Extended Abstract: Markov Game Analysis for Attack-Defense of Power Networks

Chris Y. T. Ma[♯], David K. Y. Yau[◇♯], Xin Lou[†♯], and Nageswara S. V. Rao[‡]

[♯] Advanced Digital Sciences Center, Illinois at Singapore
[◇] Purdue University, West Lafayette, IN, USA
[†] City University of Hong Kong, Hong Kong
[‡] Oak Ridge National Laboratory, TN, USA

Abstract. Electricity grids are critical infrastructures. They are credible targets of active (e.g., terrorist) attacks since their disruption may lead to sizable losses economically and in human lives. It is thus crucial to develop decision support that can guide administrators in deploying defense resources for system security and reliability. Prior work on the defense of critical infrastructures has typically used static or Stackelberg games. These approaches view network interdictions as one-time events. However, infrastructure protection is also a continual process in which the defender and attacker interact to produce dynamic states affecting their best actions. In this paper, we use zero-sum Markov games to model these interactions subject to underlying uncertainties of real-world events and actions. We solve equilibrium mixed strategies of the players that maximize their respective minimum payoffs with a time-decayed metric. Using results for a 5-bus system [1] and a WCSS 9-bus system [2], we illustrate that our game model can provide useful insights.

1 Introduction

Electricity networks are critical infrastructures. Their disruptions can have severe economic, social, and security consequences. Disruption of trading systems may lead to turmoils in financial markets. Lives may be endangered if power is lost for heating in cold weather, or it is lost for life saving procedures in hospitals. Loss of power may also prevent communication, stall work, cripple transportation, and/or lead to other major failures that can bring entire nations to a standstill. Because of their importance, power networks are credible targets for active (e.g., terrorist) attacks. On the other hand, protecting these networks is extremely challenging, due to their expansive geographical extents and complex interdependencies between system components. For example, transmission lines may run for miles in the open, and the system must maintain stable and prespecified power quality (e.g., frequency, voltage, and phase synchronization) for performance and safety of equipment.

To protect critical infrastructures from attacks, administrators need tools that support prudent decision making. In particular, administrators need to make informed decisions about where to deploy finite resources to harden a system for maximum resiliency against adversaries. Such guidance for infrastructure protection has been obtained using Markov decision processes (MDP) or game

theory. In MDP [3], the system is modeled as a set of states with Markov transitions between them. The problem is to optimize the actions of a "player" (e.g., the defender) under probabilistic outcomes of these actions. The solution optimizes the actions of a single player only. It is suitable for a defender to maximize system reliability against passive disruptors of known probabilistic behavior. Collectively, these disruptors may represent "nature," which may disrupt components by indeliberate events such as bad weather or normal wear-and-tear.

Game theoretic approaches for infrastructure protection, on the other hand, postulate a strong attacker – one capable of devising its own best counter strategies against the defender. In a *static game* [4], both players choose their moves simultaneously. In another form of leader-follower Stackelberg games [5], the optimization of the players' strategies is a bilevel problem. At the inner level, the follower maximizes its payoff *given* a leader's strategy. At the outer level, the leader chooses a strategy S to maximize its own payoff subject to the follower's solution of the inner problem defined by S. Notice that a strong attacker model may also be applicable for play against nature, in which case the analysis gives worst-case estimates of required protection for reliability.

The above kinds of games view network interdictions as one time events. However, infrastructure protection is also a continual process in which the players interact to produce dynamic states affecting their respective best actions. Markov games model these interactions subject to inherent uncertainty in the underlying physical system. They can be viewed as generalizations of MDP to an adversarial setting. For the protection of power networks, we assume that the attacker deploys resources to disrupt transmission lines in a power grid,[1] and its goal is to maximize the amount of load shedding. The defender's goal, on the other hand, is to deploy defense resources to minimize the amount of load shedding in the face of such attacks. The directly opposing goals of the attacker and defender lead to a zero-sum game formulation naturally.

Measuring damage as the amount of shed load is a natural baseline. In addition, load in practical systems may be of varying importance. For example, part of a power network may serve critical government functions predominantly, whereas another part may serve charging stations for plugged-in hybrid electric vehicles (PHEVs) predominantly. Protection of the former is arguably more important than the latter. Hence, we will also consider a more general model in which shedding load in different parts of the grid does different damage quantified by *cost functions*. In this case, the attacker's and defender's goals are to maximize and minimize the total cost of shed load, respectively.

