The rate of contact, for agent i , at any time can take values in the interval $[0, \beta^i]$ and the rate function is measurable. To be precise, agent i at

decision epoch k , i.e., at time instance τ_{k-1}^i , chooses an action $a_k^i \in L^\infty[\tau_{k-1}^i, T]$, as the acceleration process to be used till the next acquisition. Here $L^\infty[\tau_{k-1}^i, T]$ is the space of all measurable functions that are uniformly bounded by the given constant; the bounds (β^i for agent i) can be different for different agents, nevertheless we avoid notation i for simpler representation. These form a closed subset of Polish space² of essentially bounded functions, i.e., the space of functions with finite essential supremum norm:

$$\|a\|_\infty := \inf\{\beta : |a(t)| \leq \beta \text{ for almost all } t \in [\tau_{k-1}^i, T]\}.$$

Strategy: The strategy of player i is a collection of open loop policies, one for each state and lock, as given below:

$$\pi^i = \{a_k^i(\cdot; z_k^i) \in L^\infty; \text{ for all } z_k^i \text{ and all } k \in \{1, \dots, M\}\}, \quad (1)$$

where $a_1^i(\cdot; z_1^i)$ represents the open loop policy/rate function used at start, while $a_k^i(\cdot; z_k^i)$ represents the same to be used after $(k-1)$ -th contact; this choice depends upon the available information z_k^i . At times, notation z_k^i is dropped and we use L^∞ in place of $L^\infty[\tau_{k-1}^i, T]$, to simplify the explanations.

Our work majorly analyses best response (BR) of players, hence we introduce the following notations. Without loss of generality we consider BR of agent i . Let $N := \{1, 2, \dots, n\}$ be the set of players and π^j be the policies of all other players represented by $j := -i := N - \{i\}$.

Rewards/Costs: The reward of agent i is c_k^i , if it contacts all the first k locks successfully (before time T and before other agents), while the terminal reward is c_M^i . This implies, an agent unsuccessful with first contact, has no incentive to attempt for further locks; so it would remain silent henceforth.

Recall that $T \wedge \tau_k^i$ represents the time instance³ till which the k -th lock is attempted. Then the cost spent on acceleration for the k -th contact equals,

$$\bar{a}_k^i(T \wedge \tau_k^i), \text{ with } \bar{a}_k^i(t) := \int_{\tau_{k-1}^i}^t a_k^i(s) ds. \quad (2)$$

Thus the expected (immediate) utility for stage k equals:

$$r_k^i(z_k^i, a_k^i; \pi^j) = \begin{cases} c_k^i P_k^i(z_k^i; a_k^i; \pi^j) & \text{if } l_k^i = 1 \text{ and } \tau_{k-1}^i < T, \\ -\nu E[\bar{a}_k^i(T \wedge \tau_k^i) | z_k^i] & \text{else,} \\ 0 & \end{cases} \quad (3)$$

where P_k^i represents the probability of successfully contacting k -th lock, conditioned on $z_k^i = (1, \tau_{k-1}^i)$, i.e., conditioned that $(k-1)$ -th lock is acquired successfully at τ_{k-1}^i ; and $\nu > 0$ is the trade-off factor between the reward and the cost.

When $k = 1$ the probability of success P_k^i also depends upon the failure of other agents, i.e., depends upon $\pi^j := \pi^{-i}$:

$$P_1^i(z_1^i, a_1^i; \pi^j) = \int_0^T \eta_1^i(s; \pi^j) e^{-\bar{a}_1^i(s)} a_1^i(s) ds,$$

where the probability of failure of other agents before time t equals (\mathcal{X} is the indicator and see (2) for definition of \bar{a}):

$$\eta_k^i(t; \pi^j) = \mathcal{X}_{\{k>1\}} + \mathcal{X}_{\{k=1\}} e^{-\sum_{m \neq i} \bar{a}_1^m(t)}. \quad (4)$$

²Complete and separable space.

³ The contact clocks $\{\tau_k^i\}$ are free running Poisson clocks, however we would be interested only in those contacts that occurred before deadline T .

¹By convention, the start of the process commences with 0-th contact.

In the above, the indicators are introduced to have unified notation; for $k > 1$, the probability of success P_k^i conditioned on success till $(k-1)$ -th lock (now there is no opposition), equals:

$$P_k^i(z_k^i, a_k^i; \pi^j) = \int_{\tau_{k-1}^i}^T \eta_k^i(s; \pi^j) e^{-\bar{a}_k^i(s)} a_k^i(s) ds. \quad (5)$$

It is easy to observe that for any given $a_k^i(\cdot) \in L^\infty$, the expected cost equals (see (2) and with $\tau_0^i := 0$):

$$E[\bar{a}_k^i(T \wedge \tau_k^i) | z_k^i] = \bar{a}_k^i(T) e^{-\bar{a}_k^i(T)} + \int_{\tau_{k-1}^i}^T \bar{a}_k^i(s) e^{-\bar{a}_k^i(s)} a_k^i(s) ds. \quad (6)$$

If an agent fails to contact before the deadline T , it still has to pay for the entire duration $T - \tau_{k-1}^i$ and hence the first term in the above equation.

Game Formulation: This problem can be modelled as a strategic/normal form non-cooperative game,

$$G = \langle N, S, \Phi \rangle, \text{ with } S = \{S^i\}_i \text{ and } \Phi = \{\phi^i\},$$

where the set of players $N = \{1, 2, \dots, n\}$, the strategy set of player i is the class of all possible strategies as in (1), i.e., $S^i = \{\pi^i\}$, and the (overall) utility of agent i is given by

$$\phi^i(\pi^i, \pi^j) = \sum_{k=1}^M E[r_k^i(z_k^i, a_k^i; \pi^j) | z_1^i = (1, 0)]. \quad (7)$$

Our aim is to find a tuple of strategies (that depend only upon the available information) that form the Nash equilibrium (NE). We conclude this section by mentioning some of the main results of the paper.

A. Important results

We solved this problem for M locks and n players; the problem is converted to a much simplified and reduced strategic form game, such that the (unique) Nash equilibrium of the reduced game is also the Nash equilibrium of the original game. We found the reduced game with the help of following results. Before stating the results, we require the following definitions.

Threshold Policy: A threshold policy is an open loop policy, which takes value β (maximum possible value) in the interval $[s, \theta]$ (if $s \leq \theta$) and value zero in the remaining interval, where s is the starting point of the control and $0 \leq \theta \leq T$. Threshold policies are represented by $\Gamma_{\theta; s}$, where (for any $s \leq t \leq T$),

$$\Gamma_{\theta; s}(t) = \begin{cases} \beta & \text{if } t \leq \theta, \\ 0 & \text{else.} \end{cases}$$

The rate function is 0 for all t , if the starting point $s > \theta$.

Threshold (T) strategy: is a strategy made up of threshold policies. A typical T-strategy is defined by threshold functions $\{\theta_k(\cdot)\}_k$ and is defined as below:

$$\begin{aligned} \pi &= \{\theta_1(\cdot), \dots, \theta_M(\cdot)\}, \\ &= \{a_k(\cdot; z_k); a_k(\cdot; z_k) = \Gamma_{\theta_k(s); s}(\cdot) \text{ when } z_k = (1, s) \text{ and} \\ &\quad a_k(\cdot; z_k) = \Gamma_{0; s} \text{ when } z_k = (0, s), \text{ for any } k\}. \end{aligned}$$

Basically when the state of player i is $z_k^i = (l_k^i, \tau_{k-1}^i)$ with $\tau_{k-1}^i = s$ and $l_k^i = 1$, then starting from time s player i uses threshold policy with threshold equal to $\theta_k(s)$ as the open loop rate function; when $l_k^i = 0$, we equate the thresholds

to 0, i.e., the player stops trying any further. Many a times the *starting point is obvious and hence the corresponding notation is dropped from the subscript*.

