IM2DR: Incentive Based Multi-tier and Multi-agent Approach for Demand Response in Electricity Market with Reinforcement Learning

Abbasloo A., Valenzuela G., Gebhardt G.H.W., Tomar R., Piesk J.
Nuromedia GmbH, Germany
gabriel.valenzuela@nuromedia.com

Abstract. Imbalance between consumption and supply in grids causes a challenge to electricity utilities, leads to grid instability and results in failures of the grid. Traditionally, the demand for power is satisfied by increasing supply which is costly and against environmental regulations. Instead, demand response strategy harmonises the demand at the customer side e.g the utilities offer rates to customers to reduce their consumption at specific times of the day. While most research focuses on finding a strategy on just one side in this work, we approach it with an end-to-end reinforcement learning based methodology such that optimal strategies for parties are achieved with training on historical records. Agents are trained for individual customers and the electricity utilities for finding the optimal policies for the consumption, hence manifesting a multi-agents and multi-tiers approach. The new method achieves the full potential of a proactive framework for implementing the demand response strategy.

1. Introduction

Residential districts in 2019 represented 26% of final energy consumption in the EU[1] and compensating on the electrical grid seems to be possible by new innovations in renewable energy. As well, energy storages can help align peaks of renewable energy generation with peaks of consumption. However, their integration and adoption into the grid infrastructure takes time and requires extensive investigations regarding the reliability and usability. Demand response (DR) strategy aims to create stability by making the demand side flexible and shifts peak demand by providing customers with economical offers. But forcing customers to delay the usage of home appliances, undesired temperature set-points and the effort needed for acquiring information about prices and the consumption patterns create discomfort and dissatisfaction. Reinforcement learning (RL) has been successfully adopted in energy resource management such as electric vehicles, heating ventilation and air conditioning (HVAC) systems and storage management. The challenge in front of DR strategy is about its ability to minimise customer discomfort and integrate their feedback. Model-free RL seems to be the potential methodology very adaptable to its environment and can directly integrate human feedback into the decision making with least user intervention. Most of the research studies that considered human comfort focus on single-agent systems with demand-independent prices Vazquez and Nagy (2019). Moreover, modelling the electricity prices as demand-dependent variables might lead to the risk of shifting the consumption peak instead of shaving it.

We propose IM2DR, a system based on RL to coordinate multi-agent systems participating in the DR program with demand-dependent prices. As discussion concerning implementing a DR program in the EU begins, data shows that the peak reduction from DR programs in the US was only 6.6% of the peak demand in 2015. Vazquez and Nagy (2019) argued that the reason is that electricity is a commodity whose value is way higher than its price for the customers. Electricity customers generally will not give up their comfort for a lower bill.

[1] eurostat: Statistics Explained

https://doi.org/10.7146/aul.455.c201

Therefore, the implementation of DR strategy depends on offering more convenient economical savings than discomfort of customers. Moreover, the framework is required to be both automated and able to minimise user discomfort as much as possible. As Vazquez and Nagy (2019) also provided an overview of different scenarios for implementing the DR strategy with RL, districts connected to a grid require some sort of dynamic coordination. IM2DR introduces an end-to-end solution, where the regulatory-tier (Service Providers agent) coordinates the automation-tier (CUstomer multi-agents). Assuming that the market is elastic, customers receive their hourly rate from the regulatory-tier and the automation-tier schedule the appliances. As a consequence, the demand curves are flattened and the consumption peaks are shaped during high demand time of the day. The CityLearn simulation environment, empowered by EnergyPlus, provided a framework to implement the method and train the agents on historical records. We achieved the optimal policies for the individual agents by decoupling the training procedure and using the Soft Actor Critic algorithm Haarnoja (2018). To our knowledge this is the first study that models the supply and demand together.

2. Literature Review

Vazquez and Nagy (2019) discussed that a group of districts where the consumption independently is controlled under demand-independent prices is not a multi-agent approach. Moreover, they explained that if considering demand-dependent prices, the actions of any district has impacts on the price of electricity, and as the decisions taken by others. Vazquez and Detjeen (2019) demonstrated a multi-agent RL schema for load shaping in a model-free and decentralised manner where the price of electricity increases linearly with the total electricity demand of all the districts. The multi-agent framework was improved by introducing MARLISA that uses a reward with individual and collective goals, and the agents predict their own future consumption and share this with each other following a leader-follower framework Vazquez and Henze (2020). It is also unclear how RL can control a multitude of energy systems in a scalable coordinated way. Hence, Park (2019) presented LightLearn, an automating system for district lighting and HVACLearn for HVAC system Park (2020). GridLearn considers grid into the account Pigott (2021).

