



# Optimal, dynamic and reliable demand-response via OpenADR-compliant multi-agent virtual nodes: Design, implementation & evaluation

Christos Patsonakis<sup>\*</sup>, Angelina D. Bintoudi, Konstantinos Kostopoulos, Ioannis Koskinas, Apostolos C. Tsolakis, Dimosthenis Ioannidis, Dimitrios Tzovaras

Information Technologies Institute, Centre for Research and Technology - Hellas, Thessaloniki, Greece

## ARTICLE INFO

Handling editor: Cecilia Maria Villas Bóas de Almeida

### Keywords:

Demand response  
OpenADR  
Virtual power plant  
Optimisation  
Multi-agent  
Reliability

## ABSTRACT

Extracting and exploiting the flexibility of electric demand has been shown to reduce the needs of network upgrades and generation capacity increases. Demand Response (DR) is considered as one of the few available solutions for accessing the untapped energy potential of small and medium customers. Over the past decade, rigorous research has produced significant results in optimally dispatching DR in an attempt to maximize flexibility extraction. However, the vast majority of works assumes a “happy path” scenario in which DR requests are always successfully completed. Hence, there is a large gap in the literature that fails to account for non-deterministic factors that manifest in practical deployments, e.g., the stochasticity of end-user behavior that can drastically influence the DR’s outcomes. Investing on that notion, a novel, distributed, multi-agent system (MAS) that aggregates consumers and prosumers and handles automatically OpenADR-compliant DR requests is introduced, following virtual power plant (VPP) principles. Agents of the proposed MAS are able to service DR events originating from a higher level, e.g., Aggregators or Utilities, and optimally dispatch them to their assigned customers. The proposed framework ensures 100% DR success rate, compared to conventional methods, by not only optimally exploiting aggregated flexibility through a combination of clustering and optimisation engines, but also through a dynamic, bi-directional DR matchmaking process that can mitigate observed deviations both internally (intra), as well as, externally (inter) in real-time. Via experimentation, we demonstrate the framework’s efficiency in ensuring technical DR fault-tolerance along with its ability to deliver savings of up to 3 orders of magnitude to Aggregators and the customers serving the DR requests.

## 1. Introduction

Demand Response (DR) programs provide a pivotal mechanism to energy grid operators in mitigating energy shortage or excess, thus, increasing the overall reliability and stability of energy grids (Directive (eu) 2019/944 o, 2019). The ever-increasing penetration of Renewable Energy Sources (RESs) (Share of energy from rene, 2020), Energy Storage Systems (ESSs) (Dusonchet et al., 2019) and Electric Vehicles (EVs) (New passenger car registr, 2019), gives rise to new opportunities for increasing profits of key energy stakeholders. These developments led to the emergence of new business models that revolve around, e.g., aggregation and virtual power plants (VPPs) (Tracking energy integrati, 2019). However, these advances introduce additional challenges for grid operators regarding Demand-Side Management (DSM). As shown in recent work (McPherson and Cowiestoll, 1016), DR program

applications lead to noteworthy production cost reductions, which is essentially the result from the substitution of thermal generation with renewable energy sources (RES). Correlations between the impact of DR programs and CO<sub>2</sub> emissions demonstrate that, depending on the nature of each DR, direct (via peak load shaving/shifting and provision of ancillary services) and indirect (via increase in RES penetration and fuel usage decrease in thermal plants) reductions of CO<sub>2</sub> emissions are possible (Violette and Shober, 2014). In general, the goal of increased clean energy supply implies the necessity for the provision of load flexibility in terms of minutes, hours and days-ahead (Hale et al., 2018). To that end, DR programs can provide the much needed flexibility without the instalment of expensive equipment. However, there are significant challenges in unlocking the hidden potential of demand-side flexibility and being able to deliver it in a reliable, fault-tolerant manner.

In most cases around Europe where DR programs are available,

<sup>\*</sup> Corresponding author.

E-mail addresses: [cpatsonakis@iti.gr](mailto:cpatsonakis@iti.gr) (C. Patsonakis), [bintoudi@iti.gr](mailto:bintoudi@iti.gr) (A.D. Bintoudi), [kostopoulos@iti.gr](mailto:kostopoulos@iti.gr) (K. Kostopoulos), [jkosk@iti.gr](mailto:jkosk@iti.gr) (I. Koskinas), [tsolakis@iti.gr](mailto:tsolakis@iti.gr) (A.C. Tsolakis), [djoannid@iti.gr](mailto:djoannid@iti.gr) (D. Ioannidis), [Dimitrios.Tzovaras@iti.gr](mailto:Dimitrios.Tzovaras@iti.gr) (D. Tzovaras).

<https://doi.org/10.1016/j.jclepro.2021.127844>

Received 23 February 2021; Received in revised form 24 May 2021; Accepted 2 June 2021

Available online 24 June 2021

0959-6526/© 2021 Elsevier Ltd. All rights reserved.

participation is restricted to industrial or large tertiary customers. This is attributed to the lack of an information and communication technology (ICT) infrastructure capable of optimally handling small/medium residential and tertiary customers at large scales. However, the latter have been identified to possess a significant amount of flexibility which, currently, remains unexploited (Sioshansi, 2019). The potential of this untapped flexibility is further increased considering the introduction of prosumers, i.e., consumers that also produce energy. For instance, just by 2016, more than 33 GW of residential solar photovoltaics (PVs) had been installed in EU Member States, out of which 53% is exported to the grid (Study on residential pros, 2017). This is expected to rise even more due to the EU's optimistic targets that aim towards a 32% RES share by 2030 (Directive (eu) 2018/2001, 2018).

Nevertheless, great benefits come with equivalent or even greater challenges. First, the drastic increase in the number of parties participating in DR programs requires an ICT infrastructure that is horizontally scalable by design. Put simply, shifting to decentralized/localized monitoring and control strategies is required to handle this unprecedented volume of, e.g., smart meters and DERs. Second, the higher level of uncertainty, which stems from the stochasticity of smaller scale customer behavior, effectively increases the risk of reliably achieving the target(s) of DR events involving them. Finally, significant technical disparities regarding, e.g., communication protocols among DERs and various energy stakeholders exist in practice. This interoperability issue has been identified as one of the most critical roadblocks in leveraging the benefits of DR programs in energy markets (Shafie-khah et al., 2019).

This work presents the design, implementation and experimental evaluation of a novel MAS-based ICT platform that addresses all of the aforementioned roadblocks, and others, that hinder the involvement of small and medium customers in current energy markets and DR programs. The proposed MAS architecture is based on the introduction of a virtual node-based layer, where each agent is referred to as Distributed Virtual Node (DVN). DVNs are interconnected over a distributed, fault-tolerant peer-to-peer (P2P) communication network that employs industry-standard security and privacy techniques that are suitable for real-world deployments. DVNs are able to handle failures and/or deviations of their assigned customer's behavior even *during* the active period of dispatched DR events by employing a twofold mechanism. First, via an "Intra-Node Matchmaking" (IntraNM) algorithm that utilizes other available customers that the agent manages to make up for observed deviations. Second, in cases where the IntraNM approach fails, the agent employs an "Inter-Node Matchmaking" (InterNM) algorithm which, in short, involves other fellow agents to make up for the incurred deviation. This twofold approach is a first step in mitigating the risk of integrating small and medium customers in energy markets and, ultimately, delivering the requested amount of power/energy. To the author's knowledge, no prior work proposes mechanisms that are able to address such deviations. Through meticulous experimentation, it is demonstrated that the proposed framework exhibits a lucrative combination of novel attributes. First, compared to prior works, the proposed MAS ensures a 100% success rate even in the presence of multiple failures during the DR's active period, the importance of which cannot be stressed enough, considering the necessities that give birth to DR requests. Second, the suite of optimisation algorithms presented is capable of delivering savings of up to 3 orders of magnitude for customers servicing a DR request. Third, the proposed MAS facilitates interoperable participation of key energy stakeholders in energy markets and the establishment of stable revenue streams. Lastly, of independent research interest is an additional, indirect, significant outcome of the proposed matchmaking algorithms, which promote customer activity, awareness and responsiveness to DR requests, through the employment of a DR reliability index.

The manuscript is organized as follows: Section 2 presents previous research endeavours related to the submitted work, highlighting limitations and challenges identified. In Section 3, the overall system

architecture is presented, followed by the detailed implementation for the aspects explored. The experimental setup and the simulation results are elaborately documented and discussed in Section 4. Finally, in Section 5, we provide concluding remarks.

## 2. Related work

Multi-agent systems (MASs) provide several attractive properties in smart grid contexts, such as fault-tolerance, increased efficiency and reliability in the management of the underlying grid, and others (Mahela et al., Siano). Several prior works have employed MAS-based architectures to enable communications among energy stakeholders, partially distribute computational requirements for decision making processes to provide for increased adoption of DR programs (e.g., (González-Briones et al., 2018)). More recent works (e.g., (Shawon et al., 2019)) showcase that the decentralized, flexible, robust and autonomous nature of MASs positions them as ideal architectures for handling diverse sets of energy transactions and the introduction of new market roles, such as Aggregators (Woltmann et al., 2020) and VPPs (Pasetti et al., 2018).

Initially, MAS deployments in the context of DR focused heavily on market-based schemes, i.e., the ability of agents to effectively interact with marketplaces (Praça et al., 2003) without, however, measuring the amount of energy/power delivered. More recently, the authors of (Golmohamadi et al., 2019) proposed a MAS architecture that bundles industrial and smaller-scale customers into market-specific agents. This approach, whilst inclusive of various customer preferences, relies on centralized optimisation and decision making processes. However, such approaches have been proven to be sub-optimal. Karfopoulos et al. (2015) have demonstrated that highly dynamic decision making processes are required to handle, among others, the volatility of RESs and the stochasticity of end-user behavior. In (Wang et al., 2018), the authors present a MAS that is limited in handling a narrow set of devices and end-users and was only evaluated for price-based DR events. Although results are promising, the proposed MAS is unable to handle the aforementioned uncertainties.

From the tertiary perspective, the authors of (Gomes et al., 2019), to enable participation of small and medium players's in DSM programs, introduce an agent-based architecture that optimises building automation and balances demand and supply towards reducing the energy bought from the grid. Even though their MAS showcases quite interesting results, it also revealed some limitations that can be considered critical if they would take place during a real-world DR event (e.g., heavy passage of clouds), thus highlighting the need for fail-safe mechanisms in real-time operation.

While there are numerous prior works providing numerical results demonstrating the benefits of MAS-based approaches, very few tackle the issue of quantifying DR reliability. Current proposals are limited to statically assigning a reliability metric (e.g., (Silva et al., 2020)) to end-users. Such approaches are unable to capture the real-time behavior of parties participating in DR events, whose targets need to be appropriately adjusted to increase the probability of delivering the expected/requested outcome.

