# Distributed and Timely Smart Microgrid Management Through Markov Games

Leonardo Badia* and Thomas Marchioro
Department of Information Engineering, University of Padova, Italy
{leonardo.badia, thomas.marchioro}@unipd.it
*Corresponding author

***Abstract*—Energy delivery within smart microgrids often requires prompt reaction to the system state. In the presence of multiple energy sources, inefficiency may arise due to their lack of coordination. In this paper, we frame the task of efficient energy management as a dynamic program, and we further expand it to the case of multiple agents. We combine this approach with game theory and we leverage the similarity between Markov games concerning information and energy exchanges in networks. This methodological motivation allows us to identify distributed control techniques for efficient energy delivery. Specifically, we highlight how a naive distributed and selfish control of individual nodes may be inefficient from a game theoretic perspective. Yet, a decentralized strategy that combines energy availability and global network cost as shared objectives can significantly improve the outcome, approaching the performance of a centralized resource allocation still in a distributed manner.**

***Index Terms*—Smart Grid; Energy management; Distributed control; Markov games.**

## I. Introduction

Smart microgrids (SMGs) are localized networks that integrate environmental energy harvesting, battery storage, and connectivity to the main grid. Their primary objectives are to enhance energy availability and optimize the utilization of renewable sources [1], [2]. To achieve such objectives, a careful monitoring of the system conditions is required, particularly in terms of energy demand and availability in the current network capacity. This, in turn, requires intelligent management techniques to meet the energetic requirements and avoid network downtimes [3].

Even though an SMG operates on a smaller scale than a conventional power grid, it is generally better able to provide energy tailored to the user needs thanks to its combination of advanced sensing, control, and communication technologies [4], [5]. Thus, we can think of adapting the trends characterizing many cyber-physical systems pertaining to digital health [6], the Internet of things [7], or vehicular networks [8], all of them being driven towards timely monitoring and management in the latest research trends. This last point can be addressed by considering that timeliness in information systems is often quantified through age of information (AoI) [9], [10], a metric that characterize the freshness of status updates exchanged by nodes.

This work was supported by the Italian PRIN project 2022PNRR "DIGIT4CIRCLE," project code P2022788KK.

Strategies for prompt delivery of data to the end user, which is of utmost importance for real-time content in context-driven applications, can be applied to a similar extent also in the case of SMGs. While energy delivery and AoI minimization may at first seem different problems, which pertain to different kinds of intangible goods, they are in fact very similar due to their tractability as dynamic programs [11], [12].

In this work, we address the problem of scheduling a finite number of power charging opportunities within a given time window, aiming to maximize energy availability [13]. While this problem was originally conceived for an AoI framework [14], it can be transposed to the case of microgrid management by leveraging the previously mentioned similarities. In our analysis, multiple concurrent sources try to maximize an average reward depending on energy availability and outage avoidance. We consider that multiple sources act independently, which may result on inefficient redundant charging. We also include the impact of the current state of charge (SoC) of the SMG in the optimal policy so as to obtain a stateful solution [15]. While the solution with a single source present could be obtained by dynamic programming, the less efficient solution in the presence of multiple sources can be addressed as a Markov game [16].

Despite the different resulting formalization, we can still gain precious insights by leveraging such an analytical approach [17], and its extension to strategic multi-source interaction. In particular, we can investigate whether a distributed solution can be made efficient and approach, if not in the energy transfer rule, at least in the resulting performance, a fully coordinated approach [18].

Game theory can also play a key role in this context, and more advanced strategic investigations can be used to this end [19]. More in general, our reasoning for the energy provisioning in a smart grid can be extended to include multi-agent scenarios, outlining advanced distributed algorithms that can effectively handle real-time dynamics within the uncertain and complex SMG environments. This results in extending a standard Markov decision process scenario to a Markov game, where the locally optimum choice is replaced with a Nash equilibrium (NE) choice, based on best responses of the players [20]. In this context, we highlight how a carefully tailored choice of the player objectives, which reflects in a different NE, may end up being very close to the optimum allocation that can only be achieved through a centralized approach. This gives the takeaway message that distributed control of an SMG is possible if properly designed.

The remainder of this paper is arranged as follows. Section II presents some background on the scenario and Markov games application to inventory problems. The problem formalization and the proposed solution are discussed in Section III. Numerical results are shown in Section IV. Finally, we conclude in Section V.

