



Multiagent-based secure energy management for multimedia grid communication using Q-learning

Aparna Kumari¹ · Sudeep Tanwar¹

Received: 2 January 2021 / Revised: 16 July 2021 / Accepted: 19 August 2021 /
Published online: 16 September 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In smart grid infrastructure, multimedia communication plays an important role in various applications, for instance, load monitoring, automatic smart meter reading, and energy management. Energy management has gained widespread popularity with the increasing energy demand. The goal of this paper is to explore multiagent-based Reinforcement Learning (RL) for multicarrier (i.e., electricity and gas) Residential Energy Management (REM) systems with data security. It facilitates the possibility for a separate Demand Response Program (DRP) for every energy component accelerates with computational aspects of RL. This paper proposes *Q-MSEM*, a Q-learning-based Multiagent and Secure Energy Management scheme for the optimal solution of REM problems using RL and Ethereum Blockchain (EB) to reduce energy load and decrease energy costs. Then, *Q-MSEM* uses Ethereum Smart Contract (ESC) to covenants data security issues using off-chain storage, i.e., InterPlanetary File System (IPFS) to handle data storage costs issues. Experimental results disclose the effectiveness of the proposed *Q-MSEM* scheme in terms of reduction in load, energy cost (15.82%), and data storage cost.

Keywords Demand response management · Residential energy management · Reinforcement learning · Artificial intelligence · Q-learning

1 Introduction

In the REM system, smart grid productivity increases with the interaction between multiple energy carriers to increase reliability, reduce planning cost and operations cost [1].

✉ Sudeep Tanwar
sudeep.tanwar@nirmauni.ac.in

Aparna Kumari
17ftphde22@nirmauni.ac.in

¹ Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India

Energy management in multicarrier (i.e., electricity and gas) multimedia systems necessitates interaction among several resources and consumers demands without compromising the consumer comfort to handle energy consumption level [29]. Then, electricity prices are impacted by natural gas prices, hence, DRP requires energy reduction for gas and electricity both in multicarrier REM system [22]. Since residential consumers are a major contributor to energy demand and require an effective multicarrier energy system. Thus far, several past research works exist that touched the energy management issue of REM and presented a range of solutions. Though, most of these research works has mainly focused to reduce energy costs [6], decreasing greenhouse gas emissions, reducing energy load with consumer preferences, and smoothing load profile in single carrier energy system [7, 19, 27, 28].

Most of the studies use mathematical methods, specifically linear/non-linear mix integer linear programming (MIP) [1]. Then, due to data dimensionality issues and computing problem heuristic approaches have become popular using machine models in recent years. For instance, Pezeshki et al. [21] highlighted a comparison between Artificial Neural Network (ANN), and fuzzy & neural network (NN) for REM, and then Reynolds et al. [26] presented a combination of ANN and Genetic Algorithm (GA) methods for energy management of residential houses. Next, Wu et al. [14] developed the fuzzy logic, and Gutierrez-Martinez et al. [7] applied GA for energy management issues. Then, Sharifi et al. [28] presented an integrated REM approach, which includes GA with MIP method.

Most of the existing approaches use deterministic rules or abstract methods, which has certain disadvantages like it fails to guarantee optimality and could cause financial losses, application of abstract models in the actual multimedia system could be unrealistic and mostly relies on the smart grid administrator skill, and MIP/game-theoretic optimization approaches suffer from scalability issue. Motivated from the aforesaid challenges and to handle them, RL is a prominent solution, which plays a crucial role in optimal decision-making for realistic issues. It is a subset of Artificial Intelligence (AI) and proved its effectiveness by AlphaGO Zero and AlphaGO breakthroughs [15].

For REM, several RL-based approaches like Deep Q-Network (DQN), Q-learning, and others have been adopted globally by researchers to handle energy management issues for electric vehicles (EVs), electric home appliances, etc. [8, 14]. Then, day-ahead pricing mechanism also used with RL-based approach to handle energy load in REM system [10, 18, 22, 25]. Nevertheless, the utmost existing approaches use Q-learning and have been modeled as Markov Decision Process (MDP). So far, there are several limitations, for instance, (i) most studies only focused on electricity and did not consider gas consumption, then the RL algorithms would perform differently while working with different energy carriers, and (ii) REM on daily basis has been studied, though, the hourly DRP shows better prospective to balance REM associated with dynamic limitations.

