Reinforcement Learning for Jamming Games Over AWGN Channels With Mobile Players

Giovanni Perin and Leonardo Badia
Dept. of Information Engineering (DEI), University of Padova, Italy
email: giovanni.perin.2@phd.unipd.it, leonardo.badia@unipd.it

***Abstract*—We study a jamming problem in a wireless scenario, where a legitimate receiver and a jammer compete in a zero-sum game with the value to maximize/minimize being the channel capacity at the receiver's side. Classic approaches consider stationary nodes. Instead, we investigate what happens when they can move along a linear geometry, which requires expanding the game into a dynamic setup. Even though the strategy space becomes huge, we can (i) provide some analytical criteria to identify good strategies, and (ii) explore efficient strategies leading to equilibrium outcomes using reinforcement learning techniques. We analyze different scenarios of sequential and simultaneous dynamic moves, as well as perfect versus imperfect information about the position of the adversary. Our numerical evaluations show the consistency of our findings in all the considered scenarios.**

***Index Terms*—Wireless communications; Jamming; AWGN; Game theory; Zero-sum games; Reinforcement learning.**

I. Introduction

A classic scenario of application for game theory to wireless communications is that of jamming [1], [2]. This can be formalized as a zero-sum game played by two agents, i.e., a regular user, acting as a maximizing player, and a jammer being the minimizer, where the value of the game can be for example the channel capacity, or, equivalently, the signal to noise plus jamming ratio (SNJR) in the case of additive white Gaussian noise (AWGN) channels.

More precisely, a suitable scenario to explore considers a legitimate receiver R, placed in the vicinity of a base station (BS), the latter being stationary and non-strategic, i.e., not a player in the game. R wants to use the available wireless channels most efficiently, that is, receive communication from the BS with the highest possible rate. At the same time, a jammer J is present in the area to disrupt the communication, so that R obtains the lowest possible communication rate instead. To achieve this, J simply raises the noise floor of the ongoing communication by causing extra AWGN in the form of a jamming component to R's reception.

Overall, such a two-player interaction is appealing from a game-theoretic point of view, since it can be framed as a zero-sum game [3]. Notably, the usual interpretation of a mixed strategy Nash equilibrium (NE) in probabilistic terms, which is often troublesome in many contexts, is perfectly valid here if regarded as a random access probability. The setup also allows for additional theoretical considerations such as imperfect information about the role of the players [4], their channel conditions [5], or their physical location [6], which all lead to different versions of a Bayesian game analysis. In the end, whatever the theoretic framework used, strategic countermeasures are derived to be adopted against jamming.

In the literature, there are also variations on the technical premises of the communication scenario. For example, J may be assumed to possess more advanced jamming capabilities to disrupt R's communications, such as eavesdropping or spoofing [7]. Also, a reverse situation happens in the case of *friendly* jamming, where J is a legitimate network agent that wants to disturb a malicious communication by R such as stealing data or communicating in a forbidden area [8], [9].

However, it seems that the element of node mobility is rarely addressed in the literature. Up to the authors' best knowledge, most approaches for adversarial jamming focus on resource allocation or power control [1], [7], [10], whereas there seem to be no contributions discussing wireless nodes, either the jammer or the intended receiver, which are mobile; sometimes the nodes can choose their position, but this is never changed afterwards and kept as a static move or a Bayesian type [11], without any sort of dynamic update. This is kind of surprising since next-generation wireless networks strongly support mobility of the terminals also in extreme situations, not to mention that the overall performance of the wireless channel capacity is strongly affected by mobility [12]–[14]. Also, moving away from a noise source, or more in general changing location whenever the channel quality is bad, is a quite logical reaction and the "poor man's solution" to interference in wireless communications [15]. In fact, [16] proposes a game theoretic scenario for this kind of interaction, but only involving the receiving nodes as moving away from an area where jamming is present, but the jamming nodes are not strategic and do not react to that.

A possible reason for this lack of contributions lies in that, when mobile nodes are considered, the formulation transits to a *dynamic* game and the size of the problem rapidly explodes, since the number of involved strategies becomes prohibitive to be analyzed in closed form. Yet, we argue that a way to approach such an issue is the use of reinforcement learning (RL), which is commonly adopted [17] for the combination of game theory with explorations of the equilibria based on machine learning. Thus, we propose an RL methodology to solve different dynamic games resulting from the ability of nodes to move [18].