## II. Background

SMGs integrate energy harvesting, powerline technologies, and communication networks to enhance electricity generation, delivery, and usage [2]. Their key issue can be arguably related to the maximization of energy availability at any given time, so as to match peak demands and exploit energy reservoir [21]. In particular, the maximization of the energy stored can be regarded to as an instance of inventory problem.

At the same time, the seemingly unrelated research line of AoI aims at quantifying, and possibly maximize, freshness of status updates in a sensing/communication network. AoI is a metric enjoying recent popularity and representing the time elapsed since the last update [9]. Despite their apparent differences, the task of minimizing AoI and maximizing the energy storage of an SMG can be both framed as a (reverse) inventory problem. Aside from minor variations in the setup, such as the formalization as a cost minimization or reward maximization, the goal and the possible strategies to approach the problems are very similar.

As pointed out in [11], due to their nature of inventory problems, both energy availability maximization and AoI minimization can be cast as dynamic programs with properly chosen reward function. This also allows to exploit Bellman optimality conditions and recursion, so as to obtain a divide et impera approach. The inventory problem is indeed broken down into smaller problems applying analytical techniques that are well consolidated for its solution [22].

However, instead of pure dynamic programming, we include the role of multiple agents corresponding to the possible different sources of energy for the SMG system. This changes the formalization from a plain Markov decision process to a Markov game, which implies an extension to the realm of multi-agent optimization [16]. In a Markov game, agents interact with each other and the environment, so that, instead of choosing locally optimal solutions, the goal becomes to follow Nash equilibria (NEs) where each agent's choice is optimal given the strategies of the other agents. Quite logically, this implies new challenges compared to traditional single-agent MDPs.

## III. Problem Setup

The problem of energy supply from external sources in an SMG can be framed as an inventory-like problem where the energy resources are viewed as inventory items and the microgrid has a maximum storage capacity equal to $B$. The value of $B$ represents the capacity of the energy storage units, such as batteries. Since the goal is to manage the availability and utilization of resources over time, we set a reward function that simply represents the energy level available to the microgrid. In addition, we include a corrective term to penalize energy outages [23].

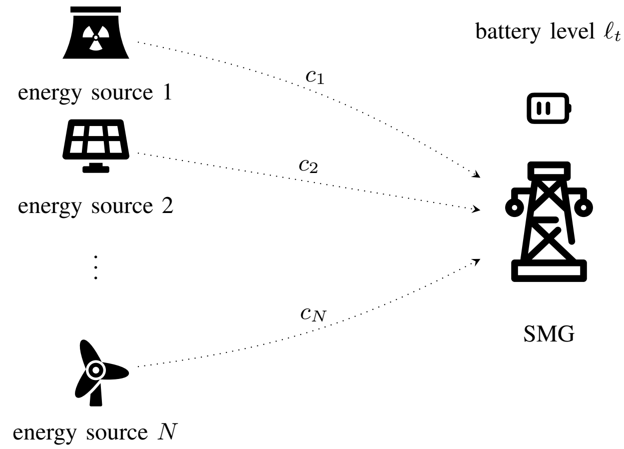


Fig. 1. Example of $N$ energy sources powering a smart microgrid (SMG). The value of $c_n$ is set to 1 if the $n$-th source handles the energy replenishment.

The energy demand within a SMG can vary according to various factors, including user consumption patterns, weather conditions, and time of day. On the supply side, the energy availability from external sources is affected by factors such as grid connectivity, grid stability, and energy purchase agreements. Solving the resulting resource allocation problem generally requires awareness of these conditions [24], while at the same time balancing energy supply and demand within the microgrid requires to efficiently manage the inventory of external energy resources [5]. For the sake of simplicity, in this analysis we consider a linearly decreasing constant usage of energy availability, in line with standard inventory problems, until energy is replenished by an external source [11].

For an SMG, energy needs to be replenished by external sources to maintain a sufficient supply to meet the energy demands. This can be achieved, e.g., by purchasing energy from the main grid or receiving energy from other connected grids. The process of replenishment ensures that the microgrid has an adequate inventory of external energy resources [25].

We consider a slotted time axis, where a time slot is the decision period of a single energy source, and in addition it corresponds to the *vulnerability interval* during which a concurrent transmission is not detected. On top of this requirements, we also introduce an externality in the form of a cost $C$ for the energy transfer. Replenishment can be made at will, even though a constraint is implicit in setting a cost that sums with the user's welfare. The objective of the management can be directly related to the energy level available to the SMG, as this translates to optimizing the utilization of energy from external sources to minimize energy loss or oversupply [21].