M-Thresholds (MT) Strategy: This is a special type of T-strategy in which the thresholds for any given lock remain the same irrespective of the time of acquisition of the previous lock. A typical MT-strategy is defined by M -thresholds:

$$\begin{aligned} \pi &= \{\theta_1, \dots, \theta_M\}, \\ &= \{a_k(\cdot; z_k); a_k(\cdot; z_k) = \Gamma_{\theta_k}(\cdot) \text{ for any } z_k \text{ and } k\}. \end{aligned}$$

In other words, MT-strategy is completely represented by M -thresholds $\{\theta_k\}_{k \leq M}$, one for each k : a) here θ_k represents the time threshold till which agent can attempt to acquire k -th lock; b) threshold θ_k is independent of τ_{k-1} , the time of acquisition of the $(k-1)$ -th lock; and c) if τ_{k-1} is bigger than θ_k (i.e., if the $(k-1)$ -th lock is acquired after the threshold for the k -th lock) then the agent would no longer attempt to acquire the k -th lock under the given MT-strategy.

To summarize, the strategy of any player is made up of open loop policies, one for every (possible) state and for each decision epoch. Thus we have infinite dimensional actions, however the following structural results about the best response strategies reduce the complexity of the game significantly:

Theorem 1: [Threshold Strategy] There exists a (threshold) T -strategy that is a best response strategy of any given player, against any given strategy profile of opponents. ■

Theorem 2: [M-thresholds strategy] There exists an MT-strategy that is a best response strategy of any given player, against any given strategy profile of opponents. ■

The proofs of both the Theorems are in the next section, section III. By virtue of the above Theorems, there exists a best response strategy represented completely using M thresholds, which is optimal among (uncountable) state dependent strategies with infinite dimensional strategy space. We will thus have a reduced game in \mathcal{R}^M which is further analyzed in section IV; we find NE of the original game among these MT strategies. The unique NE of the reduced game is characterized by Theorems 6 and 7 of section IV.

III. BEST RESPONSES

Our aim is to derive Nash equilibrium (NE) for this partial and asymmetric information stochastic game. We begin with deriving the best response (BR) of player i against any given strategy profile π^j of the opponents.

Dynamic programming equations The BR is obtained by maximizing the objective function (7) with respect to the strategies $\pi^i \in S^i$. It is easy to observe that this optimization is an example of a Markov decision process which can be solved using (M -stage) dynamic programming (DP) equations given below (see [6, Theorem 4.5.1 and Remarks after] which covers the results for Polish spaces):

$$\begin{aligned} v_k^i(z_k^i; \pi^j) &= 0 \quad \text{if } k > M \text{ or if } \tau_{k-1}^i > T \text{ or if } z_k^i = (0, \tau_{k-1}^i), \\ &\text{and other wise the value function equals,} \\ v_k^i(z_k^i; \pi^j) &= \sup_{a_k^i \in L^\infty} \{r_k^i(z_k^i, a_k^i; \pi^j) + E[v_{k+1}^i(z_{k+1}^i; \pi^j) | z_k^i, a_k^i]\}. \end{aligned} \quad (8)$$

Observe that ([6] and see equation (7))

$$v_1^i(z_1^i; \pi^j) = \sup_{\pi^i \in S^i} \phi^i(\pi^i; \pi^j),$$

and thus the BR is obtained by solving the DP equations. The k -th stage DP equation can be re-written as below:

$$v_k^i(z_k^i; \pi^j) = \sup_{a_k^i \in L^\infty} J_k^i(z_k^i, a_k^i; \pi^j), \quad (9)$$

where the cost J_k^i is defined by (see equations (3)-(6)):

$$J_k^i(z_k^i, a_k^i; \pi^j) = \int_{\tau_{k-1}^i}^T (h_k^i(t) - \nu \bar{a}_k^i(t)) a_k^i(t) e^{-\bar{a}_k^i(t)} dt \\ - \nu \bar{a}_k^i(T) e^{-\bar{a}_k^i(T)}, \text{ with} \\ h_k^i(s) := (c_k^i + v_{k+1}^i((1, s); \pi^j)) \eta_k^i(s; \pi^j). \quad (10)$$

Optimal control: From the structure of the optimization problem (9) defining the k -th stage DP equation, it is clear that one can solve it using an appropriate optimal control problem. One can write this optimization problem as:

$$v_k^i(z_k^i; \pi^j) = u(\tau, 0); \text{ when } z_k^i = (1, \tau),$$

where $u(s, x)$ (defined for any $s \in [\tau, T]$ and any x) is the value function of the optimal control problem with details:

$$u(s, x) := \sup_{a \in L^\infty[s, T]} J(s, x, a), \text{ where the objective function,} \quad (11)$$

$$J(s, x, a) := \int_s^T (h_k^i(s') - \nu x(s')) a(s') e^{-x(s')} ds' + g(x(T)),$$

with state process of the optimal control problem given by (for any $s' \in [s, T]$)

$$\dot{x}(s') = a(s'), \text{ with } x(s) = x \text{ and thus } x(s') = x + \int_s^{s'} a(\tilde{s}) d\tilde{s}$$

and with terminal cost $g(x) = -\nu x e^{-x}$.

We need to solve this optimal control problem to get the BR, and the standard technique to solve such problems is using Hamiltonian Jacobi (HJB) PDEs [3], and the one corresponding to (11) is given by:

$$u_s(s, x) + \sup_{a \in [0, \beta^i]} a \{ (h_k^i(s) - \nu x) e^{-x} + u_x(s, x) \} = 0, \text{ with} \\ u(T, x) = g(x), \quad (12)$$

where u_s, u_x are partial derivatives of the (optimal control) value function. Using the standard tools we immediately obtain the following result (Proof in Appendix):

Theorem 3: [Existence] At any stage k and any state z_k^i ,

- (i) The optimal-control value function $u(\cdot, \cdot)$ is the unique viscosity solution of the HJB PDE (12). Further, the value function $u(\cdot, \cdot)$ is Lipschitz continuous in (s, x) .
- (ii) We have a $a^*(\cdot)$ that solves the control problem. ■

Remark: Theorem 3 implies every stage of DP equation has an optimizer, i.e., for every lock k and state z_k^i , the player i has a best response policy, call it $a_k^{i*}(\cdot; z_k^i)$. Thus by [6, Theorem 4.5.1 and the following remarks about Polish space], $\pi^{i*} = \{a_k^{i*}(\cdot; z_k^i)\}_{k, z_k^i}$ is a BR strategy of player i against π^j .