Lu (2019) proposes a novel incentive based demand response algorithm for smart grid systems with RL, aiming to help the electricity utilities to purchase energy resources from its customers to balance energy fluctuations. Game theoretic methods and simulations based on RL are used to analyse electricity market equilibrium as well. Liang (2020) adopts a deep deterministic policy gradient algorithm to model the bidding strategies of utilities companies. However all of these studies just modelled the electricity market and none of them took optimization over demand (finding an optimal strategy for scheduling the appliances) into consideration.

CityLearn is an open source OpenAI Gym environment for testing RL for energy optimization Vazquez and Dey (2020). Its objective is to standardise the evaluation of RL agents such that algorithms can be easily compared. CityLearn allows the easy implementation of agents to change their demand aggregation by controlling storages of energy. Currently, CityLearn allows controlling the storage of domestic hot water, and chilled water, for sensible cooling and dehumidification. CityLearn also has models of air-to-water heat pumps, electric heaters, solar photovoltaic arrays, and the pre-computed energy loads of the districts from EnergyPlus simulations Crawley (2000) which include space cooling, dehumidification, appliances, domestic hot water, and solar generation. Refer to the CityLearn challenge Nagy (2021) for further information.

3. Problem Formulation

Modelling of the DR strategy and automating the customers electricity consumption requires understanding how different players interact in the market. Figure 1 shows an overview of the electricity market. Such, the service provider works as a middleman in the wholesale market with grid operators and in the retail market with customers. Therefore, the role of service provider can be represented as an agent, namely the SP agent which regulates the consumption of automation-tier, where CU multi-agents operate. CU multi-agents automate the energy consumption of individual districts by scheduling the appliances and storages.

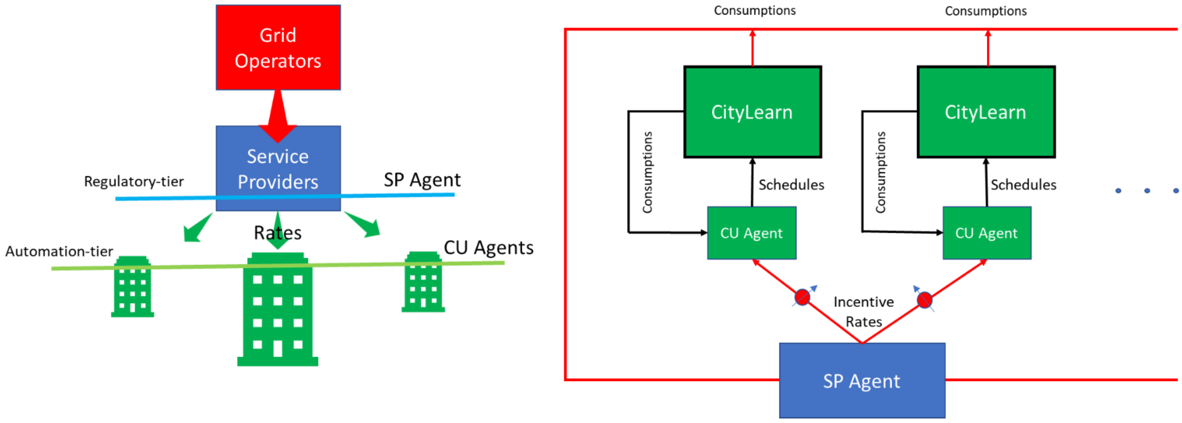

Figure 1: Left, the schematic of the electricity market. Right, demonstration of the system methodology.

DR strategy is an optimal policy in which the SP agent proactively coordinates the consumption of districts by offering some type of discounts in the form of incentive rates for convincing them to reduce the consumption and CU multi-agents schedule the entire appliances in order to satisfy the residents comfort and assuring the overall consumption is optimised using storages and domestic power generation. Lu (2019) introduced an RL based method for modelling the retail market and we will follow a similar approach in this paper. The parameters for modelling the retail market and implementing the incentive-based DR strategy is listed in Table 1. CU multi-agents are modelled in CityLearn as it is described in Vazquez and Dey (2020). The parameters for the simulation have been chosen in such a way that each district has a different attitude in the DR program.

3.1 Service Provider Model

The SP agent offer $\lambda_{n,h}$ such that the costumer n at time h needs to drop the curtailable consumption by $\Delta E_{n,h}^{curt}$ therefore the goal of the SP agent is to maximise its gain by:

$$max \sum_{n=1}^{N} \sum_{h=1}^{H} (p_h . \Delta E_{n,h}^{curt} - \lambda_{n,h} . \Delta E_{n,h}^{curt}) \text{ s.t. } \lambda_{min} < \lambda_{n,h} < \lambda_{max}$$

3.2 Customer's Regulatory-tier Model

The customers aim to maximise their outcome from the trade with the SP agent by finding a good balance between the gain from the incentive package and discomfort $\varphi_{n,h}$ by:

$$max \sum_{h=1}^{H} (\rho.\lambda_{n,h}.\Delta E_{n,h}^{curt} - (1 - \rho).\varphi_{n,h}(\Delta E_{n,h}^{curt}))$$

where $\varphi_{n,h}(\Delta E_{n,h}^{curt}) = \mu_n/2(\Delta E_{n,h}^{curt})^2 + \omega_n \Delta E_{n,h}^{curt}$.