Game theoretic analysis has typically assumed a full information setting. In practical situations, information is a valuable asset having significant effects on achieved payoffs. Control and sensing information communicated in future smart grids may be a valuable source of information for would-be attackers. For example, advanced meter infrastructures (AMIs) may give a comprehensive view of the distribution of load and resulting power flows. More sophisticated attackers may even infer the types of load based on their power signatures [6],

[1] Transmission lines can be considered particularly vulnerable targets due to the impossibility of physical isolation. However, our problem can be readily generalized to consider other system components.

leading to knowledge about the cost functions in our model. The role of on-line information gives a cyber dimension of smart grid protection as a *cyber-physical system* problem.

Our **contributions** are as follows. (i) We model the attack-defense of power networks as a Markov game. We solve equilibrium mixed strategies of the players that maximize their respective minimum payoffs by a time-decayed metric under uncertainty. (ii) We show that after our algorithm converges, the solution in each state is equivalent to that of a static game with a composite payoff matrix. Analysis of this composite matrix simplifies the interpretation of results obtained by our algorithm. (iii) We apply our solution to two realistic power systems. We contrast our numerical results with those of static games, and show that their analysis leads to useful insights in sometimes subtle situations.

The rest of this paper is organized as follows. Section 2 discusses related work. We define our system model in Section 3. We solve the Markov game in Section 4. We present and analyze numerical results for a power system in Section 5. Section 6 concludes.

2 Related Work

MDP has been used to analyze the security and vulnerability of urban infrastructures. Jones *et al.* [7] use it to analyze the actions of an intruder into transportation facilities. Jha *et al.* [3] use MDP to interpret attack graphs in communication networks, so that a minimal set of security measures can be determined that will guarantee the safety of a system. Their work optimizes the actions of the defender against a passive attacker, whose strategy is fixed and given.

Game theory has been widely used to analyze the security of critical systems. The competition between a defender and an attacker in this context has been modeled as leader-follower Stackelberg games [5], [8], and static games [4], [9]. These games analyze one move of each player only, and so they treat network interdictions as one time events. In practice, the defender may interact with the attacker in repeated plays that evolve the system state dynamically. Alpcan *et al.* model these repeated plays under uncertainty as a Markov game. They use the game model to design an intrusion detection system for a communication network [10], and compare their results with those obtained using static games.

3 Problem Formulation

In a power grid, generators supply electricity and loads consume it. They are attached to a set of buses – which we call generation and load buses, respectively – interconnected by a network of transmission lines of given capacities. Henceforth, we refer to transmission lines as *links*.

An attacker aims to disrupt the power network by bringing down one or more links, in order to cause maximum "disruption" of the load. A defender aims to minimize this disruption. It does so by reinforcing links that are up, and repairing links that are down. In a baseline case, disruption is measured simply as the amount of load (in power unit) that must be shed due to the link failures. More generally, shedding different loads may have different adverse

impact which we call *cost*. A *cost function* for a load bus, say l, is given by $u_l(x, y)$, which specifies the cost of reducing the load from x to y (in power units) on l. In this case, disruption is measured as the total cost of shed load due to the link failures.

We define a Markov game as follows. The state of the game refers to the set of links that are currently up (links that are not up are down) in the power network. The game proceeds in discrete time steps. In each time step, the players choose a pair of actions which, together with underlying probabilistic physical events, may cause state transitions in a Markov manner. For the attacker, the action is the link that it chooses to attack. For the defender, the action is the (down) link that it chooses to repair or the (up) link that it chooses to reinforce. The players have limited budgets in that in each time step, the attacker (respectively defender) can choose a limited number of links to attack (respectively repair/reinforce) only.

We use the following notations throughout the paper.

- A_p: Action set of player p, where $p = a, d$, corresponding to the attacker and defender, respectively.
- S: Set of game states, where each state is an enumeration of the status of the links in order. We use "u" and "d" to denote the up or down status, respectively.
- $PD(A)$: Set of mixed strategies over the action set A.
- p_{pf}: Probability for an up link to fail in a time step upon attack, when it is reinforced by the defender in that time step.
- p_{upf}: Probability for an up link to fail upon attack, when it is not reinforced. We have $0 \leq p_{pf} \leq p_{upf} \leq 1$.
- p_{pr}: Probability for a down link to recover (i.e., become up) in a time step, when it is repaired by the defender and not attacked by the attacker in that time step.
- p_{upr}: Probability for a down link to recover when it is not repaired by the defender and not attacked by the attacker. We have $0 \leq p_{upr} \leq p_{pr} \leq 1$.