We further show that the optimal strategy (in BR) can be a threshold strategy in the following. We begin with:

Lemma 1: For any k, z_k^i and open loop policy $a_k^i(\cdot)$, one can construct a threshold policy $\Gamma_\theta(\cdot)$ such that the contact time under threshold policy (denoted by τ_θ) stochastically dominates that (τ_a) under $a_k^i(\cdot)$, i.e., $\tau_\theta \stackrel{d}{\leq} \tau_a$, i.e., for any monotone decreasing function f ,

$$E[f(\tau_a)] \leq E[f(\tau_\theta)].$$

Further the expected costs (6) under both the policies are equal. Also, the probability of successful contact P_k^i given in (5) is better with the threshold policy, i.e.,

$$P_k^i(z_k^i; a_k^i(\cdot); \pi^j) \leq P_k^i(z_k^i; \Gamma_\theta; \pi^j).$$

Proof is in Appendix. ■

Thus at any stage k and for any state z_k^i , agent i can contact the locks faster using threshold policies and the running costs (3) are better. When the locks are contacted faster, i.e., when the next stage starts earlier, the “cost to go” (the value function) till the end is better as shown below:

Lemma 2: For any lock k , the value function with $t \leq \tau$,

$$v_k^i(z_k^i; \pi^j) \geq v_k^i(\bar{z}_k^i; \pi^j) \text{ when } z_k^i = (1, t) \text{ and } \bar{z}_k^i = (1, \tau).$$

Proof is in Appendix. ■

A. Completing the proof of Theorem 1

We need to prove that for every lock, there exists a threshold policy which is in best response against any fixed strategy of the opponents. From Theorem 3 we have the existence of the optimal policy (BR policy against any fixed π^j) for every lock. Let's denote the optimal policy for acquiring k -th lock by $a^*(\cdot)$, and, say it is not a threshold policy. Let τ_{a^*} denote the contact epoch under policy a^* . If a^* is the optimal policy, it must be the optimizer (maximizer) of the k -th stage DP equation (8) and so,

$$v_k^i(z_k^i; \pi^j) = r_k^i(z_k^i, a^*; \pi^j) + E[v_{k+1}^i((l, \tau_{a^*}); \pi^j) | z_k^i, a^*].$$

Construct a threshold policy Γ_θ using a^* as in proof of Lemma 1 and then from the same Lemma:

$$r_k^i(z_k^i, a^*; \pi^j) \leq r_k^i(z_k^i, \Gamma_\theta; \pi^j);$$

from Lemma 2 the value function is a non-increasing function of time, and hence again using Lemma 1 we have,

$$E[v_{k+1}^i((l, \tau_{a^*}); \pi^j) | z_k^i, a^*] \leq E[v_{k+1}^i((l, \tau_\theta); \pi^j) | z_k^i, \Gamma_\theta].$$

This proves that a threshold policy is among BR policies against π^j , using (8). ■

B. Proof of Theorem 2

We prove the second Theorem in two steps; in the first step we show that the optimal policies corresponding to search of k -th lock coincide in all possible time intervals, irrespective of the start of this search, τ_{k-1}^i , in the following sense:

Theorem 4: Let $\tau \geq t$. The optimal/BR policy to acquire k -th lock, $a_k^{i*}(\cdot; z)$ with $z = (1, t)$ coincides with BR policy $a_k^{i*}(\cdot; z')$ with $z' = (1, \tau)$ from τ onwards, i.e.,

$$a_k^{i*}(s; z) = a_k^{i*}(s; z') \text{ for all } \tau \leq s \leq T.$$

Proof: The optimal policy to acquire k -th lock, $a_k^{i*}(\cdot; z)$ with $z = (1, t)$ is optimizer of the value function given in equation (11), i.e., $u(t, 0) = \sup_{a \in L^\infty[t, T]} J(t, 0, a)$.

From Dynamic programming principle of optimal control problems [3, Theorem 5.12], we have:

$$u(t, 0) = \sup_{a \in L^\infty[t, \tau]} \left\{ \int_t^\tau (h_k^i(s) - \nu x(s)) a(s) e^{-x(s)} ds + u(\tau, x(\tau)) \right\} \text{ for any } t \leq \tau \leq T. \quad (13)$$

As in [3, Lemma 4.2] one needs to find the optimizer for the time interval $[t, \tau]$, considering that the optimal control from τ onwards will be the same as the one that obtains, the optimal $u(\tau, x(\tau))$, where $x(\tau)$ is the state at τ . And if both the problems have optimal policy (the existence for our case is established as in Theorem 3), then the optimal policy for the entire interval is given by (as in [3, Page 10]):

$$a_k^{i*}(s) = \begin{cases} a_1^*(s) & \text{for all } s < \tau, \\ a_2^*(s) & \text{for all } s > \tau, \end{cases} \quad (14)$$

where $a_2^*(s)$ is the optimal policy attaining $u(\tau, x^*(\tau))$, $x^*(\tau)$ is state at τ when a_1^* is used in interval $[t, \tau]$ and where a_1^* is the optimizer of (13). The optimal control from τ onwards in general depend on state $x(\tau)$ at time τ , but in our case, by Lemma 3 given in Appendix, (13) modifies to:

$$u(t, 0) = \sup_{a_k^i \in L^\infty[t, \tau]} \left\{ \int_t^\tau (h_k^i(s) - \nu x(s)) a_k^i(s) e^{-x(s)} ds + e^{-x(\tau)} [u(\tau, 0) - \nu x(\tau)] \right\}.$$

By Lemma 3, the optimal control from τ onwards is independent of the state at τ , i.e., the optimal control policies defining $u(\tau, x(\tau))$ and $u(\tau, 0)$ are the same, and the common one forms a part of a_k^{i*} (see (14)); this completes the proof. ■

Thus it suffices to optimize for every lock with $z_k = (1, 0)$ (i.e., with $\tau_k = 0$) and the rest of the optimal policies (with different starting time instances) can be constructed using this *zero-starting optimal policies*, which immediately leads to the following corollary:

Corollary 5: The optimal (BR) strategy can be completely specified by a finite (M) collection of control policies, $\pi_0^{i*}(\pi^j) := \{a_{0_k}^{i*}(\cdot)\}$, one for each lock and each of them starting at time zero and spanning till time T , and such that:

$$\begin{aligned} \pi^{i*}(\pi^j) &= \{a_k^{i*}(\cdot; z_k), \text{ for all } z_k\}, \text{ where} \\ a_1^{i*}(s; z_1) &= a_{0_1}^{i*}(s), \text{ and, for any } k > 1 \\ a_k^{i*}(s; z_k) &= a_{0_k}^{i*}(s) \text{ for all } s \geq \tau_{k-1} \text{ when } z_k = (1, \tau_{k-1}), \\ a_k^{i*}(s; z_k) &= 0 \text{ for all } s \text{ when } z_k = (0, \tau_{k-1}). \quad \blacksquare \end{aligned}$$

By further using Theorem 1, the (zero-starting) BR policies $a_{0_k}^{i*}(\cdot)$ can be chosen to be threshold policies; in other words any best response strategy can be described completely using M -thresholds, say call them $\theta_1^{i*}, \dots, \theta_M^{i*}$. This implies the existence of an MT-strategy $(\theta_1^{i*}, \dots, \theta_M^{i*})$ among the BR strategies against any given strategy profile of opponents, which completes the proof of Theorem 2. ■

IV. REDUCED GAME G

By Theorem 1 any best response includes a threshold or T-strategy and further by Theorem 2 at least one of the best response strategies is an MT-strategy. By virtue of these results, one can find a NE (if it exists) in a much reduced game; the space of strategies in the original game is infinite dimensional while that in the reduced game would be \mathcal{R}^M . We would show that there indeed exists a unique NE in the reduced game and analyze it by further reducing the dimension of the game to one.

We can reduce the problem to the following game, $G = \langle N, S, \Phi \rangle$, where N is the set of players as before, S the set of strategies of each player is simplified to a bounded set of M dimensional vectors (basically the set of MT-strategies):

$$S^i = \{\underline{\theta}^i = (\theta_1^i, \theta_2^i, \dots, \theta_M^i); \theta_k^i \in [0, T] \forall k\},$$

and the utilities $\Phi = \{\phi^i\}$ are now redefined by the following:

$$\begin{aligned} \phi^i(\underline{\theta}^i; \underline{\theta}^j) &= \phi^i(\underline{\theta}^i; \theta_1^j) = \sum_{k=1}^M E[r_k^i(z_k^i, \theta_k^i; \theta_1^j)] \text{ where} \\ \underline{\theta}^j &:= \{\underline{\theta}^m\}_{m \neq i} \text{ and } \theta_1^j := \{\theta_1^m\}_{m \neq i}; \end{aligned}$$

the above objective function depends only upon the first thresholds of the opponents (θ_1^j) because: a) once the player gets the first lock successfully, the opponents have no incentives to try for the further locks, as (and after) their first contact is unsuccessful; b) the success of any agent for first lock depends upon the failure of other agents for the same lock and hence on θ_1^j ; c) the redefined terms (e.g., r_k^i , ϕ^i etc) depend only upon the MT-strategies, the thresholds of which are defined using zero-starting optimal policies, $\{a_{0_k}^{i*}\}$ of Corollary 5; and d) thus the thresholds do not depend upon τ_1^i , the contact time of the first lock.