The consumption drops are calculated as follows defined with a set of parameters characterising the customers and the market.

$$\Delta E_{n,h}^{curt} = E_{n,h}^{curt}.\xi_h.\frac{\lambda_{n,h}-\lambda_{min}}{\lambda_{min}} \text{ s.t. } K_{min} < \Delta E_{n,h}^{curt} < K_{max}$$

3.3 Objective Function for Regulatory-tier

The SP agent needs to find an optimal policy such that it maximises the gains of both sides of the trade by:

$$max \sum_{n=1}^{N} \sum_{h=1}^{H} (p_h.\Delta E_{n,h}^{curt} - \lambda_{n,h}.\Delta E_{n,h}^{curt} + \rho.\lambda_{n,h}.\Delta E_{n,h}^{curt} - (1 - \rho).\varphi_{n,h}(\Delta E_{n,h}^{curt}))$$

3.4 Objective Function for Multi-agents Automation-tier

CU multi-agents minimise the consumption but this time integrating the SP agent contributions by introducing $\Delta E_{n,h}^{curt}$ and $\Phi(\lambda_{n,h}, \Delta E_{n,h}^{curt})$ that may reflect, e.g., the customer's discomfort in the objective:

$$min \sum_{h=1}^{H'} (p_h.(E_{n,h} - \Delta E_{n,h}^{curt}) + \Phi(\lambda_{n,h}, E_{n,h}^{curt}))$$

While CU single-agents schedule the home appliances by minimising the very first term. The action and state space are defined in CityLearn with the simulation episode of a year.

4. Reinforcement Learning for Achieving the Optimal Policy

Showing that each tier's contribution in the electricity market is a Markov Decision Process (MDP) and therefore can be modelled by RL methodology has been discussed by Lu (2019) and Vazquez and Dey (2020) in detail. Therefore, the optimal policy for the market consists of a policy for the regulatory-tier (SP agent) which implements the DR strategy as well as individual optimal policies of the CU agents that minimise the consumption and are trained against the SP agent's policy. Because of the decoupled nature of the problem, we can obtain the policies for the SP agent and the CU agents independently by training on the historical records using an off-policy algorithm like Soft Actor Critic (SAC). SAC optimises a stochastic policy to maximise a trade-off between expected return and entropy, a measure of randomness in the policy Haarnoja (2018).

4.1 Multi-tier Training Procedure

In order to train the agents from different tiers, we take advantage of the decoupled nature of the problem such that the SP agent is trained with an episode of a month (the SP agent is aware of which month it is) on the historical consumption records from the districts in which CU single-agents schedule the consumption. Training single-agents is carried out in the CityLearn environment as well. CU multi-agents are later trained by integrating already-trained SP agent in the training loop such that any district participating in the DR program needs to be aligned with the policy of the SP agent as depicted in Figure 1. We remark that the entire training and integration of the SP agent is carried on curtailable consumption.

Table 1: List of parameters for modelling the SP agent.

Parameters	Definition
p_h, $\lambda_{n,h}$, λ_{min} and λ_{max}	Electricity price, incentive rate for customer n at time h and its bounds
$E_{n,h}$, $E_{n,h}^{curt}$,$\Delta E_{n,h}^{curt}$, K_{min} and K_{max}	(curtailable) Consumption for customer n at time h, its drop and bounds
N, H, ρ and ξ_h	Number of districts, simulation length, trade-off and elasticity at time h
$\varphi_{n,h}$, μ_n and ω_n	Discomfort of customer n at time h with the flexibility and the attitude parameters

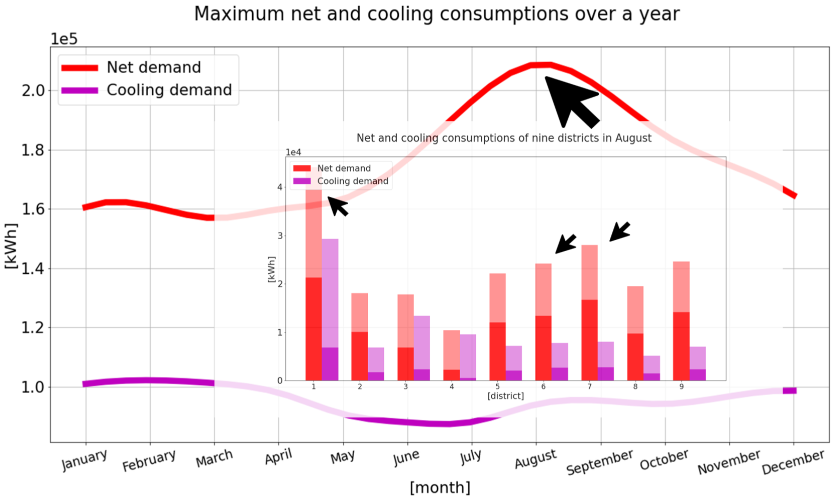

Figure 2: Consumption of all districts over a year. Bar chart shows consumption ranges over samples in August.