In the following, we assume that an *individual* reward gained by source $n \in \{1, \dots, N\}$ is equal to the energy level of the SMG, minus outage penalties and individual costs [16]. We denote with $\ell_t \in [0, B]$ the energy level at time $t$ and with $K$ the proportional reward coefficient of the energy level. We use $\Omega$ for the outage penalty constant, and $C$ for the transmission cost. The instantaneous reward of the $n$th source at time $t$ can be thus written as

$$\mathcal{R}_{n,t}(\ell_t) = K(\ell_t/B) - \Omega\mathbb{1}[\ell_t = 0] - C\mathbb{1}[c_n = 1] \quad (1)$$

where $\mathbb{1}[\cdot]$ denotes the indicator function, which is equal to 1 if the argument is true, 0 otherwise. The indicator function models the fact that transmission cost is included if $c_n = 1$, i.e., only if the source itself is handling the replenishment.

In the following, we set the normalized reward as $K/B = 10$, the outage penalty as $\Omega = 5$, and the transmission cost $C$ in the range $[0.5, 10]$. We also assume that, in the absence of a replenishment event, the energy level of the SMG decreases by 5% at each slot. All these numerical choices just relate to getting a sufficiently dense time granularity but are not

We model the evolution of the energy level as a Markov chain, where the energy level $\ell_t$ is the discretized system state and its evolution follows a transition probability $q_{ij} = \mathbb{P}[\ell_{t+1} = j | \ell_t = i]$ that depends on the probability $p_i$ that an energy source replenishes the SMG when the level $\ell_t$ is $i$. From state $i$, only states $B$ and $i-1$ are accessible in one slot. Given that all energy sources are symmetrical, the transition probabilities are

$$q_{i,j} = \begin{cases} (1-p_i)^N \;\; \forall\, i \in \mathbb{Z}_+ & j = i-1 \\ 1-(1-p_i)^N \;\; \forall\, i \in \mathbb{Z}_+ & j = B \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Then, according to [16], we can define the expected *distributed selfish* reward as $\mathbb{E}\left[\mathcal{R}_t^s(i)\right]$ when the state is $i$ and the action taken is described by $p_i$. Due to the symmetry of the problem, the expected reward is equal for all nodes, an we can thus omit the node index $n$. Since the actions of the nodes are independent, we can write

$$\mathbb{E}\left[\mathcal{R}_t^s(i)\right] = \mathbb{E}\big[Ki/B - \Omega\mathbb{1}[i = 0] - Cq_{i,B}\big]. \quad (3)$$

However, we can also consider a *centralized* policy, which implies that the nodes are coordinated, acting with the goal of avoiding multiple concurrent replenishment. This corresponds to an equivalent system with just one source, whose expected reward replaces the system transition probabilities in (2) with just $p_j$ (only one source is transmitting. Thus, the we can modify (3) to consider a centralized reward $\mathcal{R}_t^c(i)$ whose expected reward is

$$\mathbb{E}\left[\mathcal{R}_t^c(i)\right] = \mathbb{E}\big[Ki/B - \Omega\mathbb{1}[i = 0] - Cp_i\big]\,. \quad (4)$$

Finally, we can define a *distributed global* policy chosen as a kind of intermediate methodology. Here, even though nodes act as individual agents, the local costs are considered to be the ones incurred by the entire system. As before, all the nodes share the objective of maximizing the reward, i.e., keeping the energy level as high as possible, avoiding outages, and minimizing replenishment costs. In this way, we acknowledge that multiple sources are serving the SMG, while still considering a distributed approach. We expect this to lead to inefficiency due to lack of coordination [14], albeit more contained than the selfish approach. In this case, the expected reward becomes

$$\mathbb{E}\left[\mathcal{R}_t^s(i)\right] = \mathbb{E}\big[Ki/B - \Omega\mathbb{1}[i = 0] - NCq_{i,B}\big] \quad (5)$$

The addition of the coefficient $N$ in the cost function of the latter distributed policy is meant to contain the problem known as the *tragedy of the commons* [20], which affects distributed management when individual players are moved by selfish objecives.

The resulting Markov decision process built on top of this chain is easy to solve when the agents act in a coordinated fashion. However, for the distributed policies we treat it as a Markov game, where we model the simultaneous choices of the players as a static game of complete information. In this case, we consider that multiple agents make independent decision on whether to perform a replenishment, with the objective of computing the inefficiency coming from their lack of coordination [26].