Moreover, results gained from the REM agent required to be accessed securely by all stakeholders like smart grid administrators, utility suppliers, and consumers for real-time decision-making. Several RL-based methods exist for REM, yet it has not been exploited to its full potential with data security during multimedia communication. Numerous other challenges such as consumer trust, single-point failure, security, and privacy also exist. To cope up with the mentioned challenges, blockchain is a viable solution, which offers security, transparency, and trust to the stakeholders [10].

Blockchain is a decentralized database and Distributed Ledger Technique (DLT) to store data in chained blocks that protect a single-point failure issue, improve trust, privacy, and security [11]. It has been embraced in smart grid systems for secure energy management in an efficient manner [12], for example, Blom et al. [5] describe the likelihood of blockchain in a smart grid system and Li et al. [13] offered an optimal energy pricing scheme using

Stackelberg game for the nourishment of efficient energy trading. Far ahead, blockchain is adopted in REM as well [2, 9, 12]. The prevailing blockchain-based REM approaches have several limitations, like limited energy data storage ability, high energy data storage cost on ethereum, and high network bandwidth due to multimedia data redundancy.

1.1 Motivation and research contributions

Motivated from the aforesaid issues and the fact that electricity prices are impacted by natural gas prices, so, DRP requires energy consumption reduction for gas and electricity both in multicarrier REM system in order to improve smart grid reliability and Quality of Service (QoS). In this paper, we propose *Q-MSEM*, a Q-learning-based multiagent and secure energy management scheme for multimedia communication in a smart grid system. The traditional blockchain-based DRP system stores energy transactions on a blockchain, which is costly and incompetent in a real-time scenario. However, the proposed *Q-MSEM* scheme stores energy data on the off-chain storage system, i.e., IPFS, which provide low data storage cost, system scalability, and high throughput during access of multimedia energy data [12, 23].

The following are major research contributions of this paper.

- A REM scheme, i.e., *Q-MSEM* is proposed using RL methodology considering electricity and gas consumption with dissatisfaction cost to the consumer. The energy load reduction is addressed by consumer dissatisfaction cost.
- An optimal energy load curtailment strategy is embraced via finite discrete MDP and its tractability is achieved by means of the Q-learning process.
- An IPFS-based unique ESC is designed to access electricity and gas consumption multimedia data securely in real-time and attain high system throughput by scalability and low storage cost during access.
- The performance of the proposed scheme *Q-MSEM* is evaluated compared to existing approaches to shows its effectiveness in terms of different evaluation parameters like load reduction, energy cost, and data storage cost.

1.2 Organizations

The rest of the paper is organized as follows. Section 2 highlights the related work and a comparative analysis of the proposed *Q-MSEM* scheme with existing approaches. Then, Section 3 describes the system model of the proposed *Q-MSEM* scheme and Section 4 discusses the problem formulation of *Q-MSEM* scheme. Then, Section 5 shows the complete workflow of *Q-MSEM* scheme, then, experimental results and performance evaluation are highlighted in Section 6. Finally, the paper is concluded in Section 7 with future work. Then, nomenclatures and abbreviations are listed in Table 1.

2 Related work

Several RL-based approaches have been adopted globally by researchers to handle energy management issues for electric home appliances, EVs, and others. Most of the existing approaches use DQN, Q-learning, etc. to handle energy management in REM systems [8, 14]. In [18, 25], day-ahead pricing mechanism is used for RL-based approach to handle

Table 1 Nomenclatures and abbreviations

Particular	Details
t	Discrete timeslot in a Day
EC_{ξ}	Electricity consumption
EC_g	Gas consumption
US	Utility providers
ω	Critical home appliances (HA)
ν	Non-critical home appliances (HA)
γ	Total number of HAs
$\xi_{C,\omega,t}^{crit}$	Electricity consumption by critical HAs for consumer C at time t
$\xi_{C,\nu,t}^{Ncrit}$	Electricity consumption by non-critical HAs for consumer C at time t
$\phi_{C,t,\xi}$	Dissatisfaction cost function for consumer C in terms of electricity consumption reduction
$\phi_{C,t,g}$	Dissatisfaction cost function for consumer C in terms of gas consumption reduction
γ_C	Preference parameter for consumer to reduce energy load
min_{coef}	Lower bound of energy reduction
max_{coef}	Upper bound of energy reduction
$Cost_{C,t,\xi}$	Electricity cost for consumer C at timeslot t
ξ_{Price}^t	Electricity price at time t
$Cost_{C,t,g}$	Gas consumption cost for consumer C at timeslot t
g_{Price}^t	Gas price at timeslot t
i	Episode in the proposed Q -MSEM scheme for Q-learning
\mathbb{A}	Action for discrete timeslot t
$\mathbb{S}(\xi_{C,t}^{Ncrit})$	States for electricity
$\mathbb{S}(g_{C,t}^{Ncrit})$	States for gas
\mathbb{R}	Reward
ρ	Mapping states to actions
ED_{secure}^{SC}	Secured energy data storage
ED_{secure}^{SCAL}	Scalability of the proposed Q -MSEM scheme
AI	Artificial Intelligence
ANN	Artificial Neural Network
DLT	Distributed Ledger Technique
DQN	Deep Q-Network
DRP	Demand Response Program
EB	Ethereum Blockchain
ESC	Ethereum Smart Contract
EVs	Electric Vehicles
GA	Genetic Algorithm
IPFS	InterPlanetary File System
MDP	Markov Decision Process
REM	Residential Energy Management
RL	Reinforcement Learning