Muthirayan et al. (2019) propose a self-reported baseline mechanism as a means to eliminate the incentive of agents inflating their baselines to receive increased payments. While this work provides a useful framework for calculating agent reliability, it does not address failures regarding actual energy delivery and ultimately resorts in penalization. More recent works highlight the importance of accurate consumption, generation, flexibility and price forecasts in increasing the reliability of DR events (Wang et al., 2019).

When examining the success rate of DR events, the overwhelming majority of research focuses on the incentives and the optimal selection of assets and resources for successfully delivering what has been requested (Parrish et al., 2020). On the contrary, there are very few findings on anticipating and identifying participants that are expected to default (Azuma et al., 2019), and none (to the authors knowledge) that

tackle the failure during the active period of the DR event. Furthermore, interestingly enough, scientific community seems more keen to impose penalties (Ghorashi et al., 2020) towards increasing success rates than presenting with mitigation mechanisms that can salvage the risk imposed by a defaulting participant.

Another perspective that aims to increase DR success rate, is the employment of clustering solutions. Such techniques can be provided for the design of effective DR strategies (Lin et al., 2019), the management of Aggregator’s and Distribution Network Operators’ portfolio (Gouveia et al., 2016), as well as, temporal and dynamic adoption of customer segmentation models (Benítez et al., 2014). At the same time, clustering can also serve as a complementary mechanism in alleviating the computational overhead incurred while scaling up. The advent and installation of smart meters to residential houses has expanded the set of raw data sources towards sustainable energy development (Gouveia et al., 2016). In recent years, energy profile clustering has been approached from multiple perspectives and several methodologies have been proposed to decode the customer’s energy behaviour (Motlagh et al., 2019). Nonetheless, there hasn’t been sufficient bench-marking to justify the trade-off of employing clustering techniques even if the DR success rate drops.

The present work, aims to address the aforementioned challenges. By proposing a highly scalable, interoperable, and efficient decentralized MAS architecture, a highly dynamic and rapid decision making process is ensured, easily applicable under any conditions or topologies. In addition, through a range of novel software components, the proposed framework is the first study of a real-time DR-fail-safe system, that can ensure the event’s success even under uncertain conditions, either due to volatile weather conditions, highly stochastic end-user behaviour, or even sub-optimal incentive strategies. Lastly, through thorough experimentation evidence, the use of each of the implemented components is justified, discussing the trade-offs in terms of success rate and computation performance, which is frequently missing from similar findings.

### 3. MAS architecture

Firstly, the fundamental notions based on which the proposed MAS is designed. The starting point is the MAS’s ability to harness the, currently unexploited, flexibility that is readily available at smaller scales. Consequently, it immediately follows that the MAS should be horizontally and transparently scalable to handle an arbitrary amount of end-users that will provide this flexibility. To promote the large scale uptake of DR programs, there are additional properties that the MAS must provide. First, it must optimally dispatch the flexibility provided by end-users to, on the one hand, incentivize their participation and, on the other hand, promote clean energy principles. Second, it is imperative that the system provides for increased fault-tolerance during the active period of DR events to address end-user uncertainties, as was discussed previously. Lastly, OpenADR was chosen as the standard for encoding and communicating, e.g., DR events, energy-related reports and availability schedules, due to its widespread adoption from both industry and academia.

In Fig. 1, a high-level depiction of the system’s architecture is presented, a descriptive overview of which is as follows. As illustrated, the MAS is composed by a set of independent virtual agents, to which we interchangeably refer to as DVNs. One of the focal points of each DVN is the “Monitoring and Profiling” component, whose functionalities are as follows. First, this is the entry point for new customers. The registration of a new customer is accompanied with data pertaining to her consumption, generation and storage capacities (if any), geolocation, the market contexts that she is willing to participate, and others. Second, this component is responsible for collecting and maintaining data pertaining to historical and forecasted consumption, generation, power flow and storage of customers. These are used as the basis for building a dynamic, per-customer profile that encompasses metrics pertaining to the accuracy of the reported forecasts and others that capture the reliability of the customer in the context of each individual market. The reliability metrics are output/updated by an in-house developed machine learning model that incorporates marketplace-specific features.

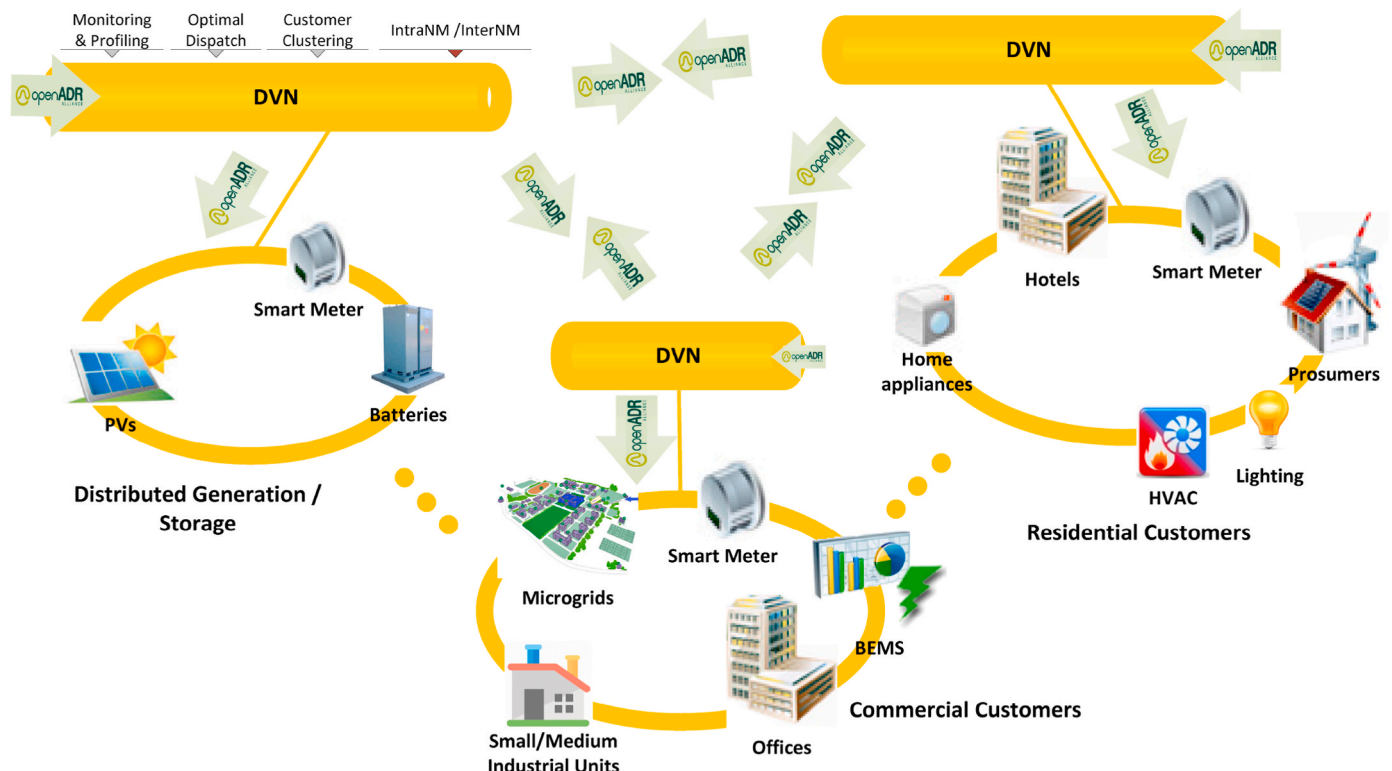


Fig. 1. High-level architectural view of the proposed MAS.

Lastly, based on the measurements and forecasts reported by its assigned customers, this component exposes aggregated historical and forecasted data.

Agents input customer measurements and profiles to their “Customer Clustering” component (Section 3.1) to cluster their assigned customers based on various features, such as geolocation, type (consumer/prosumer) and their capacities. For each cluster, the agent produces a profile, which is directly involved in the decision making process for servicing input DR events. As the system is by nature dynamic, i.e., measurements, forecasts, customer behavior and reliability in the context of DR events may change over time, customers may be reassigned among clusters. The “Optimal Dispatch” component (Section 3.2) is the focal decision point for servicing input DR events. This is a lightweight optimisation engine for consumers, prosumers and various DERs that are integrated into a unified optimisation problem and can cover multiple market contexts. Its objective function revolves around the minimization of the energy cost by taking into account parameters, such as the retail real-time price (RTP), customer cluster profiles and the virtual cost of their flexibility. Put simply, this component, on input an OpenADR request, outputs a set of optimal OpenADR requests, which are dispatched to a set of selected customers.

Communication amongst agents and their individually assigned customers is facilitated via a highly-scalable and end-to-end secure peer-to-peer network (Section 3.5). To tolerate failures and/or deviations that occur during the active period of DR events, agents employ a twofold approach to make up for incurred losses. This accomplished by the introduction of two novel algorithms, i.e., the “Intra-Node Matchmaking” (IntraNM) and “Inter-Node Matchmaking” (InterNM), which are described in Sections 3.4 and 3.3, respectively. The following subsections are dedicated in elaborating on the detailed description of the aforementioned components, as well as, the implementation details of the proposed MAS.

### 3.1. Customer clustering

The main idea of this component is to employ customer flexibility profiles as a means to convey to the “Optimal Dispatch” component the potential of a customer set to (partially) service a DR request. Additionally, this process reduces the optimisation’s computational overhead since it is input collective data of each cluster profile, instead of individual statistics.

The proposed approach in the “Customers Clustering” implementation is applied through the exploitation of flexibility measurement of each customer, which is provided from local intelligent units (George et al., 2020). This measurement can be further divided in positive and negative metrics. Positive flexibility is defined as the feasible positive power flow deviation compared to the baseline measurement at a specific time period. Similarly, negative flexibility reflects an opposing feasible negative deviation from the baseline. As illustrated in Fig. 2,

$\Delta FlexUp$  and  $\Delta FlexDown$  represent these two flexibility modes for customers that have no generation capabilities. Based on these metrics, a reliability metric is derived that provides an evaluation criterion regarding a customer’s contribution in historical DR events. This facilitates the correction of the initial measured flexibility and is expressed via the following equation:

$$cF = mF \cdot Rel \quad (1)$$

where,  $cF$  is the corrected flexibility,  $mF$  is the measured flexibility and  $Rel \in [0, 1]$  is the reliability metric.

Depending on the nature of the DR request, different customer clustering profiles will be employed by the DVN. For instance, a DR request for energy consumption reduction will harness extracted information from the *corrected downwards flexibility*, whereas a DR request for energy consumption increase utilizes the *corrected upwards flexibility*. In order to apply temporal clustering for each customer that exhibits peculiar energy behaviour alterations, the calculated baseline corrected flexibility is segmented in 24 hourly periods that are independently examined by the clustering algorithm.

