Maximizing Renewable Energy Use with Decentralized Residential Demand Response

Ivana Dusparic, Adam Taylor, Andrei Marinescu, Vinny Cahill, Siobhán Clarke
Future Cities Research Centre
Distributed Systems Group
Trinity College Dublin
{ivana.dusparic, tayloral, marinesa, vinny.cahill, siobhan.clarke}@scss.tcd.ie

Abstract*—Due to steady urbanization, the electrical grid is facing significant changes in the supply of resources as well as changes in the type, scale, and patterns of residential user demand. To ensure sustainability and reliability of electricity provision in the growing cities, a significant increase in energy generated from renewable sources (e.g., wind, solar) is required. However, renewable energy supply is much more variable and intermittent than traditional supply, as it depends on changing weather conditions. In order to optimize residential energy usage, demand response (DR) techniques are being investigated to shift device usage to the periods of low demand. Currently most DR approaches focus on traditional DR goals, e.g., reducing usage at peak times and increasing it at off-peak times. More flexible and adaptive techniques are needed that can not only meet traditional DR requirements, but enable just-in-time use of renewable energy, rather than requiring its curtailment or using expensive and inefficient storage options. This paper proposes the use of decentralized learning-based multi-agent residential DR to enable more efficient integration of renewable energy sources in the smart grid, in the presence of increased demand caused by high electric vehicle penetration. We evaluate the approach using real household usage data obtained from Irish smart meter trials and data on wind-generated energy from the Irish grid operator. We discuss advantages of the proposed decentralized approach and show that it is able to respond to multiple variable wind-generation patterns by shifting up to 35% of the overall energy usage to the periods of high wind availability.

***Index Terms*—demand response, renewable energy, decentralized learning, multi-agent systems**

I. Background and Motivation

55% of the world's population currently lives in cities, with that figure expected to increase to 66% by 2050 [1]. Such worldwide urbanization is putting increasing strains on already constrained city resources, presenting threats to the sustainability of city infrastructures including transport, energy and water supply, as well as presenting air quality and therefore health issues. To mitigate some of the effects of this expansion, European Union (EU) is looking at ways to make the European cities more climate-friendly and less energy-consuming by setting ambitious targets for reduction in CO_2 emissions. EU 2050 roadmap [2] outlines the plans for reducing CO_2 emissions to at least 80% below 1990 levels by 2050. These targets are driving the increasing penetration of renewable energy sources and electric vehicles (EVs). For example, in Ireland, it is expected that 80% of electricity will come from renewable sources by 2050 [3], and 60% of new cars sold in 2050 will be electric [4].

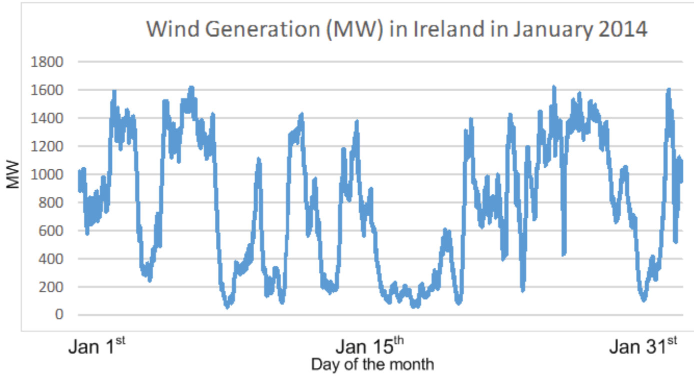

Figure 1. Intermittency of wind generation

These developments will introduce significant challenges to the way the electrical grid operates. Greater use of renewable energy and EVs will drastically reduce air pollution and associated health care costs, however, at the same time it will require adaptation of the way electrical grid operates today. Penetration of EVs will significantly increase daily energy demand, while energy supply will become more intermittent and unpredictable with a greater shift towards the use of renewables. As certain types of renewable energy (primarily wind and solar) depend on changing weather conditions, they will not necessarily follow the patterns of daily residential consumer energy usage and will vary greatly from day to day, or even from hour to hour. For example, Figure 1 shows the levels of wind-generated energy in Ireland in January 2014 [5], indicating the high diversity in levels of production over short period of time.