The simplified expressions under these special strategies are provided below. Recall the k -th component of the vector (i.e., θ_k^i) represents time threshold till which one should attempt to contact k -th lock with full intensity β^i , if the previous contact is before θ_k^i ; otherwise one would not attempt for the next lock. Hence the cost under MT strategies simplifies to (as in (10) and by rewriting the last term as an appropriate integral in equality 'a'):

$$\begin{aligned} r_k^i(z_k^i, \theta_k^i; \theta_1^j) &= 0 \text{ if } \tau_{k-1}^i \geq \theta_k^i, \text{ else it equals} \\ r_k^i(z_k^i, \theta_k^i; \theta_1^j) &= -\nu \beta(\theta_k^i - \tau_{k-1}^i) e^{-\beta^i(\theta_k^i - \tau_{k-1}^i)} \\ &\quad + \int_{\tau_{k-1}^i}^{\theta_k^i} (c_k^i \eta_k^i(s) - \nu \beta^i(s - \tau_{k-1}^i)) \beta^i e^{-\beta^i(s - \tau_{k-1}^i)} ds \\ &\stackrel{a}{=} \int_{\tau_{k-1}^i}^{\theta_k^i} (c_k^i \eta_k^i(s) - \nu) \beta^i e^{-\beta^i(s - \tau_{k-1}^i)} ds, \text{ with} \\ \eta_k^i(s) &= \mathcal{X}_{\{k=1\}} e^{-\sum_{m \neq i} \beta^m(s \wedge \theta_1^m)} + \mathcal{X}_{\{k>1\}}. \end{aligned}$$

Once again we start with BR analysis, and BR (of agent i) will be the maximizer of the following objective function

$$\Upsilon_1^{i*}(z_1; \theta_1^j) = \max_{\underline{\theta}^i} \phi^i(\underline{\theta}^i; \theta_1^j) = \max_{\{\theta_1^i, \dots, \theta_M^i\}} \sum_{k=1}^M E[r_k^i(z_k^i, \theta_k^i; \theta_1^j)].$$

By applying Theorem 2 for finding best response against MT-strategies of the opponents ($\pi^j := \{\theta^m\}_{m \neq i}$) we have:

$$\sup_{\pi^i = \{a_1(z_1)\}, \dots, \{a_M(z_M)\}} \phi^i(a_k, \dots, a_k; z_1; \pi^j) = \max_{\theta_1^i, \dots, \theta_M^i} \phi^i(\theta^i; \theta^j),$$

because the optimal strategies can be chosen to be MT-strategies. Now we apply DP equations to obtain the following, which is further simplified (in the second equation) by choosing an MT-strategy as the optimal strategy for Υ_2^{i*} (once again by Theorem 2):

$$\Upsilon_1^{i*}(z_1^i; \theta_1^j) = \max_{\theta_1^i} \gamma^i(\theta_1^i; \theta_1^j) \text{ where,} \quad (15)$$

$$\gamma^i(\theta_1^i; \theta_1^j) := \int_0^{\theta_1^i} ((c_1^i + \Upsilon_2^{i*}(t_1))\eta_1^i(t_1) - \nu) \beta^i e^{-\beta^i t_1} dt_1 \text{ with} \\ \Upsilon_2^{i*}(t) := \max_{\{\theta_2^i, \dots, \theta_M^i\}} \sum_{k \geq 2} E[r_k^i(z_k^i, \theta_k^i) | z_2^i = (1, t)]. \quad (16)$$

We now proceed with analysing the above BR and then the reduced game, as a first step, we obtain the structural properties of $\Upsilon_2^{i*}(t)$ (proof is in Appendix R):

Theorem 6: Define the following backward recursively:

$$\begin{aligned} \theta_M^{i*} &= T\mathcal{X}_{\{c_M^i > \nu\}} \text{ and} \\ \tilde{\Upsilon}_M^{i*}(t) &= (c_M^i - \nu)(1 - e^{-\beta^i(T-t)})\mathcal{X}_{\{c_M^i > \nu\}}, \\ \text{and for any } 2 \leq k < M \text{ (with } \emptyset - \text{null set)} \\ \theta_k^{i*} &:= \inf\{t \geq 0 : c_k^i + \tilde{\Upsilon}_{k+1}^{i*}(t) \leq \nu\}, \quad \inf \emptyset := T, \quad (17) \\ \tilde{\Upsilon}_k^{i*}(t) &= \mathcal{X}_{\{t < \theta_k^{i*}\}} \int_t^{\theta_k^{i*}} (c_k^i + \tilde{\Upsilon}_{k+1}^{i*}(s) - \nu) \beta^i e^{-\beta^i(s-t)} ds. \end{aligned}$$

Then, i) For any k , the function $\tilde{\Upsilon}_k^{i*}$ is strictly decreasing with t for all $t < \theta_k^{i*}$, after which it remains at 0. Further the co-efficients in (17) are uniquely defined.

ii) The function $\Upsilon_2^{i*}(\cdot)$ defined in (16) equals $\tilde{\Upsilon}_2^{i*}(\cdot)$, the former is optimized by unique optimizers $\{\theta_k^{i*}\}_{k \geq 2}$ (defined in (17)) and is strictly decreasing/remains at 0 as in (i). ■

In the view of the above discussions, the game breaks into two problems. a) an optimization problem to find the optimal thresholds from the second lock onwards, which is analysed in Theorem 6; b) a further reduced one dimensional game with utilities given by $\{\gamma^i\}$ of (15), where each player (say player i) has to choose threshold (θ_1^i) , used for searching the first-lock keeping in view of Υ_2^{i*} and the strategies of the opponents.

It is easy to observe that in the further reduced game, the utilities given in equation (15) depend upon one dimensional strategy θ_1^i and one dimensional strategies of the opponents θ_1^j . For any θ_1^j fixed, the partial derivative of γ^i with respect to θ_1^i is given by:

$$\left((c_1^i + \Upsilon_2^{i*}(\theta_1^i))\eta_1^i(\theta_1^i) - \nu \right) \beta^i e^{-\beta^i \theta_1^i}, \quad (18)$$

which is strictly decreasing by Theorem 6, by the definition of η_1^i and because $e^{-\beta^i \theta_1^i}$ is strictly decreasing. Thus the utility function is strictly concave in θ_1^i . Observe that the utility function is also continuous in θ_1^j , therefore, we have a n -player concave game. Further by strict monotonicity of the derivative (18), the reduced game satisfies *strict diagonal concavity* given by [5, equation (3.10)]. Thus by [5, Theorem

2], we have unique NE for the reduced game. It is easy to verify further details of the following Theorem:

Theorem 7: The unique NE is given by the sequence of thresholds (for second lock onwards) as given in Theorem 6 (one sequence for each player) and the first lock thresholds that simultaneously satisfy the following (for all $1 \leq i \leq n$):

$$\theta_1^{i*} = \inf \left\{ t : (\Upsilon_2^{i*}(t) + c_1^i) e^{-\sum_{m \neq i} \beta^m (t \wedge \theta_1^{m*})} \leq \nu \right\} \wedge T. \quad \blacksquare \quad (19)$$