5. Simulation and Result

The SP agent is modelled in OpenAI Gym and later integrated in the CityLearn environment for training CU multi-agents. The CityLearn environment provides nine different districts with different architecture, thermal behaviour and occupancies. Several climate zones exist as well. The thermal behaviour is modelled in EnergyPlus and is given as input to CityLearn. We run the algorithm of each stage five times to evaluate the statistical properties of the agents. In the following sections, we provided plots of medians, minimums or maximums of the generated samples at time h to explain the different expected outcomes of the DR program compared with single-agents performance.

5.1 Configuration and Scenarios

Figure 2 shows the maximum net consumption of nine districts over a year. To evaluate the DR program, we want to see if we can flatten the consumption peak in August. Moreover, for the purpose, we choose climate zone five where the curtailable consumption is the cooling demand such that the DR program requests participants drop the cooling consumption. Imitating the customers dropping their cooling consumption is implemented by subtracting the amount from the cooling demands calculated by EnergyPlus. As Figure 2 depicts the consumption range in August, we choose district one, six and seven with high consumption participating in the DR program in July, August and September. The rest will join the first and second half of the year according to Table 2 to evaluate the RL inherent bias towards maximising short term rewards Vazquez and Henze (2020).

Table 2: Participation in the DR program for simulations.

Districts	Participation program
One, six and seven	July, August and September
Two, three and four	The first half of the year
Five, eight and nine	The second half of the year

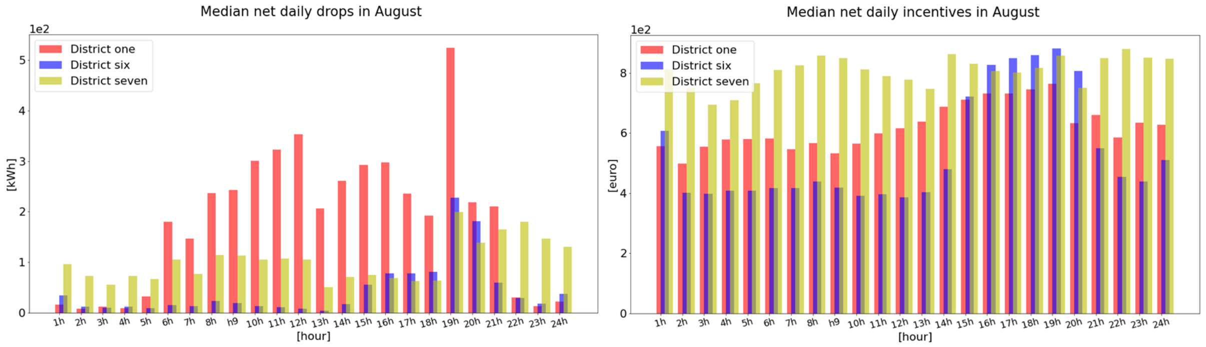

Figure 3: Daily drops and incentives statistics over several free rollouts of the SP agent.

5.2 Training and Evaluating of the SP Agent

In order to show the statistics of the optimal policy of the SP agent, we plotted the hourly net incentives and drops over several rollouts from the SP agent in Figure 3. Plots show the nontrivial statistics of the policy which has different behaviors for districts reflecting the impact of parameters characterise districts in DR programs e.g. district one has tendency to drop more for receiving less rates.

5.3 Integration of the SP Agent and Performance of CU Multi-gents

The SP agent impacts the multi-agents environment by introducing consumption drop $\Delta E_{n,h}^{curt}$ and $\Phi(\lambda_{n,h}, \Delta E_{n,h}^{curt})$ which is assumed to zero for simplicity. The SP agent takes action at time n as a form of offering incentives and will be informed about the realised drops at time n which are the state of the environment under the corresponding action. The realised drops are calculated based on the difference between the actual and expected curtailable consumption. The expected curtailable consumption needs to be predicted in the best scenario but here for simplicity, we calculated it by creating a template by taking maximum over generated samples obtained from repeating single-agents training. Figure 4 gives a good comparison of the consumption of nine districts in August and the metrics introduced in CityLearn e.g. see the arrows for changes of the range over samples. Moreover, the daily and monthly drops for district one, six and seven over a year and in August. The plots illustrate the maximums over the samples interpreted as the worst outcome.