We assume that the attacker can attack a link that is already down. Such an action will reduce the probability that the link recovers. For example, if a down link is repaired by the defender and further attacked by the attacker in a time step, then its probability of recovery is $p_{pr} \times (1 - p_{upf})$. If the down link is not repaired by the defender, then its probability of recovery under attack is $p_{upr} \times (1 - p_{upf})$.

Since the load shedding goals of the attacker and defender are directly opposing, we have a zero-sum game. A pair of player actions in a state will bring an *immediate reward* for the players. For the attacker, this reward is the expected cost of shed load due to the resulting probabilistic transitions to the possible next states. The defender's immediate reward is the negative of this number. Further to the immediate reward, each possible state transition if realized will bring the game to a new state, where the game will carry on. A further immediate reward will be obtained in the new state with further new state transitions, and so on. Hence, a pair of actions taken in a state will accrue a *long-term reward* in general.

Formally, define $R(s, a, d)$ as the expected immediate reward for the attacker when it takes action a and the defender takes action d in state s. (Reward for the defender is the negative of this number.) Further define $Q(s, a, d)$ as the expected long-term reward for the attacker when it takes action a and the defender takes action d in state s. (Expected long-term reward for the defender is the negative of this number.) The *value* of state $s \in S$ for the attacker in the Markov game is

$$V_a(s) = \max_{\pi \in PD(A_a)} \min_{d \in A_d} \sum_{a \in A_a} Q(s, a, d)\pi_a, \tag{1}$$

where π_a is the probability of action a in the optimal mixed strategy π of the attacker. The expected long-term reward, *quality*, of action a against action d in state s is

$$Q(s, a, d) = R(s, a, d) + \gamma \sum_{s'} T(s, a, d, s')V_a(s'), \tag{2}$$

where $T(s, a, d, s')$ is the state transition $T : S \times A_a \times A_d \to S$, and γ is a discount factor satisfying $0 \leq \gamma < 1$. γ gives the discount factor of future rewards on the optimal decision. Small values of γ emphasize near-term gains while large values emphasize future rewards. γ may also be interpreted as the belief of possible future interactions held by the players.

Similarly, the *value* of state $s \in S$ for the defender is

$$V_d(s) = \min_{\pi \in PD(A_d)} \max_{a \in A_a} \sum_{d \in A_d} Q(s, a, d)\pi_d. \tag{3}$$

Notice that in general, $V_a(s)$ and $V_d(s)$ computed from Eq. 1 and Eq. 3 are different. In particular, $V_a(s) \leq V_d(s)$, where Eq. 1 corresponds to the primal problem and Eq. 3 corresponds to the dual problem. The inequality expresses weak duality relating the primal and dual problems in general [11, Section 5.4]. When the Markov game is zero-sum, however, strong duality applies [11, Section 5.4] and equality holds due to the *strong max-min property*. Hence, we use $V(s)$ to denote the value of state $s \in S$, and $V(s) = V_a(s) = V_d(s)$. The optimal solutions computed individually by the two players are therefore best responses to each other and they are in Nash equilibrium. The equilibrium solutions are necessarily Pareto-optimal, because we cannot improve the payoff of one player without hurting that of the other in a zero-sum game.

4 Markov Game Solution

We now solve the Markov game defined in Sec. 3. Our goal is to compute equilibrium best *policies* for both players, where a policy is the set of per-state optimal mixed strategies of the player concerned, and an optimal strategy is one that maximizes the minimum long-term reward under the best strategy of the opponent. It is known that every Markov game has a non-empty set of optimal policies for each player, and one of them is *stationary*, i.e., it is time-independent [12]. Our solution will find this optimal stationary policy for each player. Once the

```
1. Set V(s) = 0 for all s ∈ S
2. repeat
3.     for all s ∈ S and a ∈ A_a and d ∈ A_d do
4.         Update Q according to Eq. 2
5.     end for
6.     for all s ∈ S do
7.         Update V according to Eq. 1
8.     end for
9. until V(s) → V*, i.e., V(s) converges.
```

Fig. 1. Dynamic programming algorithm for solving the Markov game.

optimal policies of the players are determined, the Markov transition probabilities are completely defined and the system will evolve as a standard Markov process.

