Supplementary of Paper #1198

Contents

1 ADL Scheduling

ADL (Activities of Daily Living) are tasks that have to be fulfilled within a certain time frame, instead of requiring power immediately. This can be further illustrated by the following example: Imagine an office goer who leaves at 9 a.m in the morning and returns at 5 p.m. in the evening. The office goer wants her clothes to be washed before she comes back. With the advent of smart washing machines, it is possible to schedule a task that has to be completed within the stipulated time frame. The washing machine then requests the microgrid to provide adequate energy to fulfill the task. Depending on other factors the microgrid decides to fulfill this request or defer it to a different time within the stipulated time frame. As time passes, the priority of the request increases and ultimately the washing machine is provided with the required energy. ADL scheduling does not reduce the total energy consumption, it merely shifts the peak load, thereby preventing the microgrid from being overloaded.
In our experiments, we have experimented with cases where the ADL demands are fixed for each day as well as ADL tasks that vary on a day to day basis. These ADL demands are specified at the start of each day to each microgrid. The configuration for the experiments we have run are as follows:

a. Configuration for fixed ADL demand: For the fixed ADL demand setting, 3 ADL tasks have to be satisfied by the microgrid. The amount of electricity required to fulfill the ADL task as well as the time period within which the task must be completed is combined together in the form of a tuple. The configuration for the fixed ADL demand is as follows : [(1,2),(1,3),(2,4)], where the first term in the tuple represents the amount of electricity required and the second term represents the time interval by which the given amount of energy has to be fulfilled.

b. Configuration for the Variable ADL demand: For the variable ADL demand, the microgrids receive 0 to 3 ADL tasks at the start of the day. The number of ADL tasks as well as the time period by which these tasks have to be completed is determined by a probability matrix. The amount of electricity required for each of these ADL tasks can be one of these 3 units: 0, 1 or 2. Similar to the above case, the amount of electricity required to fulfill the ADL task as well as the time period within which the task would have to be completed is combined together in the form of a tuple.

Binary encodings for the ADL tasks: For easy representation of the ADL tasks that have to be fulfilled, a binary representation is used. In our binary representation, 1 implies that that particular task has to be fulfilled and 0 implies that that particular task is not to be considered at that particular time step. An example of this is as follows: Let's assume that the first and third ADL tasks have to be completed. The binary representation of this would be 101 as the microgrid would be interested in fulfilling only the first and 3rd ADL task at that particular time step. This binary representation is fed into the network in a decimal format (which in the example highlighted above would be 5). Moreover, the binary representation also takes care of the time period within which the tasks would have to be fulfilled by ordering the ADL tasks in a sequential manner, i.e, the first digit of the binary tasks corresponds to the fourth time step. Similarly, the second digit corresponds to the second time step and so on.
The ADL State in the ADL Network is a decimal format representation of the ADL tasks remaining for the day. The ADL network picks an action that is present within the superset of the ADL tasks remaining throughout the day and passes that information to the ET Network in a decimal format. The ET network then uses this information for further decision making.

2 Neural Network Model

Here we explain the exact architecture of the ADL network and the Energy trading network:

2.1 ADL Network Architecture

1. Structure:
    - Zeroth Layer: Five input States: ND, D, T, ADL State, Grid Price
    - 1st Layer: 16 Neurons
    - 2nd Layer: 16 Neurons

- 3rd Layer: Outputs: 8 outputs, since the maximum ADL jobs we consider here is 3 in both fixed ADL and variable ADL case.

2. State Space
   - ND is the net demand i.e it is the cumulative sum of the battery and the energy generated
   - D is the local consumer demand
   - T is the current time step
   - ADL state signifies the number of ADL demands that have to be fulfilled for the day
   - GP is the grid price

3. Action Space
   - The model outputs the ADL loads that have to be fulfilled in the current time step

4. The network minimizes the following loss function:

$$L(\lambda) = ((R + \gamma * \max_{adl}(Q_{adl}(s_{adl+1}, adl_t^i|\lambda))) - Q_{adl}(s_{adl}, adl_t^i|\lambda))^2.$$

   We use Adam optimizer with learning rate 0.0001, β_1=0.9, β_2=0.999 and ϵ=10^{-8} to update network weights. The discount factor γ is kept at 0.9.

2.2 Energy Trading Network

1. Structure:
   - Zeroth Layer: Five input States: ND, D, T, ADL Action, Grid Price
   - 1st Layer: 32 Neurons
   - 2nd Layer: 32 Neurons
   - 3rd Layer: Outputs: The number of Neurons in the output layer is given by the following formula: *(max battery + max energy generated) * 6 + max energy that can be received + 1*

2. State Space
   - ND is the net demand i.e it is the cumulative sum of the battery and the energy generated
   - D is the local consumer demand
   - T is the current time step
   - ADL action signifies the amount of ADL load that the previous model has decided to fulfill in the current time step
   - GP is the grid price

3. Action Space
   - The model outputs the units of electricity to be bought or sold
   - The model also outputs a price between gp - 5 and gp if it is selling, else it outputs a price of 0.

4. The network minimizes the following loss function:

$$L(\theta) = ((R + \gamma * \max_{et}(Q_{et}(s_{et+1}, p_t^i, u_t^i|\theta))) - Q_{et}(s_{et}, p_t^i, u_t^i|\theta))^2.$$

   We use Adam optimizer with learning rate 0.0001, β_1=0.9, β_2=0.999 and ϵ=10^{-8} to update network weights. The discount factor γ is kept at 0.9.

3 Discussion regarding the reward formulation

The reward is computed as follows:

$$
\begin{aligned}
r_t^i = u_t^i * p_t^i - k1 * &(\text{unfulfilled Non-ADL demand}) \\
&-k1 * (\text{unfulfilled ADL demand}),
\end{aligned}
$$

Here, k1 is a positive constant. Changing the values of k1 leads to the the microgrids exhibiting different behaviors. When u_t^i is much larger than $k1$, the microgrids favour selling energy as compared to satisfying their local consumer demands (both Non-ADL as well as ADL demand). Conversely, when k1 is much larger than u_t^i, the microgrids prefer to satiate their local consumer demands (both Non-ADL as well as ADL) as compared to selling energy. This can be explained as follows: each microgrid is tasked with optimising its reward function. By changing k1, the weights given to selling energy and satisfying local consumer demand changes, which in turn changes the reward function.
To emulate a real world scenario, we have given a higher importance to satisfying local consumer demand as compared to selling energy. In our experiments, k1 is set to 30.

4 Comparing the Dynamic Pricing Policy (DPP) & the Constant Pricing Policy (CPP)

From the results of our experiments (specifically Table 1 and Table 2 in the main paper) we observe that all of the microgrids don't receive higher rewards while following the DPP. Through this subsection, we try to provide intuitive justifications for why this occurs. The justifications are as follows:

(a) **Formulation as a Stochastic Game:** We have formulated the problem of energy trading amongst microgrids as a stochastic game where each microgrid's actions end up affecting the rewards of other microgrids and each microgrid is competing to maximise its own rewards. Hence, if all the microgrids obtain a higher profit when the policies are changed, it would imply that cooperation is occurring amongst microgrids in order to optimise a conjoined reward. However that is not the case. Therefore, by using contradiction, we intuitively justify that when the inherent behaviors of the microgrids are changed (changing CPP to DPP), some microgrids would receive a benefit and some would not. This is further justified by the point below.

