Acquisition Games with Partial-Asymmetric Information

Veeraruna Kavitha, Mayank Maheshwari, Eitan Altman

Acquisition Games with Partial-Asymmetric Information

[1]Veeraruna Kavitha, [1]Mayank Maheshwari and [2]Eitan Altman
[1]IEOR, Indian Institute of Technology Bombay, India and [2]INRIA, France

***Abstract*—[1]We consider an example of stochastic games with partial, asymmetric and non-classical information. We obtain relevant equilibrium policies using a new approach which allows managing the belief updates in a structured manner. Agents have access only to partial information updates, and our approach is to consider optimal open loop control until the information update. The agents continuously control the rates of their Poisson search clocks to acquire the locks, the agent to get all the locks before others would get reward one. However, the agents have no information about the acquisition status of others and will incur a cost proportional to their rate process. We solved the problem for the case with two agents and two locks and conjectured the results for N-agents. We showed that a pair of (partial) state-dependent time-threshold policies form a Nash equilibrium.**

I. Introduction

We consider non-classical and asymmetric information (as specified in [1]) based games inspired by the full information games considered in [2]. In [2], agents attempt to acquire M available destinations; each agent controls its rate of advertisement through a Social network to increase its chances of winning the destinations, while trading off the cost for acquisition. They considered full information and no-information games, and considered discrete time policies by uniformization of the controlled Markov process. In full information games, the agents at any point of time know the number of available destinations or equivalently the number of destinations already acquired by one or the other agent. In no-information games, the agents have no information; they don't even know the number of destinations acquired by themselves.

It is more realistic to assume that the agents know the number of destinations acquired by themselves, but would not know the number acquired by others. This leads to partial, asymmetric and non-classical information games, which are the main focus of this paper. 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 knows 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.

Our approach to this problem can be summarized as "open loop control till information update". With no-information, one has to resort to open loop policies. This is the best when one has no access to information updates. With full information one can have closed loop policies, where the optimal action depends upon the state of the system at the decision epoch. 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.

We considered one and two lock acquisition problems, any agent wins reward one if it acquires all the locks and if it is the first one to acquire the locks. The agents have no access to the information/state of the others, however upon contacting a lock they would know if they are the first to contact. We obtained Nash equilibrium for these partial, asymmetric information games; a pair of (partial) state-dependent time threshold policies form Nash equilibrium. We obtained these results (basically best responses) by solving Dynamic programming equations applicable to (partial) information update epochs and each stage of the Dynamic programming equations are solved by solving appropriate optimal control problems and the corresponding Hamiltonian Jacobi equations.

A block chain network is a distributed ledger that is completely open to all nodes; each node will have a copy

[1]The work is partially sponsored by MALENA, the joint research team between IIT Bombay and Inria.

of transactions (in case of currency exchange). If a new transaction is to be added (linked) to the previously existing chain in the form of a new block, it requires the miners (designated special nodes) to search for a key (encryption), that enables it to be added to the ledger. This search of the key requires computational power and time. The first miner to get the right key, gets a financial reward. If the miners would not know the status of the search efforts of others, the resulting game is exactly as in our one lock problem. Two lock problem can be viewed as the extension of one lock problem, wherein a second key is required to gain the reward.

II. Problem Description

Two agents are competing to win a project. There are two or one lock(s) to win the project, and the aim of the agents is to reach these as quickly as possible. The agent that contacts all the locks before the other gets a reward equal to one unit. Further they need to contact the lock(s) before the deadline T.

The contact process is controllable; the agents control the rate of the contact process continuously over time and they would incur some cost for acceleration. The acquisition/contact process is modelled by a Poisson process. The rate of contact process can be time varying over the interval $[0, T]$, it can further depend upon the number of locks acquired etc. The higher the rate, the higher the chances of success, but also higher is the cost of acquisition.

Information structure: The agents have partial/asymmetric information about the locks acquired by various agents and would use available information to design their acceleration strategies optimally. The agents would know at all the times information (contacted/not contacted etc., at a given time) related to its contact attempts, however it has limited access to that of the others. When any agent contacts a lock, it would know if it is successful; we call a contact successful if the agent is the first one to contact that particular lock. If the other agent had contacted the same lock before the tagged customer, the tagged customer would have an unsuccessful contact. The agent gets an update of this information immediately after a contact, and this will also reveal some information about the status of the other agents. For example, upon a contact, if it gets aware of a successful contact, it also gets to know that this is the first one to contact.

Decision epochs: Every agent has a continuous (at all time instances) update of the status of its contact process, however there is a major update in its information only at (lock) contact instances. At these epochs it would know if the contact is successful/unsuccessful which in turn would reveal some information about the state of the other agents. Hence these form the natural decision epochs; an agent can choose a different action at such time points. Further, it is clear that the decision epochs of different agents are not synchronous.