energy load in REM system. Moreover, the existing approaches use Q-learning with MDP, yet, there are several limitations, for instance, most studies only focused on electricity and did not consider gas consumption reduction. Furthermore, RL algorithms would perform differently while working with different energy carriers in REM systems. The energy consumption reduction has been studied on daily basis, though, the hourly consumption reduction shows a better perspective to balance DRP associated with dynamic limitations in REM systems.

The authors' Lu et al. [15] presented a REM system to balance the demand response gap using RL and ANN & accelerate the stability and efficiency in the smart grid. Here, energy cost reduction has been achieved by 7.3%. Then, Xu et al. [31] proposed a data-driven mechanism for REM system using multiagent-based RL. Ahrarinouri et al. [1] presented a multiagent-based RL approach for energy management in REM systems and reduced the energy cost by 12%. Next, Rastegar et al. [24] proposed a mechanism to reduce energy consumption in REM system.

Results gained from the REM agent required to be accessed securely by all stakeholders to make a real-time decision. Several RL-based approaches exist for REM, yet it has not been exploited to its full potential for a multicarrier system with data security. Other challenges like end-consumer trust, single-point failure, security, and privacy also exist. To handle these challenges, blockchain is a prominent solution [10]. It has been embraced in smart grid systems for secure energy management in an efficient manner [12]. Blom et al. [5] describe the likelihood of blockchain in a smart grid system and Li et al. [13] proposed an optimal energy pricing scheme using the Stackelberg game for efficient energy trading. It is adopted in REM as well [2, 9, 12] and blockchain-based REM approaches have several limitations, like limited energy data storage capacity, high energy data storage cost on Ethereum, and high network bandwidth. Hence, this paper proposes *Q-MSEM* scheme for multimedia communication in a smart grid system using Q-learning and EB. The proposed *Q-MSEM* scheme stores energy data on IPFS (an off-chain storage system), which provides low data storage cost, system scalability, and high throughput during access of multimedia energy data [12]. Then, a comparative analysis of the proposed *Q-MSEM* scheme with other existing approaches are discussed in Table 2.

3 System model

Figure 1 represent the proposed *Q-MSEM* scheme, i.e., multiagent-based energy load reduction using Q-learning mechanism. Here, multiple energy carriers, i.e., electricity (EC_{ξ}) and

Table 2 A comparative analysis of the proposed *Q-MSEM* scheme with existing approaches

Approaches	Using RL	Load reduction	Cost reduction	Data storage cost	Scalability	Data security	ESC
Ahrarinouri et al. [1]	Yes	Yes	Yes	No	No	No	No
Xu et al. [31]	Yes	Yes	Yes	No	Yes	No	No
Lu et al. [15]	Yes	Yes	Yes	No	Yes	No	No
Li et al. [13]	No	No	Yes	Yes	Yes	Yes	Yes
Jindal et al. [9]	No	No	Yes	No	Yes	Yes	No
Proposed Scheme	Yes	Yes	Yes	Yes	Yes	Yes	Yes