(b) **Changing polices affects energy trading:** In both the cases (following the DPP or the CPP), the stochastic process used for simulating energy generation is same for both cases. Hence the only difference that occurs between the two scenarios is in the actions the microgrids take while selling.
When microgrids follow the CPP, all of them sell at the central grid price (GP). However, whenever they sell, all the microgrids obtain a reduced profit, due to a constraint we have imposed, i.e., if two or more microgrids decide to sell energy at the same price, the reward obtained by the selling microgrids is proportional to the energy they wish to sell. This is to ensure a fair transaction occurs amongst all selling microgrids. Therefore if the price being quoted is the same, (which majorly occurs while following the CPP), the highest profit would be obtained by the microgrid which decides to sell the maximum amount of energy. Since the reward also consists of a component that promotes satisfaction of local consumer demand (which is constant across all microgrids on an average), it is clear that the microgrid that generates the highest energy would be able to sell more energy. However, when the microgrids follow the DPP, the microgrids have the ability to quote prices in a price range of [gp-k, gp]. Therefore, each microgrid can adapt to the actions of the other microgrids. Moreover, due to the fact that the prices selected are dynamic, the profits they obtain are not reduced like they are while quoting the same price. This leads to an increase in rewards amongst certain microgrids.

4.1 Discussion regarding the difference in the magnitudes of rewards obtained by following the DPP and the CPP

The differences in the magnitudes of rewards while comparing the DPP and CPP, seem marginal in our experiments, because:

(a) The rewards formulated don't encompass the profits obtained by selling electricity to the local customers.

(b) The rewards in our formulation are scaled down versions of the actual rewards that the microgrids would receive in a real world scenario.

In reality, the energy traded amongst micrigrids is in multiples of terawatts, and the price at which electricity is traded is in multiples of thousands of dollars. Hence, due to the scale of the transaction, we believe that each decimal point becomes important.

5 Results for setting where the ADL demand is fixed

Here we present all the results and analysis of the experiments when the ADL demand is fixed. Please note that the 'λ-values Per Time Step' present in Table 1, 2 and 3 signifies the mean of the Poisson Distribution that has been fit to the data at that particular time step. (For further illustration, t_1 signifies that the mean of the Poisson Distribution that was used to fit the data in the first time step.)

5.1 Setup 1

The configurations of the different microgrids is provided in Table 1. From Figure 1 we see the judicious scheduling of ADL demands by the microgrids

Microgrids	**λ-values Per Time Step**			
	t_1	t_2	t_3	t_4
1	2.667e-07	0.541	6.5965	4.3712
2	8.8281	10.2997	9.8301	9.7514
3	8.8281	10.2997	9.8301	9.7514

Table 1: Configuration of microgrids under Setup 1

5.2 Setup 2

The configurations of the different microgrids used in this setup is provided in Table 2:

Microgrids	**λ-values Per Time Step**			
	t_1	t_2	t_3	t_4
1	5.0672	4.4150	5.0967	5.1631
2	4.5686	4.9671	4.9942	4.6172
3	3.3972	3.6215	3.3983	4.4020
4	2.6044	2.4534	3.6837	2.6472
5	0.0201	1.7532	7.8576	3.6710
6	0	0.7350	8.6901	5.7239
7	0	0.5410	6.5965	4.3712
8	0	0	6.0057	3.6077

Table 2: Configuration of microgrids under Setup 2

- From Figure 2 we can conclude that all 8 agents following our policy have converged to some good policy that is better than the constant pricing for most grids and random exploration for all grids.
- From Figure 3 we see that all the microgrids following our proposed algorithm learn to schedule the ADL jobs at different times which shows that our model is capable of shifting power consumption from the peak demand time. From Figure 4 we see the dynamic nature of the prices decided by various microgrids at different time intervals.

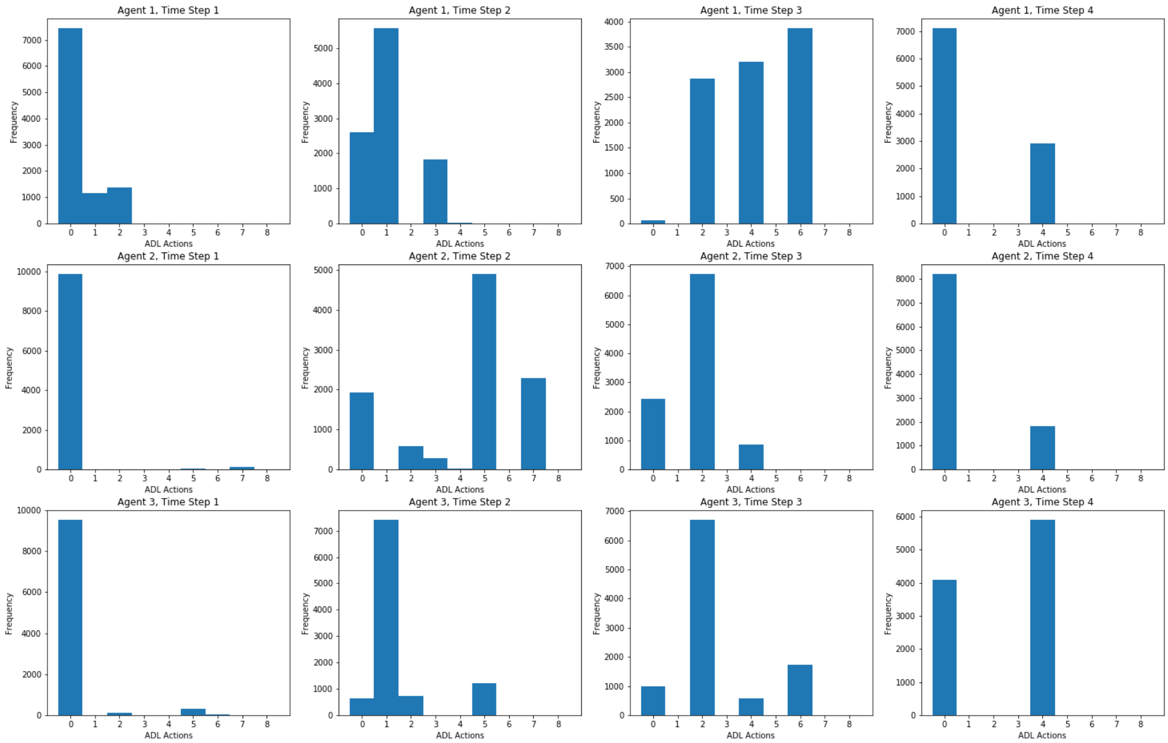

Figure 1: ADL action performed by the 3 agents at different time steps under Setup 1