Data pre-processing is the primary step of energy data analysis and consists of a sequence of individual tasks, i.e., outlier isolation, data standardization and data transformation to the frequency domain. As far as the outliers removal step is concerned, daily measurements from a specific customer that deviate from its baseline behaviour are identified and removed through a percentile analysis. More specifically, the euclidean distance of each daily time series measurements are estimated with regards to the baseline load (mean value). Next, the Signal-to-Noise-Ratio (SNR) indicator validates the fact that the power signal has been increased proportionally with the noise as an outcome of the outlier’s day removal for an initial percentile. In cases of increased noise interference, the selected percentile is configured appropriately focused on empowering the SNR indicator. The mean value of the remainder daily measurements is considered as a customer’s baseline.

The next step involves the standardization of data in a  $(-1, 1)$  range, which are subsequently transformed in the frequency domain. This transformation is achieved by applying the Continuous Wavelet Transform (CWT) method (Grossmann et al., 1989). The wavelet function of this transformation is the negative normalized second derivative of the Gaussian function (Ricker wavelet), which is typically referred to as the Mexican Hat wavelet and is represented by the following equation:

$$\psi(t) = \frac{2}{\sqrt{3}\sigma\pi^{\frac{1}{4}}} \left( 1 - \left(\frac{t}{\sigma}\right)^2 \right) e^{-\frac{t^2}{2\sigma^2}} \quad (2)$$

The fundamental advantage of CWT compared to Fast Fourier Transform (FFT) is its capability to construct time-frequency representations of a signal that exhibits exceptional time and frequency localization. Additionally, the CWT method can efficiently transform non-stationary signals and preserve time-dimension properties.

The Affinity Propagation (AP) algorithm (Zhang and Song, 2011) has

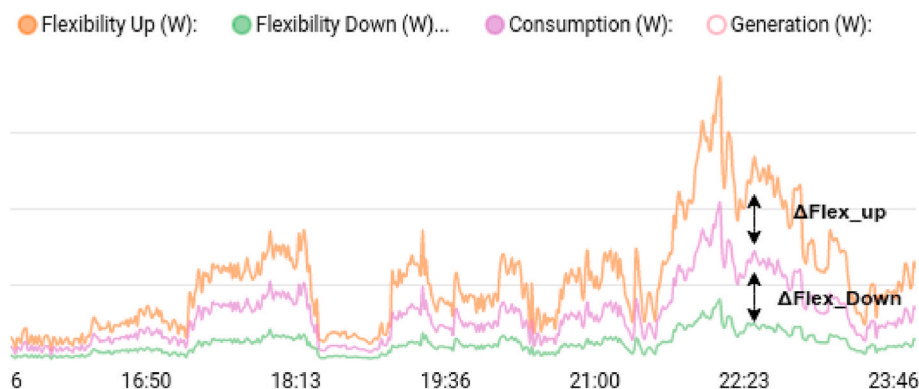


Fig. 2. Graphical representation of upwards and downwards flexibilities.

been selected to retrieve the groups of customers that share common flexibility characteristics during hourly time periods. A similarity matrix that includes the wavelet coherence between the baseline load profile of each customer compared to others is estimated and given as input to the AP algorithm. The proposed implementation calculates kernel similarities in higher dimensions to identify non-linear correlations across  $cF$  measurements that have been processed and transformed in the frequency domain. Moreover, in the proposed approach, a dynamic adaptation of the algorithm's parameters in terms of its efficiency is incorporated. Regarding the generated clusters, the AP algorithm has the property to detect the appropriate number of segmented groups autonomously. A crucial factor that affects this functionality is a "preference" parameter that reflects a point's inclination to consider itself as "exemplar". Therefore, in cases where recent customer behaviors exhibit reduced credibility, the algorithm's parameters affecting the estimation of the number of clusters are re-configured to optimally reallocate customers in different clusters. Lastly, this component outputs various statistics related to each cluster's flexibility profiles, such as mean value, variance and the slope's curve, which are consumed by other DVN components, as described in the following subsections.

### 3.2. Optimal Dispatch

Input DR signals must be broken down to a series of dynamic setpoints to be dispatched to all eligible customers of the DVN. Although this could be achieved with a simple rule-based algorithm, a novel optimisation scheme (OptiDVN) has been developed, which aims towards not only satisfying the incoming DR, but also minimising the respective actual and virtual costs associated with the DR's completion. The OptiDVN engine is a significant expansion of a recently published research work from the authors (Bintoudi et al., 2021). To facilitate reader understanding and provide for a concise description, a case study of an OpenADR load dispatch signal is discussed, where the active period is comprised by a series of time intervals, for each of which, a fixed power flow setpoint (algebraic value in Watts) is requested. The explicit case is assumed in which the DVN can directly control the customer's assets, or by interfacing with an existing Building Management System (BMS).

The basis of the optimisation algorithm is the modelling of each operational customer of the DVN as a unit of flexibility that is comprised by a series of virtual DERs (vDERs), e.g., energy storage systems (ESS, e.g., Li-ion batteries), small controllable photovoltaic units (PV). Given the different levels of abstraction introduced in this work, at the DVN level, each customer appears to have assets in an aggregated manner, meaning one overall load flexibility, one aggregated PV unit, one aggregated ESS unit etc. Implementation-wise, this component is developed in Python 3.7, whilst its core, i.e., the optimisation problem, is formulated using mixed-integer linear programming (MILP). The MILP solver selected is COIN-n-branch (CBC), an efficient and accurate open-source solver implementation for MILP optimisation problems.

The optimisation problem is derived from a variation of the classic Unit Commitment Problem and, therefore, is formulated using as optimisation variables the vDERs' energy time series setpoints, complemented by auxiliary binary variables. The objective function, expressed by Equation (3) aims towards minimising the operational cost of a VPP in the context of an explicit load dispatch DR signal. The objective function is formulated to benefit the DVN's customers and the pricing scheme selected is the real-time retail price. The associated costs for each vDER consists of the Levelised Cost of Electricity (LCoE) of the actual units (i.e., PV, ESS, WT) and a virtual cost for the upward/downward load flexibility, which is calculated according each customer's reliability. LCoEs are calculated according to typical values of investment costs, unit operation and losses.

$$OF : \min \sum_{n=1}^N \sum_{t=1}^T \left[ (E_{flex}^+(n, t) + E_{flex}^-(n, t)) \cdot C_{flex}(n) + (E_{ess}^+(n, t) + E_{ess}^-(n, t)) \cdot C_{ess} + E_{pv}(n, t) \cdot C_{pv} + (E_{pcc}^+(n, t) + E_{pcc}^-(n, t)) \cdot C_{retail}(t) \right] \quad (3)$$

where,

- $N$ , the number of available customers of the DVN, with  $n \in [1, N]$ ,
- $T$ , the DR's total duration, with  $t \in [1, T]$ ,
- $E_{flex}^+(n, t)$ ,  $E_{flex}^-(n, t)$  the energy setpoints corresponding to the  $n$ -th customer's upward (+) and downward (-) flexibility at timeslot  $t$  of the DR,
- $E_{ess}^+(n, t)$ ,  $E_{ess}^-(n, t)$  the energy setpoints for charge (+) and discharge (-) of an aggregated ESS at the  $n$ -th customer's side at timeslot  $t$  of the DR - variables equal to zero if the  $n$ -th customer does not own such assets,
- $E_{pcc}^+(n, t)$ ,  $E_{pcc}^-(n, t)$  the energy setpoints at the point of common coupling (PCC) of the  $n$ -th customer, corresponding to overall imported (+) and exported (-) energy at timeslot  $t$  of the DR,
- $E_{pv}(n, t)$  the energy setpoints for an aggregated PV unit installed at the customer's side at timeslot  $t$  of the DR, - variable equals to zero if the  $n$ -th customer does not own such assets
- $C_{flex}(n)$  a virtual energy cost associated with the  $n$ -th customer's reliability, which is used as an index to the sorted retail dynamic price.
- $C_{ess}$ ,  $C_{pv}$  the LCoE for the aggregated ESS and PV of the  $n$ -th customer's assets, assuming typical equipment installed in residential households.

For convenience, the following representations are adopted ( $\forall n \in [1, N]$ ).

$$E_{flex}(n) = \begin{bmatrix} \bar{E}_{flex}^+(n) & \bar{E}_{flex}^-(n) \end{bmatrix} = \begin{bmatrix} E_{flex}^+(n, 1) & E_{flex}^-(n, 1) \\ \dots & \dots \\ E_{flex}^+(n, t) & E_{flex}^-(n, t) \\ \dots & \dots \\ E_{flex}^+(n, T) & E_{flex}^-(n, T) \end{bmatrix} \quad (4)$$

$$E_{ess}(n) = \begin{bmatrix} \bar{E}_{ess}^+(n) & \bar{E}_{ess}^-(n) \end{bmatrix} = \begin{bmatrix} E_{ess}^+(n, 1) & E_{ess}^-(n, 1) \\ \dots & \dots \\ E_{ess}^+(n, t) & E_{ess}^-(n, t) \\ \dots & \dots \\ E_{ess}^+(n, T) & E_{ess}^-(n, T) \end{bmatrix} \quad (5)$$

$$E_{pcc}(n) = \begin{bmatrix} \bar{E}_{pcc}^+(n) & \bar{E}_{pcc}^-(n) \end{bmatrix} = \begin{bmatrix} E_{pcc}^+(n, 1) & E_{pcc}^-(n, 1) \\ \dots & \dots \\ E_{pcc}^+(n, t) & E_{pcc}^-(n, t) \\ \dots & \dots \\ E_{pcc}^+(n, T) & E_{pcc}^-(n, T) \end{bmatrix} \quad (6)$$

$$E_{pv}(n) \equiv E_{pv}(n) = \begin{bmatrix} E_{pv}(n, 1) \\ \dots \\ E_{pv}(n, t) \\ \dots \\ E_{pv}(n, T) \end{bmatrix} \quad (7)$$

A series of constraints are formulated in order to respect the limitations of each vDER and the DR's satisfaction.

Constraint 1: energy balance (supply/demand) and satisfaction of the DR's setpoint

$$\sum_{n=1}^N \left[ \vec{O}_{flex}^+(n) + \vec{O}_{flex}^-(n) + \vec{O}_{ess}^+(n) + \vec{O}_{ess}^-(n) + \vec{O}_{pcc}^+(n) + \vec{O}_{pcc}^-(n) + \vec{E}_{pv}(n) \right] = \vec{E}_L, \quad \forall n \in [1, N] \quad (8)$$

where,  $\vec{E}_L$  is the non-flexible DVN load and  $\vec{E}_{DR}$ , the DR's setpoint, expressed in energy terms that must met at every time slot (both  $T$ -sized vectors).

