

Distributed Constrained Optimization Towards Effective Agent-based Microgrid Energy Resource Management^{*}

Fernando Lezama¹[0000–0001–8638–8373], Enrique Munoz de Cote^{2,3}[0000–0002–3249–096X], Alessandro Farinelli⁴[0000–0002–2592–5814], João Soares¹[0000–0002–4172–4502], Tiago Pinto¹[0000–0001–8248–080X], and Zita Vale⁵[0000–0002–4560–9544]

¹ GECAD-ISEP, Polytechnic of Porto, Porto, Portugal
`{flzc1,jan,tcp}@isep.ipp.pt`

² PROWLER.io, Cambridge, United Kingdom

³ Department of Computer Science, INAOE, Mexico
`enrique@proowler.io`

⁴ Computer Science Department, University of Verona, Italy
`alessandro.farinelli@univr.it`

⁵ Polytechnic of Porto, Porto, Portugal
`zav@isep.ipp.pt`

Abstract. The current energy scenario requires actions towards the reduction of energy consumption and the use of renewable resources. In this context, a microgrid is a self-sustained network that can operate connected to the smart grid or in isolation. The long-term scheduling of on/off cycles of devices is a critical problem that has been commonly addressed by centralized approaches. In this work, we propose a novel agent-based method to solve the long-term scheduling problem as a distributed constraint optimization problem (DCOP) by modelling future system configurations rather than reacting to changes. Moreover, with respect to approaches based on decentralised reinforcement learning, we can directly encode system-wide hard constraints (such as for example the Kirchhoff law) which are not easy to represent in a factored representation of the problem. We compare different multi-agent DCOP algorithms showing that the proposed method can find optimal/near-optimal solutions for a specific case study.

Keywords: decentralized · distributed optimization · microgrid · multi-agent systems · smart grid · energy resource management.

^{*} This work has been developed under the MAS-SOCIETY project - PTDC/EEL-EEE/28954/2017 and has received funding from UID/EEA/00760/2019, funded by FEDER Funds through COMPETE and by National Funds through FCT. This work has been also partially supported by the project "GHOTEM" Global HOUse Thermal & Electrical energy Management for Efficiency, Lower emission and Renewables, founded by the Veneto Region through the POR FESR 2014-2020 founding scheme (Action 1.1.4), DGR n. 1139 19 July 2017.

1 Introduction

A microgrid is a small self-sustained power network, with local distribution and local generators, that uses renewable energy. Moreover, a *smart* microgrid uses a combination of the electric network, local generators, information and communication technologies (ICT), and intelligent systems to provide control in an effective and efficient way [22].

One of the main problems to tackle in a microgrid is the control and management of resources through the scheduling of on/off cycles of devices. An optimal scheduling of devices is traditionally achieved by centralized approaches [19, 12]. However, such approaches lack flexibility, since adding new devices to the system implies the recalculation of the entire scheduling, and robustness, because losing the central unit might result in a shut down for the whole system [3, 20]. On the other hand, a decentralized approach is more flexible, allowing the addition of new devices and performing the optimization in a distributed way, hence following the trend of intelligent distribution networks (i.e., the SG). Commonly, the decentralized approach has been applied to the scheduling problem without considering long-term optimization [13, 14].

Recently, significant effort has been devoted to decentralised solutions that can take into account the dynamics and sequential nature of the environment. A relevant approach in this domain is modeling the problem as a distributed constraint optimization problem (DCOP); see e.g. [9], which provides a review on the current status and future perspectives on DCOP in multiple domains. [4] presents a survey on DCOP to support multi-agent systems (MAS) in complex, real-time, and uncertain environments. The paper proposes a classification of DCOP multiple extensions, focused both on solving methods and applications. The classification identifies future evolutions for DCOP, and identifies challenges in the design of efficient algorithms by adapting approaches from diverse fields. [6] proposes an integrated approach to solve the economic dispatch and demand response problems that simultaneously maximizes the benefits of customers and minimizes the generation costs, and introduces a multi-agent-based algorithm, based on DCOPs, [5] that formalizes a smart home scheduling problem, describes a mapping of this problem to a distributed constraint optimization problem, proposes a distributed algorithm for solving the problem, and presents results from a physically distributed system capable of controlling smart devices through hardware interfaces. Additionally, [2] introduces an algorithm that adopts the optimal asynchronous partial overlay (OPTAPO) technique, which uses distributed constraint agent search to solve distributed DCOP sub-problems in a multi-agent environment. The use of the Dynamic DCOP framework has also been proposed, where the system evolution is modelled as a sequence of canonical DCOPs providing a new solution each time the system changes its configuration [17].