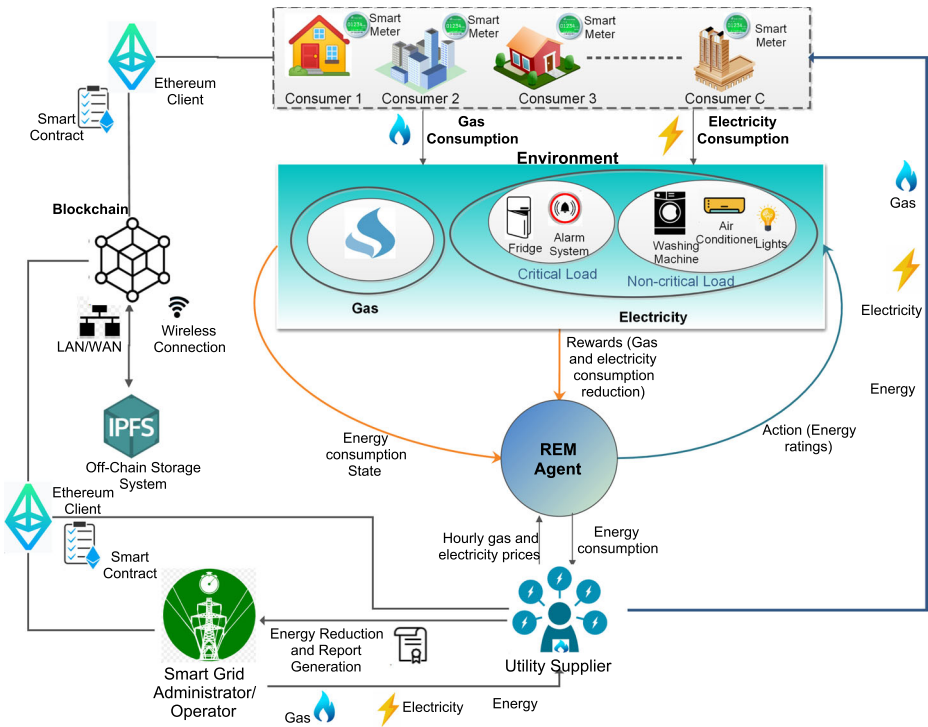


Fig. 1 System model of the proposed *Q-MSEM* scheme

gas (EC_g) consumption is reduced during peak hour and consumption is increased during a non-peak hour, which reduces load burden on smart grid *SG* that improve QoS & reliability of grid and utility supplier, i.e., $US, \forall US \in \{US_1, US_2, US_3, \dots, US_m\}$ and every US should be registered with *SG*. Here, we construct the dynamic energy load based on MDP for particular consumer C , where $C \in \{C_1, C_2, C_3, \dots, C_n\}$ in a stochastic energy consumption environment with maximum number of n consumer.

In multicarrier REM system, electricity and natural gas consumption by home appliances (e.g., fridge, water heater, lights, and others) represent the environment, which is dynamic in nature and it has a specific REM agent associated with it. Initially, we formulate the optimal energy consumption as finite MDP for a particular consumer C in a stochastic environment. Here, energy consumption is categorized into two major groups, i.e., gas consumption and electricity consumption. The electricity demand is either controlled by control signals or consumes energy throughout the day without regulating it. The first one is controllable consumption and can be curtailed based on the consumer dissatisfaction cost. The later one unchangeable and identified as a critical load that requires electricity 24-hour a day, for instance, alarm system and fridge. Then, we develop an efficient Q-learning-based model-free consumption reduction approach that does not require the knowledge of system uncertainties. Here, consumers C participate in the proposed *Q-MSEM* scheme to reduce

their energy cost and balance energy supply and demand. The *US* decides hourly energy prices, which is administered and monitored by *SG*. The REM agent for electricity and gas maximize their profit by taking action and getting rewards.

The optimal energy consumption for consumer *C* at each particular timeslot *t* is published on EB with hourly energy cost based on prices. It is incorporated with IPFS distributed and off-chain storage systems for secure and efficient data storage at a low cost with high scalability. The entire energy data is securely attainable by all stakeholders, i.e., {*SG*, *US*, and *C*} as required. Here, optimal energy consumption enables load reduction and efficient management of energy at the consumers-end, improves QoS, and maximizes SG profit.