Actions: The agents should choose an appropriate rate (function) for the contact/acceleration process. The rate of contact, for agent i, at any time point can be between $[0, \beta^i]$. The agents choose an action which specifies the acceleration process at the beginning, and change their action every time it contacts a lock (successfully or unsuccessfully). The action at any decision epoch is a measurable acceleration process (that can potentially span over time interval $[0, T]$). To be precise agent i at decision epoch k (the instance at which it contacted the $(k-1)$-customer) chooses an $a_k^i \in L^\infty[0, T]$, as the acceleration process to be used till the next acquisition. Here $L^\infty[0, T]$ is the space of all measurable functions that are essentially bounded, i.e., the space of functions with finite essential supremum norm:

$$||a||_\infty := \inf\{\beta : |a(t)| \leq \beta \text{ for allmost all } t \in [0, T]\}.$$

State: We will have two decision epochs with two lock problem, and one decision epoch with one lock problem, and have corresponding number of major state updates. Let $\mathbf{z}_k^i$ represent the information available to agent i immediately after $(k-1)$-th contact[2]. Here $\mathbf{z}_k^i$ has two components: a) first component is a flag indicating that the contact was successful; b) second component is the time of contact. The first decision epoch (the only decision epoch with one lock problem) is at time 0, and the state $\mathbf{z}_1$ is simply set to $(0, 0)$ to indicate 0 contacts and '0' contact time (which is of no importance since there is no contact yet). The state at the second decision epoch $\mathbf{z}_2 = (f, \tau)$, where flag $f = s$ implies successful contact and $f = u$ implies unsuccessful contact while τ represents the time of first contact. Let τ_k^i represent the (random) k-th contact instance of agent i. Here we view τ_k^i as a fictitious random contact instance which can take any value from $[0, \infty)$ and with $\tau_k^i > T - \tau_{k-1}^i$ indicating that the agent could not contact the k-th lock before deadline.

We distinguish the one lock problem from the two lock problem by using M to represent the number of locks.

Strategy: The strategy of player i

$$\pi^i = \left\{a_k^i(\mathbf{z}_k); \text{ for all possible } \mathbf{z}_k\right\}_{k \leq M},$$

where $a_1(\cdot)$ represents the acceleration process used at start,

[2]By convention, the start of the process commences with 0-th contact.

$a_2(\cdot)$ represents the acceleration process used after contacting one lock and this choice depends upon the available information $\mathbf{z}_2$. To keep notations simple, most of the times and unless clarity is required, state $\mathbf{z}_k$ is not used while representing the actions. One can easily extend this theory to a more general problem with M locks, but the emphasis of this paper would be on the case with $M = 2$. We briefly discuss the extension to a larger M towards the end of the paper.

Rewards/Costs: The reward of any agent is one, only if it succeeds to contact all M locks, and further, if it is the first one to contact the first lock. Let $d_1^i = T \wedge \tau_1^i$ ($\wedge$ implies minimum of the two) and $d_2^i = ((T \wedge \tau_2^i) - \tau_1^i)^+$ respectively represent the durations[3] of the first and the second contact process. The cost spent on acceleration for the k-th contact equals,

$$\bar{a}_k^i(d_k^i), \text{ with } \bar{a}_k^i(s) := \int_0^s a_k^i(t)dt. \tag{1}$$

The paper mainly considers two agent problem. The N agent problem is discussed in section V and the results are conjectured using the two agent results. *Thus for simpler notations, we represent the tagged customer by i while the other customer is indexed by j.*

The expected utility for stage k equals:

$$r_k^i(\mathbf{z}_k, a_k; \pi^j) = \begin{cases} P_M^i 1_{k=M} - \nu E\left[\bar{a}_k^i(d_k^i)\right] & \text{if } \tau_{k-1}^i < T \\ 0 & \text{else,} \end{cases} \tag{2}$$

where P_M^i represents the probability of eventual success (when all the M locks are contacted before T and when the first contact is successful) and $\nu > 0$ is the trade-off factor between the reward and the cost. Note here that the reward (probability of success) is added only to the last stage, i.e., only when $k = M$.

For one lock problem, i.e., when $M = 1$, the probability of success equals

$$P_1^i = \int_0^T P_f^j(s|\pi^j) f_{\tau,1}^i(s)ds, \tag{3}$$

where $P_f^j(s|\pi^j)$ is the probability that the other agent has not contacted the lock before the agent i, i.e., before time $\tau_1^i \approx s$ and (see equation (1) for definition of $\bar{a}_k^i(\cdot)$)

$$f_{\tau,1}^i(s) \quad := \quad \exp\left(-\bar{a}_1^i(s)\right) a_1^i(s) \tag{4}$$

is the density[4] of the associated contact process. Recall the contact process for k-th contact[5], is a Poisson process with time varying rate given by a_k^i. Note simply that the probability of agent j not contacting the first lock before agent i, for the given strategy pairs, equals (more details in the proof of Theorem 1)

$$P_f^j(s|\pi^j) = P\left(\tau_1^j > s | \tau_1^i \approx s\right) = \exp\left(-\bar{a}_1^j(s)\right).$$

In a similar way for the two lock problem,

$$P_2^i = \int_0^T P_f^j(s|\pi^j) \left(\int_s^T f_{\tau,2}^i(u)du\right) \quad f_{\tau,1}^i(s)ds, \tag{5}$$

where $P_f^j(s|\pi^j)$ is the probability that the other agent (agent j) has not contacted the first lock before agent i (which is the same as the one discussed in the one lock problem) and

$$f_{\tau,2}^i(\tau_1^i + s) \quad := \quad \exp\left(-\bar{a}_2^i(s)\right) a_2^i(s) \tag{6}$$

is the density of the second-lock contact process of agent i.