On the other hand, a series of works aims at considering the system dynamics by modelling the transition function for system configurations by using reinforcement learning techniques that can perform computation in a decentralised fashion [16, 8, 21, 1], including cooperative [7] and non-cooperative distributed

reinforcement learning approaches [18]. A core component of these approaches is to exploit the structure of the domain, so to have a factored representation of the problem that can be conveniently solved in a decentralized fashion by message passing techniques.

In this paper, we propose a mathematical model for the long-term microgrid control. This is a modification of a former model presented in [11, 10], which is optimised for distributed settings. After that, we characterize the mathematical formulation as a DCOP. In the agent-based optimization, to keep the model tractable for long-time horizons, we split the problem into time windows and solve it using different off-the-shelf approaches.

In contrast to approaches based on Dynamic DCOPs, our method explicitly performs long-term optimization by modelling future system configurations rather than reacting to changes. Moreover, with respect to approaches based on decentralised reinforcement learning, we can directly encode system-wide hard constraints (such as for example the Kirchhoff law) which are not easy to represent in a factored representation of the problem⁶.

The results show that, even when the agent-based distributed approach provides optimal and near-optimal solutions for small window sizes, it pays a large computational cost associated with the interaction of agents for large window sizes.

2 Problem formulation for distributed settings

Let $G = \{1, \dots, n\}$ be the set of all generators and $type_i$ be the type of energy produced by each generator $i \in G$. Furthermore, let $E_{G_i}(t)$ be the instantaneous (at timestep t) energy produced by each generator $i \in G$ and $E_{B_j}(t)$ be the instantaneous energy being stored ($E_{B_j}(t) > 0$) or used ($E_{B_j}(t) < 0$) by each battery $j \in B = \{1, 2, \dots, m\}$ at time t . Because a microgrid is never completely balanced, let $E_{dis}(t)$ be the amount of undelivered ($E_{dis} > 0$) or exceeded ($E_{dis} < 0$) energy at each time. Finally, we use C to refer to costs, where C_{G_i} is associated with the cost of energy production through $type_i$ generator, $C_{B_j^+}/C_{B_j^-}$ with the cost of charging/discharging battery j , and $C_{E_{dis}^-}/C_{E_{dis}^+}$ with the cost of undelivered/exceeded energy. Now that energy generation and storage is defined, let $Load(t)$ captures an aggregated quantity of energy consumption.

The objective is to minimize the cost of generated energy (first term of Eq. (1)) and storage/used battery energy (second term of Eq. (1)), while minimizing the unbalance between production and consumption (third term of Eq. (1)) in the microgrid. This can be formulated as follows:

$$\text{Min } f = \sum_{t=1}^T \left(\frac{\sum_{i \in G} E_{G_i}(t) C_{G_i}(t) - \sum_{j \in B} E_{B_j}(t) \mathbf{C}_B(E_{B_j}(t)) + E_{dis}(t) \mathbf{C}_{dis}(E_{dis}(t))}{\sum_{j \in B} E_{B_j}(t) \mathbf{C}_B(E_{B_j}(t)) + E_{dis}(t) \mathbf{C}_{dis}(E_{dis}(t))} \right) \quad (1)$$