Part of this work was supported by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016).

In more detail, in this paper we focus on a linear topology. As discussed in the following, this allows for a simpler analytical formulation, while still keeping into account all the relevant aspects of any two-dimensional geometry, since in the end the computations of the SNJR can be related to physical distance between the terminals. Indeed, R wants to be as close as possible to the BS while at the same time escaping J, which can instead be considered as chasing R [19].

In this context, we study three different scenarios of dynamic games. In the first one, nodes move by dynamically changing their position, which is done *sequentially* with perfect information about each other's positions; individual payoffs are cumulated after each player's move. In the second formulation, we iterate *simultaneous* changes of positions by both players, once again with perfect information. Finally, we consider a third and last game, where each player has *imperfect information* about the opponent's placement, it just observes the channel capacity changing as a consequence of the other player's position.

For all these games, the complexity of the strategies makes it prohibitive to derive a closed-form solution of the game. However, we are still able to identify some general principles of strictly dominant strategies, especially for what concerns the jammer J. Shortly put, this agent can play an efficient strategy which is to always move in R's direction. Also, we adopt an adversarial RL approach to determine the NE strategies through a Q-learning with greedy exploration-exploitation [20]. Finally, we derive numerical results that confirm a good matching with the theoretical principles, and also hint at possible further investigations.

The rest of this paper is organized as follows. In Section II we present the location scenario and the mobility model that the nodes use to play. Section III details the game-theoretic analysis and its solution via reinforcement learning. Section IV presents the performance evaluation and, finally, Section V concludes the paper.

II. Physical Environment and Game Scenarios

We begin our analysis by making some assumptions on the physical geometry of the environment where the nodes are placed and move. If we assume R and J to be terrestrial devices we can limit their positions onto a two-dimensional surface. The extension to drones or devices capable of moving alongside a third vertical coordinate is left for future investigations. R and J's positions can be represented in polar coordinates (ρ, ϑ) with the origin in the location of the BS. It is immediate to remark that, while the distance from the BS ρ really influences the SNJR perceived by the devices, the angle ϑ just implies a rotational symmetry of the positions.

Motivated by this reasoning, in this paper we limit the investigations to a linear geometry, which means that all nodes can occupy positions on a positive coordinate ρ, taken as the only parameter of interest. This model, which would be perfectly appropriate for nodes placed alongside a road [14], still allows for a descriptive setup that contains all of the characteristics that are relevant to our analysis, namely the ability for the players to change position, and do so dynamically.

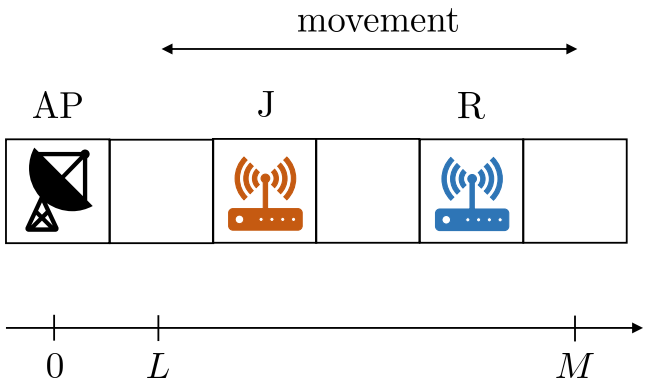

Fig. 1. Graphical representation of the considered game, with the access point located at the origin. In the depicted case, J is in position $y = 2$, and R in $x = 4$; the players can move in both directions between $L = 1$ and $M = 5$.

While using a single coordinate offers the advantage of notation simplicity, we will see that the resulting setup already offers a challenging complexity. Naturally, our findings can be expanded to more general topologies (e.g., with more directions available). Such an extension could for example explore the choice of an angle ϑ alongside the value of ρ for both players. Even though this is not reported here for space limitations, this would actually be a straightforward extension, as will be clear from the following analysis.

Thus, we consider a half-line starting from an origin $\rho = 0$, where three devices are placed. The first one is a transmitter, which can be thought of as a BS, which is stationary and not controlled by any player of the game. This BS is placed in the origin, i.e., position $\rho = 0$, which is in turn forbidden for the other nodes to take. The second and third nodes in the geometric placements are a receiver R and a jammer J, each of them controlled by a player that determines the position, denoted as $\rho_{\mathrm{R}} = x$ and $\rho_{\mathrm{J}} = y$, for R and J, respectively.