We consider the case in which both players have complete information about the game. The solution is a generalization of *value iteration*, a common dynamic programming technique for solving MDPs [12], [10], to a game-theoretic setting.

Recall from Sec. 3 that the value of state $s \in S$ in the game is given by Eq. 1 for the attacker, and by Eq. 3 for the defender. The optimal mixed strategy π of the attacker can be obtained by solving the following linear program:

$$\begin{aligned} &\max_{\pi \in PD(A_a)} V(s), \\ \text{s.t.} \quad &\sum_{a \in A_a} Q(s,a,d)\pi_a \geq V(s), \\ &\sum_{a \in A_a} \pi_a = 1, \\ &\pi_a \geq 0. \end{aligned}$$

The optimal π of the defender can be obtained by the above formulation with the order of the maximization and minimization swapped.

The value iteration algorithm to compute the optimal Q and V for given s, a, d is specified in Fig. 1. The algorithm iteratively estimates the values of V and Q by treating the equal signs in Eqs. 2 and 1 as assignment operators for updating the estimates. These estimates will converge to their correct values [13]. Notice that each iteration of the algorithm produces a mixed strategy for one player in state s by linear programming (Line 7). These mixed strategies will similarly converge to the optimal one, and hence we obtain one player's optimal policy when the algorithm terminates. We then use the converged Q's to solve for V's by linear programming from the perspective of the other player, and obtain the optimal policy of the other player.

Notice that we initialize $V(s) = 0$. As a result, the mixed strategy of the player after the first iteration is its optimal mixed strategy in a *static* game that does not consider rewards in future time steps, and the obtained $V(s)$ corresponds to the payoff in state s of this static game. For instance, for the 5-bus system shown in Fig. 2, Table 1(a) shows the payoff matrix of the static game for state {u,u,u,u,u}. Notice that the matrix shows the payoff to an attacker, and hence, the attacker prefers an action that returns a larger number, while the defender prefers an action that returns a smaller number. As we consider future rewards, the payoff matrix will evolve during the iterative process of

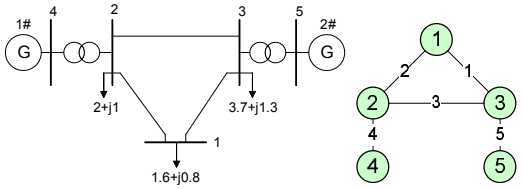
Link protected by defender	Link attacked by attacker: 1	2	3	4	5
1	61.05	0	112.1	472.1	230
2	122.1	0	112.1	472.1	230
3	122.1	0	56.05	472.1	230
4	122.1	0	112.1	236.1	230
5	122.1	0	112.1	472.1	115

(a)

Link protected by defender	Link attacked by attacker: 1	2	3	4	5
1	210.6	152.4	279.7	674.4	445.5
2	292.4	140.6	279.7	674.4	445.5
3	292.4	152.4	204.2	674.4	445.5
4	292.4	152.4	279.7	401.6	445.5
5	292.4	152.4	279.7	674.4	287.2

(b)

Table 1. Quality of actions of the two players in state {u,u,u,u,u} for (a) a static game that does not consider future rewards, (b) the full Markov game. Numbers are payoffs for the attacker. Hence, the attacker prefers larger numbers, while the defender prefers smaller numbers. $p_{pf} = 0.5, p_{upf} = 1, p_{pr} = 0.6, p_{upr} = 0$.



(a) Bus diagram (b) Link diagram

Fig. 2. 5-bus system.

the algorithm. When their effects are fully considered, Table 1(b) shows the "composite" payoff matrix for the same state {u,u,u,u,u} after the convergence of V in Line 9. For any state s, the optimal mixed strategy of the player in the Markov game is equivalent to the optimal mixed strategy solved for an equivalent static game with the composite matrix as payoffs. This view facilitates the interpretation of results obtained for the Markov game.