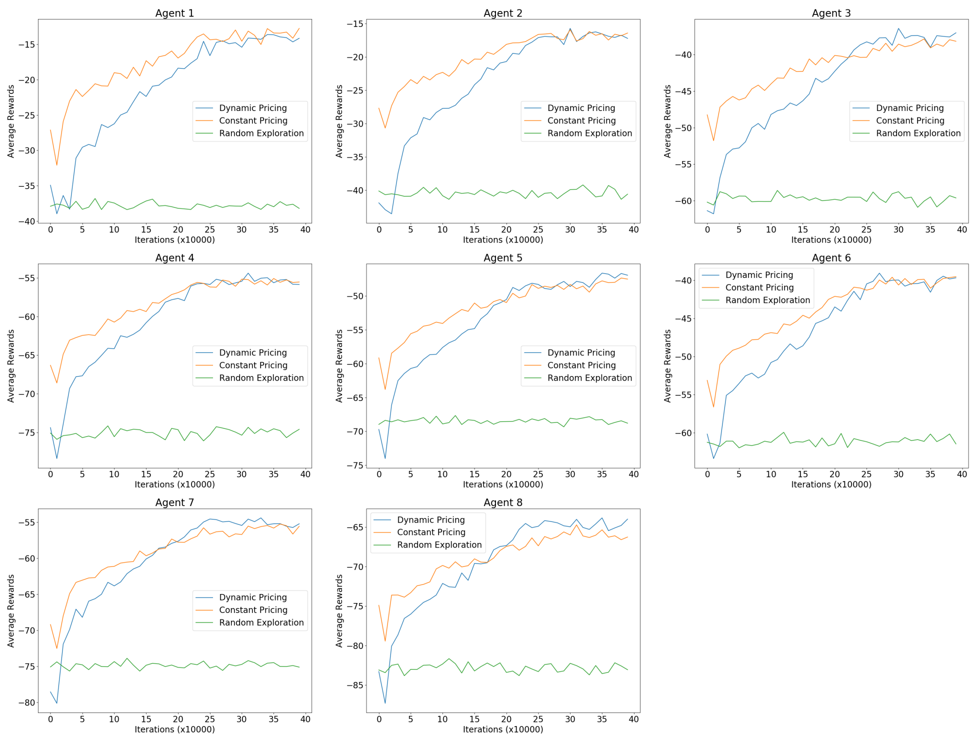

Figure 2: Convergence of all the agents under Setup 2

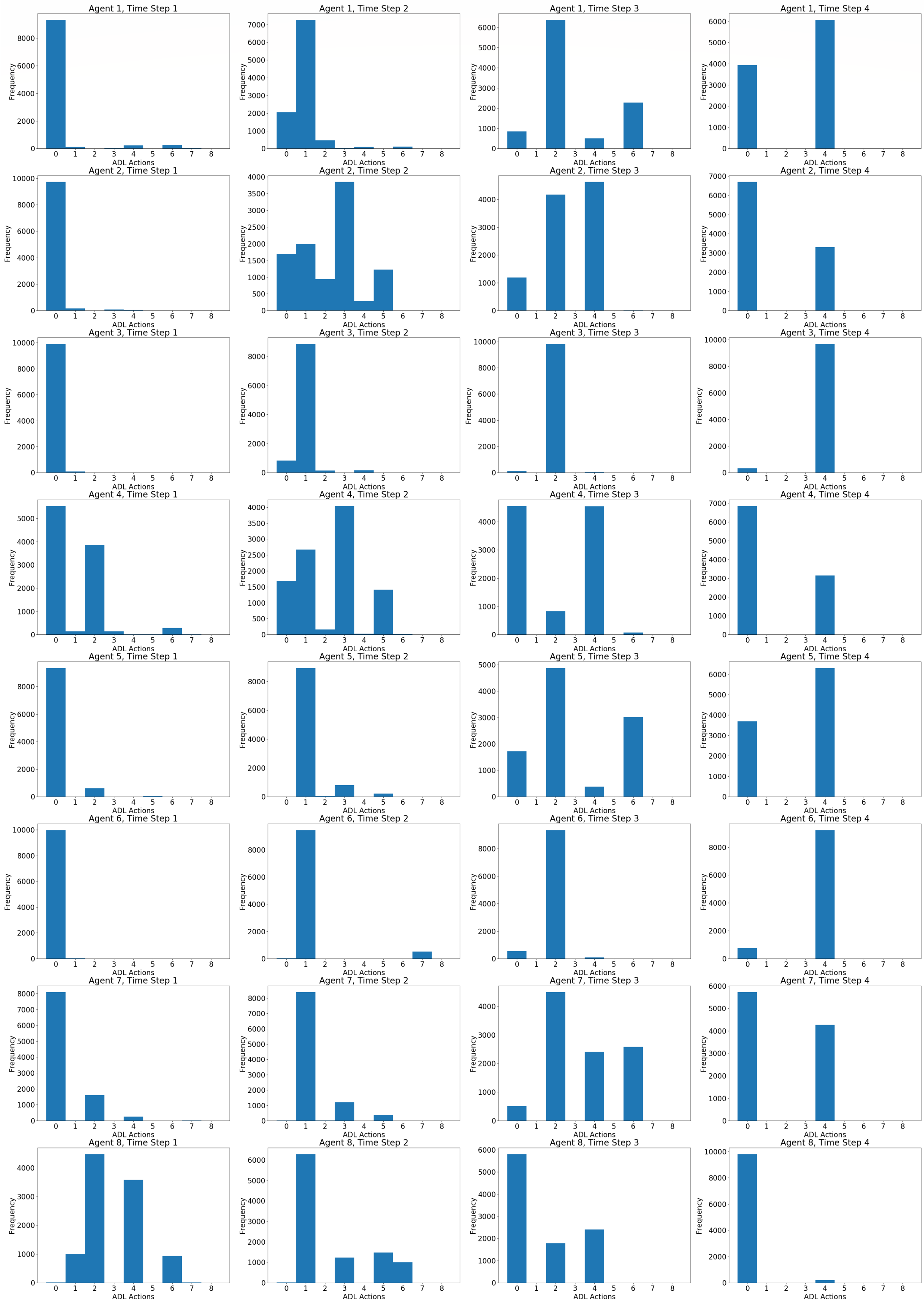

Figure 3: ADL action performed by the 8 agents at different time steps under Setup 2

Figure 4: Dynamic Pricing by all the 8 agents at different time steps under Setup 2

5.3 Setup 3

The configurations of the different microgrids used in this setup is provided in Table 3:

Microgrids	λ-values Per Time Step			
	t_1	t_2	t_3	t_4
1	5.0861	5.6614	5.1954	4.9762
2	5.2135	5.1766	6.3450	5.6715
3	6.4785	6.3222	6.1513	6.1295
4	6.0991	5.8984	6.7075	7.5607
5	3.3972	3.6215	3.3983	4.4020
6	0.0201	1.7532	7.8576	3.6710
7	0	0.7350	8.6901	5.7239
8	4.5686	4.9671	4.9942	4.6172

Table 3: Configuration of microgrids under Setup 3

- From Figure 5 we can see that all the 8 agents have converged to some good policy in both constant pricing and dynamic pricing case and it is clear that in most of the cases the microgrids which follow our dynamic pricing model perform better than those following the constant pricing one.
- From Figure 6 it is clear that all the agents learn to schedule the demands judiciously.
- From Figure 7 we again see the dynamic nature of pricing decided by various microgrids. The microgrids learn to sell judiciously at different time steps. From the graphs, we observe that solar grids sell at time steps 3 and 4 which is intuitive as the solar generation is maximum at these time steps.