4 Problem formulation

We consider a REM system, where houses are equipped with two energy carriers, i.e., electricity and gas, and consuming energy using various kinds of home appliances *HA*, with the need to optimize energy consumption as shown in Fig. 1. The REM is connected to the *US* and *SG* through a bi-directional communication network, which enables the exchange of multimedia energy consumption information and energy prices. Then, REM agents manage electricity and gas consumption in response to energy prices. In REM system, *HA* are categories as critical appliances and controllable/non-critical appliances based on their priorities and characteristics [15, 17]. The mathematical formulations of critical appliances are described as follows.

$$\xi_{C,\omega,t}^{crit} = \mathbb{E}_{\omega,t}^{crit}, \tag{1}$$

where ω is the $HA \in \{1, 2, 3, \dots, \alpha\}$, α is maximum number of critical appliances, timeslot of a day is represented as $t \in \{1, 2, 3, \dots, 24\}$. Hence, critical load for specific consumer *C* at timeslot *t* can be represented as.

$$\xi_{C,t}^{crit} = \sum \xi_{C,\omega,t}^{crit} \tag{2}$$

Moreover, another type of appliances consist flexible energy load from minimum load to maximum energy load, for example, washing machine, lights, Air Conditioner (AC), etc., which is calculated as follows.

$$\xi_{C,v,t}^{Ncrit} = \mathbb{E}_{v,t}^{Ncrit}, \tag{3}$$

where, $\forall v \in \{1, 2, 3, \dots, \beta\}$, β is maximum number of non-critical appliances. Hence, non-critical load for specific consumer *C* at timeslot *t* can be represented as.

$$\xi_{C,t}^{Ncrit} = \sum \xi_{C,v,t}^{Ncrit}, \tag{4}$$

$$\min(\xi_{\beta,t}) \leq \mathbb{E}_{C,t}^{Ncrit} \leq \max(\xi_{\beta,t}) \tag{5}$$

So, total electricity consumption is calculated using (2) and (4) as follows.

$$\xi_{C,t} = \xi_{C,t}^{Ncrit} + \xi_{C,t}^{crit} \tag{6}$$

Here, only non-critical consumption could be reduced as critical appliances consume electricity 24-hour. Though reduction in electricity consumption cause dissatisfaction to consumer. So, the dissatisfaction cost function for consumer C in terms of electricity consumption reduction is defined as follow [32].

$$\phi_{C,t,\xi} = \frac{\gamma_C}{2} \left\{ \xi_{C,t}^{Ncrit} - \xi_{C,t}^{\Omega Ncrit} \right\}^2 + \delta \cdot \left\{ \xi_{C,t}^{Ncrit} - \xi_{C,t}^{\Omega Ncrit} \right\}. \tag{7}$$

$\gamma_C, \delta > 0$

where, γ_C is preference parameter to consumer to reduce energy load that has greater importance in dynamic load reduction. In case of high value of γ_C , dissatisfaction cost to consumers will be high in terms of load reduction [32]. Here, $\xi_{C,t}^{\Omega Ncrit}$ represents dynamic energy load and range of load reduction is identified by $min_{coef} \cdot \xi_{C,t}^{Ncrit}$ and $max_{coef} \cdot \xi_{C,t}^{Ncrit}$ when energy pricing is in effect, where min_{coef} , max_{coef} , and δ are system dependent parameter.

Furthermore, the main aim is to reduce energy consumption for both the energy carrier EC_{ξ} and EC_g , so dissatisfaction cost in case of gas consumption reduction could be calculated as follows (considering gas consumption in case of room heating as critical consumption).

$$\phi_{C,t,g} = \frac{\gamma_C}{2} \left\{ g_{C,t}^{Ncrit} - g_{C,t}^{\Omega Ncrit} \right\}^2 + \delta \cdot \left\{ g_{C,t}^{Ncrit} - g_{C,t}^{\Omega Ncrit} \right\}. \tag{8}$$

$\gamma_C, \delta > 0$

4.1 Objective function

The proposed Q -MSEM scheme focuses on energy consumption reduction and energy cost $Cost_{C,t}$ reduction to consumer C by employing dynamic energy consumption policy for efficient energy management. Therefore, electricity cost can be mathematically formulated as follows using (6).

$$Cost_{C,t,\xi} = \xi_{Price^t} * \xi_{C,t}. \tag{9}$$

$\exists \gamma = \alpha + \beta$

where, ξ_{Price^t} represent electricity price at time t and γ describe total number of HAs. Similarly, gas consumption cost can be mathematically calculated as follows.

$$Cost_{C,t,g} = g_{Price^t} * g_{C,t}. \tag{10}$$

$\exists \gamma = \alpha + \beta$

where, g_{Price^t} represent gas price at time t and γ describe total number of HAs consuming gas at time t . Here, both electricity and gas prices are definable hourly by US that is varying based on time and monitored by SG to handle any kind of discrepancies. Their contributions in the consumer objectives are stated in the following relation using (7), (8), (9), and (10):

$$\theta_{\xi} = \xi_{Price^t} * \xi_{C,t}^{\Omega Ncrit} + \phi_{C,t,\xi} \tag{11}$$

$$\theta_g = g_{Price^t} * g_{C,t}^{\Omega Ncrit} + \phi_{C,t,g} \tag{12}$$

Hence, the objective function is defined as follows using (11) and (12).