It is easy to observe that for any given $a \in L^\infty$, the expected cost equals (see (1) and with $\tau_0^i := 0$):

$$\begin{aligned} E[\bar{a}_k^i(d_k^i)] &= E\left[\int_0^{d_k^i} a_k^i(s)ds\right] \\ &= \bar{a}_k^i(T - \tau_{k-1}^i)\exp(-\bar{a}_k^i(T - \tau_{k-1}^i)) \\ &\quad + \int_0^{T-\tau_{k-1}^i} \bar{a}_k^i(s)\exp\left(-\bar{a}_k^i(s)\right) a_k^i(s)ds. \end{aligned} \tag{7}$$

If the contact occurs after the deadline T, one has to pay for the entire duration $T - \tau_{k-1}^i$ (with zero reward) and hence the first term in the above equation.

Game Formulation: The overall utility of agent i, when player j chooses the strategy π_j is given by

$$J^i(\pi_i, \pi_j) = \sum_{k=1}^M E[r_k^i(\mathbf{z}_k, a_k;\ \pi_j)]. \tag{8}$$

Thus we have a strategic/normal form non-cooperative game problem and our aim is to find a pair of policies (that depend only upon the available (partial) information) that form the Nash equilibrium (NE).

We begin with a one lock problem in the next section.

[3]As already mentioned, 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.

[4]It is clear that the complementary CDF of τ_1^i (time till agent i contacts the destination, immaterial of whether it is the first one or not and whether it is before T or not), under deterministic policy π^i, is not influenced by the strategies of others and is given by:

$$\begin{aligned} P(\tau_1^i > t) &= \exp\left(-\int_0^t a_1^i(s)ds\right) \text{ and thus its PDF is given by} \\ f_{\tau,1}^i(t) &= \exp\left(-\bar{a}_1^i(t)\right) a_1^i(t) \text{ with } \bar{a}_1^i(s) := \int_0^t a_1^i(s)ds. \end{aligned}$$

[5]We would like to emphasise here that k represents the number of the contacts.

III. One lock problem

We specialize to one lock problem in this section, while the two lock problem is considered in the next section. Here the agent that first gets the lock, gets the reward. At any time, the agents are aware of their own state, but they would know the information of the others only when it contacts the lock. At that time, the contact can be successful or unsuccessful, with the former event implying the other agent has not yet contacted the lock. In this case the agents have to choose one contact rate function $a^i(\cdot)/a^j(\cdot)$ at the start, and would stop either after a contact or after the deadline T expires. There is no major information update at any time point before the (only one) contact and hence this control process is sufficient.

We prove that a pair of (time) threshold policies form an NE for this game. We prove this by showing that the best response against a threshold policy is again threshold. Towards this, we first discuss the best response against any given strategy of the opponent. Given a policy $\pi^j = a^j$ with $a^j \in L^\infty[0,T]$ of agent j, it is clear that the failure probability of the other agent j in equation (3) equals:

$$P_f^j(s|\pi^j) = \exp(-\bar{a}^j(s)),$$

where $\bar{a}^j(\cdot)$ is as defined in (1). The best response for this case is obtained by (see equations (1)-(7)):

$$\begin{aligned} v_1^i(\mathbf{z}_1;\pi_j) &= \sup_{a^i\in L^\infty} J(a^i;a^j) \text{ with} \\ J(a^i;a^j) &= \int_0^T \Big(\exp(-\bar{a}_1^j(s)) - \nu\bar{a}^i(s)\Big)\exp(-\bar{a}^i(s))a^i(s)ds \\ &\quad -\nu\bar{a}^i(T)\exp(-\bar{a}^i(T)). \end{aligned} \tag{9}$$

For any given policy $\pi^j = a^j(\cdot)$ of the other agent, the above best response is clearly a finite horizon (T) optimal control problem as J can be rewritten as

$$J(a^i) = \int_0^T \big(h^j(s) - \nu x(s)\big)\exp(-x(s))a^i(s)ds + g(x(T)), \tag{10}$$

with state process

$$\dot{x}\,(t) = a^i(t) \text{ and thus } x(t) = \int_0^t a^i(t)dt = \bar{a}^i(s),$$

a given function

$$h^j(s) := \exp(-\bar{a}^j(s)),$$

and with terminal cost

$$g(x) = -\nu x\exp(-x). \tag{11}$$

Here we need to note that $x(t)$ represents the state process for the optimal control problem that is used as a tool to solve the original best response problem and is not the state process of the actual/original problem. Further in two lock problem (considered in the next section), we will have one such optimal control problem for each stage and for each state and each of those optimal control problems will have their own state processes.