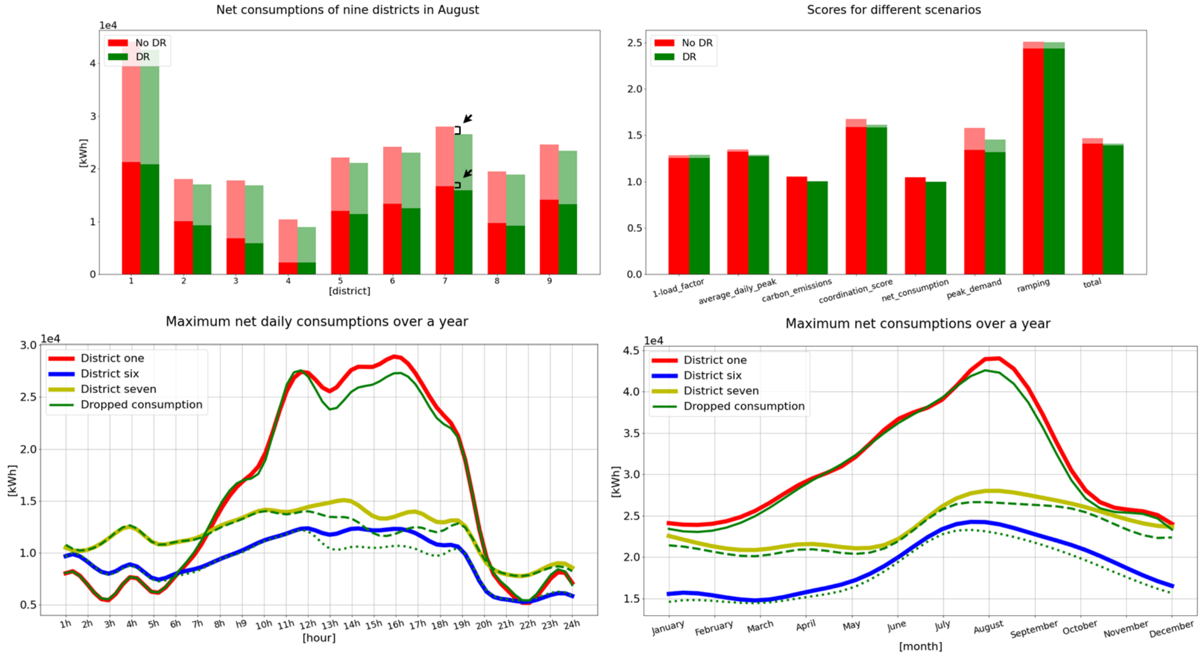

Figure 4: CU multi-agents performance compared to single-agents. Top, depicts some indicator’s ranges. Bottom, presenting the consumption in different time frames.

5.4 Statistics and Insights

To have more insights about the SP agent, we calculated the histogram of free rollouts drawn from the SP agent. We also calculated the same free rollouts statistic when the states for all districts are forced to zero except district one (shown by arrow in Figure 5). Comparing them shows that the policy for each district proactively changes based on the contributions of the others. Figure 5 also shows those statistics when the SP agent is integrated with CU multi-agents. The statistics completely changed and the minimums over the samples tended to cover the whole range meaning the SP agent adapted to the task. We evaluate CU multi-agents by comparing the unrealised and realised daily drops presented in Figure 5. The unrealised drops are obtained by the free rollout of the SP agent and the consumption scheduled by CU single-agents. We can interpret it as such that CU multi-agents’ policies seem to integrate well with the SP agent’s policy after going through a round of training.

Figure 6 shows that the districts in the second half of the year program are able to maintain the consumption reduction despite pressures from previous months therefore, IM2DR seems not to be inherently biased towards consuming less energy now at the expense of consuming more later. For those in the first half of the year program, the reduction propagates to the entire year supporting the same claim as well. Even though the consumption are characterised to be less elastic in the middle of the day the net daily consumption over the year show a significant reduction during this time across all participants showing that the IM2DR tends to flatten the high peak hours, daily plots in Figure 4 and 6.