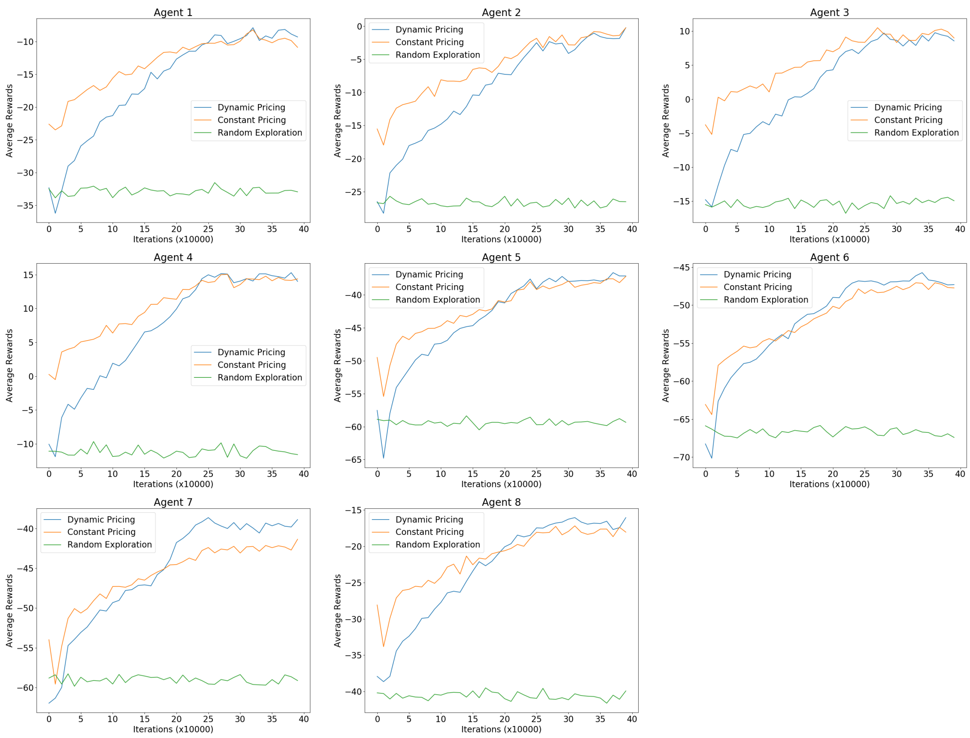

Figure 5: Convergence of all the agents under Setup 3

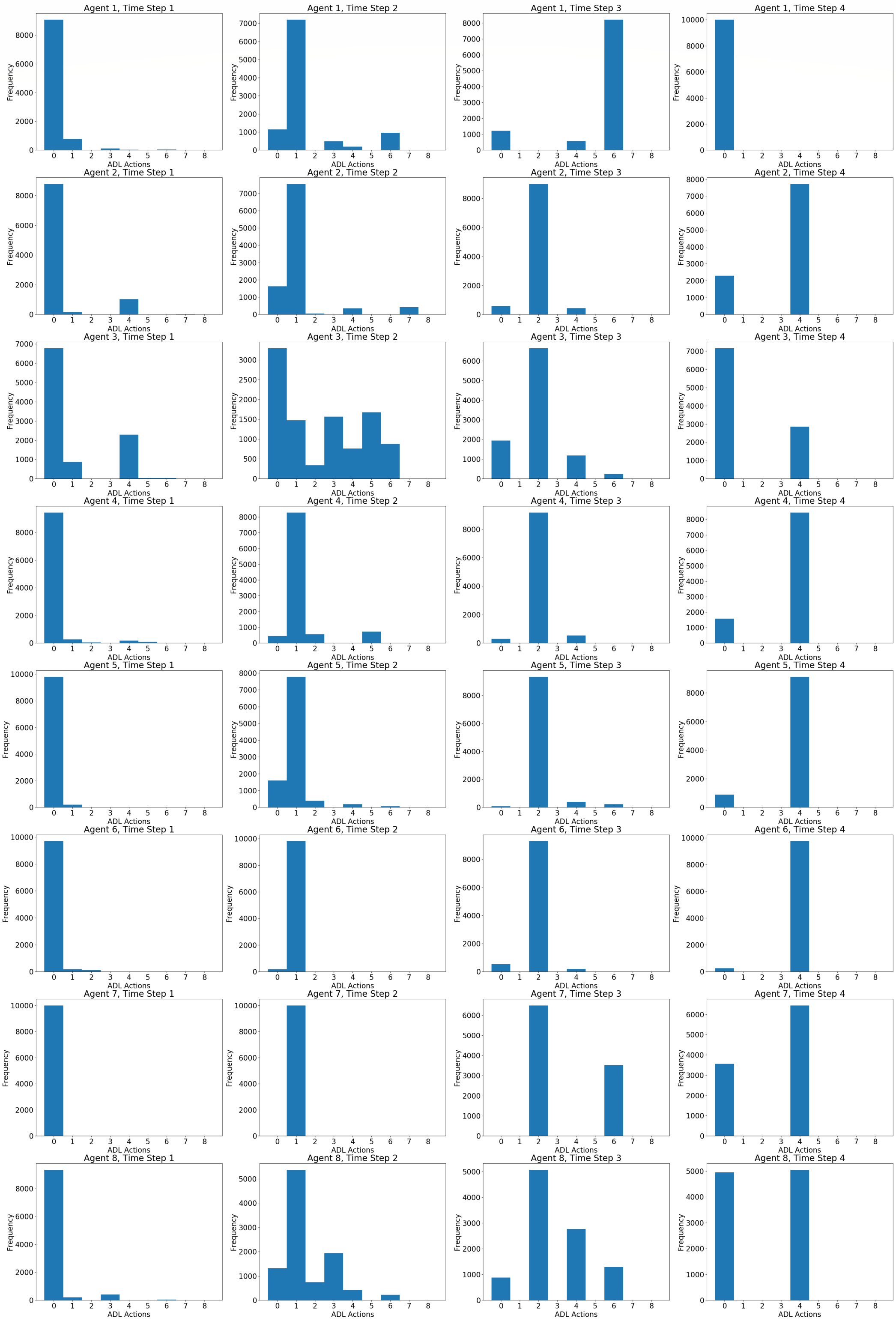

Figure 6: ADL action performed by the 8 agents at different time steps under Setup 3

Figure 7: Dynamic Pricing by all the 8 agents at different time steps under Setup 2

6 Results for setting where the ADL demand is variable

Here we present all the results and analysis of the experiments conducted when the ADL demand was variable. Please note configurations of the microgrids are the same as those used in the case when the ADL demand was fixed.

6.1 Setup 1

The configurations of the different microgrids are provided in Table 1.

- From Figure 9 we conclude that the 3 microgrids converge to a policy that gives higher rewards than random exploration.
- We also observe that even in the case of variable ADL demand, the microgrid that follows the dynamic pricing policy performs better than the one following the constant pricing policy.
- From Figure 8, we observe that the microgrids learn to schedule their ADL demands.
- From Figure 10, we observe the dynamic nature of prices decided by the microgrids.