Some examples

Symmetric case with large T : When $M = 2$ and consider the case with large T . All symmetric agents. By symmetry and uniqueness, $\theta_k^{i*} = \theta_k^*$ for all $i \leq n$. As $T \rightarrow \infty$, we have that

$$\Upsilon_2^*(t) = (c_2 - \nu)(1 - e^{-\beta(T-t)})\mathcal{X}_{\{c_2 > \nu\}} \approx (c_2 - \nu)^+ \text{ for } t \ll T.$$

All the functions defining infimum in (19) are continuous and hence infimum is achieved and hence,

$$\begin{aligned} \theta_2^* &= T\mathcal{X}_{\{c_2 > \nu\}} \text{ and} \\ \theta_1^* &\approx \max \left\{ 0, -\frac{1}{(n-1)\beta} \log \left(\frac{\nu}{(c_2 - \nu)^+ + c_1} \right) \right\} \end{aligned}$$

In fact when you substitute the approximate $\Upsilon_M^*(t)$ in $\Upsilon_{M-1}^*(t)$, we obtain (again with approximation as $\theta_M^* = T\mathcal{X}_{\{c_M > \nu\}}$):

$$\begin{aligned} \theta_{M-1}^* &= T\mathcal{X}_{\{(c_M - \nu)^+ + c_{M-1} - \nu > 0\}}, \\ \Upsilon_{M-1}^*(t) &= ((c_M - \nu)^+ + c_{M-1} - \nu)(1 - e^{-\beta(T-t)}) \\ &\approx ((c_M - \nu)^+ + c_{M-1} - \nu)^+ \text{ for } t \ll T. \end{aligned}$$

Progressing this way for all $k > 1$, define $\bar{c}_l^k := \sum_{l'=l}^k c_{l'}$.

$$\begin{aligned} \theta_k^* &= T\mathcal{X}_{\{\bar{c}_k^{M^o_{k+1}} \geq (M_{k+1}^o - k + 1)\nu\}} \text{ where } M_{k+1}^o = M - \mathcal{X}_{\{c_M < \nu\}}, \\ M_k^o &:= M_{k+1}^o \mathcal{X}_{\{\bar{c}_k^{M_{k+1}^o} \geq (M_{k+1}^o - k + 1)\nu\}} + (k-1)\mathcal{X}_{\{\bar{c}_k^{M_{k+1}^o} < (M_{k+1}^o - k + 1)\nu\}} \\ \text{and} \\ \Upsilon_k^*(t) &\approx (\bar{c}_k^{M_{k+1}^o} - (M_{k+1}^o - k + 1)\nu)^+ \text{ for } t \ll T, \end{aligned}$$

and then

$$\theta_1^* \approx \max \left\{ 0, -\frac{1}{(n-1)\beta} \log \left(\frac{\nu}{\bar{c}_1^{M_2^o} - (M_2^o - 1)\nu} \right) \right\}.$$

Monotone case with large T : We consider an asymmetric case, in which the costs are monotone, i.e., without loss of generality assume $c_k^i \geq c_{k+1}^i$ for each i , and that $c_M^i \geq M\nu$. Also assume $\beta^i = \beta$ for all i . We would derive the results that would be accurate for large T and verify the same using numerical results.

With significantly large T , it is easy to observe (for all i): $\Upsilon_M^{i*} = (c_M^i - \nu)(1 - e^{-\beta(T-t)}) \approx (c_M^i - \nu) \quad \forall t \ll T$ and $\theta_M^{i*} = T$.

By substituting this approximation in $\Upsilon_{M-1}^*(t)$ ($\theta_{M-1}^{i*} = T$): $\Upsilon_{M-1}^{i*}(t) \approx \sum_{k=M-1}^M (c_k^i - \nu)(1 - e^{-\beta(T-t)}) \approx \sum_{k=M-1}^M (c_k^i - \nu)$ for $t \ll T$.

Progressing similarly for all $k > 1$, (with $\bar{c}_k^i := \sum_{l=k}^M c_l^i$):

$$\theta_k^* \approx T, \text{ and } \Upsilon_k^{i*}(t) \approx (\bar{c}_k^i - (M - k + 1)\nu) \text{ for } t \ll T,$$

Since the rewards of the players are monotone, we conjecture the corresponding θ_1^{i*} defining the NE (19) are also

decreasing with i . We will derive the solution of fixed point equations given in (19), which eventually verifies the above conjecture. Further since the functions defining infimum in (19) are continuous (and strictly monotone), the infimum is achieved and hence,

$$\theta_1^{n*} = -\frac{1}{(n-1)\beta} \log\left(\frac{\nu}{\bar{c}_1^{n-1} - (M-1)\nu}\right), \text{ if } \theta_1^{n*} \leq \theta_1^{i*} \text{ for all } i.$$

Now consider the player $n-1$, repeating the same logic and substituting θ_1^{n*} derived in previous step we have:

$$\begin{aligned} \theta_1^{(n-1)*} &= -\frac{1}{(n-2)\beta} \log\left(\frac{\nu}{(\bar{c}_1^{n-1} - (M-1)\nu)e^{-\beta\theta_1^{n*}}}\right) \\ &= -\frac{1}{(n-2)\beta} \log\left(\frac{\nu}{(\bar{c}_1^{n-1} - (M-1)\nu)}\right) - \frac{\theta_1^{n*}}{n-2}. \end{aligned}$$

As $\bar{c}_1^{n-1} > \bar{c}_1^n$, it is indeed true that $\theta_1^{(n-1)*} \geq \theta_1^{n*}$. Progressing in exactly similar way, and verifying at every step the required monotonicity of $\{\theta_1^{j*}\}$, we obtain:

$$\theta_1^{(n-i)*} = \frac{-\log\left(\frac{\nu}{\bar{c}_1^{n-i} - (M-1)\nu}\right) - \beta \sum_{j=n-i+1}^n \theta_1^{j*}}{(n-i-1)\beta}. \quad (20)$$

Remarks: This solution is exact when $M = 1$, thus we solved the problem completely for this case (solution matches with that in [4] when $n = 2$); here the agents attempt to acquire the lock/destination without knowing if it already taken by others. For other cases it is an approximation, the accuracy is verified in the next section.

V. NUMERICAL EXAMPLES

We consider some numerical examples with an aim to reinforce the theoretical results. We also demonstrate that the approximation (20) is good even for moderate T . The first example considers symmetric case with moderate $T = 8$ and the results are in Figure 1. The other details are: $M = 5$, $n = 4$, $\beta = 1$, $c_1 = 1$, $c_2 = 3$, $c_3 = 3$, $c_4 = 3$, and $c_5 = 3$. We plot the NE for varying values of ν . Theoretical results imply $\theta_k^* = T$ (for all $k > 1$) and we observe the same, and, thus we plot only θ_1^* . We computed the NE by solving the fixed point equation (19), using fixed point iterates, we also plot the theoretical approximation (20), and the two curves are indistinguishable (see Figure 1).