⁶ Nonetheless, we believe that using decentralised learning approaches for scheduling on/off cycles of devices in the microgrid deserves further investigation, and is definitely an interesting direction for future work

where $\mathbf{C}_B(\cdot)$ and $\mathbf{C}_{E_{\text{dist}}}(\cdot)$ are functions returning the battery and disbalance cost depending on the conditions (i.e., battery charge/discharge status and undelivered/exceeded energy) defined as:

$$\mathbf{C}_B(E_{B_j}(t)) = \begin{cases} C_{B_j^+} & \text{if } E_{B_j}(t) > 0 \quad (\text{battery charging}) \\ C_{B_j^-} & \text{otherwise} \end{cases} \quad (2)$$

$$\mathbf{C}_{\text{dis}}(E_{\text{dis}}(t)) = \begin{cases} C_{E_{\text{dis}}^-} & \text{if } E_{\text{dis}}^- > 0 \quad (\text{demand} > \text{generation}) \\ C_{E_{\text{dis}}^+} & \text{otherwise} \end{cases} \quad (3)$$

Subject to the following constraints:

- Kirchhoff law or power balance:

$$\sum_{i \in G} E_{G_i}(t) - \sum_{j \in B} E_{B_j}(t) + E_{\text{dis}}(t) = \text{Load}(t) \quad \forall t \quad (4)$$

where $\text{Load}(t)$ is the sum of energy required by all consumers at time t . $E_{\text{dis}}(t)$ is a variable that balances above equation when *demand* \neq *generation*, meaning that:

$$E_{\text{dis}}(t) = \text{Load}(t) - \left(\sum_{i \in G} E_{G_i}(t) - \sum_{j \in B} E_{B_j}(t) \right) \quad (5)$$

$E_{\text{dis}}(t)$ takes positive values when *demand* $>$ *generation* and negative values otherwise.

- Energy type production limits at time t :

$$0 \leq E_{G_i}(t) \leq \text{cap}_G(\text{type}_i) \quad \forall i \in G, t \quad (6)$$

where $\text{cap}_G(\text{type}_i)$ is the limit capacity if generator type_i .

- Storage, charge and discharge battery limits at each time t :

$$CH_j(t) \leq \text{cap}_B \quad \forall j \in B, t \quad (7)$$

$$\Delta \text{cap}_{B^-} \leq E_{B_j}(t) \leq \Delta \text{cap}_{B^+} \quad \forall j \in B, t \quad (8)$$

where CH_j is the current battery $j \in B$ charge level, cap_B is the maximum energy storage capacity of any battery, and $\Delta \text{cap}_{B^+}/\Delta \text{cap}_{B^-}$ are the maximum instantaneous (per timestep) charge/discharge rate of any battery.

- State balance of the battery:

$$CH_j(t) = E_{B_j}(t) + CH_j(t-1) \quad \forall j \in B, t \quad (9)$$

- Charge and discharge limits at time t considering period $t-1$:

$$0 \leq E_{B_j}(t) + CH_j(t-1) \leq \text{cap}_B \quad j \in B \quad (10)$$

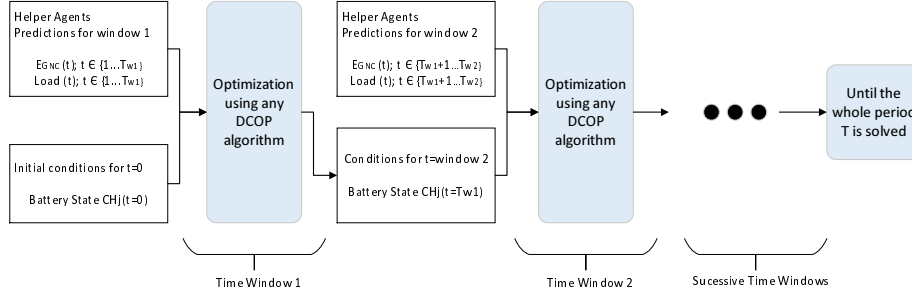


Fig. 1: Optimization using time windows.

The new formulation can also be optimally solved using Mixed-integer Linear Programming (MILP) in a centralized fashion as in [15]. In the next section, we present a long-term decentralized approach using this new formulation that allows solving the problem using agent-based theory and DCOP algorithms.

3 Agent-based long-term optimization

We consider a microgrid with four types of elements: *i*) consumers (i.e., *Load*), *ii*) non-controllable generators ($G_{NC} \subseteq G$), *iii*) controllable generators ($G_C \subseteq G$) and *iv*) storage devices (B).