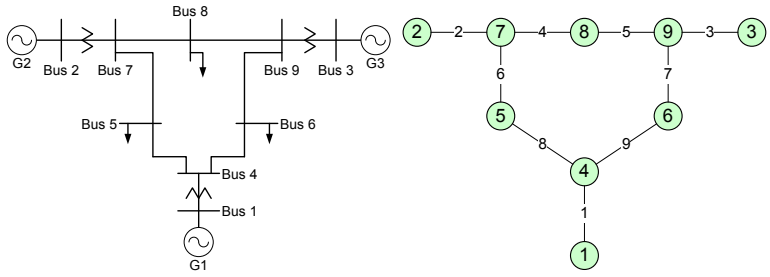
5 Evaluations

We present numerical results to illustrate solutions of the Markov games, which include the static games as a special case ($\gamma = 0$), using the failure and recovery probabilities as follows, $p_{pf} = 0.5, p_{upf} = 1, p_{pr} = 0.6, p_{upr} = 0$, unless stated otherwise. We assume that both players have complete information of the game. The cost function of load shedding is the amount of load shed. We have results for a 5-bus system [1] and a WCSS 9-bus system [2]. Their bus and link diagrams are given in Figures 2 and 3, respectively, and their per-bus aggregate generation and load are listed in Tables 2 and 3, respectively. We will focus on the 5-bus system for illustration of the more detailed results, since its relative simplicity facilitates the exposition.

Notice that certain links in a power system are particularly important, in that interdicting such a link by itself will already cut off a large amount of power flow from generation to load. In the 5-bus system, links l_4 and l_5 are particularly

Bus ID	1	2	3	4	5
Load (MW)	160	200	370	0	0
Supply (MW)	0	0	0	500	257.8

Table 2. Load and supply distribution of 5-bus system.



(a) Bus diagram (b) Link diagram

Fig. 3. Standard WCSS 9-bus system.

Bus ID	1	2	3	4	5	6	7	8	9
Load (MW)	0	0	0	0	125	90	0	100	0
Supply (MW)	71.64	163	85	0	0	0	0	0	0

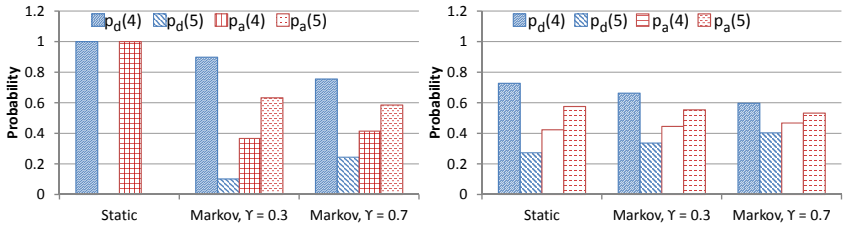
Table 3. Load and supply distribution of WCSS 9-bus system.

important, with l_4 being more so. In the 9-bus system, links l_1, l_2, and l_3 are particularly important, with l_2 being the most. These important links usually form the focus of the player strategies.

Fig. 4 shows the player strategies in selected states of the Markov game for the 5-bus system. In the figure, a bar labeled $p_a(x)$ gives the probability that the attacker will attack link x, and a bar labeled $p_d(x)$ gives the probability that the defender will repair link x (if x is down) or reinforce link x (if x is up). For example, $p_a(5)$ represents the probability for the attacker to attack l_5. Only actions with non-zero probabilities are included in the figure. The defender and the attacker have budgets to affect one link only in a time step. The results show that the optimal policies of the players may change significantly as we vary γ from zero (static game) to 0.7.