Conjecture: We aim to prove using Hamilton Jacobi (HJB) equations (which provide solution to the above mentioned optimal control problems of each stage), that the best response against any policy of agent j would be a time-threshold type policy as discussed below. We are yet to prove this. However from the nature of the HJB equations one can conclude that the best response policies are of bang-bang type. Currently we continue with deriving the best response against time-threshold policies.

A. Best response against silent opponent

We begin with best response of an agent, when the other agent is silent, i.e., when $a^j(t) = 0$ for all t. This particular result will be used repeatedly (also for the case with two locks) and hence is stated first. Let $\mathcal{C} := \{a \in L^\infty : ||a|| \le \beta^i\}$.

Theorem 1. [Best response, against silent opponent] *When an agent attempts to acquire a lock for any given time U, and when there is no competition and if it receives a reward c upon success: i) the best response of (9) is derived by solving the following optimal control problem:*

$$\begin{aligned} v(x) := \sup_{a(\cdot)\in\mathcal{C}} \Bigg\{ &\int_0^U \big(c - \nu x(s)\big)\exp(-x(s))a(s)ds \\ &-\nu x(U)\exp(-x(U))\Bigg\} \\ \textit{with } \dot{x}\,(t) = a(t) &\textit{ and } x(0) = x. \end{aligned}$$

ii) The solution of the above problem is the following:

$$a^*(t) = \begin{cases} \beta^i & \textit{for all } t \quad \textit{if } \nu \le c \\ 0 & \textit{for all } t \quad \textit{if } \nu > c \end{cases} \quad \textit{and}$$

$$v(x) = \begin{cases} \big[(c-\nu)\big(1-\exp(-\beta^i U)\big) - \nu x\big]\exp(-x) & \textit{if } \nu \le c \\ -\nu x\exp(-x) & \textit{if } \nu > c. \end{cases} \tag{12}$$

Proof: We drop the superscript i in this proof, for simpler notations. Using density (4), the expected reward (against

silent opponent) equals

$$E[R] = cE[\tau_1 < U] = c\int_0^U \exp(-\bar{a}(t))a(t)dt = c\int_0^U \exp(-x(t))a(t)dt. \quad (13)$$

The cost does not depend upon the existence of other players, hence remains the same as in (7), reproducing here for clarity:

$$E[\bar{a}(d_1)] = \exp(-x(U))x(U) + \int_0^U x(t)\exp\left(-x(t)\right)a(t)dt$$

Then the overall problem is to maximize $E[R] - \nu E[\bar{a}(d_1)]$ and hence we consider

$$\sup_{a(.)}\left\{\int_0^U L(t,x(t),a(t))dt + g(x(U))\right\} \text{ with}$$

$L(t,x,a) = (c-\nu x)\exp(-x)a$ and $g(x) = -\nu x\exp(-x)$.

Thus we need the solution of the following (Hamiltonian Jacobi) HJB PDE, with $v(t,x)$ representing the value function and with v_t, v_x its partial derivatives

$$v_t(t,x) + \max_{a\in[0,\beta]}\{L(t,x,a) + av_x(t,x)\} = 0$$

or in other words

$$v_t(t,x) + \max_{a\in[0,\beta]}\{(c-\nu x)\exp(-x)a + av_x(t,x)\} = 0$$

with boundary condition $v(U,x) = g(x) = -\nu x\exp(-x)$.

Claim: We claim the following is a solution[6] satisfying the above boundary valued problem (when $\nu < c$):

$$W(t,x) = (-\nu x - \nu + c)\exp(-x) + \kappa\exp(-x+\beta t) \text{ with}$$
$$\kappa = \exp(-\beta U)(\nu - c).$$

Note that its partial derivatives are:

$$W_x(t,x) = (\nu x - c)\exp(-x) - \kappa\exp(-x+\beta t),$$
$$W_t(t,x) = \beta\kappa\exp(-x+\beta t), \text{ and clearly for any } a,$$
$$L(t,x,a) + aW_x(t,x) = -\kappa\exp(-x+\beta t)a.$$

Thus if $\kappa \le 0$, the maximizer in HJB PDE is β^i and then $W(.,.)$ satisfies the HJB PDE,

$$W_t(t,x) + \kappa\exp(-x+\beta t)\beta = 0,$$

and also satisfies the boundary condition

$$W(T,x) = (-\nu x - \nu + c)\exp(-x) + \exp(-\beta U)(\nu - c)\exp(-x+\beta U) = -\nu x\exp(-x) \text{ for any } x.$$

[6]We derived the above solution by solving the HJB PDE after replacing the maximizer with β.

It is further easy to verify that $a^*(t) = \beta^i$ for all t and $x^*(t) = \beta^i t$ satisfy equation (5.7) of [3, Theorem 5.1]. Thus when $\kappa \le 0$ or equivalently when $\nu \le c$ then the optimal policy is to attempt with highest possible rate all the times.