Analysis of wind-generated energy in Germany shows that variation between hour to hour in wind generation can be more than 10% of the installed capacity, and the deviation between day-ahead prediction of wind-generated energy and actual generation can reach over 30% [6]. Unpredictability and variability resulting from integration of wind energy into electricity systems requires increased efforts to balance and control the power system. This costs anywhere between 1 and

978-1-4673-6552-9/15/$31.00 ©2015 IEEE

10 euro per MWh hour of wind-generated energy [7].

Traditionally, achieving balance in the electrical grid between energy supply and demand was achieved by controlling the supply side - increasing and decreasing production and using energy storage. For example, currently, the level of renewable (mostly wind-generated) energy in Ireland averages at 20% [8], and wind generation has to be curtailed once it reaches 50% of the overall energy generation in order to maintain the stability of the network [9], resulting in significant amounts of wind energy being wasted.

Recently, the onus on maintaining the balance has been shifting to the demand side, with demand-response (DR) programs enabling supply to remain steady while encouraging consumers to shift their demand from peak times to off-peak times. Traditionally, DR focused on incentivizing greater usage during the off-peak times (e.g., during the night), and reducing the usage during the peak times (e.g., 5-7pm).

To make DR techniques relevant in the scenarios of high penetration of renewable energy, they not only need to focus on shifting usage from peak to off-peak times, but they need to be more flexible and adaptive in order to respond to variable wind generation patterns. Device usage should be shifted to the times of high renewable energy availability, while also taking into account traditional patterns of consumer usage. Even though they introduce additional demand on the grid, EVs also present an opportunity for introducing DR programmes, as their charging times are flexible, as long as user requirements are met (i.e., as long as the vehicle is charged sufficient amount by the desired departure time). However, residential DR programmes introduce additional complexity arising from the need to manage a large number of devices with a diverse range of user preferences and requirements. Additionally, implementation of calculated DR actions needs to be actuated by the devices promptly, as studies show that even a 30 minute lag in response results in 72% loss of DR value for the system [10].

In this paper, we propose that in order to be sufficiently flexible and adaptive to support efficient integration of renewable energy into a smart grid, DR system needs to be decentralized, where each device is controlled by an intelligent agent in charge of its own preferences and schedules, responding to incentives and signals from the network and cooperating with other devices to ensure everyone's requirements are met. This will enable accurate and timely response to changes in the wind generation patterns therefore maximizing DR effects. In our initial work [11] we showed that multiple devices acting independently using learning techniques can respond to traditional DR request to shift usage from peak to offpeak times. In this paper we extend the DR system to include collaboration between the devices to improve the grid-wide effect and show that devices can learn to react to availability of wind-generated energy and modify the usage to follow that of the highly fluctuating wind-generation pattern.

The rest of this paper is organized as follows: Section 2 surveys the existing approaches to DR. Section 3 introduces Distributed W-Learning (DWL), the learning algorithm that underpins our DR approach. Section 4 presents the details of DWL-based DR approach to maximizing renewable energy use, while Section 5 discusses advantages of decentralized DR. Section 6 presents experimental results and their analysis. Finally, Section 7 concludes the paper, discussing avenues for further exploration of the proposed technique.