To model the dispatch problem from Sect. 2 as a DCOP, we need to define agents, variables, domains, and functions according to the mathematical formulation.

First, we assign each type of element to a specific type of Agent. So, in our model we have four types of agents: *i*) \mathcal{A}_{Load} , *ii*) $\mathcal{A}_{G_{NC_k}}$, *iii*) $\mathcal{A}_{G_{C_i}}$ and, *iv*) \mathcal{A}_{B_j} corresponding to each type of element.

These agents are in charge of variables related to energy production or consumption, i.e., *i*) $Load(t)$, *ii*) $E_{G_{NC_k}}(t)$, *iii*) $E_{G_{C_i}}(t)$, and $E_{B_j}(t)$.

The domains of each variable are defined as: *i*) $[0, cap_{Load}]$, *ii*) $[0, cap_{G_{NC_k}}]$, *iii*) $[0, cap_{G_{C_i}}]$, and *iv*) $[\Delta cap_{B-}, \Delta cap_{B+}]$.

Finally, the agents must coordinate their actions to minimize the cost function from Eq. (1), subject to the constraints defined through Sect. 2.

In this model, agents assigned to non-controllable elements (i.e., \mathcal{A}_{Load} and $\mathcal{A}_{G_{NC}}$) are considered helper agents, and their functions are simply to provide the forecast of energy production and consumption in a determined period. That information is taken as an input for agents assigned to controllable elements (i.e., \mathcal{A}_{G_C} and \mathcal{A}_B) to perform optimization in a distributed way.

The distributed agent-based model enables to treat the problem as a DCOP, allowing the use of distributed agent algorithms. However, optimally solving a DCOP is known to be an NP-complete problem, hence solving the long-term optimization problem directly will be impractical even for short optimization horizons.

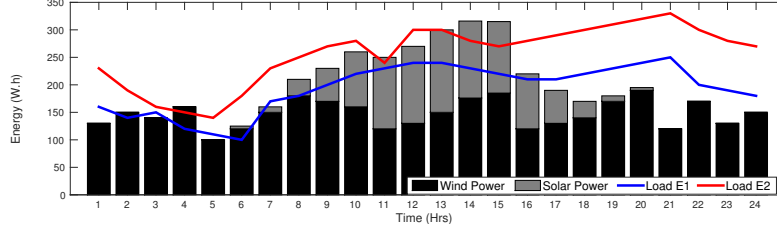


Fig. 2: Forecast of distributed energy sources (DES) and Load for scenarios E1 and E2.

We proposed in our previous work to split the problem in time windows [10]. In this way, for a period T , we can solve the problem by dividing such period T in $N_{windows} = T/n$, where n is the size of the window. This means that for a size $n = 1$ we will solve T windows, for $n = 2$ we will solve $T/2$ windows and so on. Figure 1 shows the scheme of optimization in time windows. The input for the first window corresponds to initial conditions of the microgrid (e.g., the battery could start with an initial charge of 100 W, and a fuel-cell generator in off state ready for being activated).

Then, optimization for that window is done by using any of the algorithms to solve the DCOP. The result obtained in that window is used as input for the next window. The process is repeated sequentially until a solution for the long-term period T is obtained.

One disadvantage of this approach is that the optimal solution cannot be guaranteed and depends directly on the size of the window chosen. Small time windows (e.g., size 1) are solved fast since the optimization is done without considering the available knowledge on the future, keeping the information and variable relations low. However, the quality of the solution may not be good enough because of the lack of global vision for the conditions of the successive periods (e.g., wasting all the resources at the current time without considering the demand for the successive time).

In general in this model for an arbitrary window size (WS) and considering only 2 variables (i.e., variable FC owned by agent \mathcal{A}_{FC} and variable B owned by agent \mathcal{A}_B), the number of variables grows as a function of i) $2 * WS$; the number of relations between variables is given by ii) $WS + WS * (WS - 1)/2$; finally, the domain size grows according to iii) $|D_{FC}|^{WS} * |D_B|^{WS}$, where $|D_i|$ is the cardinality of the domain of variable i .