$$\theta = \theta_{\xi} + \theta_g \tag{13}$$

Further, the entire energy consumption multimedia data like, energy cost to consumer, electricity consumption (in kWh), electricity price, gas consumption (in kWh), gas price for particular consumer C at timeslot t require to be accessible by all stakeholders $\in \{SG, UP, C\}$. The entire energy data (ED) is stored on IPFS and hash-key ($hkey$) is generated to published on EB at timeslot t for real-time access, which is securely accessible ED_{secure} by $\{SG, UP, C\}$ using their hash-key ($SG_{ipfs \rightarrow hkey}, UP_{ipfs \rightarrow hkey}, C_{ipfs \rightarrow hkey}$) (generated at EB). The ED_{secure} storage cost is high on EB , therefore a unique access mechanism, i.e., IPFS is uses for low data storage cost ED_{secure}^{SC} and make proposed scheme scalable ED_{secure}^{SCAL} .

$$ED_{secure} \leftarrow \{\mathcal{E}, \xi_{C,t}, Cost_{C,t,\xi}, g_{C,t}, Cost_{C,t,g}\}. \tag{14}$$

Therefore, the objective of the proposed Q -MSEM scheme is defined as follows using (13).

$$\begin{aligned} \mathbb{O} &= \min\{\theta, ED_{secure}^{SC}\} + \max\{ED_{secure}, ED_{secure}^{SCAL}\}, \tag{15} \\ &\leftarrow \min\{\xi_{Price^t} * \xi_{C,t}^{\Omega_{Ncrit}} + \phi_{C,t,\xi} \\ &\quad + g_{Price^t} * g_{C,t}^{\Omega_{Ncrit}} + \phi_{C,t,g}, ED_{secure}^{SC}\} \\ &\quad + \max\{ED_{secure}, ED_{secure}^{SCAL}\}. \end{aligned}$$

Subject to the following constraints,

- $C1 : \xi_{C,t}, g_{C,t} \neq 0, \forall C$
- $C2 : t \neq 0$
- $C3 : ED_{secure} \Leftrightarrow EB_{IPFS}$
- $C4 : \{ipfs \rightarrow hkey\} = valid_{hkey}, \forall \{SG, UP, C\}$

where, first term represents the minimisation of energy consumption and energy multimedia data storage cost at EB . Then, second term illustrate the maximization of security of ED , i.e., ED_{secure} and maximize the REM system scalability in the proposed scheme. The real-time availability of ED_{secure} over EB makes the approach fairly efficient and handle single-point failure issue. Here, constraint $C1$ shows energy consumption $\xi_{C,t}$ and $g_{C,t}$ should have non-zero value for all consumer C and constraint $C2$ represents timeslot start from 1. Then, constraint $C3$ shows that ED_{secure}^{SC} and ED_{secure}^{SCAL} , which is calculated in case ED_{secure} is stored on EB associated with IPFS mechanism and consisting valid $\{ipfs \rightarrow key\}$ for all stakeholders in constraint $C4$.

5 Q-MSEM: The proposed scheme

This section highlight the workflow of the proposed Q -MSEM scheme using RL methodology. Figure 2 illustrate the workflow of Q -MSEM, where US comprises two REM agent (one for gas and another for electricity) for energy consumption reduction and decision-making for DRP. Energy consumption at residential houses creates stochastic environment, which is dynamic in nature. Q -MSEM start its execution from beginning of a day and initialises various parameters like Q-value to 0, timeslot t to 1, and episodes i to 1. Then, other simulation parameters are set in the beginning.