II. Related Work

The need for automated, flexible and dynamic DR techniques has been recognized by the research community and various techniques have been investigated to shift residential demand from high demand periods to those of lower demand. Some of the techniques focus on optimization and shifting of energy use within a single household, using, for example, neural networks [12], expert systems theory [13], and linear programming [14]. A number of approaches aim to centrally manage the consumption of multiple households within a community, e.g., [15]. Recently, multi-agent systems, which are inherently decentralized, have been identified as a promising DR control approach due do their flexibility, extensibility and fault tolerance [16]. Several multi-agent approaches have been proposed in the literature (e.g., [17]), and have been shown to successfully achieve the desired device energy usage shift. Increasingly, the role and impact of DR in integration and penetration of renewable energy is being investigated (e.g., [18]). However, currently proposed DR approaches rarely address integration of renewables and the consequent need for highly dynamic device rescheduling. [19] proposes a DR mechanism for integration of solar power based on scheduling of both traditional generation sources as well as demand scheduling. [20] proposes that large-scale integration of renewable sources can be achieved by aggregation of defferable loads, however, in their DR solution deferrable load usage is scheduled and controlled directly by a centralized aggregator. We propose a DR technique which is implemented only on demand side, i.e., it does not depend on generation side, and is decentralized, whereby each device/household is in charge of its on schedule rather than being controlled by the aggregator. The following sections present this approach, first by presenting the background algorithm (DWL) used by our approach, and then the details of the DR technique itself.

III. Distributed W-Learning (DWL)

DWL [21] is a learning-based algorithm for multi-agent optimization that enables collaboration between heterogeneous entities in order to simultaneously satisfy multiple system goals. DWL is based on Reinforcement Learning (RL) [22]. In DWL, each agent uses a single Q-Learning [23] process to implement each of its own local goals. In each state of the environment that an agent can be in, the suitability of each action that an agent can take is learnt over time, and expressed as a Q-value. To arbitrate between different policies, an agent uses W-Learning [24] which learns the relative importance of agent's policies. In W-Learning, for each of the states that each agent's policy can be in, an agent learns how is the performance of that policy affected if its preferred action does not get executed. This difference between the reward that the agent would receive if its preferred action is executed and the reward agent receives when another policy's action is executed, is expressed as a W-value.

Q-values for state-action pairs, $Q(s, a)$, are updated using the Formula 1:

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \underset{a'}{max} Q(s', a')) \quad (1)$$

and W(s) are updated according to Formula 2:

$$W_i(s) = (1-\alpha)W_i(s) + \alpha(Q_i(s, a_i) - (r_i + \gamma \max Q_i(s', a_i'))) \quad (2)$$

where r_i is the immediate reward received, s is the current state of the policy, s' is the next possible state, a_i is the current action, and a_i' is the possible next action for policy i. α and γ are learning parameters which assign weight to new agent's experiences and the rate at which old experiences are discounted.

In DWL, agents learn Q-values and W-values for all of their local policies, but all agents also learn Q-values and W-values for all of the policies that their immediate neighbours implement (so-called remote policies), i.e., they learn how their local actions affect their neighbours' performance. At each time step, each agent considers the W-values for the current state of each of its local and remote policies. Neighbours' W-values can be multiplied by a cooperation coefficient C, to enable a local agent to give a varying degree of importance to the neighbours' action suggestions. A winning action is selected based on formula 3:

$$W_{win} = \max(W_{il}, C \times W_{ijk}). \quad (3)$$

Through use of Q-Learning, W-Learning, local policies, remote policies, and collaboration coefficient, DWL simultaneously optimizes multiple policies on multiple agents while priorities of the policies are respected both locally on the device, on other devices and requests from the grid are respected. In the next section we present how is DWL used in the design of decentralized DR approach.

IV. DWL-based DR for maximizing renewable energy use

In our proposed approach, each device enrolled in a DR program is controlled by an intelligent agent which learns how to meet its own multiple goals (e.g., charge an EV sufficiently for the daily journey, save energy costs), grid requests (e.g., maximize the usage during high wind-generated energy availability), and collaborates with other devices to meet their own goals. This is enabled by each agent implementing a number of policies using the DWL algorithm. We implement a case study in a small residential community with deferrable loads, in this case EVs.