It is expected that the quality of the solution improves by increasing the size of the windows. However, as shown in Fig. 3, the number of variables to handle for each agent grows significantly with the size of the windows, hence resulting into an exponential grow of the number of possible configuration to explore.

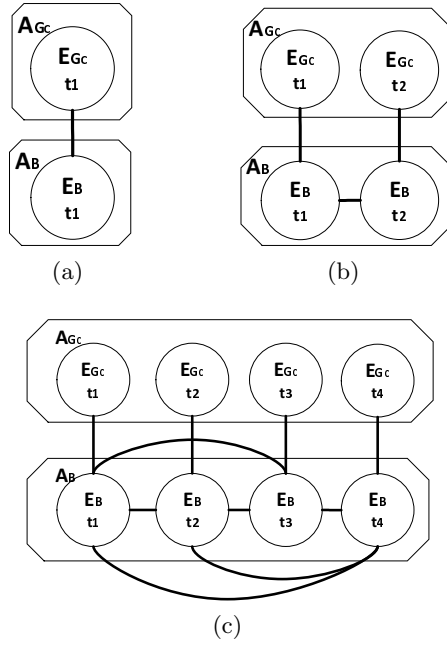


Fig. 3: Agents and variable relations with window sizes of: a) 1, b) 2 and c) 4.

4 Results and Discussion

The reported results consider the Budapest Tech case study presented in [15]. This case study considers a microgrid with one wind generator (G_W), one solar generator (G_S), one fuel cell generator (G_{FC}), and one battery (B). Energy costs are considered constants for simplicity. Such costs and power limits are presented in Table 1.

Table 1: Costs and limits of energy generation [15].

Costs (W/h)	production limits (W)
$C_{G_W} = 0.4$	$cap_{G_W} = 400$
$C_{G_S} = 0.4$	$cap_{G_S} = 150$
$C_{G_{FC}} = 0.9$	$cap_{G_{FC}} = 80$
$C_{BC} = 0.4$	$\Delta cap_{B^+} = [0 - 200]$
$C_{BD} = 0.6$	$\Delta cap_{B^-} = [0 - 50]$
$C_{E^-} = 1.5$	$cap_B = 200$
$C_{E^+} = 1.5$	$CH_{t_0} = 100$

The optimization of on/off cycles was done considering a period of 24 hours (i.e., $T = 24$) in intervals of 1 hour. Figure 2 presents the forecast of wind energy

(E_W), solar power (E_S), and the joint *Load* for two different scenarios (Scenario 2 presents a very high load compared to Scenario 1) at each time t used as an input of the optimization task (all values given in Wh). Also, we consider that the battery starts charged with $CH_{t_0} = 100$ W.

We present the results using DCOP algorithms for the case study with the new proposed model. The optimization procedure was done for different window sizes as explained in Sect. 3. We present results for each DCOP algorithm (i.e., DPOP, MB-DPOP, AFB, and SBB) in terms of percent error (regarding the optimal solution) and number of exchanged messages⁷. The execution times were in the order of 70 mS with a standard deviation of 10 mS. We used a PC with Processor Intel(R) Core(TM) i7-4770 @ 3.40GHz and 64GB of RAM.

The first thing to point out is the dependency of the search space and the window size considering variables and domains of the case study. Figure 4 presents a comparison against the model used in [11, 10]. It can be noticed that the number of variables (Fig. 4a) grows linearly, while the number of relations and the domain size present an exponential behavior (Figs. 4b and 4c). The new model reduces the impact of such exponential growth, making it more suitable for agent-based optimization and allowing the exploration of large window sizes for optimization.

Table 2: Percentage error and number of exchanged messages for scenario 1.