Constraint 2: minimum-maximum operation limits  $\forall n \in [1, N]$

$$\begin{aligned} \vec{E}_{flex-min}(n) \odot \vec{O}_{flex}(n) &\leq \vec{E}_{flex}(n) \leq \vec{E}_{flex-max}(n) \odot \vec{O}_{flex}(n) \\ \vec{E}_{ess-min}(n) \odot \vec{O}_{ess} &\leq \vec{E}_{ess}(n) \leq \vec{E}_{ess-max}(n) \odot \vec{O}_{ess}(n) \\ \vec{E}_{pcc-min}(n) \odot \vec{O}_{pcc} &\leq \vec{E}_{pcc}(n) \leq \vec{E}_{pcc-max}(n) \odot \vec{O}_{pcc}(n) \\ \mathbf{0} &\leq \vec{E}_{pv}(n) \leq \vec{E}_{pv-mpp}(n) \end{aligned} \quad (9)$$

where,

- $\vec{E}_{flex-min}(n) = \begin{bmatrix} \vec{0} & \vec{E}_{flex-min}^-(n) \end{bmatrix}$ ,  $\vec{E}_{flex-max}(n) = \begin{bmatrix} \vec{E}_{flex-max}^+(n) & \vec{0} \end{bmatrix}$ ,
- $\vec{E}_{ess-min}(n) = \begin{bmatrix} \vec{0} & \vec{E}_{ess-min}^-(n) \end{bmatrix}$ ,  $\vec{E}_{ess-max}(n) = \begin{bmatrix} \vec{E}_{ess-max}^+(n) & \vec{0} \end{bmatrix}$ ,
- $\vec{E}_{pcc-min}(n) = \begin{bmatrix} \vec{0} & \vec{E}_{pcc-min}^-(n) \end{bmatrix}$ ,  $\vec{E}_{pcc-max}(n) = \begin{bmatrix} \vec{E}_{pcc-max}^+(n) & \vec{0} \end{bmatrix}$ ,
- $\vec{E}_{pv-mpp}(n)$  the maximum-power-point aggregated energy PV production vector of the  $n$ -th customer.

In 9,  $\vec{O}_{flex}(n)$ ,  $\vec{O}_{ess}(n)$ ,  $\vec{O}_{pcc}(n) \in \mathbb{Z}_2$  represent the decision variables modeling the operation of the complementary aggregated units, which are defined as follows:

$$\vec{O}_{flex}(n) = \begin{bmatrix} \vec{O}_{flex}^+(n) & \vec{O}_{flex}^-(n) \\ \dots & \dots \\ \vec{O}_{flex}^+(n, t) & \vec{O}_{flex}^-(n, t) \\ \dots & \dots \\ \vec{O}_{flex}^+(n, T) & \vec{O}_{flex}^-(n, T) \end{bmatrix} \quad (10)$$

$$\vec{O}_{ess}(n) = \begin{bmatrix} \vec{O}_{ess}^+(n, 1) & \vec{O}_{ess}^-(n, 1) \\ \dots & \dots \\ \vec{O}_{ess}^+(n, t) & \vec{O}_{ess}^-(n, t) \\ \dots & \dots \\ \vec{O}_{ess}^+(n, T) & \vec{O}_{ess}^-(n, T) \end{bmatrix} \quad (11)$$

$$\vec{O}_{pcc}(n) = \begin{bmatrix} \vec{O}_{pcc}^+(n, 1) & \vec{O}_{pcc}^-(n, 1) \\ \dots & \dots \\ \vec{O}_{pcc}^+(n, t) & \vec{O}_{pcc}^-(n, t) \\ \dots & \dots \\ \vec{O}_{pcc}^+(n, T) & \vec{O}_{pcc}^-(n, T) \end{bmatrix} \quad (12)$$

$$??? \quad (13)$$

$\vec{O}_{flex}(n)$ ,  $\vec{O}_{ess}(n)$ ,  $\vec{O}_{pcc}(n) \forall n \in [1, N]$  are bound by:

$$\vec{O}_{flex}^+(n) + \vec{O}_{flex}^-(n) = \vec{O}_{ess}^+(n) + \vec{O}_{ess}^-(n) = \vec{O}_{pcc}^+(n) + \vec{O}_{pcc}^-(n) = \vec{1} \quad (14)$$

Finally, a final set of constraints is needed for the case of customers with ESSs installed in order to bound  $\forall n \in [1, N]$ ,  $t \in [1, T]$  the ESS's operation to the limitations deriving from its technical characteristics, i. e., its capacity ( $C(n)$ ), charge and discharge C-rates ( $R_C^+$ ,  $R_C^-$ ) and Depth-of-Discharge ( $DoD$ ):

$$C(n) \cdot SoC_{min}(n) \leq \sum_{\tau=1}^{t-1} [E_{ess}^+(n, \tau) + E_{ess}^-(n, \tau)] + E_{ess}^+(n, t) + E_{ess}^-(n, t) \quad (15)$$

$$C(n) \cdot SoC_{max}(n) \geq \sum_{\tau=1}^{t-1} [E_{ess}^+(n, \tau) + E_{ess}^-(n, \tau)] + E_{ess}^+(n, t) + E_{ess}^-(n, t) \quad (16a)$$

$$0 \leq E_{ess}^+(n, t) \leq R_C^+(n) \cdot C(n) \cdot DoD(n) \cdot r_t \quad (16b)$$

$$0 \geq E_{ess}^-(n, t) \geq R_C^-(n) \cdot C(n) \cdot DoD(n) \cdot r_t \quad (16b)$$

where,  $r_t = 60/m_r$  and  $m_r$  is the measurement sampling resolution (e.g., 1 or 15).

Given the different levels of abstraction introduced in this work,  $R_C^+$ ,  $R_C^-$  take common values found in typical residential ESS installations, namely,  $R_C^+ = 0.75$ ,  $R_C^- = 1$ . Similarly,  $DoD = 0.8$  if the usage of Li-ion batteries is taken into account.

The fact that all optimisation expressions are linear and their variables are either continuous, or binary, is very beneficial because LP/MILP problems are solved much faster than their non-linear equivalents. Moreover, re-running the optimisation algorithm can be achieved quickly, thus, facilitating real-time applications.

### 3.3. Intra-Node Matchmaking

Mitigating failures during the active period of a DR event imposes, the following time-related constraints. First, as OpenADR allows for intervals that lie in the order of seconds, we must employ a procedure that, apart from accurate, is computationally lightweight. Second, there needs to be enough time for the agent to engage in OpenADR's DR issuance protocol with the candidates that will cover the excess/shortage. Given the high variability of consumption and generation forecasts, especially in cases of residential households, failures should be expected frequently. Therefore, the proposed MAS is equipped with two "safety valves" (matchmaking algorithms) to ensure the fulfillment of the optimally dispatched DR events in a timely manner, even when high deviations between forecasts and actual measurements are observed. For this reason, deterministic optimisation problems and simple clustering algorithms are applied. Both matchmaking components are optimisation problems solving two variations of the OptiDVN module. Their differences concern the various levels of abstraction, i.e., what is considered as an "asset" in each case and additional dynamic input time series adjustments.

The Intra-Node Matchmaking (IntraNM) algorithm is the first "safety valve" in cases where one or more customers deviate from the setpoints output by OptiDVN, based on a tunable threshold. IntraNM reallocates the remainders of the setpoints of the deviating customers to unallocated

customers of the DVN. This algorithm is input a time series of setpoints that is computed as the algebraic difference of the OptiDVN's originally output setpoints minus the sum of the customer setpoints that have not deviated from their targets, which is necessary to respect the latter's ongoing operation. To ensure fast response times and to minimize the number of lost timeslots, IntraNM considers only customers that are configured to automatically accept DR requests, i.e., without requiring explicit human consent. To mitigate further deviations, IntraNM adjusts customer time series based on actual collected reports since the beginning of the current day. Next, the error between the forecasted and the actual values is computed. Forecasted and actual values are partitioned into two clusters via k-means clustering. The lower value clusters isolate the forecasted and the actual base load respectively. Their difference represents the current day's deviation from the forecasted values. The remaining forecasts are calibrated based on the calculated deviation. Put simply, the clustering of the collected measurements attempts to capture the customers' behavior from the beginning of the day and modifies accordingly the load forecast in order to adapt to current day conditions. The remainder of the IntraNM algorithm is identical to OptiDVN, i.e., it involves the same objective function (Equation (3)) and constraints (Equations ((4) and (8)–(10) and (14) and (15)).

### 3.4. Inter-Node Matchmaking

In cases where the IntraNM fails to compensate for the observed deviations, the DVN activates its second "safety valve", i.e., Inter-Node Matchmaking (InterNM). Conceptually, this algorithm attempts to rectify the situation by engaging with other DVNs of the MAS which, from the originating DVN's point of view, are modeled as "customers". In order for a DVN to be able to assess which other agents can be considered as candidates for selection, it requires data pertaining to their individual aggregated forecasted flexibility. However, communicating this information during the active period of a DR event would impose additional time constraints. We eliminate this by having the agent preemptively query its other fellow agents prior to the start of the DR event's active period. To evaluate to which other agents DR requests will be dispatched, we follow a similar strategy as before. The optimization problem follows the same principles as OptiDVN with the difference that all variables and constraints discussed in Section 3.2 refer to other DVNs, as already noted. Each DVN maintains a local perception of the reliability of other fellow agents in the context of DR events. This metric is updated in a similar fashion as that of DVN customers. Assuming the InterNM algorithm finds a feasible solution to handle the excess/shortage, it dispatches the DR requests to the selected fellow agents. From hereon in, the selected agents follow the same flow that we have described up to this point.

### 3.5. Agent peer-to-peer (P2P) network

Open Automated Demand Response (OpenADR (Alliance, 2013)) is an open, interoperable communication standard that facilitates smart grid information exchange among various energy stakeholders and end-users. To provide for interoperability, the presented MAS architecture builds upon OpenADR to allow the exchange of information regarding a plethora of report types, forecasts, availability schedules, market prices, DR events and others. In addition to OpenADR's expressive data model, one of its main benefits stem from its inherent support for hierarchical deployment architectures of arbitrary complexity that can be scaled in real-time. The advantages of OpenADR have, on the one hand, led to its adoption by relevant standardization bodies, e.g., the International Electrotechnical Commission (IEC) and, on the other hand, its embrace from both industry and research institutions to provide, among others, DER management, ancillary and a wide variety of demand response services.