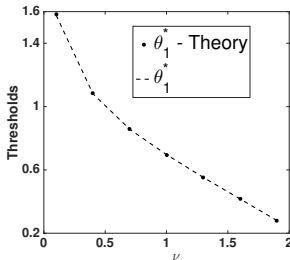


Fig. 1. Theoretical and numerical θ_1^* against ν (for large T)

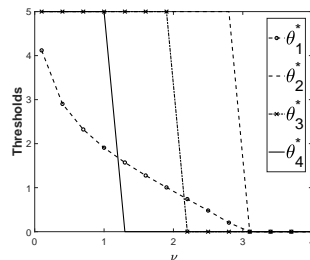


Fig. 2. The NE thresholds against ν (for small T)

Further, to analyse the NE with small $T = 5$, we consider a second case with $M = 4$ and $n = 2$ in Figure 2; we also set $\beta = 1$, $c_1 = 4$, $c_2 = 3$, $c_3 = 2$ and $c_4 = 1$. For large values of ν , the optimal thresholds are less than T even for $k > 1$, thus even after acquiring the first (or a consecutive)

lock the agent will not continue further if any of latter locks are not acquired before the corresponding thresholds. Also observe that, as the ν increases, the optimal threshold for $k = 4$ becomes zero, while others are positive implying the agents will only attempt for three locks. As ν increases further, the optimal threshold for $k = 3$ also becomes zero, while others are positive implying the agents will only attempt for two locks.

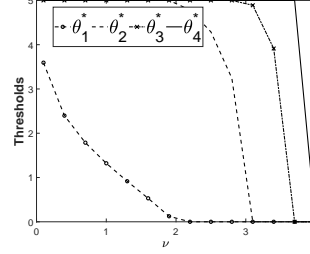


Fig. 3. The NE thresholds for different values of ν when $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $c_4 = 4$

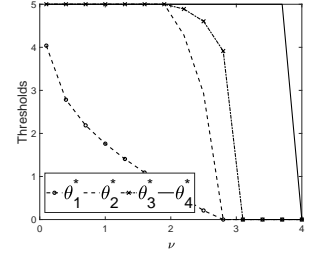


Fig. 4. The NE thresholds for different values of ν when $c_1 = 4$, $c_2 = 2$, $c_3 = 2$, $c_4 = 4$

Further, to analyse the NE with small $T = 5$, we consider more examples with $M = 4$ and $n = 2$ and $\beta = 1$. In Figure 3 we set $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $c_4 = 4$ and we see, when ν approaches 2.5, $\theta_1^* = 0$, which implies the agents won't attempt the locks even though other thresholds are positive. In Figure 4 we took the rewards associated with locks to be $c_1 = 4$, $c_2 = 2$, $c_3 = 2$, $c_4 = 4$. In all the examples, we observe that the optimal thresholds are decreasing (non-increasing) with ν .

CONCLUSIONS

We considered acquisition games with partial, asymmetric information. Agents attempt to acquire M destinations, the first one to contact a destination acquires it; destinations can be acquired only in a given order. When an agent acquires a destination and if it has also acquired all the previous ones, it gets some reward. The agents are not aware of the acquisition status of others. It is possible that an agent continues its acquisition attempts, while the destination is already acquired by another agent. Thus we have a partial and asymmetric information game. The agents control the rate of their Poisson clocks to contact the destinations; they also incur a cost proportional to their rates of contact. We found NE of this asymmetric game by reducing it to a much simpler game such that the NE of the reduced game would also be the NE of the original game; the original game has infinite number of state and has infinite dimensional actions. We proved that a tuple of time-threshold policies form the unique NE of the reduced game. We also provided an algorithm to compute this NE. We further have approximate closed form expressions for the NE.

REFERENCES

- [1] T. Basar and J. B. Cruz Jr, "Concepts and methods in multi-person coordination and control." ILLINOIS UNIV AT URBANA DECISION AND CONTROL LAB, Tech. Rep., 1981.

- [2] A. Eitan et al., "A stochastic game approach for competition over popularity in social networks," *Dynamic Games and Applications*, vol. 3, no. 2, pp. 313–323, 2013.
- [3] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*. Springer Science & Business Media, 2006, vol. 25.
- [4] Veeraruna Kavitha, Mayank Maheshwari, and Eitan Altman "Acquisition Games with Partial-Asymmetric Information," *Allerton 2019, USA*, also downloadable at <https://arxiv.org/abs/1909.06633>.
- [5] J Ben Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.
- [6] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [7] Emilio Roxin et al., "The existence of optimal controls," *The Michigan Mathematical Journal*, 9(2):109–119, 1962.

APPENDIX

Proof of Theorem 3: The DP equation for any k can be rewritten as:

$$u(s, x) = \sup_{a(\cdot)} \left\{ \int_s^T L_k(s', x(s'), a(s')) ds' + g(x(T)) \right\} \text{ with}$$

$$L_k(s', x, a) = (c_k^i + v_{k+1}^i((1, s'); \pi^j)) \eta_k^i(s'; \pi^j) - \nu x) a e^{-x},$$

$$\eta_k^i(s', \pi^j) = e^{-\sum_{m \neq i} \bar{a}^m(s')} \mathcal{X}_{\{k=1\}} + \mathcal{X}_{\{k>1\}}.$$

It is easy to solve (8) with $k = M$, because v_{M+1}^i is defined to be 0 (details in [4], also discussed in Theorem 6):

$$v_M^i(1, s) = (c_M^i - \nu) \left(1 - e^{-\beta^i(T-s)} \right) \mathcal{X}_{\{c_M^i > \nu\}}.$$

Observe that $v_M^i(1, s)$ is a differentiable function of s , hence clearly L_{M-1} is Lipschitz continuous on $[t, T] \times [0, \beta T] \times [0, \beta]$. It is also bounded. Further clearly the RHS of the ODE and the terminal cost g are all bounded and Lipschitz continuous. Thus by [3, Theorem 10.1 and the following Remark 10.1] the value function $u(\cdot, \cdot)$ is unique viscosity solution which is Lipschitz continuous when $k = M-1$. This implies $v_{M-1}^i(1, s)$ is Lipschitz continuous in s , further it is also bounded. This implies L_{M-2} is also Lipschitz continuous and bounded which proves the same for $v_{M-2}^i(1, s)$. By backward induction on k , the part (i) is true.

For part (ii) we apply the results of [7]. Towards this the optimal control problem can be converted into Mayer-type (finite horizon problem with only terminal cost) by usual technique of augmenting a new component to state which represents

$$y(s') := \int_s^{s'} L_k(\tilde{s}, x(\tilde{s}), a(\tilde{s})) d\tilde{s}$$

and equivalently maximizing $y(T) + g(x(T))$. By part (i) all the required assumptions [7, Assumptions (i) to (vii)] are satisfied, with compact control space $U = [0, \beta]$, compact state space $\hat{X} = [0, \beta T]$ (it is easy to verify that the state variable could be confined to this range); assumptions (i)-(ii) are trivially satisfied; assumption (iii) follows by part (i); for assumption (iv) one can actually bound by a constant independent of $(s', (x, y))$; easy to verify convexity requirement of (vii), because for any given $(s', (x, y))$ the set in mention is an interval.

This proves part (iii). \blacksquare

Proof of Lemma 1: For any lock k , given any state z_k^i , consider the open loop policy $a_k^i(\cdot)$, we denote it as $a(\cdot)$ for the ease of notation. If $a(\cdot)$ is already a threshold type,

we will have nothing to prove. If not, choose two intervals $[t_1, t_1 + \delta_1]$ and $[t_2, t_2 + \delta_2]$ with $t_2 \geq t_1 + \delta_1$ such that

$$\int_{t_1}^{t_1+\delta_1} a(t) dt < \beta \delta_1 \text{ and } \int_{t_2}^{t_2+\delta_2} a(t) dt > 0.$$

One can further ensure (by choosing appropriate end points) that

$$\int_{t_1}^{t_1+\delta_1} a(t) dt + \int_{t_2}^{t_2+\delta_2} a(t) dt = \beta \delta_1.$$

Now we construct another policy, $a'(t)$ such that,

$$\int_{t_1}^{t_1+\delta_1} a'(t) dt = \beta \delta_1 \text{ and } \int_{t_2}^{t_2+\delta_2} a'(t) dt = 0,$$

and on rest of the intervals the policy $a'(\cdot)$ matches completely with policy $a(\cdot)$. This new policy is basically constructed by shifting the mass from a later interval $[t_2, t_2 + \delta_2]$ to a former interval $[t_1, t_1 + \delta_1]$ in policy $a(\cdot)$, note that if one can't find such intervals, it implies that the policy $a(\cdot)$ itself is a Threshold policy and we will have nothing to prove.