Percentage error				
Window size	DPOP	MB-DPOP	AFB	SBB
size 1	0.42	0.42	0.42	0.42
size 2	0.42	0.42	0.42	0.42
size 3	0	0	-	-
size 4	0.42		-	-
Messages				
Window size	DPOP	MB-DPOP	AFB	SBB
size 1	192	216	9999	8282
size 2	168	192	8534589	23468014
size 3	160	184	-	-
size 4	156	180	-	-

Tables 2 and 3 show the percent error (i.e., the percent error between the DCOP approach and the optimal solution found with a centralized approach as in [11]) for both scenarios. The experiments confirm that all the algorithms present an exponential growth in the time required to find a solution when the window size increases. For that reason, we limited our analysis to window sizes of 4 since above that window sizes all the algorithms required more than 24 hours to find a solution (in fact, AFB and SBB were not able to find a solution with

⁷ We used the available implementations of DCOP algorithms by FRODO2 and Ja-CoP. Both from <http://frodo2.sourceforge.net> and <http://www.jacop.eu> respectively

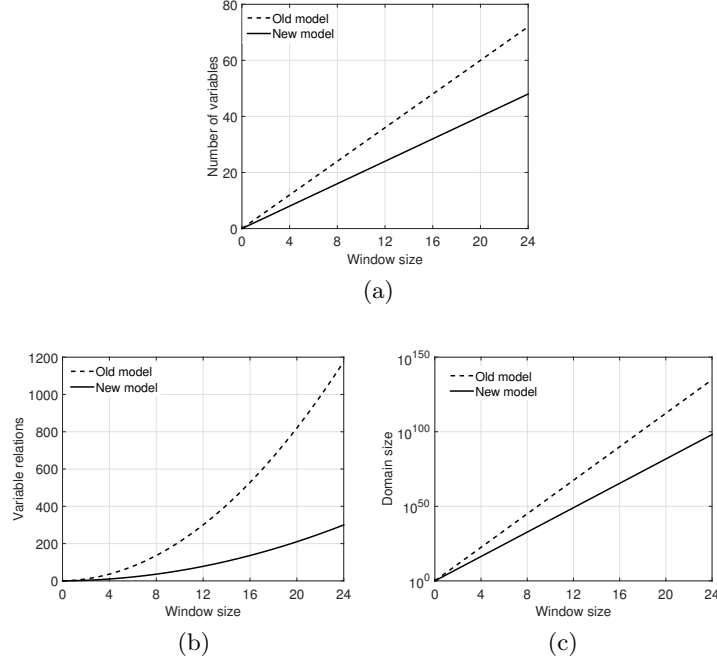


Fig. 4: Comparison of models. Window size and increase of: a) Variables, b) Relations, and c) Domains.

window sizes greater than 2 for both scenarios shown with a “-” mark in Tables 2 and 3).

Notice that in both scenarios the percent error improves along with the window size to the point of reaching the optimal value in scenario 1 with a window size of 3. However, a window size of 4 does not present an improvement of the percent error. In fact, the percent error increases for both scenarios. This behavior is explained in Fig. 5, in which we present the configurations found with window sizes of 3 and 4 for specific time slots (i.e., from time slot 17 to 24 for scenario E1 (Figs. 5a and 5b) and time slot 13 to 20 for scenario E2 (Figs. 5c and 5d)). For scenario 1, the battery status at time slot 17 is 200 W for both window sizes (i.e., 3 and 4). This means that in both horizons the battery can be discharged 200 W. Both horizons makes the same use of the battery in time slots 17 to 19 (using 120 W of the battery, letting the battery status in 80 W). However, since the horizon of window size 4 does not care about the next four time slots, it decides to use the battery in time slot 20 minimizing its cost for that time slot (a cost of 105 compared to 109.5 using window size 3) but leaving a battery status of 35 W for the next horizon of optimization. As consequence, the battery status is not sufficient to satisfy the demand of time slot 21 (which is the higher demand due to a high load in that specific time slot), given as a

Table 3: Percentage error and number of exchanged messages for scenario 2.