When $\nu > c$, using similar logic one can show that $W(t,x) = -\nu x\exp(-x)$ (for all x,t) is the solution of the HJB PDE and $a^*(t) = 0$ for all t. ■

Using similar techniques one can find best response against any given time-threshold policy, which is next considered.

B. Best response against a time Threshold policy

Assume now player j uses the following time-threshold policy, represented by:

$$\Gamma(\psi^j) : a^j(t) = \beta^j\mathcal{X}_{[0,\psi^j]}(t), \text{ with } \psi^j \le T.$$

Basically agent j attempts with maximum acceleration β^j till time ψ^j and stops completely after that. In this case the failure probability of agent j in equation (3) simplifies to:

$$P_f^j(s|\pi^j) = \exp(-\beta^j(s\wedge\psi^j)), \text{ when } \pi^j = \Gamma(\psi^j),$$

and so

$$h^j(s) = \begin{cases} \exp(-\beta^j s) & \text{if } s \le \psi^j \\ \exp(-\beta^j\psi^j) & \text{if } s > \psi^j. \end{cases} \quad (14)$$

The best response against such a Threshold policy of agent j is obtained in the following. From Theorem 1, it is clear that the best responses against any strategy would be to remain silent when $\nu \ge 1$ and when the reward equals one. Thus the Nash equilibrium strategies would be to remain silent by both the agents for $\nu \ge 1$. From now on, we consider $\nu < 1$.

Theorem 2. [Best response] *Assume $\nu < 1$. The best response of agent i against Γ (ψ) policy of agent j is given by:*

$$BR_i(\Gamma(\psi)) = \begin{cases} \Gamma(T) & \text{if } \nu < \exp\left(-\beta^j\psi\right) \\ \Gamma(\theta_\nu^i) & \text{else.} \end{cases}$$

where, $$\theta_\nu^i = \min\left\{-\frac{\ln(\nu)}{\beta^j}, T\right\}.$$

Proof: The details of this proof are in Appendix A. ■

Thus when agent j uses threshold strategy with small ψ, best response of agent i is to attempt till the end and if the threshold of agent j is larger, $\psi \ge \theta_\nu^i$ then the best response of agent i is to try till θ_ν^i (irrespective of the actual value of ψ).

C. Nash Equilibrium

We observe from the above result that the best response against a threshold policy is again a threshold policy. Thus one can get the Nash equilibrium if one can find two thresholds one for each agent, such that

$$\Gamma(\psi^i) \in BR_i(\Gamma(\psi^j)) \text{ and } \Gamma(\psi^j) \in BR_j(\Gamma(\psi^i)).$$

From Theorem 2, it is easy to find such a pair of thresholds and is also easy to verify that this pair of thresholds is unique. We have the following:

Theorem 3. ***[Nash Equilibrium]*** *Assume $\nu < 1$, and without loss of generality $\beta^j \geq \beta^i$. For a two agent partial information game, we have a Nash equilibrium among (time) threshold policies, as defined below:*

$$\begin{array}{ll} \big(\Gamma(\theta^i_\nu), \quad \Gamma(\theta^j_\nu)\big) & \text{if } \beta^j = \beta^i \\ \big(\Gamma(\theta^i_\nu), \quad \Gamma(T)\big) & \text{if } \beta^j > \beta^i \end{array}$$

where threshold θ^i_ν is as in Theorem 2, while θ^j_ν is given by:

$$\theta^j_\nu = \min\left\{-\frac{\ln(\nu)}{\beta^i}, T\right\}. \qquad \blacksquare$$

Proof: The first line is easily evident from Theorem 2. For the second one, observe the following: when $\psi^i = \theta^i_\nu$:

$$\exp(-\beta^i\theta^i_\nu) = \exp(\beta^i \ln(\nu)/\beta^j) \geq \nu,$$

because $\beta^i \ln(\nu)/\beta^j \geq \ln(\nu)$ (note $\ln(\nu) < 0$) and thus

$$BR_j(\theta^i_\nu) = \Gamma(T).$$

Now if $\exp(-\beta^j T) \leq \nu$, then clearly $BR_i(T) = \Gamma(\theta^i_\nu)$. On the other hand if $\exp(-\beta^j T) > \nu$, then $\theta^i_\nu = T$. $\blacksquare$

It is further clear that (simple calculations using Theorem 2) we have unique Nash Equilibrium among time threshold policies. It would be more interesting if we can show this is the unique NE, but that would be a part of future work.

Thus when one has no access to the information of the other agent till their own contact with the lock, the NE are given by open loop policies. But this is true only for one lock ($M = 1$) problem. With large M, we will have closed loop policies but the policies change only at major information change epochs. In all, we will see that the NE will again have a group of open loop policies, each of which is used till a major change in the information.

IV. Two Lock Problem

Before we proceed with the analysis we would summarize the protocol again. Any agent succeeds only if it contacts lock one, followed by lock two and only the agent that gets both the locks receives reward one. If a particular agent contacts the lock one, we say it had an unsuccessful contact if it is not the first one to contact the lock. If an agent's contact is unsuccessful, there is no incentive for the agent to try any further. On the other hand when an agent is successful, it knows it would be the only one to chase the second lock. We can use the previous ($M = 1$ case) analysis, Theorem 1, to compute the best response against silence opponent (for second lock).