For tractability reasons, most games consider a discrete state space [21]; we take a similar approach, and we assume that N positions are available to the mobile nodes, within extremes $\rho = L$ (closest to the BS) and $\rho = M$ (furthest), all equally spaced. Thus, the positions are quantized with a step

$$\Delta = \frac{M - L + 1}{N}.$$

In the numerical evaluations of Section IV we will take the following simplifying choices: $L = \Delta = 1$, $M = N = 5$, which implies 25 possible position pairs of the mobile nodes. However, it is evident that these choices can be extended to more states and a finer granularity, with just increased complexity but little additional insight. A graphical representation of the resulting physical environment is depicted in Fig. 1.

Another related assumption that we exploit in the dynamic game relates to the mobility of the nodes since we assume that, at each round, they can only move to adjacent positions, or stay still. In other words, from position ρ at time t, both players can only move to $\rho' \in \{\rho - \Delta, \rho, \rho + \Delta\}$ at time $t + 1$.

We assume a narrowband AWGN wireless channel between the BS and R, with bandwidth B. The role of the jammer is to cause additional noise at the receiver's end. Thus, R experiences a data rate quantified through Shannon's capacity as $C = B\log_2(1+\Gamma)$ where Γ is the SNJR computed as

$$\Gamma = \frac{g_{\mathrm{R}} P_{\mathrm{tx}}}{\nu_0 + g_{\mathrm{J}} P_{\mathrm{J}}} \tag{1}$$

with terms P_{tx} and P_{J} being the powers transmitted by the BS and the jammer, respectively, while g_{R} and g_{J} are gain terms between the BS and R, and the jammer and R, respectively, and ν_0 is a noise term.

We apply this physical setup to different games where the sets of players always consist of R and J, treated as having contrasting objectives, which is captured by all the games being *zero-sum* [1], [3]: the *value* of the game is defined as the channel capacity, so that player R has utility $u_{\mathrm{R}} = C$, while J has utility $u_{\mathrm{J}} = -u_{\mathrm{R}} = -C$. This means that R tries to obtain a position pair (x, y) that maximizes the channel capacity C, while J wants to minimize it.

If this happens, we can simplify the equations related to the telecommunication scenario by observing that, in game theory, utility functions just represent the preferences of the users but have no strict specific physical meaning in themselves. Thus, certain monotonic rescaling would leave the preferences unaltered since they still respect the principle that the higher the utility, the more preferred the outcome. Especially, if the utility transformation is a linear rescaling, also the set of mixed NEs found is the same in the transformed utilities [21].

Thus, in the following, we make these assumptions. First, we neglect the noise by setting $\nu_0 = 0$, which is reasonable as the jamming effect is expected to be preponderant. At this point, we can also ignore the bandwidth B, which only causes a proportional rescaling of capacity; in other words, we can set a unit value for B and just treat the capacity as the spectral efficiency of the resulting AWGN channel between the BS and R. Moreover, we can ignore the differences between the transmitted powers (either assuming them comparable or remarking that they cause another linear rescaling anyways) and set $P_{\mathrm{tx}} = P_{\mathrm{J}} = 1$.

Since we are left with a value that is a logarithmic function of $\log_2(1+\Gamma)$, under the above assumptions, we can exploit the equivalence $\log(1+a) \sim a$ for small a. After another proportional rescaling due to the change of base for the logarithms, we ultimately obtain that value of the game just depends on the ratio between g_{R} and g_{J}. Finally, if we assume that the channel gain just contain a path loss term, we can ultimately connect the value of the game to the positions x and y as

$$\text{value} = u_{\mathrm{R}} = -u_{\mathrm{J}} = \frac{|x-y|^\alpha}{x^\alpha} \tag{2}$$

where α is the path loss exponent, that is, the channel gain is proportional to $d^{-\alpha}$ where d is the geometric distance.