The inherent distributed deployment of smart meters, IoT devices and DERs brought on by the integration of small/medium customers to

DR programs necessitates the employment of advanced ICT to allow for, e.g., monitoring, control and reports. Put simply, regardless of the data model, a fundamental requirement of any MAS that aims for practical, real-world deployments, is the establishment of a distributed and scalable communications network that can provide for service liveness, even in the presence of failures. Furthermore, OpenADR's compliance rules impose a set of security requirements that agents of the MAS need to abide by. Moreover, it is necessary to acknowledge additional security constraints that, e.g., Aggregators and utilities are subject to, e.g., data privacy.

To address these requirements, the introduced MAS, from a technical standpoint, employs OpenFire (Realttime, 2020) as its main communication broker, which allows for both HTTP and XMPP communications (the latter is an OpenADR requirement). OpenFire's built-in security features, e.g., support for Transport Layer Security (TLS (Dierks and RescorlaRFC, 2008)) and X.509 digital certificates (Housley et al., 1999), provide the necessary security and privacy features for practical deployments (Gelenbe et al., 2013). To provide for scalability, decentralization and fault-tolerance, a cluster of OpenFire servers is deployed that can be, in real-time, scaled accordingly. From the viewpoint of participating customers, the cluster acts as a unified, virtual communication broker that, transparently, load balances communications and message routing.

### 3.6. Agent back-end

Fig. 3 presents the physical diagram of a DVN, which is comprised by: 1) an HTTP server that exposes the agent's interfaces to the P2P network, 2) the agent's back-end, which "glues" together its sub-components and, 3) a SQL-based, relational database, which stores all the data that are relevant to the agent's operation. We integrate parts (1) and (2) via a standard Web Server Gateway Interface (WSGI (Eby, 2010)) approach and parts (2) and (3) by employing the SQLAlchemy (Bayer et al., 2012) toolkit.

The agent's back-end is a multi-threaded, asynchronous, event-based application developed in Python (Van Rossum and Drake, 2009). The agent encompasses timer threads that execute on predefined time intervals. The "Model Training" thread is responsible for training the agent's machine learning models. The "Customer Clustering" thread updates the profiles of the agent's customer clusters.

The agent also encompasses threads whose execution is triggered by external inputs, e.g., the "Optimal Dispatch" thread is activated when the agent receives a DR request. Furthermore, the execution of events may lead to the internal scheduling of other events, which is typical in, e.g., the context of the IntraNM and InterNM algorithms. To reliably synchronize the concurrent execution of all these events, we implemented, from scratch, an asynchronous and persistent scheduler for time-based events. This scheduler is comprised by a master thread that assigns the handling of events to a pool of worker threads. This approach allows events to be executed in parallel and provides for increased throughput.

Finally, a high-level description of a DVN's decision making process to optimally manage an input DR request originating from, e.g., an aggregator, as well as, its monitoring process, which takes place during the event's active period, is provided (Fig. 4). On input an OpenADR event, a DVN first examines whether it's capable of servicing it or not. Assuming that is the case, the DVN identifies clusters of customers that can service the DR's targets, based on the nature of the request and, subsequently, proceeds on invoking OptiDVN. If OptiDVN is able to provide a feasible solution, a list of OpenADR sub-DR requests is constructed and dispatched to the selected customers. Depending on the nature of the DR event, the DVN may wait for explicit confirmation from the selected customers. Following the receipt of approvals from all selected customers, the DVN communicates to the aggregator its ability to service the input DR request. At the beginning of a DR event's active period, the DVN begins to monitor both the power flow from the

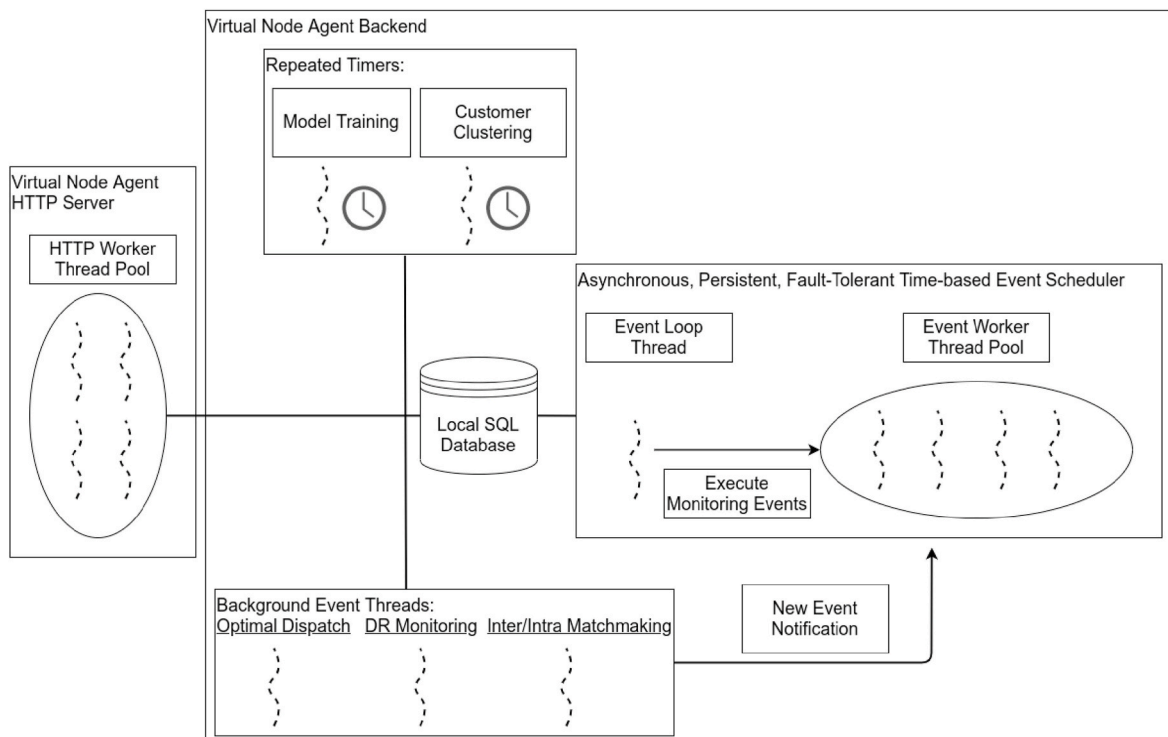


Fig. 3. Physical (implementation) diagram of a DVN.

engaged customers with the sub-DR requests, as well as the aggregated power flow of the entire DVN towards the Aggregator. This dual monitoring aims towards ensuring that the expected behaviour will be delivered, since there is always a possibility that other customers, not engaged, will alter their behaviour, hence altering the aggregated power flow and leading to the DR failure even through the sub-DR requests are still on track. If a deviation in either power flows is identified, an ancillary process is triggered. Upon a second consecutive deviation, the DVN executes the matchmaking process to mitigate the issue that has occurred. First within the DVN with IntraNM, and then with other DVNs through InterNM. If a new feasible solution is identified by either IntraNM or InterNM, a new set of sub-DR requests is issued, whereas the sub-DR to the customer that did not deliver the agreed targets is canceled. The monitoring process, along with the ancillary ones, continues until all DR sub-signals are completed.

#### 4. Experimental evaluation

##### 4.1. Experimental setup, test cases and evaluation metrics

We begin this section by elaborating on the experimental setup that was designed and implemented to assess the proposed framework. We assume an Aggregator with a portfolio of 50 customers that are assigned across 4 DVNs. DVN\_test is assigned 25 customers and DVN1, DVN2 and DVN3 are assigned 10, 10 and 5 customers, respectively. For cost and profit estimation, we employ the RTP of Stadtwerk Haßfurt GmbH, a notable German utility provider.

In Table 1, the customer portfolio of DVN\_test is provided. A customer's profile consists of capacity (in kW) information regarding her installed assets (PVs, ESSs), an index corresponding to her reliability as calculated by the DVN following the customer's participation in DR events. The "DR type" attribute signifies whether the customer participates in explicit ("EXP" - direct control of her assets, including controllable loads, such as HVACs), or implicit ("IMP" - via incentives provided by the DR framework, which imply actions from the customer's side whenever a DR arrives) DR events. The column "Customer

Type" signifies whether the customer is a pure consumer ("CONS") or a prosumer ("PROS"). Finally, the contracted power (in kW) is specified to ensure that this value is never surpassed in the context of a DR, thus, mitigating customer penalization.

A total of 8 scenarios are considered, which are summarized in Table 2. The time horizon of all scenarios is the same (15:00–15:30, UTC+3) for the same day (27<sup>th</sup> of April 2020) in order to ensure that the consumption and generation time series are the same. DR setpoints and their respective measurements are dispatched and collected every 1 min. In order to provide a benchmark for all scenarios, a baseline case has been created to calculate the profits deriving from the proposed DR application framework. The baseline scenario is essentially a rule-based algorithm, which calculates setpoints by, first, dispatching power from prosumers (PVs are prioritized over ESSs) and, subsequently, employs customers in descending order of reliability until the demand is met. We note that this rule-based, greedy algorithm is unable to produce setpoints across all examined scenarios, which showcases the limited solution space of such approaches compared to more sophisticated schemes, such as the OptiDVN presented in this work. Clearly, any observed deviation in such scenarios immediately leads to the overall failure of the DR signal.

All scenarios (1–8) essentially progress by activating consecutively the components and functionalities of the proposed DR framework. Each of these scenarios incorporates one or more failures in order to prove the resilient nature of the proposed DR framework, which is considered as one of the key innovative contributions of this work. Scenarios 1 to 4 do not involve customer clustering, while scenarios 5–8 do. The clustering component affects the feasible solution space of the 3 optimisation engines and the meaning of the incoming DR signal. On the one hand, without clustering, all active customers of the DVN are being fed to OptiDVN, regardless of DR type, therefore, the incoming load dispatch DR signal demands the particular setpoint to be the overall constant power flow of the entire DVN, i.e., it is a  $P$  setpoint (in kW). In case of explicit load dispatch, only explicit customers can be used to satisfy the DR, whilst implicit customers continue their normal operation as forecasted. This information is used by OptiDVN and IntraNM in order to