This is a two stage problem, as the utility of agent i is given by:

$$J^i(\pi_i, \pi_j) = \sum_{k=1}^{2} E[r^i_k(\mathbf{z}_k, a_k;\ \pi_j)].$$

For this two lock case, the best response of agent i against any given strategy of player j can be solved using (two stage) dynamic programming (DP) equations as below

$$\begin{aligned} v^i_k(\mathbf{z}_k; \pi_j) &= 0 \text{ if } k = 3 \text{ or if } \tau^i_{k-1} > T, \text{ and else,} \\ v^i_k(\mathbf{z}_k; \pi_j) &= \sup_{a_k}\left\{r^i_k(\mathbf{z}_k, a_k; \pi_j) + E[v^i_{k+1}(\mathbf{z}_{k+1}; \pi_j))|\mathbf{z}_k, a_k]\right\}, \end{aligned} \tag{15}$$

with stage wise costs as defined in equations (2)-(7). Note these DPs hold even when the action spaces are Banach spaces, as in our case.

Like in one-lock case, we obtain a NE, by finding best response against appropriate threshold strategies.

Threshold strategy for two-lock problem: Our conjecture is that the strategy constructed using state dependent time-threshold policies will form a part of the NE. At contact instance of the first lock, the contact could be successful or unsuccessful. Thus we have two types of states immediately after the first contact, i.e., the state after the first contact is either given by $\mathbf{z}_2 = (s, \tau)$ or by $\mathbf{z}_2 = (u, \tau)$. We compactly represent Threshold policy by $\Gamma_2(\psi)$ which means the following:

$$\begin{array}{rl} \text{at start, use} & \Gamma(\psi) \text{ policy,} \\ \text{if } \mathbf{z}_2 = (s,\tau), \text{ i.e., when successful use} & \Gamma(T-\tau) \text{ and} \\ \text{if } \mathbf{z}_2 = (u,\tau), \text{ i.e., when unsuccessful use} & \Gamma(0) \text{ policy.} \end{array}$$

Theorem 1, inspires us to conjecture that this kind of a threshold strategy becomes a part of the NE and the same is proved in Theorem 4. We begin with the best response.

A. Best response against a Threshold strategy

Say agent j uses threshold strategy $\Gamma_2(\psi^j)$. We obtain the best response by solving the DP equations (15) using backward induction. When $k = 2$ in (15) and if $z_2 = (u, \tau)$ it is immediately clear that (see (2))

$$v_2(\mathbf{z}_2; \Gamma_2) = 0 \text{ for any } \tau,$$

as failure with first lock implies zero reward. If $\mathbf{z}_2 = (s, \tau)$, i.e., if the player i is successful with first lock and the contact was at τ, the agent j will either have unsuccessful contact or may not even contact the first lock before the deadline T.

Further because agent j uses $\Gamma_2(\psi)$ policy it would not try for the second lock. Thus agent i will attempt for second lock, while the other agent is silent with respect to second lock. Thus the optimization problem corresponding to this stage from equations (2)-(7) is given by:

$$\sup_{a(\cdot)\in\mathcal{C}} \Big\{ \int_0^U \big(1-\nu x(s)\big)\exp(-x(s))a(s)ds \\ -\nu x(U)\exp(-x(U))\Big\}$$
$$\text{with } \dot{x}(t) = a(t) \text{ and } x(0) = x, \text{ with } U = T-\tau.$$

This is exactly the optimization problem considered in Theorem 1 with $U = T-\tau$ and hence the best response (with $\nu < 1$) is given by: $\Gamma(T-\tau)$ (attempt with maximum for the rest of the period). Thus from Theorem 1 with $U = T-\tau$ and $x = 0$ we have:

$$\begin{aligned} v_2(s,\tau;\Gamma_2) &= v(0) = (1-\nu)\big(1-\exp(-\beta^i(T-\tau))\big) \text{ and} \\ v_2(u,\tau;\Gamma_2) &= 0. \end{aligned}$$

Now solving the DP equations for $k = 1$, it is easy to verify that the corresponding optimization problem is (with $x(\cdot)$ as before and see (2), (15)):

$$\sup_{a(\cdot)} \Big\{ -\nu\int_0^T x(\tau)\exp(-x(\tau))a(\tau)d\tau - \nu x(T)\exp(-x(T)) \\ + \textstyle\int_0^T \exp(-\bar{a}^j(\tau))v_2(s,\tau;\Gamma_2)\exp(-x(\tau))a(\tau)d\tau\Big\}.$$

This optimization problem is once again solved using optimal control theory based tools and we directly obtain the following. When $\nu \geq 1/2$ it is easy to verify that, both agents being silent is the Nash equilibrium. This result can easily be derived (by finding the best responses as in Theorem **??** of Appendix B, which provides the best response against the silent opponent).