Each EV agent implements three DWL policies:

- P1: **Renewable Energy Policy** - specifies the maximum aggregate energy load that the whole community of agents should not exceed. This maximum load is variable and follows the pattern of wind-based energy generation, encouraging agents to use only renewable energy where possible. EV agents are given a high negative reinforcement for exceeding this limit, thereby enabling EVs to learn to charge only during the periods where the difference between current transformer load and the maximum specified load is sufficient to enable additional EV load.
- P2: **Battery Charge Policy** - specifies the minimum charge that EV should achieve. EV agents were given high positive reward for achieving the minimum battery charge of 60%. Depending on the length of the daily journey and required battery charge, this limit can be set lower or higher. Additionally, agents are given a positive reward each time a battery charge increased.
- P3: **Baseload Policy** - which provided agents with the current baseload information and 24-hour prediction of the baseload, encouraging EVs to learn to postpone the charging if the period of low baseload is coming up. The baseload level maps directly to the price of energy; high load means price of energy is high, and agents are encouraged to charge at the lowest possible cost. Instead of using absolute values for the baseload, we classified it as "low", "medium", and "high", with respect to the 24-hour load prediction. If agents detect that the current baseload is "high", that means that load is high comparing to what it will be for the remainder of the available charging period, and agents learn to wait for the lower price period.

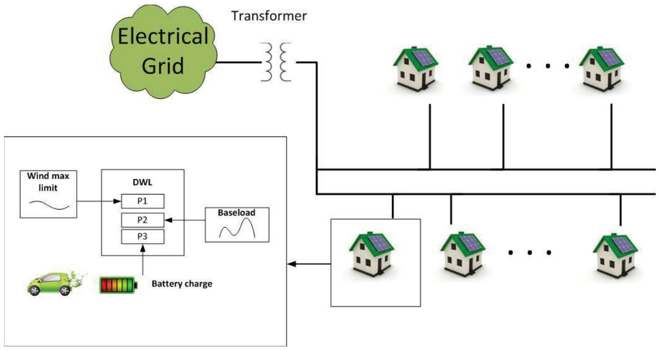

Figure 2. DWL-based residential DR

Relative priorities of these policies are learnt based on the scale of the rewards given in each specific policy. For example, failing to achieve the required battery charge was punished the most severely, as that would mean that the EV would run out of battery on its planned daily trip. It is deemed preferable to charge at a higher energy price, than not to achieve the desired battery charge. Due to the flexibility of DWL implementation of policies, each agent can implement all of these policies, a subset of them, or any additional ones. For example, different types of deferrable loads can be supported - a water heater device might have a policy specifying desired water temperature, or a washing machine device might have a policy specifying desired finish time.

Collaboration

In DWL devices collaborate by each device learning how its actions affect those of its collaborators. For example, a device learns that it charging at a specific time period might affect other agents negatively, as it increases overall load in the neighbourhood and exceeds maximum transformer load, resulting in all agents receiving negative reward. Therefore, agents learn to implicitly synchronize their energy usage. In small scale scenarios, it is feasible for all agents to collaborate with each other, however in larger communities, alternative models of collaboration need to be explored. In our approach, we divide a community in smaller collaborating sub-communities, where only agents within a community cooperate with each other.

In the next section we discuss benefits of our proposed decentralized DWL-based DR.

V. Benefits of Decentralized DR

DWL-based DR has a number of benefits arising from its decentralization:

1) **Scalability** – In centralized approaches a set of instructions or a schedule for energy usage is calculated centrally and sent to end-user devices periodically (e.g., [15]), therefore local devices/households are not in charge of their own schedule. If some of the devices do not respect the assigned schedule (i.e., their use is manually overridden by the user), a new schedule needs to be recalculated and resent to all the devices. Similarly, a new schedule needs to be recalculated ny time a new request comes from the grid, or a change in predicted load/renewable energy supply occurs. DWL-based approach offers significantly increased scalability as the decisions are not made at a central point. This is crucial for timely response to changes in renewable energy generation, due to financial loses associated with delayed DR as already discussed.
2) **Synchronization of end-user energy usage** – A number of decentralized, cost-saving approaches are being proposed where each household calculates its own schedule without collaboration with other users (e.g., [25], [26]). However, aggregation effects of all customers shifting their demand to low price/low demand periods, potentially resulting in a sudden increase in demand, is not considered. Agent cooperation and coordination enabled by DWL can ensure that devices are aware of each other's performance as well as ensure grid requests are satisfied.
3) **Addressing multiple and flexible end-user and grid policies** – Individualized goals in DWL approach can easily be added or removed, activated or deactivated, and their relative priority changed over time, periodically, or per user/device. For example, some users will at certain times prefer cost-saving over comfort. Modularity and flexibility is increased by expressing the goals individually rather than as a single optimization problem with numerous constraints (as in the case in, e.g., [25]).
4) **Privacy** - Numerous techniques are being proposed in literature to ensure collection and analysis of fine-grained energy usage data does not infringe on users' privacy (e.g., [27] and [28]). Decentralized design of DWL approach ensures consumer privacy is respected, as the end-user preferences and schedules do not need to be transferred to a centralized scheduling unit but are only required locally on each end-user device, removing the potential for interception and misuse. Only non-identifying information on the learning process (e.g., rewards received at a particular timestep) is shared between a small number of neighbouring households/devices.

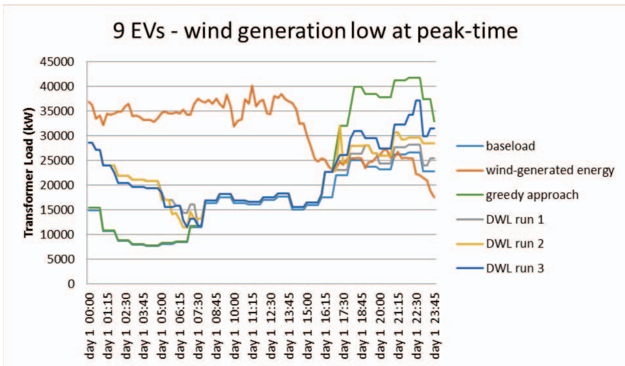

Figure 3. Scenario 1: 9 EVs, wind decreasing at peak-time

In the next section we present the design of the experiments performed to empirically evaluate the suitability of the proposed DWL-based automated residential DR in maximizing the use of renewable energy by shifting the flexible device usage to the times of the high availability of the wind-generated energy.

VI. Evaluation, Results and Analysis

This section presents the simulation environment in which our experiments were performed, scenarios that have been simulated, results of the evaluation and analysis of the results.

A. *Use Case Design*

Simulations were performed in two different sets, with 9 and 90 households enrolled in DWL-based DR, each with a non-shiftable baseload, and with a schedulable EV controlled by a DWL agent. Experiments were run for 30 simulated days and split in 2 parts: learning/exploration period, in which agents were exploring the quality and results of their actions with respect to their policies, and exploitation period, during which all agents selected only actions which they have learnt to be the most suitable for the performance of their own local policies as well as policies of the agents with which they were collaborating. Suitable duration of the exploration period and the learning parameters was determined experimentally,