Observe that for all $t < t_1$ we have,

$$a(t) = a'(t) \text{ and hence } \bar{a}(t) = \int_{\tau_{k-1}}^t a(s) ds = \bar{a}'(t)$$

similarly for all $t_1 < t < t_2$, $\bar{a}(t) < \bar{a}'(t)$ and for all $t_2 < t < t_2 + \delta$ we have $\bar{a}(t) \leq \bar{a}'(t)$ and for all $t > t_2 + \delta$ we have $\bar{a}(t) = \bar{a}'(t)$.

This implies, the time to contact the k -th lock with policy $a(\cdot)$ denoted as τ_a is stochastically dominated by that with policy $a'(\cdot)$ denoted as $\tau_{a'}$ as explained below; consider the CDFs with both the policies; i) for any $x < t_1$,

$$F_a(x) = \text{Prob}(\tau_a \leq x) = 1 - e^{-\bar{a}(x)} = 1 - e^{-\bar{a}'(x)} = F_{a'}(x),$$

ii) for any $x \in (t_1, t_1 + \delta_1)$, we have $\bar{a}(t) < \bar{a}'(t)$ and so

$$F_a(x) = 1 - e^{-\bar{a}(x)} < 1 - e^{-\bar{a}'(x)} = F_{a'}(x),$$

iii) for any $x \in (t_2, t_2 + \delta_2)$, we have $\bar{a}(t) \leq \bar{a}'(t)$

$$F_a(x) = 1 - e^{-\bar{a}(x)} \leq 1 - e^{-\bar{a}'(x)} = F_{a'}(x),$$

and iv) for any $x \in [t_2 + \delta_2, T]$, we have $\bar{a}(t) = \bar{a}'(t)$

$$F_a(x) = 1 - e^{-\bar{a}(x)} = 1 - e^{-\bar{a}'(x)} = F_{a'}(x).$$

This proves the required stochastic dominance, $\tau_a \stackrel{d}{\leq} \tau_{a'}$. Now, let $\tau^j = \min_{m \neq i} \tau^m$ where τ^m denotes the contact epoch of m -th agent, and observe

$$\{\omega : \tau_a(\omega) \leq \tau^j(\omega)\} \subset \{\omega : \tau_{a'}(\omega) \leq \tau^j(\omega)\},$$

which is same as saying,

$$\{\text{success with policy } a\} \subset \{\text{success with policy } a'\}$$

and hence,

$$P_k^i(z_k^i; a; \pi^j) \leq P_k^i(z_k^i; a'; \pi^j) \text{ for any } k, z_k.$$

Further observe that the expected cost with policy $a(\cdot)$ (by change of variables)

$$\begin{aligned}
E[\bar{a}(\tau_a); \tau_a < T] + \bar{a}(T)e^{-\bar{a}(T)} \\
&= \int_0^T \bar{a}(t)e^{-\bar{a}(t)} a(t) dt + \bar{a}(T)e^{-\bar{a}(T)} \\
&= \int_0^{\bar{a}(T)} x e^{-x} dx + \bar{a}(T)e^{-\bar{a}(T)} = 1 - e^{-\bar{a}(T)},
\end{aligned}$$

which is the same as that using a' because $\bar{a}(T) = \bar{a}'(T)$.

$$E[\bar{a}(\tau_a); \tau_a < T] + \bar{a}(T)e^{-\bar{a}(T)} = 1 - e^{-\bar{a}(T)},$$

which is the same as that using a' because $\bar{a}(T) = \bar{a}'(T)$.

One can keep on improving the policy until it becomes a threshold policy. This completes the proof. ■

Proof of Lemma 2: For any lock k , given any state z_k^i , the k -th stage DP equation can be re-written as the following optimal control problem (see equation (11))

$$u(t, 0) = \sup_{a(\cdot) \in L^\infty} \left\{ \int_t^T (h_k^i(s') - \nu x(s')) a(s') e^{-x(s')} ds' + g(x(T)) \right\}.$$

From the Dynamic programming principle [3, Theorem 5.1], we can rewrite it as follows;

$$\begin{aligned}
u(t, 0) &= \sup_{a_k^i \in L^\infty[t, \tau]} \left\{ \int_t^\tau (h_k^i(s) - \nu x(s)) a_k^i(s) e^{-x(s)} ds \right. \\
&\quad \left. + u(\tau, x(\tau)) \right\}, \text{ where,}
\end{aligned}$$

$$u(\tau, x(\tau)) = \sup_{a_k^i \in L^\infty[\tau, T]} J(\tau, x(\tau), a_k^i).$$

Observe that the function $J(\tau, x(\tau), a_k^i)$ (see equation (11)) has same structure as function $J_h(t, x, a)$ in Lemma 3 with $h = h_k^i$, $t = \tau$, $x = x(\tau)$ and $a = a_k^i$ and continuity follows by Theorem 3. Hence we have,

$$\begin{aligned}
u(\tau, x(\tau)) &= e^{-x(\tau)} [u(\tau, 0) - \nu x(\tau)], \text{ and so} \\
u(t, 0) &= \sup_{a_k^i \in L^\infty[t, \tau]} \left\{ \int_t^\tau (h_k^i(s) - \nu x(s)) a_k^i(s) e^{-x(s)} ds \right. \\
&\quad \left. + e^{-x(\tau)} [u(\tau, 0) - \nu x(\tau)] \right\}.
\end{aligned}$$

In the above equation if we consider zero policy, i.e., if we consider policy $a_k^i([t, \tau]) \equiv 0$ (basically $a_k^i(s) = 0$ for all $s \in [t, \tau]$), then clearly $x(\tau) = 0$, the first term (integral) in the above supremum is zero and hence:

$$u(t, 0) \geq u(\tau, 0).$$

This implies, $v_k^i(z_k^i; \pi^j) \geq v_k^i(\bar{z}_k^i; \pi^j)$, as $u(\tau, 0)$ is the value function of optimal control problem when the control starts in the state \bar{z}_k^i . ■

Lemma 3: Let $J_h(t, x, a)$ be a function of the form

$$J_h(t, x, a) = \int_t^T (h(s) - \nu x(s)) a(s) e^{-x(s)} ds - \nu x(T) e^{-x(T)},$$

defined using continuous function $h(\cdot)$ and state process

$$\dot{x}(s) = a(s), \text{ with initial condition, } x(t) = x.$$

Define $u(t, x) := \sup_{a \in L^\infty} J_h(t, x, a)$, then we have:

$$(i) J_h(t, x, a) = e^{-x} [J_h(t, 0, a) - \nu x]$$

$$(ii) u(t, x) = e^{-x} [u(t, 0) - \nu x]$$

(iii) The optimal policy $a^*(\cdot)$ is independent of x .