Some remarks are now in order. First of all, more complex propagation models can be employed, but this would just make the analysis more complicated (especially in the learning part), while the meaning of the communication performance would still be valid on average. Moreover, the dependence on α is also marginal, since it will not alter the ordinal meaning of the utilities. Other results [6], [11] in the field of games with variable position of the users hint at this principle by confirming it numerically. For the sake of choosing a value, in the following, we will consider $\alpha = 2$, which would correspond to a free-space propagation, but as we argued the obtained results are general. Finally, we point out that the eventual formulation of (2) implies a zero-sum game on the unit square with polynomial utilities. According to [22], this is an overall tractable scenario, for which some analytical conclusions can be drawn in the static case. Still, we complicate it by introducing a dynamic evolution of the players' moves due to their mobility.

III. Game Analysis

First of all, we study a static version of the game, where R and J just choose, independently and unbeknownst to each other, a location in $L, L+\Delta, \ldots, M$ and stay there forever. This game does not involve any mobility, but the free choice of the location might shed light on how R and J are supposed to behave in a dynamic game. It is immediate to prove that this game cannot have a NE in pure strategies [23], since all of R's choices of a specific location with probability 1 trigger J's best response of choosing the same location. It turns out that the best strategy for R is to choose a mixed strategy with a combination of locations L and M, to which the best response by J is to place itself in a specific position (even though, as a result of the quantization, the actual strategy played by J is a mixture of the two values surrounding that position). The exact details of this solution are omitted here for brevity, but they are not difficult to derive. However, it is also evident that such a mixed strategy for R is impractical in a dynamic context, due to our assumption that nodes can only move to adjacent positions, instead of teleporting at each round on either side of the road. As we will see, this gives a significant advantage to J in the dynamic version.

This reasoning strongly supports the need for our analysis and possibly casts some doubt over the validity of most of the literature where game-theoretic approaches are just considered in a static context, whereas (as discussed above) dynamic games develop very differently.

As a matter of fact, when we investigate a dynamic setup, we can obtain several different formulations of the game, especially revolving around the order of move of the players, which can be either sequential or simultaneous, and the information about the opponent position. Thus, we formalize the following three dynamic games.

Game $\mathcal{G}_1$: players move sequentially, i.e., taking turns with alternating moves. They start from a uniformly random position and the first player to move is also chosen at random. After each move, the player collects its payoff for that round. Players are always informed of each other positions, so the state of the game (x, y) can be seen as a common type of the players [21].

Game $\mathcal{G}_2$: similar to the previous case in that the positions are known to both players and they start from a position chosen

at random. However, moves are now simultaneous, which can be alternatively seen as the information set of each player i at time t comprising the positions of both players up to time $t-1$, but only that of player i at time t, while the current position of the opponent $-i$ is not distinguishable. After each move, both players collect their payoff for that round.

Game $\mathcal{G}_3$: the players have no information on the position of the opponent at any past, present, or future time. Even though the order of moves is kind of irrelevant here, just to set the ideas we assume that they play simultaneously as per $\mathcal{G}_2$. Importantly, while the position of the opponent is not known, the fact that there is an opponent and must be in a specific position between L and M, and it can move one step at a time, and propagation effects are as described in Section II are all common knowledge among the players. From the point of view of the following analysis, this scenario is configured as a partially-observable system [14].

While these games have an excessively high computational complexity to be analyzed precisely, we can formalize some general principles, formalized by Remarks 1 and 2 below.

Remark 1. *Moving towards R's position is a strictly dominant strategy for player J.*

Proof: It is intuitive to realize that J should follow (and possibly catch) R [19]. Actually, whenever J is in the same position as R, the value of the game goes to 0. Since for any other choice of position, the value is positive, as per (2), J always has an incentive to move towards R. ■

Remark 2. *If J chooses to stay around $\rho = L$, this guarantees to attain an upper bound on the value of the game, regardless of R's actual position.*

Proof: If J is located at position $\rho = L$, R is forced to stay sufficiently far, and move towards M. Indeed, moving further away causes the value of the game to approach 1, which is still better than moving towards L where the value would be 0. Notably, this can be a useful building block for a practical strategy to play in the case of imperfect information. However, in such a case it may also be convenient for J to keep exploring, to further locally reduce the value. ■