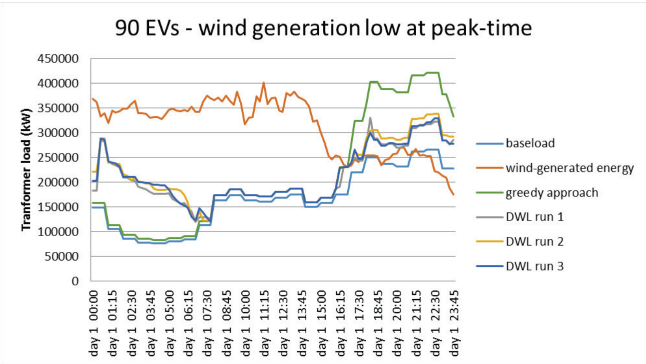

Figure 4. Scenario 2: 90 EVs, wind decreasing at peak-time

and it was observed that in the scenarios selected, results of the learning stabilize after ~20 days of learning data. Each agent was collaborating with 8 of its nearest neighbours (i.e., in 9-agent scenario, all agents were collaborating with each other, while in 90-agent scenarios, agents were divided in sub-groups of 9). The results presented are from the exploitation phase. Batteries of EVs simulated have a capacity of 30 kWh and charge at a rate of 1.4kW. The required daily mileage is 50 miles. Baseload in each household ranges from 0.8kW (during the night) to 3kW (at evening peak time), and is taken from the data recorded in Smart Metering Electricity Customer Behaviour Trials in Ireland [29]. All experiments were performed in GridLAB-D [30].

We have run 4 experimental scenarios:

- **Scenarios 1 and 2**: Devices were given a maximum load limit which followed the shape of wind-generated energy and significantly *decreased* at peak-time. Scenario 1 had 9 households enrolled, and Scenario 2 had 90 households.
- **Scenarios 3 and 4**: Devices were given a maximum load limit which followed the shape of wind-generated energy and significantly *increased* at peak-time. Scenario 3 had 9 households enrolled, and Scenario 4 had 90 households.

All scenarios were also compared to a so-called greedy, non-DR approach, where all vehicles were plugged in and started charging immediately after arriving home, and were charged continuously until the battery was fully charged.

B. Experimental Results

Results of the evaluation of DWL-based DR in Scenarios 1 and 2 are presented in Figures 3 and 4 respectively, while results of Scenarios 3 and 4 are presented in Figures 5 and 6. The graphs show recorded transformer load resulting from all enrolled houses' baseload and DWL-controlled EVs. Results of 3 runs of DWL are presented. Each graph also shows the non-schedulable baseload and the quantity of the wind-generated energy. Wind generation patterns are based on values obtained from Eirgrid [5] (Irish transmission system operator and market operator) and are scaled down to be comparable to household usage; they are not representing absolute wind generation values but the wind generation pattern.

Wind-generated energy decreasing at peak time: We first examine the results of evaluation in Scenarios 1 and 2, where wind energy generation decreases in the afternoon, coinciding with the afternoon peak. DWL agents are tasked with charging the vehicles fully while using only wind-generated energy. Agents should therefore learn not to charge at peak-time as that time the wind limit is almost fully met by the houses' baseload alone. In 9-household Scenario 1 (Figure 3), a non-adaptive greedy approach, which charges an EV as soon as it arrives home, starts charging during the peak baseload and low availability of wind energy, going over the assigned load limit by over 35%, fully charging during the low wind availability, and not utilizing the high wind availability that follows during the night-time where baseload uses only about 15-20% of the wind-generated energy. On the other hand, in 9-household Scenario 1, DWL-based devices successfully shift the usage of all devices to off-peak times, reducing the overall peak-time low-wind usage by 35% and increasing the usage during the night-time when wind generation is higher. The pattern is similar in 90-household Scenario 2 (Figure 4): greedy approach increases the peak-time low-wind usage by 42%. DWL-based approach also goes over the wind-generated limit at times (especially DWL run 3), but still successfully shifts ~20-25% of usage to off-peak high-wind times. The loss of performance in 90-household scenario is potentially due to the higher number of devices needing to cooperate with each other and should be further investigated by exploring different collaboration settings.