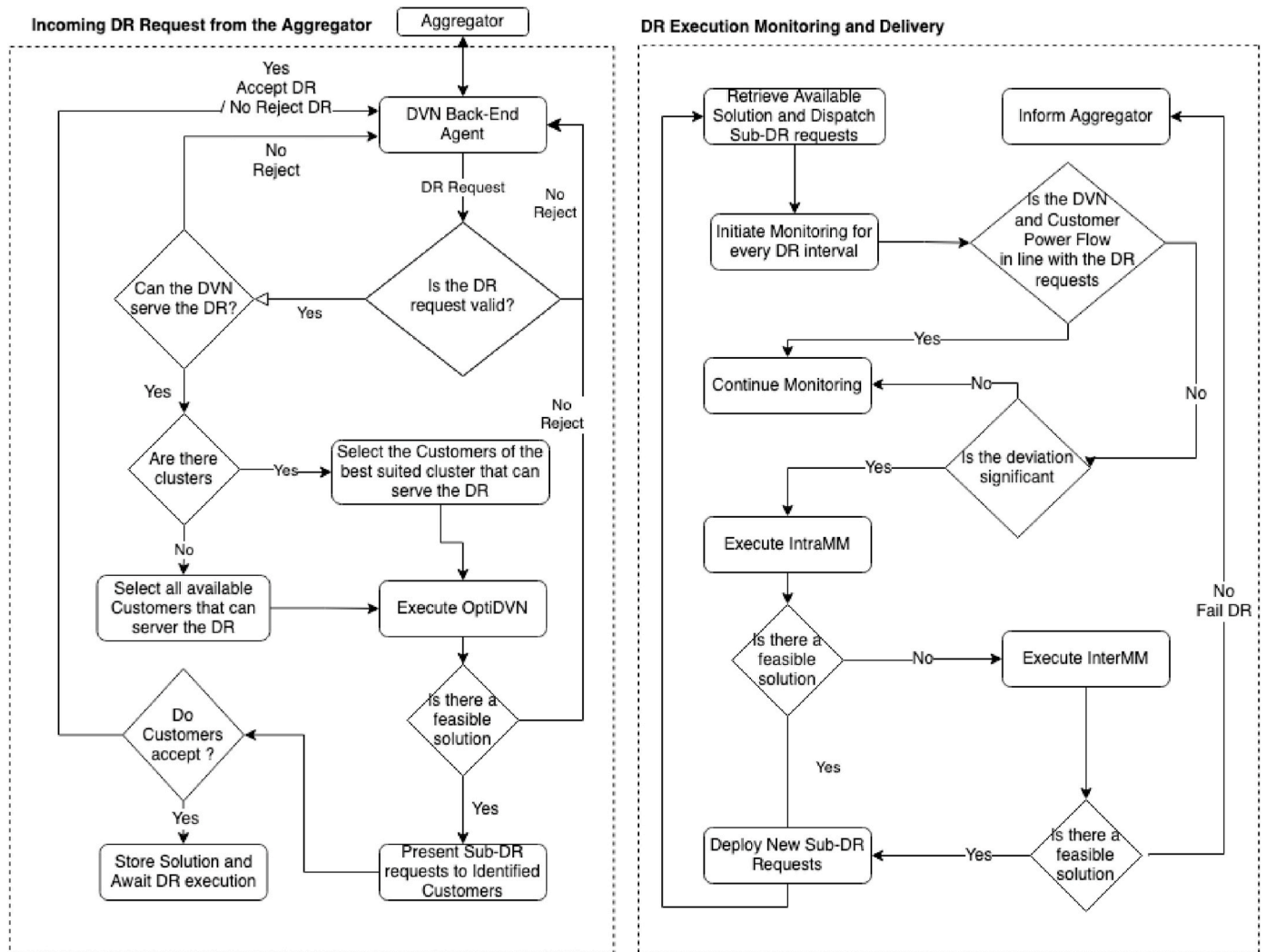


Fig. 4. High-level diagrams depicting a DVN’s decision making flow on input a DR event originating from, e.g., an aggregator (left), and its monitoring process, which takes place during the active period of a DR event (right).

calculate an equivalent augmented  $E_L$  in Equation (8), which includes the baseline demand of both the explicit and implicit users. In case of implicit prosumers, their PVs are considered as negative loads and are algebraically added to their consumption. On the other hand, when customer clustering is employed, OptiDVN and IntraNM are fed only customers of the calculated cluster corresponding to the necessary DR customer type. This implies that the DR signal setpoint corresponds to a relative increase or decrease of the DVN’s aggregated power flow, thus, it is a  $\Delta P$  setpoint (in kW). This differentiation of the interpretation of the DR signal is useful in order to extract valuable outcomes, as we discuss in the next paragraph.

As seen in Table 2, the quantifiable evaluation metrics are the DR’s implementation costs, both expected and actual, the execution time of the enabled DVN components and the number of customers participating in the DR. First, regarding costs, they have all been calculated for both the running and the baseline scenarios to produce comparable numerical results regarding the financial aspects of the application of a DR scheme from the point of view of customers. As previously noted, in many cases, the rule-based engine cannot produce results to service the requested DR and, therefore, in these cases, the base cost cannot be calculated. We note that negative values correspond to monetary values that customers would get paid, whilst positive values correspond to customer penalties. Second, the number of customers that service the DR has been selected as an evaluation metric in order to assess whether the

proposed optimisation algorithms are greedy, meaning whether they have the tendency of selecting as much customers as possible. Note that, in general, the number of participating customers should be the minimum that provides a viable solution for the DR. Finally, the execution time of the three optimisation engines has been included as a performance indicator since, in the proposed DR framework, the time resolution is as low as 1-min long and, therefore, has to be computationally efficient to avoid introducing delays on the real-time decision-making and monitoring processes. To measure the execution time of the respective DVN algorithms, i.e., OptiDVN, IntraNM and InterNM, which are noted in the table as “Opt”, “iNM” and “INM”, respectively, and be able to compare them, the experiments were executed on the same computer, which is equipped with 12 GB RAM, an Intel core i7 (2.3 GHz) and a SSD hard-drive. We do not include metrics regarding the execution time of the clustering algorithm as its invocation is not in conflict with the aforementioned algorithms.

#### 4.2. Experimental results

Table 2 also provides the results of the 8 validation scenarios. To facilitate the reader’s comprehension of the examined scenarios and in the interest of space, we elaborate, in detail, for two scenarios. In both cases, the DR specifies a single setpoint in kW, which must be constantly maintained by the DVN (either as a  $P$  or  $\Delta P$  setpoint) for the DR’s entire

**Table 1**  
DVN\_test customer portfolio overview.

ID	PV [kW]	ESS [kW]	Reliability	DR Type	Customer Type	Contract [kW]
001	0	0	0.979	EXP	CONS	8
002	0	0	0.726	EXP	CONS	8
003	0	0	0.869	IMP	CONS	8
004	0	0	0.890	EXP	CONS	8
005	0	0	0.694	EXP	CONS	8
006	0	0	0.631	EXP	CONS	8
007	0	0	0.793	EXP	PROS	8
008	8	5	0.975	EXP	PROS	21
009	0	0	0.884	IMP	CONS	8
010	0	0	0.966	IMP	CONS	8
011	0	0	0.947	IMP	CONS	8
012	0	0	0.869	IMP	CONS	8
013	10	2.5	0.798	EXP	PROS	21
014	8	0	0.965	EXP	PROS	8
015	0	0	0.948	EXP	CONS	8
016	0	0	0.678	EXP	CONS	8
017	0	5	0.793	EXP	PROS	21
018	0	0	0.788	IMP	CONS	8
019	0	0	0.814	EXP	CONS	8
020	10	0	0.883	EXP	PROS	21
021	0	0	0.944	IMP	CONS	8
022	0	0	0.841	EXP	CONS	8
023	0	0	0.711	IMP	CONS	8
024	0	0	0.849	IMP	CONS	8
025	0	0	0.854	EXP	CONS	8

duration. Failures are presented according to the customer (C-id) that failed to maintain the requested setpoint at a specific timestamp (T). Each failure is associated with the DVN’s mitigation actions.

4.2.1. Scenario 3: OptiDVN → IntraNM → InterNM → IntraNM

As our first reference scenario, we chose “Scenario 3” (Fig. 5) because it demonstrates the increased level of reliability of the proposed DR framework stemming from its multiple safety valves. The input DR event requests a setpoint of −20kW between 15:00–15:30. OptiDVN selected five customers to service it, i.e., C-id007, C-id008, C-id013, C-id014, C-id017 and C-id020.

C-id020 fails at 15:11, thus, triggering IntraNM which, however, fails to provide a feasible solution. Consequently, InterNM is triggered, which succeeds and delegates part of the DR to DVN1 and DVN2. We note that, since DVN1 and DVN2 are now engaged in a DR, they cannot participate

**Table 2**  
Overview of the proposed DR framework’s evaluation scenarios.

ID	DVN Modules				DR [kW]	Failure		Action	No of Pax	Outcome	Base Cost	Exp. Opt Cost [euro ]	Achieved Cost [euro ]	Execution	
	Opt	iNM	I-NM	C		C-id	T							Time (sec)	
1	✓	–	–	–	–1	–	–	–	3	Completed	–0.1232	–8.67994	–8.67994	Opt	2.571
2	✓	✓	–	–	6	017	15:11	iNM	13	Completed	None	3.87007	3.11885	Opt iNM	2.374
3	✓	✓	–	–	–20	020	15:11	iNM I-NM	6	Completed	None	–40.1264	–35.47882	Opt iNM I-NM iNM	2.395 2.095 1.508 2.004
4	✓	✓	✓	–	6	020	15:02	iNM	13	Completed	None	3.87007	3.01488	Opt iNM iNM	2.683 1.109 0.597
5	✓	–	–	✓	–1	–	–	–	2	Completed	–0.1232	–2.73694	–2.74607	Opt	1.927
6	✓	✓	–	✓	–5	013	15:11	iNM	3	Completed	–0.2374	–9.4956	–9.86199	iNM	1.115
7	✓	✓	✓	✓	–5	014	15:02	iNM	3	Completed	–0.2374	–9.4956	–10.38941	Opt iNM iNM	1.187 1.088 0.752
8	✓	✓	✓	✓	2	014	15:02	iNM	6	Completed	–0.0375	–3.63442	–2.88741	Opt iNM iNM I-NM iNM I-NM	1.799 0.530 No sol 1.159 No sol 0.994
						008	15:06	iNM							
						017	15:18	iNM I-NM iNM							

in another DR for the same time period. Therefore, if InterNM is triggered again for some reason, only DVN3 will be considered as a candidate. As illustrated by the power flows depicted in Fig. 5, the event in DVN\_test does not fail, as only one time slot has not been met, thus, monitoring progresses as expected. At 15:20, C-id017, who is allocated a “small” part of the DR compared to other customers, fails. This failure triggers IntraNM again, which is now able to provide an optimal solution to compensate. Individual customer power flows are illustrated in Fig. 5, which depicts in bold lines the failures of customers C-id020 and C-id017 (marked in red in the legend). Customers marked with green in the legend show the ones selected for compensating the failures of the two aforementioned customers. At the end of the DR interval, DVN\_test is able to complete the DR with the assistance of DVN1 and DVN2.