Beyond these findings, for tractability reasons it may be convenient to take an alternative approach beyond that of purely analytical computations to assess how the system behaves in dynamic cases. Especially, we resort to reinforcement learning (RL) [18]. We investigate an adversarial scenario involving players R and J as the learning agents. We assume the two agents learn their policy concurrently, and three different cases $\mathcal{G}_1$–$\mathcal{G}_3$, are considered. Q-learning with ϵ-greedy exploration-exploitation policy is used to train the agents' policies online. Specifically, the algorithm relies on the Bellman recursive update of the so-called state-action value function

$$Q(a_t, s_t) = (1-\ell)\, Q(a_t, s_t) + \ell\Big(r + \gamma \max_a Q(a, s_{t+1})\Big) \quad (3)$$

where $s \in \mathcal{S}$ and $a \in \mathcal{A}$ are the observed state of the system and the action chosen by the agent, whereas t refers to the time index, ℓ is the learning rate, and γ the discount factor accounting for the relevance of the future. Both players, hence, keep a table where they update this function, which corresponds to the expected long term reward of taking action a_t while in state s_t. Note that, besides the divergence between simultaneous and sequential moves, the three analyzed cases differ also because of the input state information. In games $\mathcal{G}_1$ and $\mathcal{G}_2$ the state for each agent is the tuple (x, y) (perfect information), while the input in game $\mathcal{G}_3$ (imperfect information) is represented by either element of the position pair, the one corresponding to the position of the moving player.

Since the general goal of the players is to maximize their own payoff, their best policy is to choose the action maximizing the Q-function in a certain state, i.e.,

$$a_t^* = \operatorname{argmax} Q(a, s_t). \quad (4)$$

Motivated by this, one can define the state value function, which corresponds to the average long-term reward obtained following the best policy in a given state, i.e.,

$$V(s_t) = \max_a Q(a, s_t). \quad (5)$$

Rule (4) is named *greedy* policy. However, because the agent must also *explore* the environment's responses, it is convenient to add a certain amount of random behavior. Hence, an ϵ-greedy policy uses (4) to select the action with probability 1-ϵ, and picks a random action otherwise (prob. ϵ). Possibly, the value ϵ is decayed during the time, as it is more important to explore the environment at the beginning.

The considered task is made challenging by the presence of an opponent which modifies the environment's response in a non-stationary way, as it concurrently learns its policy. While several works, exploiting both traditional and deep RL, have shown that Q-learning can be outperformed by more complex algorithms in adversarial contexts modeled through Markov games [20], [24], we chose it for its simplicity, providing evidence, in Section IV-B, that it is good enough for reaching convergence in a proof-of-concept scenario.

IV. Numerical Results

A. Simulation settings

The considered scenario is analogous to the static case discussed previously, also making use of a discretized version of the positions in the range $[L, M] = [1, 5]$. Specifically, we take 5 positions in $\{1, 2, \ldots, 5\}$ and step $\Delta = 1$. Under these numerical assumptions, there are $N^2 = 25$ possible states when players act in the joint space (namely, $\mathcal{G}_1$ and $\mathcal{G}_2$), and $N = 5$ states for $\mathcal{G}_3$. Moreover, there are 3 available actions for $\rho \in \{L + \Delta, \ldots, M - \Delta\}$, i.e., stay still, move left, and move right, while only 2 if $\rho \in \{L, M\}$. This means that each player has to explore 65 values in the joint space, and 13 values, instead, in the case $\mathcal{G}_3$ with imperfect information. The simulation environment is run for 750×10^3 iterations, with learning rate $\ell = 0.01$, and discount factor $\gamma = 0.99$. The first 100×10^3 iterations are performed with a pure exploration policy ($\epsilon = 1$), while for the remaining part ϵ is decayed exponentially up to 0.01. The last 10^4 plays are used for evaluation, with a residual $\epsilon = 0.01$.