Whether distributed or centralized, the problem can be solved via value iteration [27]. This requires the introduction of discount factor $\gamma \in [0, 1[$ to obtain a finite expected reward. In the following, we set $\gamma = 0.99$, but other values provide analogous results. Then, we estimate the discounted value of the present state $\ell_t = i$ under an optimal policy $\pi^* : \{1, \dots, B\} \to [0, 1]$ as

$$v^h(i) \doteq \mathbb{E}\left[\sum_{k=0}^{+\infty} \gamma^k R^h_{t+k+1} \,\middle|\, \ell_t = i, p_k = \pi^*(i)\right] \quad (6)$$

for states $i = 1, \dots, B$ and reward type $h \in \{s, g, c\}$. In (6), $R_t$ denotes the reward obtained at time $t$ by following the optimal policy $\pi$. The state value is estimated by iterating the following update rule until convergence:

$$\begin{aligned} v^h_{k+1}(i) &\doteq \max_{p_i} \mathbb{E}\left[R^h_{t+1} + \gamma v^h_k(\ell_{t+1}) \middle| \, \ell_t = i\right] \\ &= \max_{p_i} \mathbb{E}\left[\mathcal{R}^h_\infty(i) + \sum_{j=1}^{B} \gamma\, q_{i,j} v^h_k(j)\right]. \end{aligned} \quad (7)$$

## IV. Numerical Results

We assess energy provision from $N = 5$ sources in an SMG by assuming battery of $B = 20$ units and linear energy usage of 1 energy unit per time epoch. We compare three possible strategies to apply VI and derive the replenishment policies by individual nodes, as well as the following performance metrics: (i) average probability of replenishment under the optimal policy; (ii) average reward under the optimal policy. Both (i) and (ii) are evaluated for variable costs of replenishment.

Fig. 2 shows the probability distribution of the three policies. Interestingly, the centralized policy is threshold-based, with a sudden decrease to 0 in the probability of replenishment when the state is high. Moreover, this transition point depends on the cost (the higher $C$, the lower the threshold).

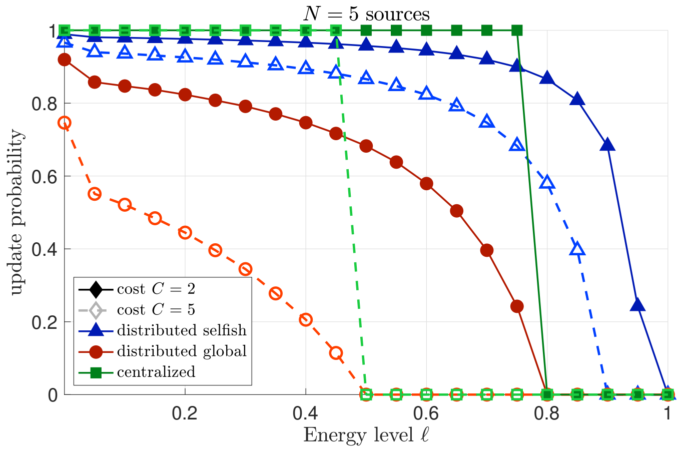


Fig. 2. Comparison of different policies for $N$=5 energy sources for different values of the energy transfer cost.

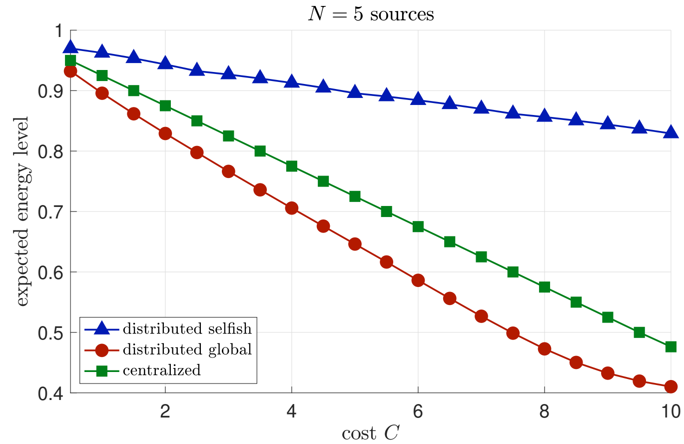


Fig. 4. Average energy level of the SMG vs. energy transfer cost, $N = 5$ energy sources.