6.2 Setup 2

The configurations of the different microgrids used in this setup is provided in Table 2:

- From Figure 13 we can conclude that all 8 agents following our policy have converged to some good policy that is better than the constant pricing for most grids and random exploration for all grids.
- From Table 4, we observe that the proposed dynamic pricing model performs better the constant pricing model for the majority (6 out of 8) of microgrids.
- From Figure 11 we see that all the microgrids following our proposed algorithm learn to schedule the ADL jobs at different times which shows that our model is capable of shifting power consumption from the peak demand time.
- From Figure 12 we see the dynamic nature of the prices decided by various microgrids at different time intervals.

Microgrid	**Rewards Obtained By Following:**		**Difference in Rewards**	**Winning Policy**
	Dynamic Pricing Policy (DPP)	**Constant Pricing Policy (CPP)**		
1	-10.047576428852192	-10.142673518187	0.095097	DPP
2	-13.703830385917806	-12.634650821952459	-1.06918	CPP
3	-33.41034390308725	-34.25077351809638	0.84043	DPP
4	-52.0459081898873	-50.5368173295583	-1.509091	CPP
5	-43.28976021341345	-43.84382733650302	0.554067	DPP
6	-37.19913599827946	-37.24948439537201	0.050348	DPP
7	-51.68732739565454	-52.16803273614265	0.480705	DPP
8	-60.42121748490767	-60.43654034418823	0.015323	DPP

Table 4: Average Rewards comparison under Setup 2

6.3 Setup 3

The configurations of the different microgrids used in this setup is provided in Table 3:

- From Figure 16 can see that all the 8 agents have converged to some good policy in both constant pricing and dynamic pricing models and it is clear that in most of the cases the microgrids which follow our dynamic pricing model perform better than those following the constant pricing one.