Wind-generated energy increasing at peak time: In Scenarios 3 and 4 wind generation increases at peak time and drops again at off-peak night-time (see Figures 5 and 6). Traditionally, DR would aim to reduce the device usage at peak time, however, when the aim is to maximize the renewable energy usage, which in scenario coincides with peak-time energy usage, devices use should not be reduced but ever further encouraged to maximize the available energy. As in the previous scenarios, greedy non-adaptive approach charges EVs immediately upon arriving home, and since wind production is coincidentally high at that time it initially does not go over the assigned limits in previous scenarios. However, as the wind starts dropping, the greedy approach still continues to charge the vehicles, going over desired load limits by as much as 18%. DWL, on the other hand, in both 9-household and 90-households scenarios adjusts to the reduced wind-generated energy and leaves the remaining battery charge for low-load period. This period also has low-wind production, however base-load only utilizes less than 50% of the available wind-generated energy, leaving enough room for the EVs to finish charging. Overall, DWL-based DR shifts about 25% of the load to off-peak period to ensure wind-energy limits are not exceeded.

In summary, the results show that DWL-based DR is able to follow the wind generation patterns and shift device usage to the periods of high wind generation, or to low baseload times. It is important to note that DWL agent settings in all 4 scenarios were exactly the same; no adjustment to

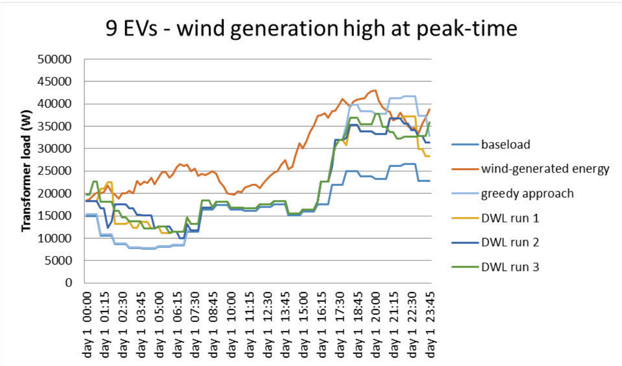

Figure 5. Scenario 3: 9 EVs, wind increasing at peak-time

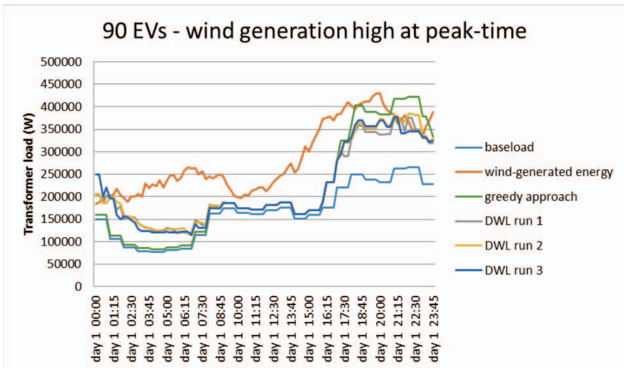

Figure 6. Scenario 4: 90 EVs, wind increasing at peak-time

policies, rewards, goals, collaboration or learning parameters was needed, as the approach was able to autonomously adapt to the change in number of devices and wind-patterns supplied.

VII. Conclusion and Future Work

This paper proposed a decentralized residential DR approach which is able to respond to changes in availability of wind-generated energy and reschedule user devices to the times of high renewable energy penetration. The approach has been shown to successfully adapt to multiple wind generation patterns in both small and large scale scenarios. However, there is a number of further research issues that should be addressed. One example is, how can the approach address devices others than EVs which have different constraints. In addition, the results of large-scale Scenario 2 suggest that further performance improvements could be obtained by exploring different collaboration mechanisms. Agents could limit or extent the number of other agents to cooperate/coordinate with, in order to ensure further renewable energy is utilized.