Proof: By change of variables $x(s) = x + \tilde{x}(s)$,

$\dot{\tilde{x}}(s) = a(s)$, i.e., $\tilde{x}(s) = \int_t^s a(\tilde{s}) d\tilde{s}$, and $\tilde{x}(t) = 0$, we get,

$$\begin{aligned}
J_h(t, x, a) &= \int_t^T (h(s) - \nu(x + \tilde{x}(s))) a(s) e^{-(x + \tilde{x}(s))} ds \\
&\quad - \nu(x + \tilde{x}(T)) e^{-x - \tilde{x}(T)} \\
&= e^{-x} \left(\int_t^T (h(s) - \nu(x + \tilde{x}(s))) a(s) e^{-\tilde{x}(s)} ds \right. \\
&\quad \left. - \nu(x + \tilde{x}(T)) e^{-\tilde{x}(T)} \right) \\
&= e^{-x} \left(\int_t^T (h(s) - \nu \tilde{x}(s)) a(s) e^{-\tilde{x}(s)} ds - \nu \tilde{x}(T) e^{-\tilde{x}(T)} \right. \\
&\quad \left. - \nu x \int_t^T a(s) e^{-\tilde{x}(s)} ds - \nu x e^{-\tilde{x}(T)} \right) \\
&= e^{-x} (J_h(t, 0, a) - \nu x).
\end{aligned}$$

The last equality follows because the sum of two terms (probabilities) is 1. This completes part (i). For part (ii), from the above equation we have,

$$u(t, x) = \sup_{a \in L^\infty} J_h(t, x, a) = \sup_{a \in L^\infty} e^{-x} [J_h(t, 0, a) - \nu x]$$

and hence we have, $u(t, x) = e^{-x} [u(t, 0) - \nu x]$. This proves part (ii). Further it is clear from above that the optimal policy $a^*(\cdot)$ remains the same for all initial conditions x and this proves part (iii). ■

APPENDIX R: REDUCED GAME

Proof of Theorem 6: Define the following, for any $2 \leq l \leq M$ and any MT-strategy $\pi = (\theta_l \dots \theta_M)$ (see (1)):

$$\gamma_l(\theta_l \dots, \theta_M; t) := \sum_{k=l}^M E[\tau_k^i(z_k^i, \theta_k) | z_l^i = (1, t)].$$

Note these objective functions (with $l \geq 2$) do not depend upon the strategies of opponents. Define

$$\Upsilon_l^*(t) := \sup_{\theta_l, \dots, \theta_M} \gamma_l(\theta_l \dots, \theta_M; t),$$

and observe that

$$\Upsilon_2^{i*}(t) = \Upsilon_2^*(t) = \sup_{\theta_2, \dots, \theta_M} \gamma_2(\theta_2 \dots, \theta_M; t).$$

Further, by definition, for any given $(\theta_{k-1}, \dots, \theta_M)$ we have:

$$\begin{aligned}
&\gamma_{k-1}(\theta_{k-1} \dots, \theta_M; t) \\
&= \int_t^{\theta_k} (c_{k-1}^i - \nu + \gamma_k(\theta_k \dots, \theta_M; s)) \beta^i e^{-\beta^i(s-t)} ds \\
&\leq \int_t^{\theta_k} (c_{k-1}^i - \nu + \sup_{\theta'_k, \dots, \theta'_M} \gamma_k(\theta'_k \dots, \theta'_M; s)) \beta^i e^{-\beta^i(s-t)} ds,
\end{aligned} \tag{21}$$

with $\gamma_{M+1} \equiv 0$. Observe that $\sup_{\theta'_k, \dots, \theta'_M} \gamma_k(\theta'_k \dots, \theta'_M; s)$ is the problem of finding the best response against silent opponent (i.e., when none of the opponents are attempting)

with $M - k$ locks. By applying Theorem 2 one can choose the MT-strategy as a best response, i.e.,

$$\sup_{\pi=\{a_k(z_k)\}\cdots,\{a_M(z_M)\}} \gamma_k(a_k\cdots, a_k; s) = \max_{\theta'_k\cdots, \theta'_M} \gamma_k(\theta'_k\cdots, \theta'_M; s),$$

also optimal $(\theta_k^*\cdots, \theta_M^*)$ do not depend upon s , i.e.,

$$\Upsilon_k^*(s) := \max_{\theta'_k\cdots, \theta'_M} \gamma_k(\theta'_k\cdots, \theta'_M; s) = \gamma_k(\theta_k^*\cdots, \theta_M^*; s) \text{ for all } s. \quad (22)$$

And hence we have,

$$\begin{aligned} & \gamma_{k-1}(\theta_{k-1}\cdots, \theta_M; t) \\ & \leq \int_t^{\theta_{k-1}} (c_{k-1}^i - \nu + \Upsilon_k^*(s)) \beta^i e^{-\beta^i(s-t)} ds. \\ & \leq \sup_{\theta'_{k-1}} \int_t^{\theta_{k-1}} (c_{k-1}^i - \nu + \Upsilon_k^*(s)) \beta^i e^{-\beta^i(s-t)} ds. \end{aligned} \quad (23)$$

Consider the case with $k = M$. In this case clearly

$$\begin{aligned} \Upsilon_M^*(s) &:= \max_{\theta'_M} \gamma_M(\theta'_M; s) \\ &= \max_{\theta'_M} \int_t^{\theta'_M} (c_M^i - \nu) \beta^i e^{-\beta^i(s-t)} ds. \end{aligned}$$

The integrand is a strictly decreasing function, which implies the integral is strictly concave and hence has a unique maxima θ_M^* , as given below along with optimal Υ_M^* :

$$\begin{aligned} \Upsilon_M^*(s) &= \gamma_M(\theta_M^*) \text{ with } \theta_M^* = T \mathcal{X}_{\{c_M > \nu\}} \\ &= (c_M^i - \nu) \left(1 - e^{-\beta^i(T-s)}\right) \mathcal{X}_{\{c_M > \nu\}} \text{ for all } s. \end{aligned}$$

Observe that the function $\Upsilon_M^*(\cdot)$ is a strictly decreasing for $s < \theta_M^*$ or remains at zero for all t if $\theta_M^* = 0$ and further the coefficient θ_M^* is unique. In other words, the function is strict decreasing for all $s \leq \theta_M^*$ and remains at 0 after (unique) θ_M^* . Assume the same holds true for all $k = M, M-1 \dots p+1$ (for any p) and then consider $k = p$. From equation (23)

$$\gamma_p(\theta'_p, \dots, \theta'_M; t) \leq \sup_{\theta'_p} \int_t^{\theta'_p} (c_p^i - \nu + \Upsilon_{p+1}^*(s)) \beta^i e^{-\beta^i(s-t)} ds.$$

Since $\Upsilon_{p+1}^*(\cdot)$ is non-increasing function, the integrand in the above inequality is strictly decreasing with s . This implies the upper bounding integral is strictly concave and hence has a unique maximizer, in fact the maximizer equals:

$$\theta_p^* = \inf\{t \geq 0 : c_p^i + \Upsilon_{p+1}^*(t) \leq \nu\} \wedge T, \text{ with } \inf \emptyset := 0.$$

Thus we have for any $\theta'_p, \dots, \theta'_M$:

$$\gamma_p(\theta'_p, \dots, \theta'_M; t) \leq \gamma_p(\theta_p^*, \dots, \theta_M^*; t) \text{ and hence}$$

$$\begin{aligned} \Upsilon_p^*(t) &= \gamma_p(\theta_p^*, \dots, \theta_M^*; t) \\ &= \int_t^{\theta_p^*} (c_p^i - \nu + \Upsilon_{p+1}^*(s)) \beta^i e^{-\beta^i(s-t)} ds \end{aligned}$$

From the above it is clear that the function $\Upsilon_p^*(t)$ is strictly decreasing with t for all $t \leq \theta_p^*$ and $\Upsilon_p^*(t) = 0$ for all $t > \theta_p^*$. The proof is complete by backward induction, with $\Upsilon_k^* = \Upsilon_k^*$ and $\theta_k^{i*} = \theta_k^*$ for all k . ■