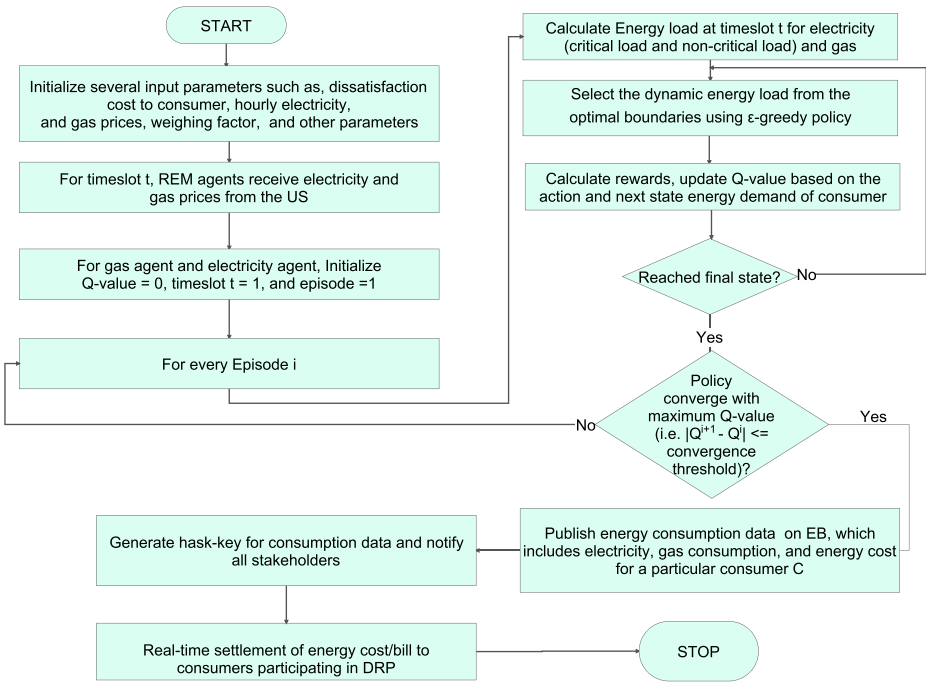


Fig. 2 Workflow of the proposed Q-MSEM scheme

Algorithm 1 Energy management using Q-Learning.

Input: $\xi_{C,t}^{Ncrit}$, $g_{C,t}^{Ncrit}$ Consumer, C , q_t

Output: Reward \mathbb{R} for timeslot t

procedure REM_Agent_Proc($\xi_{C,t}^{Ncrit}$, C , $\phi_{C,t}$, q_t)

for t timeslot **do**

 Initialize Q-values to 0

for Every Episode i **do**

 Initialize state S_k

for Each steps in i **do**

 Select energy rating using ϵ -greedy policy for $A_{k,t}$

 Execute an action $A_{k,t}$ for state $S_{k,t}$

 Calculate reward $r(S_{k,t}, A_{k,t})$

 Observe new state $S_{k,t+1}$

 Calculate Q-value $Q(S_{k,t}, A_{k,t})$

 Calculate Q-value for next episode $Q(S_{k,t+1}, A_{k,t+1})$

if $|Q^{(t+1)} - Q^t| \leq \varsigma$ **then**

 Output Q-value and \mathbb{R}

 exit()

end if

end for

end for

end for

end procedure

Initially, $\xi_{C,t}^{crit}$, $\xi_{C,t}^{Ncrit}$, $g_{C,t}^{crit}$, and $g_{C,t}^{Ncrit}$ is calculated based on the HAs used in a residential building. Then, dynamic energy load is articulated using MDP in distinct finite-step in stochastic environment. Here, energy demand and reward depends on dynamic load $\xi_{C,t}^{\Omega Ncrit}$ and $g_{C,t}^{\Omega Ncrit}$ at the specific timeslot t . It consists various elements, for example : action \mathbb{A} , discrete timeslot t , states $\mathbb{S}(\xi_C, t^{Ncrit})$ for electricity, $\mathbb{S}(g_{C,t}^{Ncrit})$ for gas, and Reward \mathbb{R} . Then, dynamic load $\{\xi_{C,t}^{\Omega Ncrit}, g_{C,t}^{\Omega Ncrit}\}$ are executed as \mathbb{A} . The energy load demand before applying dynamic load is $\xi_{C,t}^{Ncrit}$ and after is $\xi_{C,t}^{\Omega Ncrit}$, then, reward \mathbb{R} at timeslot t is obtained from the objective function using (13).