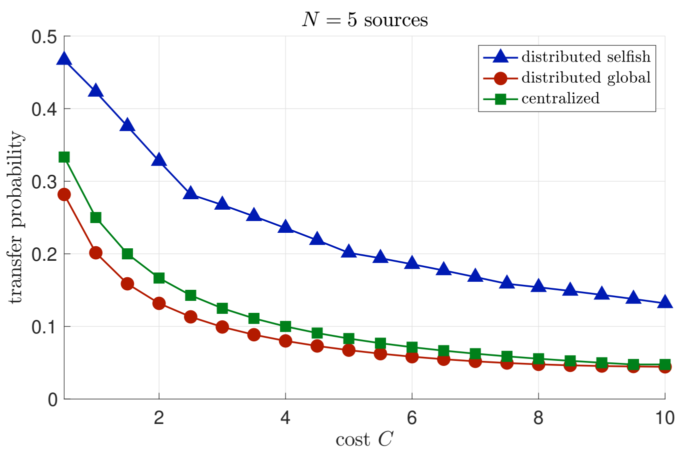


Fig. 3. Average energy transfer probability vs. energy transfer cost, $N = 5$ energy sources.

Conversely, the distributed policies can only approach this with a smoother behavior, as they have no control on who is going to perform the replenishment since they have no direct form of coordination.

Fig. 3 considers the average replenishment probability obtained by the three policies, versus the replenishment cost. One can see that the three policies offer a similar trend, i.e., replenishment actions become less frequent as they are more expensive. It is worthwhile noting that the distributed selfish approach is more aggressive than the centralized optimum. A distributed global approach appears to closely mimic the centralized optimum and qualifies as a better approach for a centralized-like management.

To better elaborate on this trend, we consider in Fig. 4 the expected energy level. Notably, the distributed selfish approach is the one with highest level of energy but this also comes at a cost. Indeed, this confirms that the distributed selfish approach is at risk of overloading (a possible limitation to remove in further versions of the approach is that the types are common knowledge).

## V. Conclusions and Future Work

In this paper, we applied the rationale of Markov games to the problem of energy replenishment in an SMG with multiple available sources [16]. This is meant to provide NE allocations that come close to a centralized optimal strategy for delivering energy to the SMG.

In general, the extension of dynamic programming to Markov games for the case of multiple agents and the exploitation of recursive procedures such as value iteration [27] makes for a convincing distributed approach. Our work assumptions, such as requiring full knowledge on the system state, are also easy to extend to more complicated scenarios.

For this reason, future work will involve the extension to energy replenishment problems within more elaborate game theoretic scenarios, possibly including a Bayesian approach [19]. In addition, game theory can also be applied to security issues for SMGs, due to their susceptibility to malicious data injection and attacks [28], [29].

In general, game theory provides a tool to model and analyze the strategic interactions between multiple agents, either collaborative as in the present paper, but also driven by malicious intents [30]. The advantage of a game theoretic investigation lies in its analytical character and its direct implementation through distributed dynamic management [12].

## References


[1] D. Roana, S. Boscolo, L. Crosara, L. Badia, and E. Gindullina, "Strategic energy trading among prosumers in a smart grid," in *Proc. icSmartGrid*. IEEE, 2023.

[2] M. Ghiasi, Z. Wang, M. Mehrandezh, S. Jalilian, and N. Ghadimi, "Evolution of smart grids towards the Internet of energy: Concept and essential components for deep decarbonisation," *IET Smart Grid*, vol. 6, no. 1, pp. 86–102, 2023.

[3] E. Giglio, G. Luzzani, V. Terranova, G. Trivigno, A. Niccolai, and F. Grimaccia, "An efficient artificial intelligence energy management system for urban building integrating photovoltaic and storage," *IEEE Access*, vol. 11, pp. 18 673–18 688, 2023.

[4] M. A. Judge, A. Khan, A. Manzoor, and H. A. Khattak, "Overview of smart grid implementation: Frameworks, impact, performance and challenges," *J. En. Stor.*, vol. 49, p. 104056, 2022.

[5] G. A. Pagani and M. Aiello, "Generating realistic dynamic prices and services for the smart grid," *IEEE Syst. J.*, vol. 9, no. 1, pp. 191–198, 2014.

[6] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of EEG and EMG signals for wireless biometrics," in *Proc. IEEE Globecom*, 2018.

[7] Y. Gu, H. Chen, Y. Zhou, Y. Li, and B. Vucetic, "Timely status update in Internet of things monitoring systems: An age-energy tradeoff," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5324–5335, 2019.