Percentage error				
Window	DPOP	MB-DPOP	AFB	SBB
size 1	3.1	3.1	3.1	3.1
size 2	2.9	2.9	3.1	2.9
size 3	2.2	2.2	-	-
size 4	2.9	2.9	-	-

Messages				
Window	DPOP	MB-DPOP	AFB	SBB
size 1	192	216	13269	10550
size 2	168	192	31647420	65315732
size 3	160	184	-	-
size 4	156	180	-	-

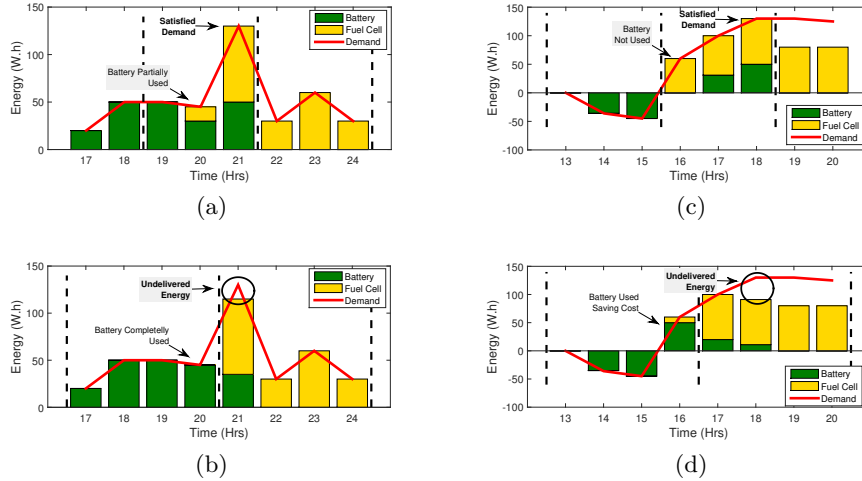


Fig. 5: Explaining the increase in percent error with a larger window size. a) Dispatch with window size 3 for E1. b) Dispatch with window size 3 for E2. c) Dispatch with window size 4 for E1. d) Dispatch with window size 4 for E2.

result a higher cost due to undelivered energy (remember that nondelivery energy has the greatest cost). That situation does not happen for a window size of 3, in which the battery status for time slots 20-21 is 80 W, and most of that energy can be used in time slot 21 leading to a complete satisfaction of the demand (time slot 21 of Fig. 5a). The same situation is presented in scenario 2 for window size 4, in which the use of the battery in time slot 16 (and the saves for that time slot) leads to a higher use of the fuel cell unit in time slot 17 and insufficient energy in time slot 18 giving as a result a higher cost overall.

5 Conclusions

In this paper, we proposed an agent-based method for decentralized microgrid control. By doing that, the problem of scheduling on/off cycles can be treated as a DCOP, and different multi-agent algorithms can be used to find the solution. The results open research directions for the use of multi-agent systems in the long-term optimization of microgrids. For instance, it is clear the necessity of faster procedures of search, allowing agents to handle large window sizes and improving the quality of the solutions. Also, it has been shown that the quality of the solution when time horizons are used for the optimization procedure does not depend completely on the size, but in an adequate size depending on the scenario. Our method, with respect to approaches based on decentralised reinforcement learning, directly encode system-wide hard constraints (such as for example the Kirchhoff law) which are not easy to represent in a factored representation of the problem. Nonetheless, we believe that using decentralised learning approaches for scheduling on/off cycles of devices in the microgrid deserves further investigation, and is definitely an interesting direction for future work. On the other hand, heuristics and pre-processing strategies can be applied to reduce the configuration search space significantly, thus reducing the run time of the approach in the average case and open another interesting venue for future work.