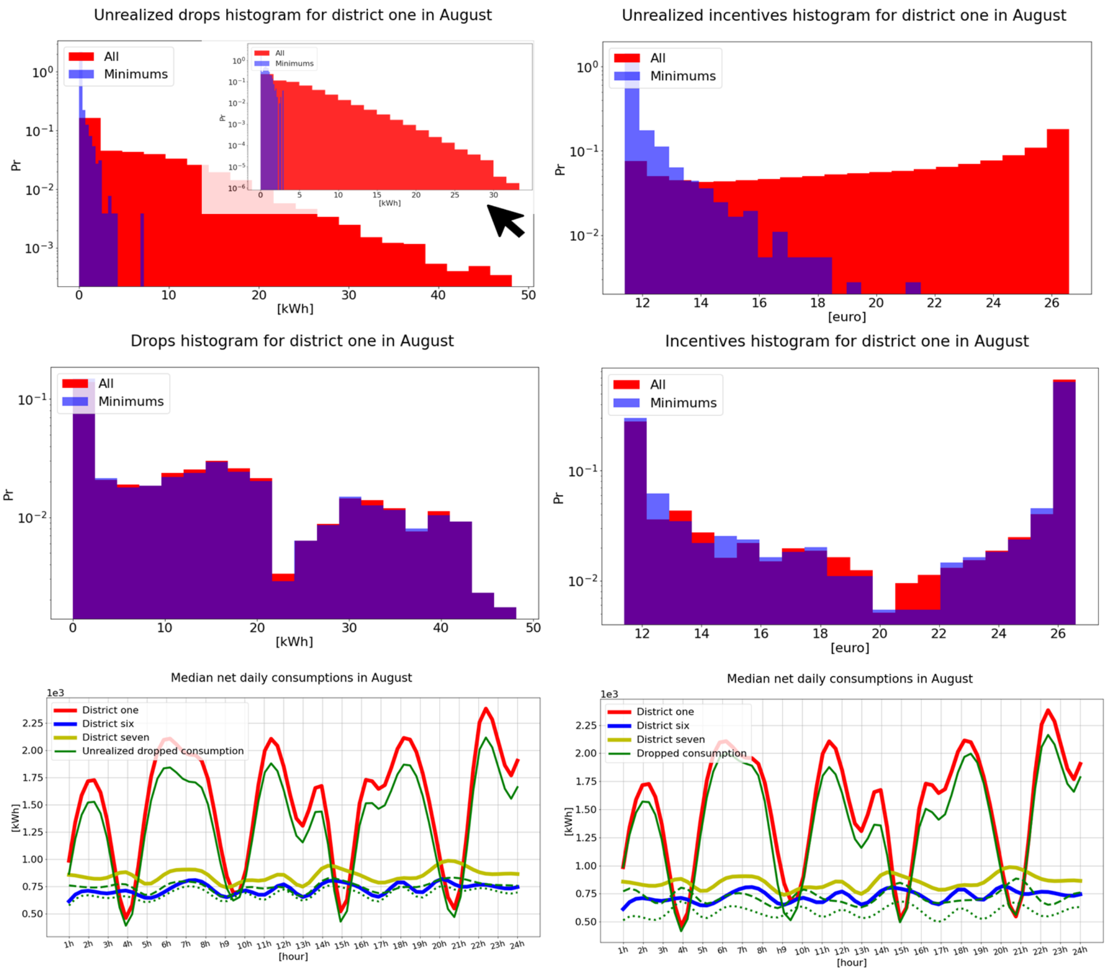

Figure 5: Top, histograms of drops and incentives for the SP agent in different scenarios. Bottom, consumption with CU single and multi-agents.

6. Guideline for Real Life System Implementation

We proceed now to the real life implementation of IM2DR. As depicted in Figure 7, the service provider collects the consumption records of all districts over time and (re)trains the SP agent. At the beginning of the year, each customer receives a deal concerning the offers via the system as a message shown in the UI. Individuals can see the possible options of the consumption reductions in the UI e.g. shifting the cooling setpoints with two degrees can fulfil the deal. After agreeing partially on the offer, the service provider receives them. Then CU multi-agents will go through (re)training with considering the agreement, and schedule the appliances for the year.