Theorem 4. [Nash Equilibrium] *Let $\nu < 1/2$ and assume $\beta^j \geq \beta^i$. The NE is given by the following, under the conditions:*

$$\begin{cases} \Big(\Gamma_2(0),\Gamma_2(0)\Big), & \text{if } \exp(-\beta^j T) > \frac{1-2\nu}{1-\nu}, \\ \Big(\Gamma_2(0),\Gamma_2(\psi_0^{j*})\Big), & \text{with } \psi_0^{j*} = T + \frac{1}{\beta^j}\ln\Big(\frac{1-2\nu}{1-\nu}\Big) \\ & \text{if } \exp(-\beta^i T) > \frac{1-2\nu}{1-\nu} > \exp(-\beta^j T). \end{cases}$$

The above $\psi_0^{j*} > 0$. Now consider that $\exp(-\beta^i T) \leq \frac{1-2\nu}{1-\nu}$. Let ψ^{i*} satisfy

$$\exp(-\beta^j\psi^{i*}) = \min\Big\{1, \exp(-\beta^i(T-\psi^{i*}) - \beta^j\psi^{i*}) + \frac{\nu}{(1-\nu)}\Big\}.$$

and ψ^{j*} satisfy the following equation

$$\exp(-\beta^i\psi^{j*}) = \min\Big\{1, \exp(-\beta^j(T-\psi^{j*}) - \beta^i\psi^{j*}) + \frac{\nu}{1-\nu}\Big\}.$$

If the following two conditions are satisfied

$$\begin{aligned} \exp(-\beta^i(T-\psi^{j*})) &> \frac{\exp(-\beta^j\psi^{j*}) - 2\nu}{\exp(-\beta^j\psi^{j*}) - \nu} \text{ and} \\ \exp(-\beta^j(T-\psi^{i*})) &> \frac{\exp(-\beta^i\psi^{i*}) - 2\nu}{\exp(-\beta^i\psi^{i*}) - \nu}, \end{aligned} \tag{16}$$

then the pair $\Big(\Gamma_2(\psi^{i*}),\Gamma_2(\psi^{j*})\Big)$ forms a Nash equilibrium. The above $\psi^{i*},\psi^{j*} < T$. The conditions (16) are immediately satisfied with $\beta^j = \beta^i = \beta$, in which case the common

$$\exp(-\beta\psi^*) = \min\Big\{1,\ \exp(-\beta T) + \frac{\nu}{1-\nu}\Big\}. \ \blacksquare$$

Remarks: Few interesting observations for the cases that we derived the result: a) in two lock problem none of the agents at an NE would try till T (in contrast to one lock problem); b) the agents either remain silent or attempt for a time period that is strictly less than T; and c) we obtained NE for all the values of the parameters for the case when $\beta^j = \beta^i$.

V. Extensions and Future work

One can easily extend the results to N-player game with symmetric parameters, i.e., to the case when $\beta^i = \beta$ for all i. For one lock problem it is not difficult to conjecture that the Nash equilibrium among time-Threshold policies is given by,

$$\Big(\Gamma(\theta^*),\Gamma(\theta^*),\cdots,\Gamma(\theta^*)\Big) \text{ with}$$
$$\theta^* := \begin{cases} -\frac{\ln(\nu)}{(N-1)\beta} & \text{if } \exp(-\beta(N-1)T) \leq \nu \\ T & \text{else.} \end{cases}$$

In a similar way the two-lock Nash equilibrium for symmetric agents, could probably be obtained using Theorem 4; with the parameter of the opponent as $\beta^j = (N-1)\beta$ and with $\beta^i = \beta$. We conjecture the NE for this case to be, $\Big(\Gamma_2(\psi^{i*}),\Gamma_2(\psi^{i*}),\cdots,\Gamma_2(\psi^{i*})\Big)$. These are only conjectures and we need to verify and prove the same. Further we would like to work with asymmetric agents.

It would be equally interesting to work with M-lock problem with $M > 2$. We anticipate that the silence theorem (like Theorems 1 and **??**) should be extended and then the analysis would follow easily. It would be more interesting to work with the problem in which each lock fetches a reward. For all these and more general problems, the methodology would be the same; One needs to consider open loop control till a new information update. Thus these partial information problems would span from completely open loop policies (no information) to completely closed loop policies (or full information).

CONCLUSIONS

We considered lock acquisition games with partial, asymmetric information. Agents attempt to control the rate of their Poisson clocks to acquire two locks, the first one to get both would get the reward. There is a deadline before which the locks are to be acquired, only the first agent to contact the lock can acquire it and the agents are not aware of the acquisition status of others. It is possible that an agent continues its acquisition attempts, while the lock is already acquired by another agent. The agents pay a cost proportional to their rates of acquisition. We proposed a new approach to solve these asymmetric and non-classical information games, "open loop control till the information update". With this approach we have dynamic programming equations applicable at state change update instances and then each stage of the dynamic programming equations is to be solved by optimal control theory based tools (HJB equations). We showed that a pair of (available) state dependent time threshold policies form Nash equilibrium. We also conjectured the results for the games with N-agents.