[8] U. Michieli and L. Badia, "Game theoretic analysis of road user safety scenarios involving autonomous vehicles," in *Proc. IEEE PIMRC*, 2018, pp. 1377–1381.

[9] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[10] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Trans. Mobile Comput.*, vol. 19, no. 12, pp. 2903–2915, 2019.

[11] L. Badia and A. Gandelli, "A comparison of status update optimization and microgrid management," in *Proc. IEEE CGEE*, 2023, pp. 6–10.

[12] P. Zeng, H. Li, H. He, and S. Li, "Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4435–4445, 2018.

[13] P. Rafiee, Z. Ju, and M. Doroslovački, "Adaptive on/off scheduling to minimize age of information in an energy harvesting receiver," *IEEE Sensors J.*, vol. 24, no. 3, pp. 3898–3911, Feb. 2024.

[14] L. Badia, "Age of information from two strategic sources analyzed via game theory," in *Proc. IEEE CAMAD*, 2021.

[15] N. Michelusi, K. Stamatiou, L. Badia, and M. Zorzi, "Operation policies for energy harvesting devices with imperfect state-of-charge knowledge," in *Proc. IEEE ICC*, 2012, pp. 5782–5787.

[16] M. Pagin, L. Badia, and M. Zorzi, "A Markov game of age of information from strategic sources with full online information," in *Proc. IEEE ICC*, 2023, pp. 76–81.

[17] V. Tripathi and E. Modiano, "An online learning approach to optimizing time-varying costs of AoI," in *Proc. ACM MobiHoc*, 2021, pp. 241–250.

[18] N. Bastianello and L. Badia, "Decentralized intersection control using Bayesian game theory," in *Proc. Int. Seminar Machine Learning Optimiz. Data Sc. (ISMODE)*, 2022, pp. 442–447.

[19] A. Tolio, D. Boem, T. Marchioro, and L. Badia, "A Bayesian game framework for a semi-supervised allocation of the spreading factors in lora networks," in *Proc. IEEE UEMCON*, 2020, pp. 0434–0439.

[20] L. Prospero, R. Costa, and L. Badia, "Resource sharing in the Internet of Things and selfish behaviors of the agents," *IEEE Trans. Circuits Syst. II*, vol. 68, no. 12, pp. 3488–3492, 2021.

[21] Á. F. Gambin, E. Gindullina, L. Badia, and M. Rossi, "Energy cooperation for sustainable iot services within smart cities," in *Proc. IEEE WCNC*, 2018.

[22] D. Bertsekas, *Dynamic programming and optimal control*. Athena scientific, 2012, vol. 1.

[23] R. Garmabdari, M. Moghimi, F. Yang, E. Gray, and J. Lu, "Multi-objective energy storage capacity optimisation considering microgrid generation uncertainties," *Int. J. El. Pow. En. Syst.*, vol. 119, p. 105908, 2020.

[24] L. Badia, A. Baiocchi, A. Todini, S. Merlin, S. Pupolin, A. Zanella, and M. Zorzi, "On the impact of physical layer awareness on scheduling and resource allocation in broadband multicellular IEEE 802.16 systems," *IEEE Wireless Commun.*, vol. 14, no. 1, pp. 36–43, 2007.

[25] R. Ashrafi, M. Amirahmadi, M. Tolou-Askari, and V. Ghods, "Multi-objective resilience enhancement program in smart grids during extreme weather conditions," *Int. J. El. Pow. En. Syst.*, vol. 129, p. 106824, 2021.

[26] D. A. Temesgene, M. Miozzo, D. Gündüz, and P. Dini, "Distributed deep reinforcement learning for functional split control in energy harvesting virtualized small cells," *IEEE Sustain. Comput.*, vol. 6, no. 4, pp. 626–640, 2020.

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

[28] J. Liu, Y. Xiao, S. Li, W. Liang, and C. P. Chen, "Cyber security and privacy issues in smart grids," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 981–997, 2012.

[29] M. Borgo, B. Principe, L. Spina, L. Crosara, L. Badia, and E. Gindullina, "Attack strategies among prosumers in smart grids: A game-theoretic approach," in *Proc. icSmartGrid*. IEEE, 2023.

[30] T. Marchioro, A. Kazlouski, and E. P. Markatos, "User identification from time series of fitness data." in *SECRYPT*, 2021, pp. 806–811.