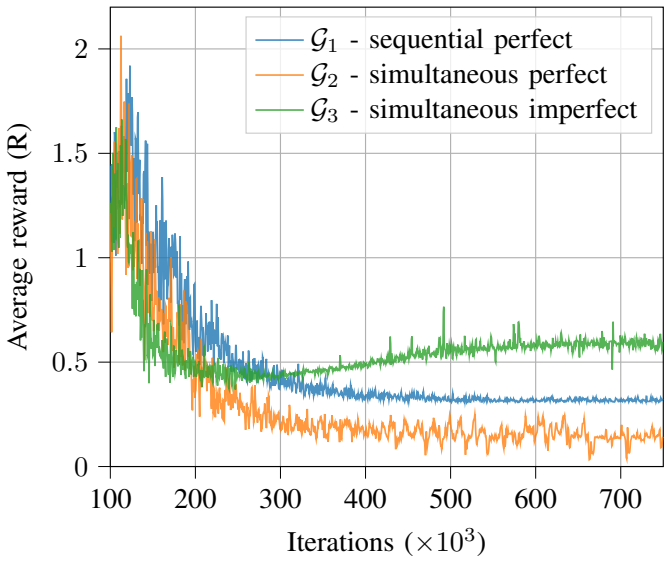

Fig. 2. Receiver's average reward of the tabular Q-learning in the three cases. The result is filtered with a moving average of width $W = 5000$ samples.

B. Stability convergence

In Fig. 2, the receiver's instantaneous reward is plotted as a function of time (iterations) for the three cases, starting from when the exploitation policy is gradually enabled (iteration 100×10^3). The plot shows post-processed values after a moving average (MA) filtering, using a window of size $W = 5000$. As stated before, the three curves stabilize, which provides evidence of convergence towards a stable solution: this happens at around mid simulation for perfect information games, while more time is required under imperfect information.

Notably, among the three cases, the one with imperfect information is the most advantageous for R, as J does not know R's position, and is therefore prevented from following it closely, see Remark 1. On the other hand, the simultaneous game with perfect information is advantageous for J, because it knows R's position and can forecast its moves, keeping always close to it. The sequential game places itself in the middle: J observes R's position and action, but can only react to it as soon as possible, allowing thus a slight improvement in terms of average payoff for R.

It is also worth noting that, while the perfect information games show a monotonic convergence, if the opponent's position is unknown, the jammer is the first player finding a near-optimal policy, while the receiver learns its best response as a consequence. Therefore, a minimum is observable at around iteration 250×10^3.

C. Players policies in the three dynamic games

The logarithmic heatmaps of Figs. 3 and 4 show, on the left, the joint probability of finding R and J in positions x and y, respectively, and, on the right, R's state value function (5), for the perfect information games with sequential ($\mathcal{G}_1$) and simultaneous moves ($\mathcal{G}_2$), respectively. The sequential game $\mathcal{G}_1$ presents a simple equilibrium: R "bounces" between positions 1 and 2, while J can only follow it reactively, see

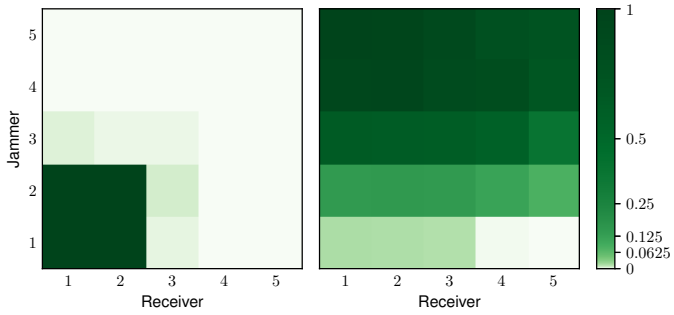

Fig. 3. Game $\mathcal{G}_1$, sequential game with perfect information. Joint probability for R and J to find themselves in position (x, y) (left). State value function of R, i.e., expected long-term reward starting from state (x, y) (right).

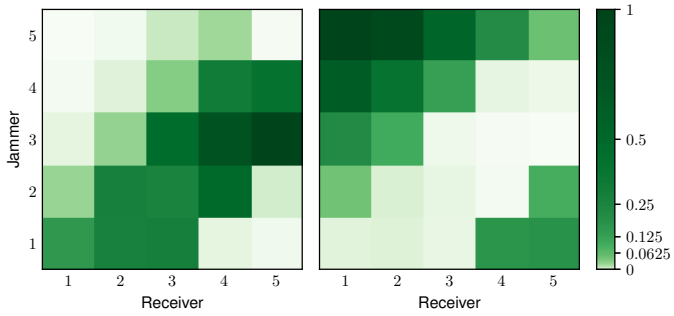

Fig. 4. Game $\mathcal{G}_2$, simultaneous game with perfect information. Joint probability for R and J to find themselves in position (x, y) (left). State value function of R, i.e., expected long-term reward starting from state (x, y) (right).