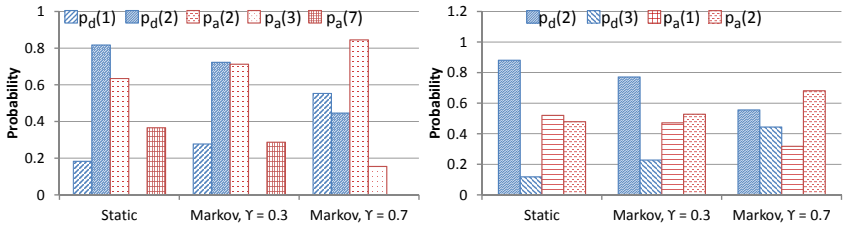
For instance, Fig. 4(a) shows that in state {u,u,u,u,u}, the defender progressively shifts its focus from reinforcing l_4 to reinforcing l_5, while the attacker also attacks l_5 apart from l_4, as γ increases. This observation can be explained using the payoff matrix of the static game (Table 1(a)) and the composite payoff matrix of the Markov game when $\gamma = 0.3$ (Table 1(b)). Notice that the numbers shown are the costs of load shedding and hence represent payoffs for the attacker – the attacker prefers higher numbers while the defender prefers lower numbers. Table 1(a) shows that in the static game, the payoff of attacking l_4 is always higher than that of attacking l_5, i.e., both l_4 and l_5 are important but l_4 is even more so. Hence, the attacker will only attack l_4, and the defender will always defend l_4 to minimize its cost. However, Table 1(b) shows that in the Markov game, the payoff of attacking l_4 is always higher than attacking l_5, *except* in the case that the defender is also reinforcing l_4. Hence, when l_4 is being reinforced with sufficiently high probability, the attacker begins to use a mixed strategy that includes l_5. This illustrates a subtle interplay between the players: Although a *successful* attack on l_4 will bring higher benefit for the attacker, it is also more difficult if l_4 is also reinforced by the defender. Hence, the attacker shifts some of its focus to the easier target l_5 since that link is also important.

Fig. 5 shows selected strategies of the players in the Markov game for the 9-bus system.



(a) State {u,u,u,u,u} (b) State {d,u,u,u,u}

Fig. 4. Player strategies in selected states of the Markov game for the 5-bus system. Both players have budgets to affect one link only in a time step.



(a) State {d,u,u,u,u,u,u,u,u} (b) State {u,u,d,u,d,u,u,u,u}

Fig. 5. Selected player strategies for the 9-bus system. Both players have budgets to affect one link in a time step. $p_{pf} = 0.5, p_{upf} = 1, p_{pr} = 0.6, p_{upr} = 0$.

6 Conclusion

We have presented a Markov game analysis of attack-defense in power systems. Our results complement related results using static games or Stackelberg games. We show that consideration of repeated plays under Markov-type uncertainties will in general modify the strategies of the players relative to games with single plays. This is because the players will need to consider the impact of a current action on the future plays, although the future rewards are generally discounted by a factor γ. We have applied our analysis to a 5-bus system that has been studied in the literature and a WCSS 9-bus system. The relative simplicity of the 5-bus system has allowed us to analyze its results in detail. Our analysis exposes subtle features of the game solutions, considering the *values* of different game states to the players and the intricate interplay between their strategies. It is also interesting to apply our analysis to other critical infrastructures.

References

1. : Calculation of The Electrical Power System. Hydro-electricity Press (1978)
2. Anderson, P.M., Fouad, A.A.: Power System Control and Stability. Galgotia (1981)
3. Jha, S., Sheyner, O., Wing, J.: Two formal analysis of attack graphs. In: Proc. of the IEEE workshop on Computer Security Foundations. (2002)
4. Holmgren, A., Jenelius, E., Westin, J.: Evaluating strategies for defending electric power networks against antagonistic attacks. IEEE Trans. Power Syst **22**(1) (2007)
5. Salmeron, J., Wood, K., Baldick, R.: Analysis of electric grid security under terrorist threat. IEEE Trans. Power Syst **19**(2) (2004)

6. Laughman, C., Lee, K., Cox, R., Shaw, S., Leeb, S., Norford, L., Armstrong, P.: Power signature analysis. IEEE Power & Energy Magazine **1**(2) (2003)
7. Jones, D.A., Davis, C.E., Turnquist, M.A., Nozick, L.K.: Physical security and vulnerability modeling for infrastructure facilities. In: Proc. of the Hawaii International Conference on System Sciences. (2006)
8. Brown, G., Carlyle, M., Salmeron, J., Wood, K.: Defending critical infrastructure. Interfaces **36**(6) (2006)
9. Chen, G., Dong, Z.Y., Hill, D.J., Xue, Y.S.: Exploring reliable strategies for defending power systems against targeted attacks. IEEE Trans. Power Syst **26**(3) (2011)
10. Alpcan, T., Basar, T.: Network Security: A Decision and Game Theoretic Approach. Cambridge University Press (2010)
11. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
12. Littman, M.: Markov games as a framework for multi-agent reinforcement learning. In: Proc. of the International Conference on Machine Learning. (1994)
13. Owen, G.: Game Theory: Second edition. Academic Press (1982)