4.2.2. Scenario 8: OptiDVN → IntraNM → InterNM featuring customer clustering

As our second reference scenario, we chose “Scenario 8” (Fig. 6), which is similar to “Scenario 3”, the only difference being that customer clustering is now involved, which is executed at the beginning of each calendar day. The setpoint here corresponds to a relative increase of the DVN’s power flow ΔP by 2 kW from 15:00 to 15:30. OptiDVN selects C-id022, C-id015, C-id020, C-id017, C-id008 and C-id014 to service it. At 15:02, C-id014 fails and IntraNM is triggered, who’s optimal solution is the selection of C-id013 to compensate for the deviation. Thanks to the fast execution time of both IntraNM and the DVN’s monitoring, only one DR interval is lost, with the DR resuming normal operation at 15:03, as illustrated in Fig. 6. Next, C-id018 fails at 15:08. IntraNM is triggered again, however, due to the limited portfolio provided by the customer clustering component, no solution is feasible. Therefore, InterNM is invoked and DVN1 and DVN2 are selected to compensate for the incurred deviation. Finally, C-id017 fails at 15:18 triggering IntraNM which again, however, is unable to provide for a feasible solution. As a result, InterNM is invoked, with only DVN3 available for compensating the customer failure, which is indeed selected and, ultimately, leads to the successful completion of the original DR event.

4.2.3. Experimental outcomes, discussion and limitations

From the 8 scenarios examined, quite interesting and valuable outcomes can be derived. Starting with the straightforward comparison between the optimisation engines and the rule-based decision making, it is evident that the former present consistently at least 3 orders of magnitude larger profit margins and faster execution times. In fact, these

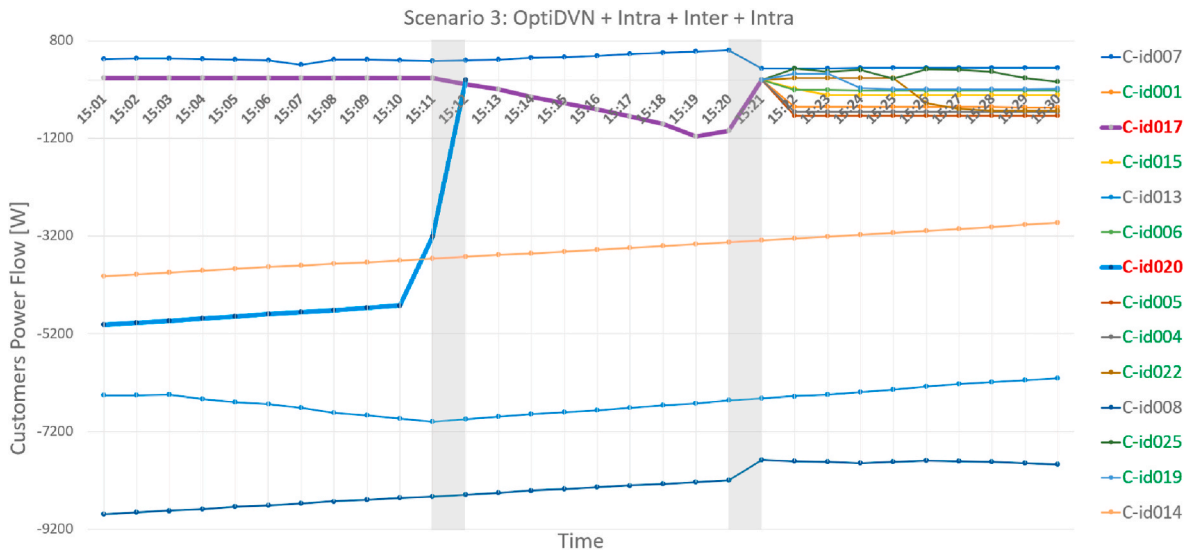


Fig. 5. Scenario 3: Power flow graph of customers servicing a load dispatch signal. Gray areas indicate customer failures.

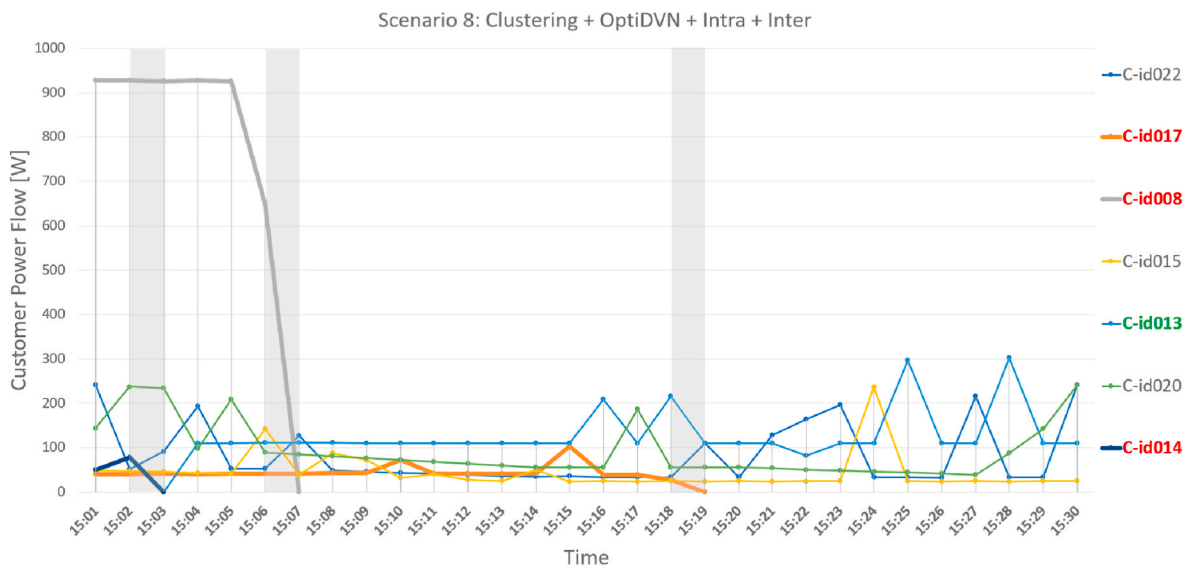


Fig. 6. Scenario 8: Gray areas indicate the time of C-id failure. C-id014 fails at 15:02, C-id008 fails at 15:08 and C-id017 fails at 15:18.

algorithms (OptiDVN, IntraNM, and InterNM) are so fast (sub-second up to few seconds) that they can be considered most pertinent for such MAS-oriented resilient applications, especially when scaling up to hundreds, or even thousands of customers. Furthermore, it is also easily observed that the rule-based approach fails frequently to produce a solution, as its solution space is extremely limited compared to the algorithms proposed in this work. This point is of extreme relevance to, e.g., Aggregators, who can reap significant benefits from the expanded solution space of the proposed algorithms which, in turn, allows trading reliably increased amounts of flexibility in energy markets.

From another perspective, the proposed optimisation engine can be said to introduce a more fair and balanced utilisation of customer assets, compared to greedy algorithms. Indeed, by limiting the number of customers to an optimally selected subset that can deliver the requested amount of power, the number of engaged customers is limited. In addition, through the various constraints and restrictions introduced per type of asset (e.g., PV, ESS), their inclusion also remains within some reasonable bounds.

Combining customer clustering with the optimisation engine, conveys even clearer the meaning of the statement above. Nevertheless,

even though this combination provides for faster customer selection latency, we observe cases where the optimisation engine is unable to provide a feasible solution. This may be due to the fact that both flexibility estimation and clustering introduce certain errors that marginally may not allow for a feasible solution to be found. Such issues could perhaps be resolved if clusters are recomputed prior to the execution of OptiDVN. Nevertheless, we conclude that clustering the customer portfolio is appropriate for larger DVNs (or equivalently VPPs), compared to smaller entities, such as the DVNs used in the scenarios evaluated.

Overall, the proposed MAS offers re-configurable DR application despite of failures that may happen within the DR’s active period. Based on our experimentation, it is evident that InterNM servers as a “last-resort solution”, i.e., it manages to compensate even extreme deviations. This implies that we are not going to be led necessarily to reduced DR application costs as shown in Table 2, where InterNM operation may affect differently the actually achieved DR application cost (e.g., Scenario 5 versus 8). The optimally reduced DR applications costs are a result of the first two optimisation engines, i.e., OptiDVN and IntraNM.

As a general remark, it has been proven through well-designed and

deployed case studies that through the additional two countermeasures (IntraNM and InterNM), it is not only possible to complete an otherwise failed DR, but, under circumstances, it is even possible to increase savings for customers participating in the DR. This is quite an important feat given the importance of customer activity, awareness, and responsiveness to DR requests. By providing for more merits during a critical situation, that could lead to serious losses (both technical and financial), it can be expected that quite high delivery rates can be achieved. Such aspects become more and more important when scalability is considered. Hence, it is not only possible to cost-efficiently distribute large portfolios through the proposed virtual agents that automatically resolve critical instances, but also to deliver further added-value functionalities that increase customer awareness, responsiveness and overall participation in DR programs.

Finally, given the fact that the proposed MAS-based framework was tested with a relative small portfolio on 50 customers, in order to prove further the scalability of the proposed methodology, the authors are currently working on validating the framework's functionalities on a bigger dataset of 300 customers, which will be a mixture of residential, commercial and industrial sites.

## 5. Conclusions

Modern energy markets have started to change into more dynamic and distributed schemes, introducing a variety of advantages, opportunities and challenges. Demand response has proven to play a key role for unlocking the true potential of flexibility offered within these markets, not only for large, but also for small and medium customers. In parallel, given the urgency of increasing significantly the RES penetration in the upcoming years, the need for unlocking the true potential of flexibility offered by all consuming entities and ensuring its reliable delivery when requested is becoming more and more pertinent. Most of related research focuses mainly on decision making optimisation, without addressing the hidden potential of small and medium customers, generally assuming that DRs will be completed or penalised in case of failure, or very rarely adopting adaptive methodologies that learn from failures and present future improvements.

Focusing on resolving DR failure in real-time operation, a novel dynamic multi-agent system has been presented, following the VPP paradigm with virtual nodes of customers. Each node leverages state-of-the-art clustering and optimisation techniques and a two step methodology for mitigating a potential DR failure, first, internally (IntraNM) and, then, in-between nodes (InterNM).

The work presented has shown that the designed MAS framework offers a valuable fail-safe mechanism that can ensure the completion of a DR request, even under very challenging circumstances. This is quite an important achievement, given the volatile and intermittent nature of both RESs and end-user behavior. Furthermore, interestingly enough, there are also cases where additional profits can be achieved during the DR for involved customers (up to 3 orders of magnitude higher), presenting quite an interesting engagement strategy towards a more flexible and reliable portfolio.

By successfully servicing otherwise destined to fail DR requests, the proposed architecture can be considered a "must-have" asset for key energy stakeholders, such as Aggregators, DSOs and TSOs. This is of utmost technical importance, considering the necessity that initially gives birth to such requests. Finally, for Aggregators, as well as, other market players, this could prove to be of significant financial added-value, since flexibility bids (translated partially or completely to DR requests to customers) to markets are ensured to succeed, thus, facilitating ease of participation and the establishment of a stable revenue stream.