Fig. 3. This means that 50% of the times, R will be prevented from transmitting (when it is caught by J), but the remaining 50% of the times it is allowed to transmit close to the BS (positions 1 or 2). This translates into a payoff of $5/16$ for R.

Interestingly, in this case, the learned state value function is almost flat concerning x, and only changes as a function of y. Now, we consider $\mathcal{G}_2$, where, instead, the players move simultaneously. As can be seen from the state value function of Fig. 4, R's favorite states are the ones where it is far from J, and possibly also close to the BS, i.e., near to L. However, since J knows R's position, it forces R to visit the complementary states to its favorites (figure on the left). This confirms the very low average payoff reached by R in these conditions: concerning the sequential game analyzed previously, the payoff is halved (Fig. 2).

Finally, in Fig. 5, the game $\mathcal{G}_3$ with simultaneous moves and imperfect information is considered. The long-term value of a state cannot be represented anymore, because the game is not intended for a joint space, and the Q value is only a function of x, and not of y. Therefore, together with the joint probability of the position pair (x, y), the expected instantaneous reward is plotted, i.e., R's payoff in state x while J is in state y, weighted by the probability that R is indeed in x. Since J does not know R's position, it more often chooses to be close to the BS, to prevent R from being there too, coherently with Remark 2. As visible from the heatmap on the left, J lingers around position 1, visiting positions 2-4 with low probability. Consequently, R learns that it must stay

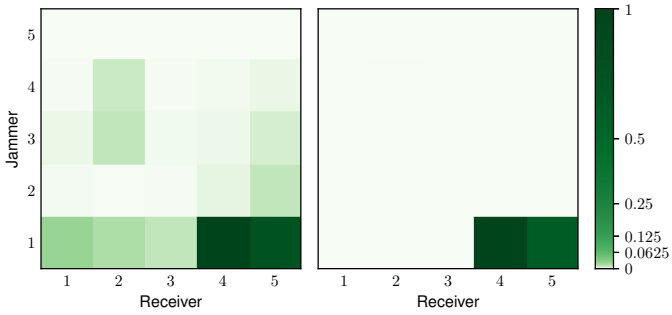

Fig. 5. Game $\mathcal{G}_3$, simultaneous game with imperfect information. Joint probability for R and J to find themselves in position (x, y) (left). Expected instantaneous reward, weighted by the joint position probability (right).

away from J, going toward the opposite side of the geometry. Actually, it positions itself on positions 4 and 5, which give a very similar payoff, with almost equal probablity. As visible from the heatmap on the right, the states where R and J are very far apart are responsible for the vast majority of R's payoff. The average value for the joint positions $(4, 1)$ and $(5, 1)$ is approximately 0.6, which is the convergence point observable in Fig. 2.

V. Conclusions and Future Work

We discussed a jamming scenario where we allow for the mobility of the involved nodes over subsequent rounds. This represents an original extension over the existing literature which may lead to interesting findings.

We considered a dynamic setup with three possible variations related to the timing of the moves of the players and the information available to them. We found out that all the dynamic games obtain interesting results that differ, for the most, from the characterization with a static approach. Nevertheless, some general principles can be formulated and they happen to be well verified and confirmed by our results.

Our results possibly just scratch the surface of an elephantine problem, since it is likely that the approaches to reinforcement learning adopted here can be improved. For example, [25] proposed a federated RL for the similar problem of avoiding jamming, but without involving mobility, and therefore an extension alongside these lines can be interesting. More in general, even though our approach based on Q-learning with greedy exploration-exploitation has proven itself as effective, we conjecture that some original developments are possible for a dedicated technique for this specific problem.

Especially, we remark that the reinforcement learning approach of the scenario with imperfect information just tries to estimate the optimal policy by blind iterations. However, mutual assumptions of the rationality of the players would suggest that some strategic choices can be anticipated, and also a joint estimate of multiple environmental parameters (despite the tremendous growth in the search space) would be more effective to reach the equilibrium. At the same time, different mobility scenarios and patterns can also be adopted. All of these extensions appear to be interesting for future work in this line of research.

References

[1] E. Altman, K. Avrachenkov, and A. Garnaev, "A jamming game in wireless networks with transmission cost," in *Proc. Int. Conf. Netw. Control Optimiz.* Springer, 2007, pp. 1–12.