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] "Acquisition Games with Partial-Asymmetric Information," *Technical report downloadable at https://arxiv.org/abs/1909.06633*.

APPENDIX A: PROOFS RELATED TO ONE LOCK PROBLEM

Proof of Theorem 2: The best response against a threshold policy can be obtained by solving the optimal control problem (see equation (10) with h^j as in (14))

$$v(x) := \sup_{a(\cdot)\in\mathcal{C}} \left\{ \int_0^T L(t,x(t),a(t))ds - \nu x(T)\exp(-x(T)) \right\}$$

with state update equation given by

$\dot{x}\,(t) = a(t); x(0) = x$ and with running cost,

$$L(t,x,a) = \begin{cases} \Big(\exp\left(-\beta^j t\right) - \nu x\Big)\exp\left(-x\right)a & \text{for } t \le \psi \\ \Big(\exp\left(-\beta^j \psi\right) - \nu x\Big)\exp\left(-x\right)a & \text{else.} \end{cases}$$

Further the terminal cost is $g(x) = -\nu x \exp(-x)$. Thus the HJB (PDE) equation that needs to be solved as in the proof of Theorem 1 is given by the following:

$$\frac{\partial}{\partial t}v(t,x) + \sup_{a\in[0,\beta^i]} \left\{ L(t,x,a) + a\frac{\partial v}{\partial x} \right\} = 0, \qquad (17)$$

$$v(T,x) = -\nu x \exp(-x).$$

Let $v_t := \frac{\partial v}{\partial t}$ and $v_x := \frac{\partial v}{\partial x}$. We conjecture that the optimal control for this problem is a threshold policy $\Gamma(t_1)$ for some appropriate $0 \le t_1 \le T$.

Claim:We further claim the following to be the solution of the above PDE[7], we prove this claim alongside computing t_1 (we would actually show that $t_1 = \theta^i_\nu$ or T):

$$W(t,x) = \begin{cases} -\nu x\exp(-x) - \nu\exp(-x) + \frac{\beta^i}{\beta^i+\beta^j}\exp(-x-\beta^j t) & \\ \quad +\kappa_1\exp(-x+\beta^i t) & \text{if } t \le (t_1 \wedge \psi) \\ -(\exp(-\beta^j\psi)-\nu)\exp(-\beta^i t_1)\exp(-x+\beta^i t) & \\ +(\exp(-\beta^j\psi)-\nu x-\nu)\exp(-x) & \text{if } \psi \le t \le t_1 \\ -\nu x\exp(-x) & \text{else.} \end{cases}$$

where,

$$\kappa_1 = \begin{cases} \nu\exp(-\beta^i t_1) - \frac{\beta^i}{\beta^i+\beta^j}\exp(-(\beta^i+\beta^j)t_1) & \\ & \text{if } \nu \ge \exp(-\psi\beta^j) \\ \nu\exp(-\beta^i t_1) - \frac{\beta^i}{\beta^i+\beta^j}\exp(-(\beta^i+\beta^j)\psi) & \\ +\exp(-\beta^j\psi)\Big(\exp(-\beta^i\psi) - \exp(-\beta^i t_1)\Big) & \text{else.} \end{cases}$$

The partial derivatives of the above are:

$$W_x(t,x) = \begin{cases} \nu x\exp(-x) - \frac{\beta^i}{\beta^i+\beta^j}\exp(-x-\beta^j t) & \\ \quad -\kappa_1\exp(-x+\beta^i t) & \text{if } t \le t_1 \wedge \psi \\ +(\exp(-\beta^j\psi)-\nu)\exp(-\beta^i t_1)\exp(-x+\beta^i t) & \\ -(\exp(-\beta^j\psi)-\nu x)\exp(-x) & \text{if } \psi \le t \le t_1 \\ (\nu x-\nu)\exp(-x) & \text{else.} \end{cases}$$

$$W_t(t,x) = \begin{cases} -\frac{\beta^i\beta^j}{\beta^i+\beta^j}\exp(-x-\beta^j t) + \beta^i\kappa_1\exp(-x+\beta^i t) & \\ & \text{if } t \le t_1 \wedge \psi \\ -\beta^i(\exp(-\beta^j\psi)-\nu)\exp(-\beta^i t_1)\exp(-x+\beta^i t) & \\ & \text{if } \psi \le t \le t_1 \\ 0 & \text{else.} \end{cases}$$

We verify that the above partial derivatives verify PDE (17), when t_1 is set equal to θ^i_ν (defined in the hypothesis of the theorem) and the details are in [4]. In all, one can verify that the following values of t_1 satisfy all the required conditions ([3, Theorem 5.1]) and we will have the best response as $\Gamma(t_1)$ with:

$$t_1 = \begin{cases} \theta^i_\nu & \text{if } \exp(-\beta^j\psi) \le \nu \\ T & \text{else.} \end{cases}$$

■

APPENDIX B: TWO LOCK PROOFS

These proofs are provided in [4].

[7]We compute the following solutions, replacing the maximizers in HJB PDEs $a^* = \beta^i$. One of them is for the case when $t_1 \le \psi$ and one for the other case.