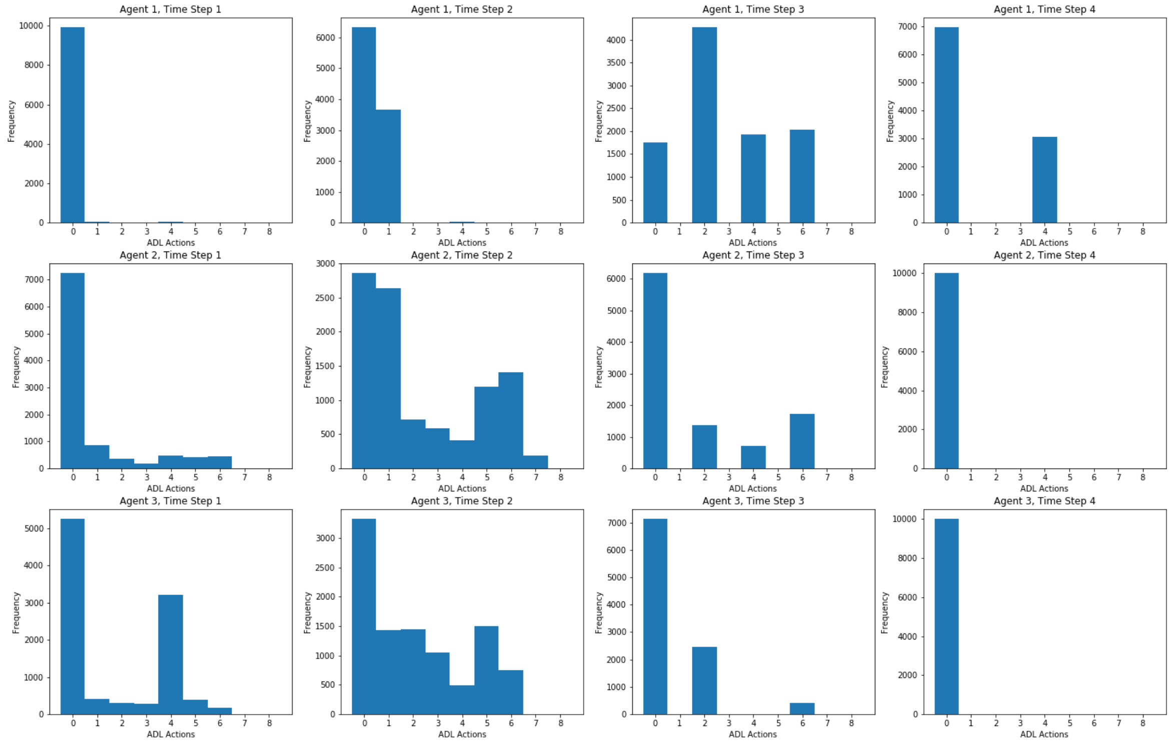

Figure 8: Dynamic Prices scheduling under Setup 1

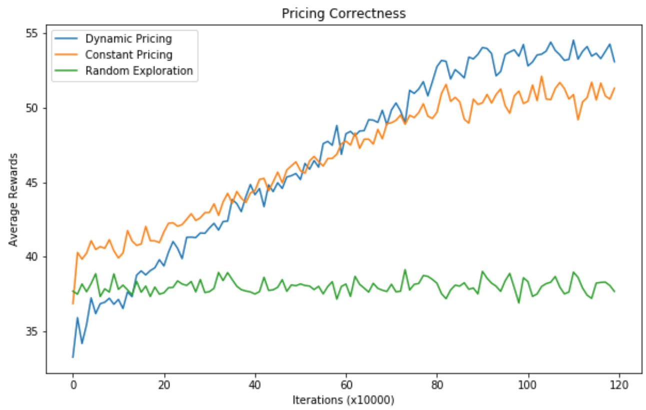

Figure 9: Comparison of all models under Setup 1

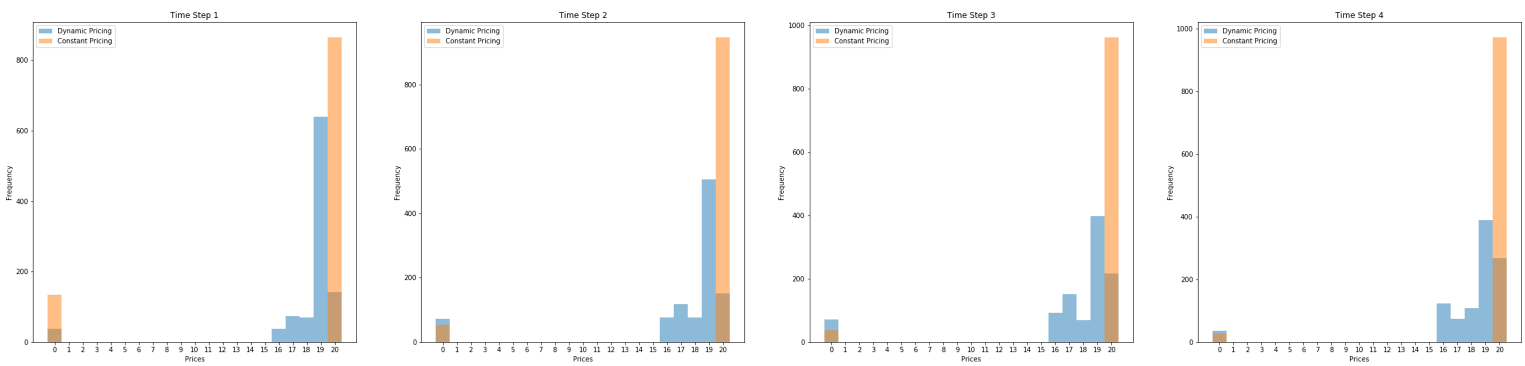

Figure 10: Dynamic Pricing scheduling under Setup 1

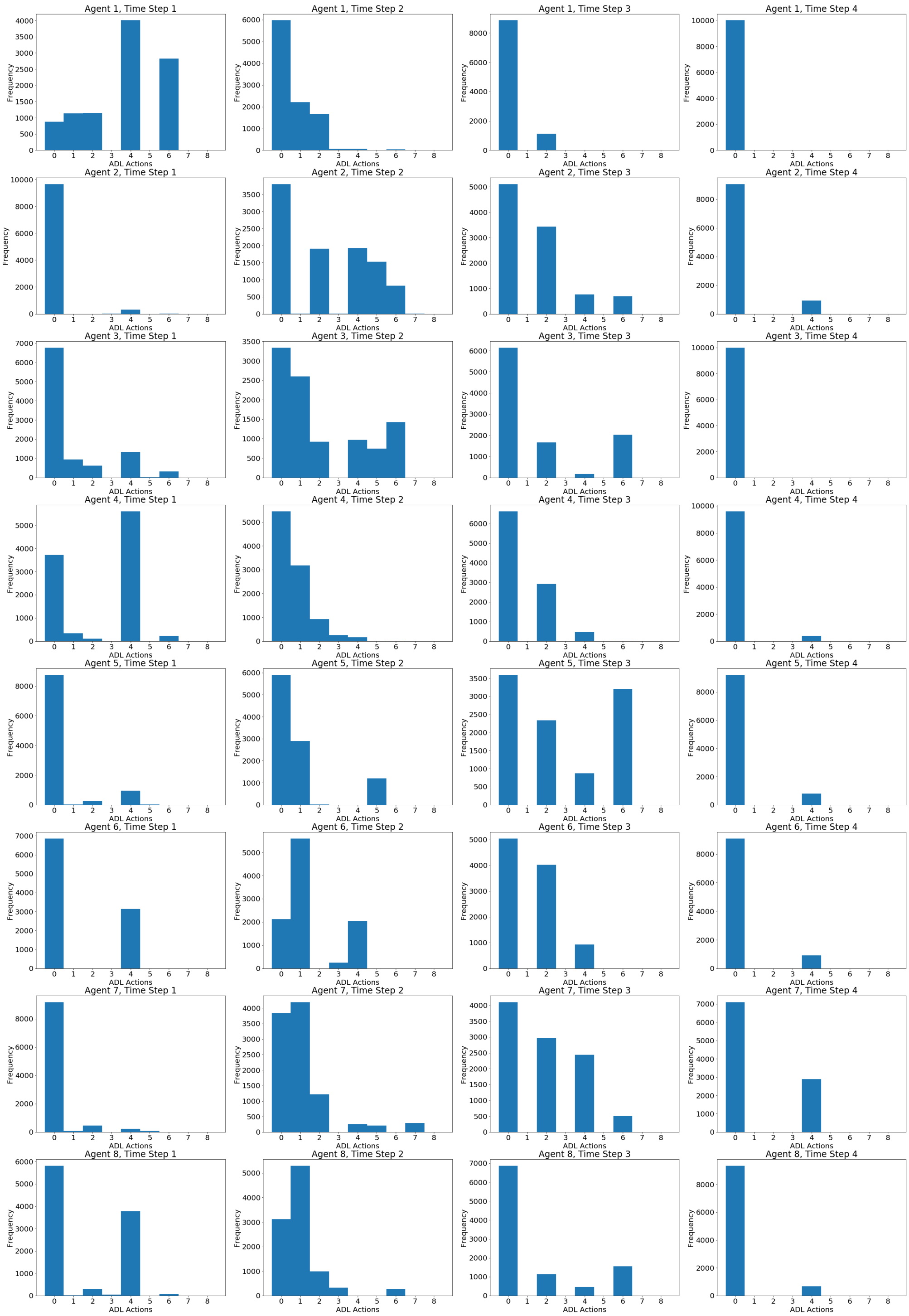

Figure 11: ADL demands scheduling by all 8 microgrids under Setup 2

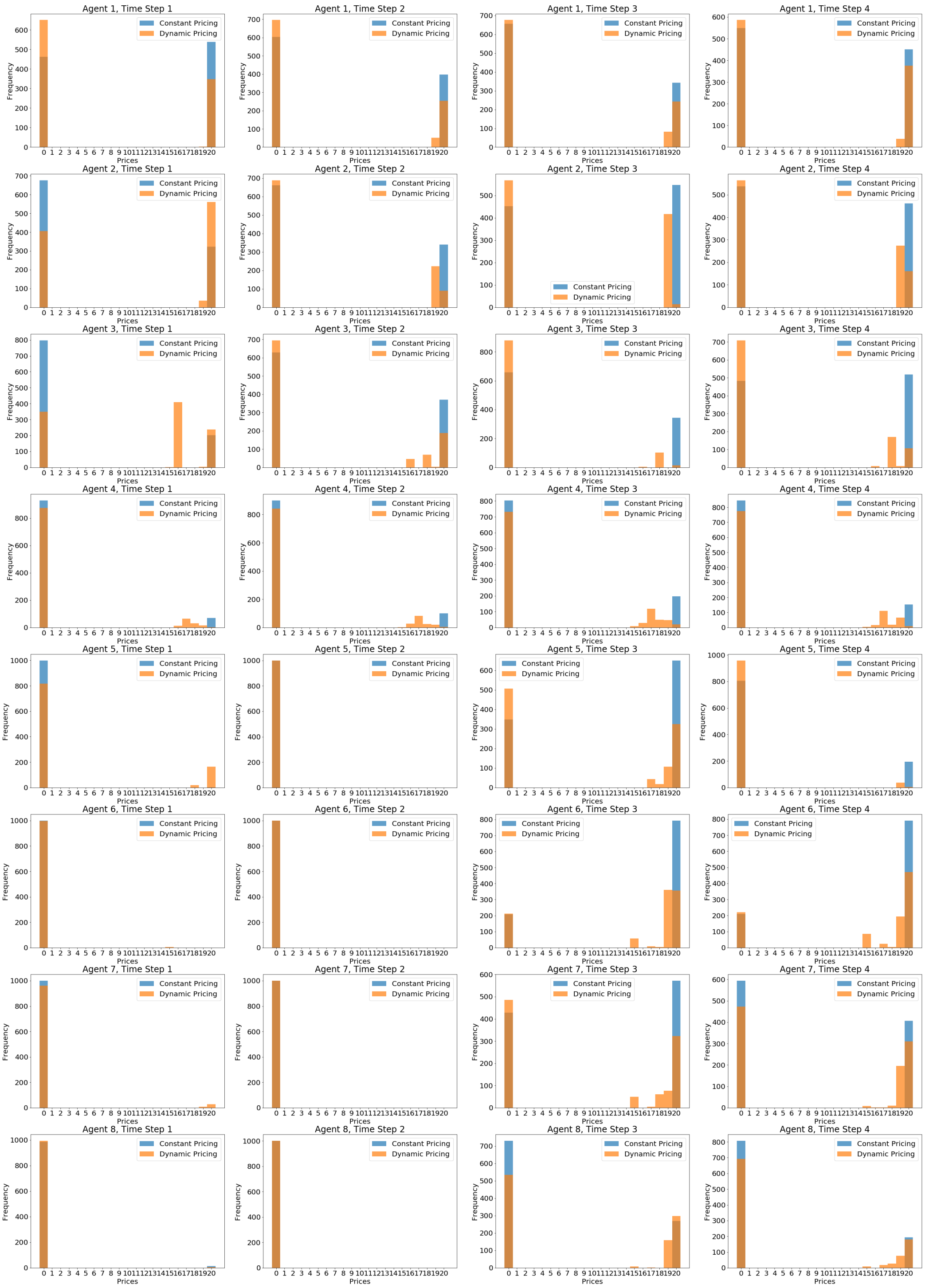

Figure 12: Dynamic Prices scheduling by all 8 microgrids under Setup 2

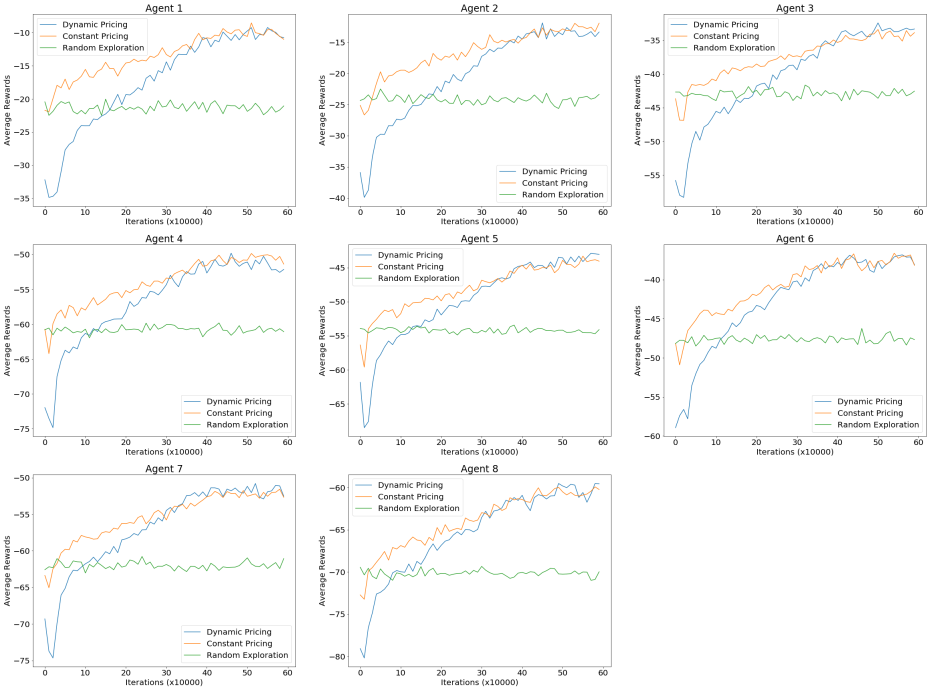

Figure 13: Convergence of all 8 microgrids under Setup 2

Microgrid	**Rewards Obtained By Following:**		**Difference in Rewards**	**Winning Policy**
	Dynamic Pricing Policy (DPP)	**Constant Pricing Policy (CPP)**		
1	-5.6601477396286	-5.94341006389091	0.283262	DPP
2	1.3235604338389502	0.7678307735577367	0.55573	DPP
3	13.49873328662527	12.623290377198161	0.875443	DPP
4	17.44889743859347	17.64793207242157	-0.199035	CPP
5	-33.328549718099254	-33.976077954010954	0.647528	DPP
6	-42.403029888740306	-44.715955245008125	2.312925	DPP
7	-35.66002260071537	-37.44014765870661	1.780125	DPP
8	-13.914241211874241	-13.378362301560838	-0.535879	CPP

Table 5: Average Rewards comparison under Setup 3

- From Figure14 it is clear that all the agents learn to schedule the ADL demands judiciously.
- From Table 5, we observe that the proposed dynamic pricing model performs better than the constant pricing model for the majority (6 out of 8) of microgrids.
- From Figure 15 we again see the dynamic nature of pricing decided by various microgrids. The microcrids learn to sell judiciously at different time steps. From the graphs, we observe that solar grids sell at time steps 3 and 4 which is intuitive as the solar generation is maximum at these time steps.