References

[1] The Population Division of the Department of Economic and Social Affairs, "World urbanization prospects," United Nations, Tech. Rep., 2014.
[2] European Commission, "A roadmap for moving to a competitive low carbon economy in 2050," 2011.
[3] SEAI, "Renewable energy in ireland," 2012.
[4] ——, "Electric vehicles roadmap," 2011.
[5] Eirgrid, "System performance data, wind generation," http://www.eirgrid.com/operations/systemperformancedata/windgeneration/, last accessed July 2014.
[6] M. Klobasa, "Analysis of demand response and wind integration in germany's electricity market," *Renewable Power Generation, IET*, vol. 4, no. 1, pp. 55–63, January 2010.
[7] M. Klobasa and M. Ragwitz, "Demand response - a new option for wind integration," in *European Wind Energy Conference and Exhibition*, 2006.
[8] Sustainable Energy Authority of Ireland, "Energy in ireland: Key statistics," 2013.
[9] Eirgrid, "Curtailment report," 2012.
[10] S. Madaeni and R. Sioshansi, "Measuring the benefits of delayed price-responsive demand in reducing wind-uncertainty costs," *Power Systems, IEEE Transactions on*, vol. 28, no. 4, pp. 4118–4126, Nov 2013.
[11] I. Dusparic, C. Harris, A. Marinescu, V. Cahill, and S. Clarke, "Multi-agent residential demand response based on load forecasting," in *SusTech 2013*. IEEE, 2013, pp. 90–96.
[12] H. Jiang and Z. Tan, "Load forecasting in demand response," in *APPEEC*, 2012.
[13] Q. Dam, S. Mohagheghi, and J. Stoupis, "Intelligent demand response scheme for customer side load management," in *Energy 2030 Conference, 2008. ENERGY 2008. IEEE*, nov. 2008, pp. 1 –7.
[14] S.-J. Kim and G. Giannakis, "Efficient and scalable demand response for the smart power grid," in *CAMSAP*, 2011.
[15] A. Dimeas and N. Hatziargyriou, "Multi-agent reinforcement learning for microgrids," in *PES General Meeting, 2010 IEEE*, July 2010.
[16] S. McArthur, E. Davidson, V. Catterson, A. Dimeas, N. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-agent systems for power engineering applications 2014;part i: Concepts, approaches, and technical challenges," *Power Systems, IEEE Transactions on*, vol. 22, no. 4, Nov 2007.
[17] S. Vandael, N. Boucke, T. Holvoet, and G. Deconinck, "Decentralized demand side management of plug-in hybrid vehicles in a smart grid," in *ATES*, 2010.
[18] A. Pina, C. Silva, and P. Ferrao, "The impact of demand side management strategies in the penetration of renewable electricity," *Energy*, vol. 41, 2011.
[19] N. Cicek and H. Delic, "Demand response for smart grids with solar power," in *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2014 IEEE*, May 2014, pp. 566–571.
[20] A. Papavasiliou and S. Oren, "Large-scale integration of deferrable demand and renewable energy sources," *Power Systems, IEEE Transactions on*, vol. 29, no. 1, pp. 489–499, Jan 2014.
[21] I. Dusparic and V. Cahill, "Autonomic multi-policy optimization in pervasive systems: Overview and evaluation," *TAAS*, vol. 7, no. 1, 2012.
[22] R. S. Suton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts: A Bradford Book. The MIT Press, 1998.
[23] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 3, May 1992.
[24] M. Humphrys, "Action selection methods using reinforcement learning," in *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. MIT Press, 1996.
[25] Z. Wen, H. R. Maei, and D. ONeill, "Optimal demand response using device based reinforcement learning," Stanford University, Tech. Rep., 2012.
[26] R. Fazal, J. Solanki, and S. Solanki, "Demand response using multi-agent system," in *NAPS*, 2012.
[27] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. Shen, "Eppdr: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 8, Aug 2014.
[28] A. Cardenas, S. Amin, and G. A. Schwartz., "Privacy-aware sampling for residential demand response programs," in *HiCoNS*, 2012.
[29] Irish Social Science Data Archieve. Comission for Energy Regulation, "CER Smart Metering Project, Smart Metering Electricity Customer Behaviour Trials 2009-2011 ," http://www.ucd.ie/issda/data/commissionforenergyregulationcer/, last accessed July 2014.
[30] GridLAB-D. U.S. Department of Energy at Pacific Northwest National Laboratory. http://www.gridlabd.org/.