[2] C. W. Commander, P. M. Pardalos, V. Ryabchenko, S. Uryasev, and G. Zrazhevsky, "The wireless network jamming problem," *J. Comb. Optimiz.*, vol. 14, no. 4, pp. 481–498, 2007.

[3] L. A. DaSilva, H. Bogucka, and A. B. MacKenzie, "Game theory in wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 110–111, 2011.

[4] A. Garnaev, A. P. Petropulu, W. Trappe, and H. V. Poor, "A jamming game with rival-type uncertainty," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5359–5372, 2020.

[5] Y. E. Sagduyu, R. A. Berry, and A. Ephremides, "Jamming games in wireless networks with incomplete information," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 112–118, 2011.

[6] M. Scalabrin, V. Vadori, A. V. Guglielmi, and L. Badia, "A zero-sum jamming game with incomplete position information in wireless scenarios," in *Proc. European Wireless*, 2015, pp. 1–6.

[7] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Comput. Surv.*, vol. 45, no. 3, pp. 1–39, 2013.

[8] J. Kim, P. K. Biswas, S. Bohacek, S. J. Mackey, S. Samoohi, and M. P. Patel, "Advanced protocols for the mitigation of friendly jamming in mobile ad-hoc networks," *J. Netw. Comp. Appl.*, vol. 181, 103037, 2021.

[9] L. Badia and F. Gringoli, "A game of one/two strategic friendly jammers versus a malicious strategic node," *IEEE Netw. Lett.*, vol. 1, no. 1, pp. 6–9, 2019.

[10] Y. Gao, Y. Xiao, M. Wu, M. Xiao, and J. Shao, "Game theory-based anti-jamming strategies for frequency hopping wireless communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5314–5326, 2018.

[11] F. Chiariotti, A. Signori, F. Campagnaro, and M. Zorzi, "Underwater jamming attacks as incomplete information games," in *Proc. IEEE Infocom Workshops*, 2020, pp. 1033–1038.

[12] R. Agüero, B.-L. Wenning, Y. Zaki, and A. Timm-Giel, "Architectures, protocols and algorithms for 5G wireless networks," *Mobile Netw. Appl.*, vol. 23, no. 3, pp. 518–520, 2018.

[13] R. F. El Khatib, N. Zorba, and H. S. Hassanein, "Rapid sensing-based emergency detection: A sequential approach," *Comp. Commun.*, vol. 159, pp. 222–230, 2020.

[14] M. Hussain, M. Scalabrin, M. Rossi, and N. Michelusi, "Mobility and blockage-aware communications in millimeter-wave vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 072–13 086, 2020.

[15] W. Xu, W. Trappe, and Y. Zhang, "Channel surfing: defending wireless sensor networks from interference," in *Proc. IPSN*, 2007, pp. 499–508.

[16] S. Misra, A. Mondal, P. Bhavathankar, and M.-S. Alouini, "M-jaw: Mobility-based jamming avoidance in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5381–5390, 2020.

[17] Y. Arjoune, F. Salahdine, M. S. Islam, E. Ghribi, and N. Kaabouch, "A novel jamming attacks detection approach based on machine learning for wireless communication," in *Proc. IEEE ICOIN*, 2020, pp. 459–464.

[18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[19] G. Perin, A. Buratto, N. Anselmi, S. Wagle, and L. Badia, "Adversarial jamming and catching games over AWGN channels with mobile players," in *Proc. WiMob*, 2021.

[20] W. Uther and M. Veloso. (2003) Adversarial reinforcement learning. [Online]. Available: http://www.cs.cmu.edu/~mmv/papers/03TR-advRL.pdf

[21] S. Tadelis, *Game Theory: An Introduction*. Princeton University Press, October 2012.

[22] I. Glicksberg and O. Gross, "Notes on games over the square," in *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, 2016, ch. 9, pp. 173–182.

[23] T. Parthasarathy, "On games over the unit square," *SIAM Journal on Applied Mathematics*, vol. 19, no. 2, pp. 473–476, 1970.

[24] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, "Robust adversarial reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2817–2826.

[25] N. I. Mowla, N. H. Tran, I. Doh, and K. Chae, "Afrl: Adaptive federated reinforcement learning for intelligent jamming defense in fanet," *J. Commun. Netw.*, vol. 22, no. 3, pp. 244–258, 2020.