Although the presented work successfully provided a proof of concept, it does not come without limitations. One of the core challenges that need to be further explored is the customers' dataset. In order to assess scalability, as well as, real-life potential, a more diverse and large

dataset should be introduced. The need to cover both residential, tertiary and other commercial customers (either prosumers or consumers) is evident, while the inclusion of ancillary assets (e.g., ESS banks) specifically for DR services, should also be introduced to the examined portfolio. Finally, the accuracy of the forecasting engine used in the experiments, as well as, the stochastic behaviour of end-customers, have not been thoroughly examined. These challenges warrant further investigation that are part of ongoing research activities.

## CRedit authorship contribution statement

**Christos Patsonakis:** Supervision, design and implementation of MAS, P2P network, Intra & Inter matchmaking algorithms, manuscript drafting and critical revision, Visualization, experimental planning, implementation and data gathering. **Angelina D. Bintoudi:** Design and implementation of optimisation engine, manuscript drafting and critical revision, Visualization, experiment implementation and data gathering, Data curation. **Konstantinos Kostopoulos:** Software development and testing, experiment data gathering, Visualization, manuscript drafting, Data curation. **Ioannis Koskinas:** Design and implementation of clustering algorithms, manuscript drafting, Visualization. **Apostolos C. Tsolakis:** Project administration, Conceptualization, manuscript drafting and critical revision, Visualization. **Dimosthenis Ioannidis:** Funding acquisition, Conceptualization, Supervision, editing assistance and general support. **Dimitrios Tzovaras:** Funding acquisition, editing assistance and general support.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work is partially funded by the European Union's Horizon 2020 Research and Innovation Programme through DELTA project under Grant Agreement No. 773960.

## References

- Alliance, O., 2013. Openadr 2.0 Specifications. <https://www.openadr.org/specification>.
- Azuma, S.-I., Sato, D., Kobayashi, K., Yamaguchi, N., 2019. Detection of defaulting participants of demand response based on sparse reconstruction. *IEEE Trans. Smart Grid* 11 (1), 368–378.
- Bayer, M., 2012. SQLAlchemy. In: Brown, A., Wilson, G. (Eds.), *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks*, aosabook.Org, pp. 1–350. <http://aosabook.org/en/sqlalchemy.html>.
- Benítez, I., Quijano, A., Díez, J.-L., Delgado, I., 2014. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int. J. Electr. Power Energy Syst.* 55, 437–448. <https://doi.org/10.1016/j.ijepes.2013.09.022>.
- Bintoudi, A.D., Zyglakis, L., Tsolakis, A.C., Gkaidatzis, P.A., Tryferidis, A., Ioannidis, D., Tzovaras, D., 2021. OptiMEMS: an adaptive lightweight optimal microgrid energy management system based on the novel virtual distributed energy resources in real-life demonstration. *Energies* 14 (10). <https://doi.org/10.3390/en14102752>.
- Dierks, T., Rescorla, E., August 2008. In: RFC (Ed.), *The Transport Layer Security (TLS) Protocol*, vol. 5246. RFC. <https://doi.org/10.17487/RFC5246>.
- Directive (EU) 2018/2001 on the Promotion of the Use of Energy from Renewable Sources, 2018. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2001&from=EN>.
- Directive (EU) 2019/944 on common rules for the internal market for electricity and amending Directive 2012/27/EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0944&from=EN>, 2019.
- Dusonchet, L., Favuzza, S., Massaro, F., Telaretti, E., Zizzo, G., 2019. Technological and legislative status point of stationary energy storages in the EU. *Renew. Sustain. Energy Rev.* 101, 158–167.
- Eby, P., 2010. Pep 3333 – python Web Server Gateway Interface v1.0.1. <https://www.python.org/dev/peps/pep-3333/#copyright>.
- Gelenbe, E., Görbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., Lyberopoulos, G., 2013. Nemesys: enhanced network security for seamless service provisioning in the smart mobile ecosystem. In: Gelenbe, E., Lent, R. (Eds.), *Information Sciences and Systems*. Springer International Publishing, pp. 369–378.

- George, Pavlidis, Tsolakis, Apostolos C., Ioannidis, Dimosthenis, Tzovaras, Dimitrios, 2020. Demand flexibility estimation based on habitual behaviour and motif detection. In: 25th International Conference on Pattern Recognition, p. ICPR2020.
- Ghorashi, S.M., Rastegar, M., Senemmar, S., Seifi, A.R., 2020. Optimal design of reward-penalty demand response programs in smart power grids. *Sustain. Cities Soc.* 60, 102150.
- Golmohamadi, H., Keypour, R., Bak-Jensen, B., Pillai, J.R., 2019. A multi-agent based optimization of residential and industrial demand response aggregators. *Int. J. Electr. Power Energy Syst.* 107, 472–485.
- Gomes, L., Spínola, J., Vale, Z., Corchado, J.M., 2019. Agent-based architecture for demand side management using real-time resources' priorities and a deterministic optimization algorithm. *J. Clean. Prod.* 241, 118154.
- González-Briones, A., De La Prieta, F., Mohamad, M.S., Omatu, S., Corchado, J.M., 2018. Multi-agent systems applications in energy optimization problems: a state-of-the-art review. *Energies* 11 (8), 1928.
- J. Gouveia, J. Seixas, A. Mestre, Daily electricity consumption profiles from smart meters - proxies of behavior for space heating and cooling, *Energy* 141.. doi:10.1016/j.energy.2017.09.049.
- Grossmann, A., Kronland-Martinet, R., Morlet, J., 1989. Reading and understanding continuous wavelet transforms. In: *Wavelets: Time-Frequency Methods and Phase Space -1*, p. 2. [https://doi.org/10.1007/978-3-642-75988-8\\_1](https://doi.org/10.1007/978-3-642-75988-8_1).
- Hale, E., Bird, L., Padmanabhan, R., Volpi, C., 2018. Potential Roles for Demand Response in High-Growth Electric Systems with Increasing Shares of Renewable Generation.
- Housley, R., Ford, W., Polk, W., Solo, D., January 1999. In: RFC (Ed.), *Internet x.509 Public Key Infrastructure Certificate and Crl Profile*, RFC 2459. <https://doi.org/10.17487/RFC2459>.
- Karfoopoulos, E., Tena, L., Torres, A., Salas, P., Jorda, J.G., Dimeas, A., Hatzigiorgiouri, N., 2015. A multi-agent system providing demand response services from residential consumers. *Elec. Power Syst. Res.* 120, 163–176.
- Lin, S., Li, F., Tian, E., Fu, Y., Li, D., 2019. Clustering load profiles for demand response applications. *IEEE Trans. Smart Grid* 10 (2), 1599–1607. <https://doi.org/10.1109/TSG.2017.2773573>.
- O. P. Mahela, M. Khosravy, N. Gupta, B. Khan, H. H. Alhelou, R. Mahla, N. Patel, P. Siano, Comprehensive overview of multi-agent systems for controlling smart grids, *CSEE J. Power Energy Syst.*
- M. McPherson, B. L. Cowiestoll, Demand response for variable renewable energy integration: a proposed approach and its impacts, *Energy*:10.1016/j.energy.2020.117205.
- Motlagh, O., Berry, A., O'Neil, L., 2019. Clustering of residential electricity customers using load time series. *Appl. Energy* 237, 11–24. <https://doi.org/10.1016/j.apenergy.2018.12.063>.
- Muthirayan, D., Kalathil, D., Poolla, K., Varaiya, P., 2019. Mechanism design for demand response programs. *IEEE Trans. Smart Grid* 11 (1), 61–73.
- New Passenger Car Registrations by Fuel Type in the european union, 2019. <https://tinyurl.com/rkto7pa>.
- Parrish, B., Heptonstall, P., Gross, R., Sovacool, B.K., 2020. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. *Energy Pol.* 138, 111221.
- Pasetti, M., Rinaldi, S., Manerba, D., 2018. A virtual power plant architecture for the demand-side management of smart prosumers. *Appl. Sci.* 8 (3), 432.
- Praça, I., Ramos, C., Vale, Z., Cordeiro, M., 2003. Mascem: a multiagent system that simulates competitive electricity markets. *IEEE Intell. Syst.* 18 (6), 54–60.
- Realttime, I., 2020. Openfire. <https://www.igniterealttime.org/projects/openfire>.
- Shafie-khah, M., Siano, P., Aghaei, J., Masoum, M.A., Li, F., Catalão, J.P., 2019. Comprehensive review of the recent advances in industrial and commercial dr. *IEEE Trans. Ind. Inf.* 15 (7), 3757–3771.
- Share of energy from renewable sources. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg\\_ind\\_ren&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_ind_ren&lang=en), 2020.
- Shawon, M.H., Muyeen, S., Ghosh, A., Islam, S.M., Baptista, M.S., 2019. Multi-agent systems in ict enabled smart grid: a status update on technology framework and applications. *IEEE Access* 7, 97959–97973.
- Silva, C., Faria, P., Vale, Z., 2020. Rating the participation in demand response programs for a more accurate aggregated schedule of consumers after enrolment period. *Electronics* 9 (2), 349.
- Sioshansi, F.P., 2019. *Consumer, Prosumer, Prosumer: How Service Innovations Will Disrupt the Utility Business Model*. Academic Press.
- Study on Residential Prosumers in the European Energy Union, 2017. [https://ec.europa.eu/info/sites/info/files/executive-summary-study-residential-prosumers\\_en.pdf](https://ec.europa.eu/info/sites/info/files/executive-summary-study-residential-prosumers_en.pdf).
- Tracking energy integration. <https://www.iea.org/reports/tracking-energy-integration/demand-response>, 2019.
- Van Rossum, G., Drake, F.L., 2009. *Python 3 Reference Manual*. CreateSpace.
- Violette, D., Shober, M., 2014. *Carbon Dioxide Reductions from Demand Response Impacts in Three Markets Prepared for: Advanced Energy Management Alliance*.
- Wang, Y., Lin, H., Liu, Y., Sun, Q., Wennersten, R., 2018. Management of household electricity consumption under price-based demand response scheme. *J. Clean. Prod.* 204, 926–938.
- Wang, Z., Paranjape, R., Chen, Z., Zeng, K., 2019. Multi-agent optimization for residential demand response under real-time pricing. *Energies* 12 (15), 2867.
- Woltmann, S., Coordes, A., Stomberg, M., Kittel, J., 2020. Using multi-agent systems for demand response aggregators: a technical implementation. In: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), vol. 1. IEEE, pp. 911–918.
- Zhang, Huan, Song, Kun, 2011. Research and experiment on affinity propagation clustering algorithm. In: 2011 Second International Conference on Mechanic Automation and Control Engineering, pp. 5996–5999. <https://doi.org/10.1109/MACE.2011.5988401>.