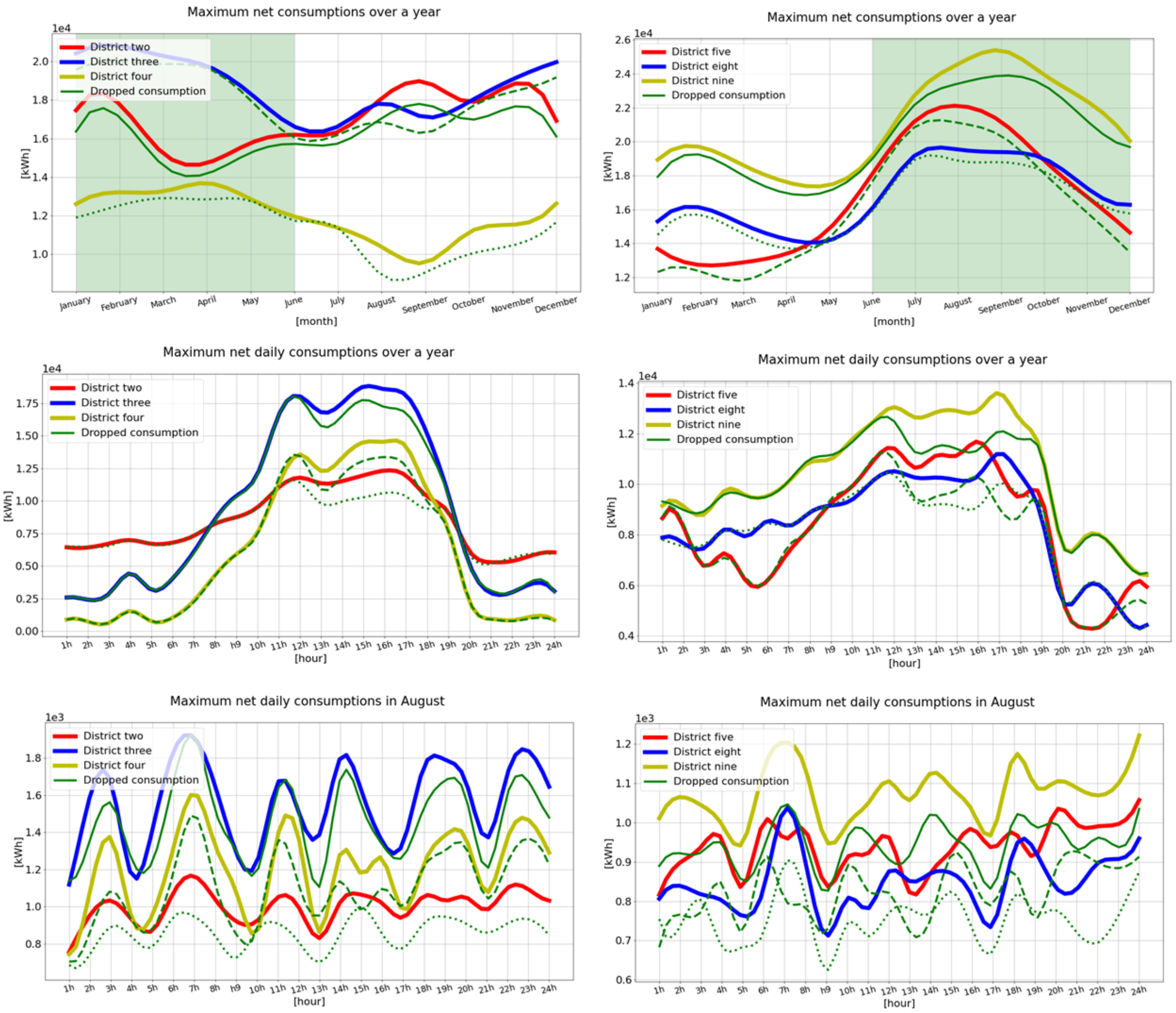

Figure 6: Consumption comparison of districts participating in the half of the year DR program.

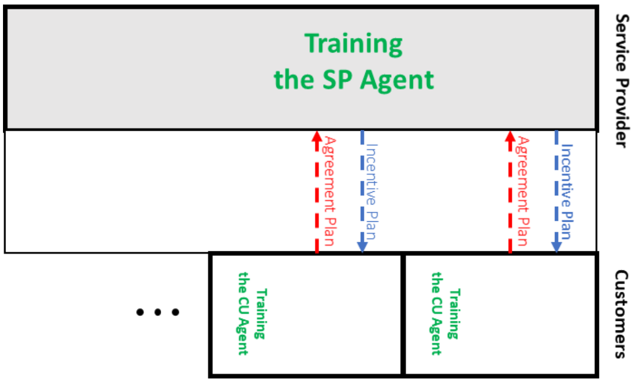

Figure 7: Real life implementation of the system.

Meanwhile the service provider needs to do some maths over the deals for estimating the expectations. It becomes challenging when dealing with agents with stochastic policies. If you collect a good amount of run sessions, by looking at consumption range over all samples at a specific time, following the procedure in Figure 8, one gets a rough estimate of the consumption expectation bounds e.g. see the arrows for August. In this way, a baseline can be estimated for the DR program.