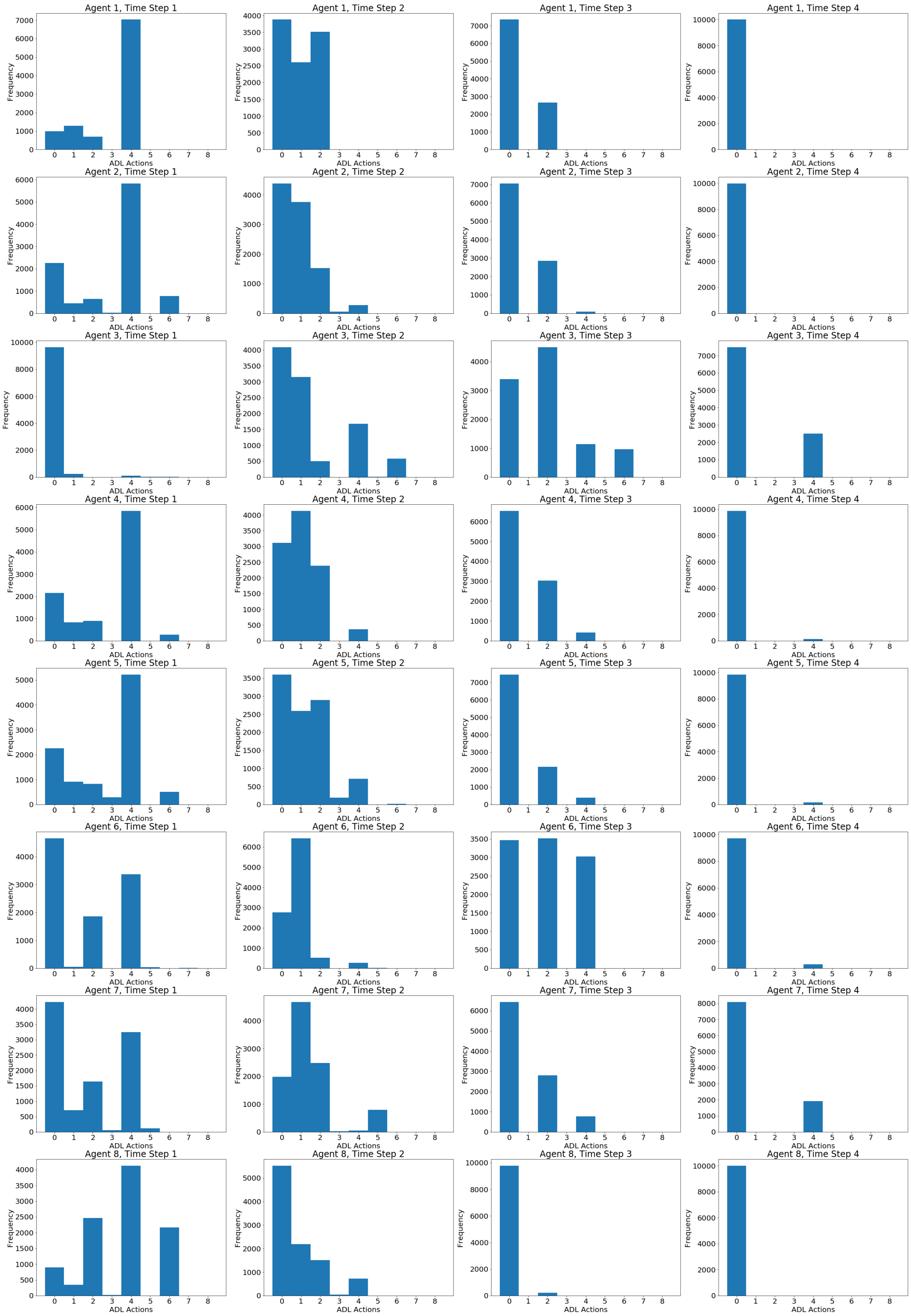

Figure 14: ADL demands scheduling by all the microgrids under Setup 3

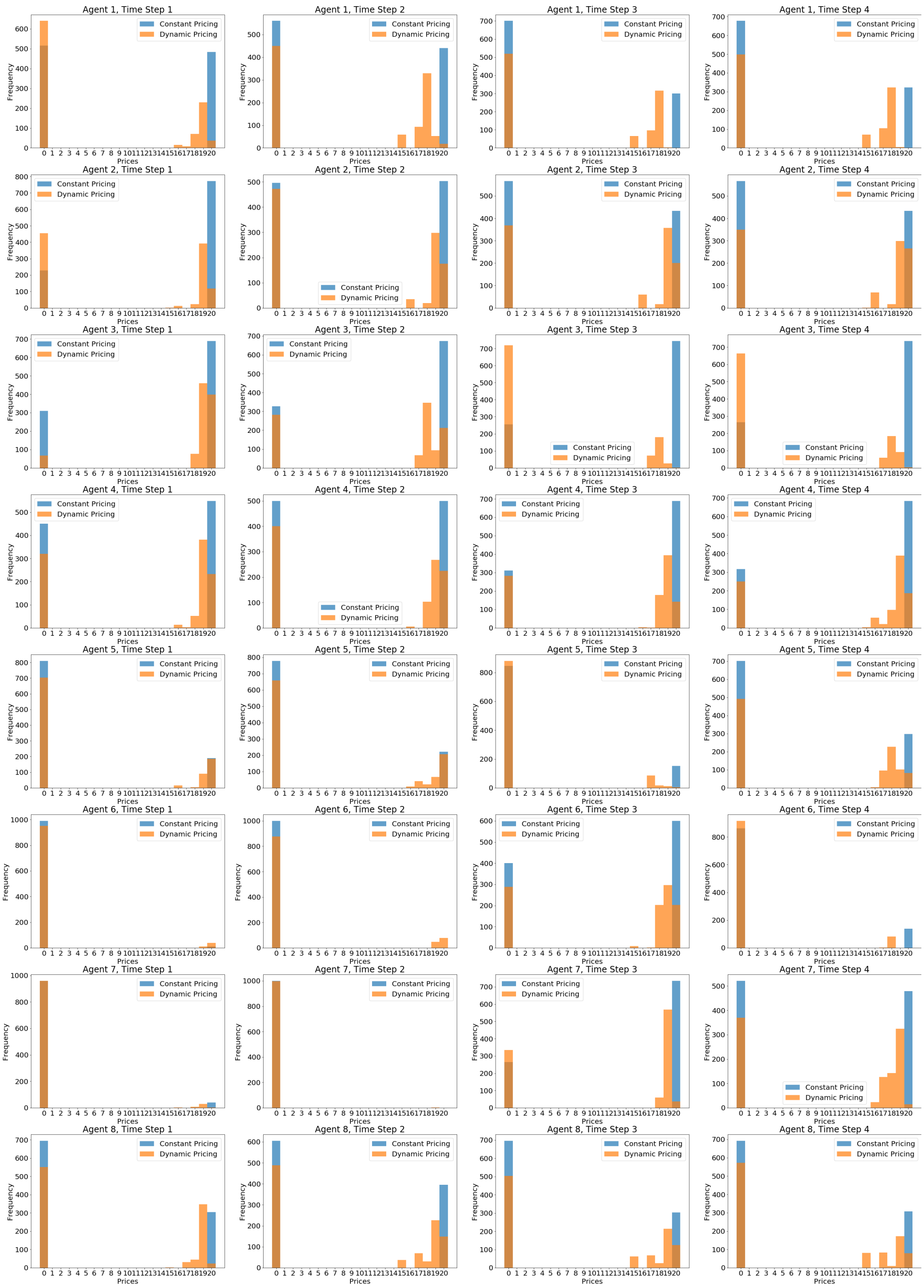

Figure 15: Dynamic Prices scheduling by all the microgrids under Setup 3

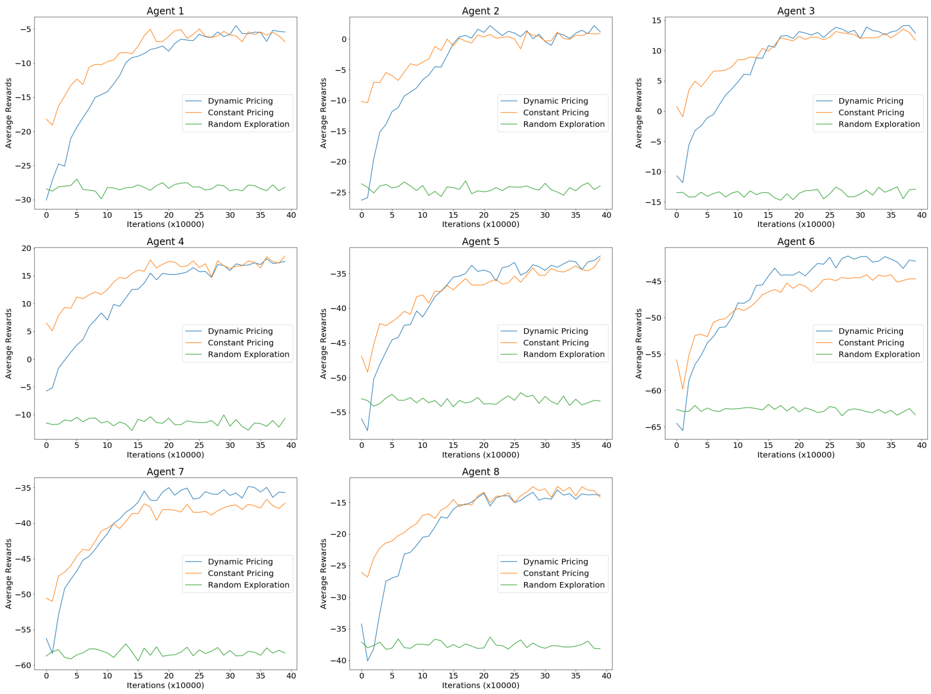

Figure 16: Convergence of all the microgrids under Setup 3