References

1. Amato, C., Oliehoek, F.A., et al.: Scalable planning and learning for multiagent pomdps. In: AAAI. pp. 1995–2002 (2015)
2. Awasthi, S.K., Vij, S.R., Mukhopadhyay, D., Agrawal, A.J.: Multi-strategy based automated negotiation: Bgp based architecture. In: International Conference on Computing, Communication and Automation (ICCCA). pp. 588–593. IEEE (2016)
3. Chaouachi, A., Kamel, R.M., Andoulsi, R., Nagasaka, K.: Multiobjective intelligent energy management for a microgrid. *IEEE Transactions on Industrial Electronics* **60**(4), 1688–1699 (2013)
4. Fioretto, F., Pontelli, E., Yeoh, W.: Distributed constraint optimization problems and applications: A survey. *Journal of Artificial Intelligence Research* **61**, 623–698 (2018)
5. Fioretto, F., Yeoh, W., Pontelli, E.: A multiagent system approach to scheduling devices in smart homes. In: Conference on Autonomous Agents and MultiAgent Systems. pp. 981–989 (2017)
6. Fioretto, F., Yeoh, W., Pontelli, E., Ma, Y., Ranade, S.J.: A distributed constraint optimization (dcop) approach to the economic dispatch with demand response. In: Conference on Autonomous Agents and MultiAgent Systems. pp. 999–1007 (2017)
7. Kofinas, P., Dounis, A., Vouros, G.: Fuzzy q-learning for multi-agent decentralized energy management in microgrids. *Applied Energy* **219**, 53–67 (2018)
8. Kok, J.R., Vlassis, N.: Collaborative multiagent reinforcement learning by payoff propagation. *Journal of Machine Learning Research* **7**(Sep), 1789–1828 (2006)
9. Leite, A.R., Enembreck, F., Barthes, J.P.A.: Distributed constraint optimization problems: Review and perspectives. *Expert Systems with Applications* **41**(11), 5139–5157 (2014)

10. Lezama, F., Palominos, J., Rodríguez-González, A.Y., Farinelli, A., de Cote, E.M.: Agent-based microgrid scheduling: An ict perspective. *Mobile Networks and Applications* pp. 1–17 (2017)
11. Lezama, F., Palominos, J., Rodríguez-González, A.Y., Farinelli, A., de Cote, E.M.: Optimal scheduling of on/off cycles: A decentralized iot-microgrid approach. In: *Applications for Future Internet*, pp. 79–90. Springer (2017)
12. Lezama, F., Soares, J., Hernandez-Leal, P., Kaisers, M., Pinto, T., Vale, Z.: Local energy markets: Paving the path towards fully transactive energy systems. *IEEE Transactions on Power Systems* (2018)
13. Logenthiran, T., Srinivasan, D., Khambadkone, A.M.: Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system. *Electric Power Systems Research* **81**(1), 138–148 (2011)
14. Miller, S., Ramchurn, S.D., Rogers, A.: Optimal decentralised dispatch of embedded generation in the smart grid. In: *International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. pp. 281–288 (2012)
15. Morais, H., Kádár, P., Faria, P., Vale, Z.A., Khodr, H.: Optimal scheduling of a renewable micro-grid in an isolated load area using mixed-integer linear programming. *Renewable Energy* **35**(1), 151–156 (2010)
16. Nguyen, D.T., Yeoh, W., Lau, H.C., Zilberstein, S., Zhang, C.: Decentralized multi-agent reinforcement learning in average-reward dynamic DCOPs. In: *AAAI Conference on Artificial Intelligence*. pp. 1341–1342 (2014)
17. Petcu, A., Faltings, B.: Superstabilizing, fault-containing distributed combinatorial optimization. In: *Proceedings of the National Conference on Artificial Intelligence*. p. 449 (2005)
18. Pourpeighambar, B., Dehghan, M., Sabaei, M.: Non-cooperative reinforcement learning based routing in cognitive radio networks. *Computer Communications* **106**, 11–23 (2017)
19. Soares, J., Pinto, T., Lezama, F., Morais, H.: Survey on complex optimization and simulation for the new power systems paradigm. *Complexity* pp. 1–32 (aug 2018)
20. Su, W., Wang, J., Roh, J.: Stochastic energy scheduling in microgrids with intermittent renewable energy resources. *IEEE Transactions on Smart Grid* **5**(4), 1876–1883 (2014)
21. Teacy, W.L., Chalkiadakis, G., Farinelli, A., Rogers, A., Jennings, N.R., McClean, S., Parr, G.: Decentralized bayesian reinforcement learning for online agent collaboration. In: *International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. pp. 417–424 (2012)
22. Vega, A., Santamaria, F., Rivas, E.: Modeling for home electric energy management: a review. *Renewable and Sustainable Energy Reviews* **52**, 948–959 (2015)