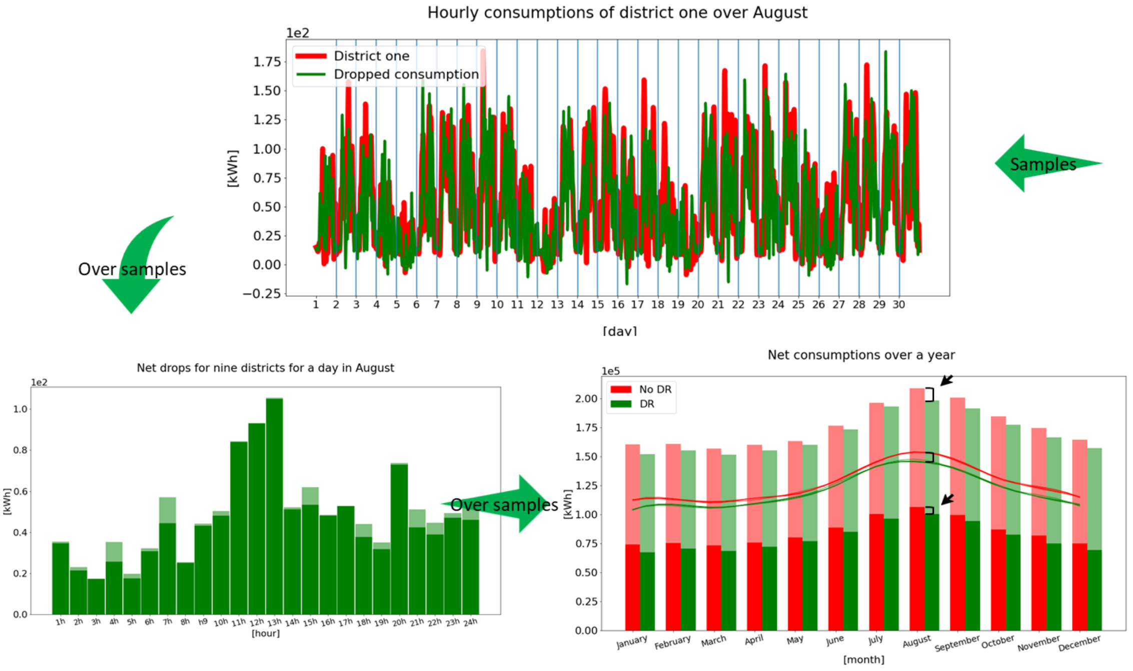

Figure 8: Calculating DR program expectation's outcomes for the service provider.

7. Conclusion and Future Work

We have introduced an end-to-end system for implementing DR strategy that lets the customer's automation agents schedule the appliances while the service provider agent harmonises the market by creating balance between demand and response. On the other hand we obtained a set of optimal policies with which the service provider proactively coordinates the consumption and customers in the program schedule their appliances. The customers are modelled by a set of parameters in the DR program therefore, for future work, it is the very first step to investigate the external realisation of those parameters related to the comfort of the residents in terms of heating, cooling and air conditioning. Overall such an automated decision making is more sensible if customers also receive the offers during the year in form of monthly, weekly or daily offers. This leads to retraining CU multi-agents for the rest of the year achieving a policy that is suboptimal for the entire year but is an optimal policy for the rest of the year. And integrating different SP agents optimised with different time scales e.g. one for daily and one for weekly offers. Moreover, looking into alternative rewards can show more insights since the incentive and reductions contain the MDP states of the other agents so it helps the agents to reach a better policy during training.

Acknowledgement

This research has been financially supported by PRECEPT, an EU Horizon 2020 project registered with the topic: LC-EEB-07-2020, the type of action: IA and the proposal number: 958284. The first author would like to thank the ML team in Nuromedia GmbH and other industrial partners in the project for their collaboration.

References

Vázquez-Canteli, J.R. and Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modelling techniques. Applied energy, 235, pp. 1072-1089.

Park, J.Y., Dougherty, T., Fritz, H. and Nagy, Z. (2019). LightLearn: An adaptive and occupant centred controller for lighting based on reinforcement learning. Building and Environment, 147, pp. 397-414.

Park, J.Y. and Nagy, Z. (2020). HVACLearn: A reinforcement learning based occupant-centric control for thermostat set-points. In Proceedings of the Eleventh ACM International Conference on Future Energy Systems, pp. 434-437.

Vazquez-Canteli, J., Detjeen, T., Henze, G., Kämpf, J. and Nagy, Z. (2019). Multi-agent reinforcement learning for adaptive demand response in smart cities. In Journal of Physics: Conference Series, Vol. 1343, No. 1, p. 012058. IOP Publishing.

Vazquez-Canteli, J.R., Henze, G. and Nagy, Z. (2020). MARLISA: Multi-agent reinforcement learning with iterative sequential action selection for load shaping of grid-interactive connected buildings. In Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, pp. 170-179.

Pigott, A., Crozier, C., Baker, K. and Nagy, Z. (2021). GridLearn: Multiagent Reinforcement Learning for Grid-Aware Building Energy Management. arXiv:2110.06396 [cs.MA].

Vázquez-Canteli, J.R., Dey, S., Henze, G. and Nagy, Z. (2020). CityLearn: Standardising Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management. arXiv:2012.10504 [cs.LG].

Nagy, Z., Vázquez-Canteli, J.R., Dey, S. and Henze, G. (2021). The citylearn challenge 2021. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, pp. 218-219.

Wang, J., Guo, C., Yu, C. and Liang, Y. (2022). Virtual power plant containing electric vehicles scheduling strategies based on deep reinforcement learning. Electric Power Systems Research, 205, p.107714.

Lu, R. and Hong, S.H. (2019). Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Applied Energy, 236, pp. 937-949.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P. and Levine, S. (2018). Soft Actor-Critic Algorithms and Applications. arXiv:1812.05905 [cs.LG].

Crawley, D.B., Lawrie, L.K., Pedersen, C.O. and Winkelmann, F.C. (2000). EnergyPlus: Energy simulation program. ASHRAE Journal, 42(4), pp. 49-56.