For adopting optimal action decision, the proposed scheme considered a discrete finite horizon MDP that contains the Markov property, where state transitions are solely dependent only on the current action and state, resulted independent of all previous states and actions are taken by the agent. Here, the policy ρ mapping states to actions are defined as follows.

$$\rho : \mathbb{A}_{k,t} = \rho(\mathbb{S}_{k,t}) \tag{16}$$

Further, the main aim is to discover an optimal policy ρ for each state $\mathbb{S}_{k,t}$, so that selected action $\mathbb{A}_{k,t}$ maximizes the reward $\mathbb{R}(\mathbb{S}_{k,t}, \mathbb{A}_{k,t})$ [15]. The reward maximization based on optimise policy ρ can be calculated using *Bellman Equation* as follows:

$$Q_{\rho}^*(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) = \mathbb{R}(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) + \zeta \cdot \max Q(\mathbb{S}_{k,t+1}, \mathbb{A}_{k,t+1}) \tag{17}$$

where, ζ is convergence parameter. Algorithm 1 is used for Q-value convergence, where after executing various episodes, Q-value converges to maximum Q-value. Here, trial and error mechanism is used by REM agents to store and update Q-values. REM agent perform action in every hour and Q-value in Q-table is updated using (17) as follows.

$$\begin{aligned} Q(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) \leftarrow & Q(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) + \chi \cdot \left[\mathbb{R}(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) \right. \\ & \left. + \zeta \cdot \max Q(\mathbb{S}_{k,t+1}, \mathbb{A}_{k,t+1}) - Q(\mathbb{S}_{k,t}, \mathbb{A}_{k,t}) \right] \end{aligned} \tag{18}$$

Where, χ is learning rate $\in [0, 1]$ and ζ is the future reward discount factor, which is set as 0.95 for accumulation of high reward. Q-value is convergence to a maximum Q-value using optimal policy that is formulated as follows:

$$\rho = \operatorname{argmax}(Q(\mathbb{S}_{k,t}, \mathbb{A}_{k,t})). \tag{19}$$

Furthermore, ϵ -greedy approach is applied within dynamic energy consumption boundaries (maximum and minimum consumption) and agent select actions randomly. Then, reward is calculated and repeated the process till the end of a particular day. Next, REM agent converge to maximum Q-value using convergence parameter value ζ with terminating condition $|Q^{(i+1)} - Q^i| \leq \varsigma$, where ς is a system dependent parameter.

Once Q-value converges, the details of optimal energy consumption, its cost, and energy costs for specific consumer C at timeslot t are required to be accessible in real-time by all stakeholders $\in \{C, SG, UP, \}$. So, ESC and Algorithm 2 is proposed to maximize the security of ED_{secure} compare to the existing approaches like Li et al. [13] & Jindal et al. [9] and published energy multimedia data on EB , which is incorporated with IPFS mechanism. Here, payment of energy $Cost_{C,t}$ depends on the consumer by various means like using EB , online payments or others.

The energy data storage cost ED_{secure}^{SC} is high on EB , so it necessitates to develop a unique access mechanism, i.e., IPFS protocol for low ED_{secure}^{SC} and make proposed scheme

more scalable ED_{secure}^{SCAL} . In conventional EB , one word contains 256 bits $\approx 2^5$ Bytes. Next, a word is stored in EB using ethereum gas (EG) and SSTORE, where $SSTORE = 20000$ gas, i.e., $EB_{store} = 20000$ gas [30].

$$1word = 20000gas.$$

$$1KB = (2^{10})/(2^5) \times 20000gas.$$

Then, EG price EG_{price} (in Gwei) and Ethereum cryptocurrency (Ether) ETH_{price} (in USD) are dynamic in nature. EG_{price} is 10.928 Gwei as of 12th January, 2020.

$$EG_{price} = \mathbb{X}Gwei,$$

$$EG_{price} = \mathbb{X}/10^9 Ether. \tag{20}$$

Where, \mathbb{X} changes as per cryptocurrency market, So, one word storage cost in EB is calculated as follows.

$$Storage_{cost}^1 = EG_{price} * 1word, \tag{21}$$

Now, calculated $Storage_{cost}^W$ for W -words using (20) as follows [30].

$$Storage_{cost}^W = (\mathbb{X} * W * 1word)/10^9 Ether. \tag{22}$$

Based on the ETH_{price} in USD, conversion of the storage cost in USD is calculated as follows.

$$Storage_{cost}^W USD = USD\{Storage_{cost}^W \times ETH_{price}\},$$

$$\Rightarrow \{(EG_{price} * W * 1word) \times ETH_{price}\},$$

$$\Rightarrow \{(\mathbb{X} * W * 1word)/10^9 \times ETH_{price}\}. \tag{23}$$

Therefore, the energy data storage cost ED_{secure}^{SC} for W -words is calculated in USD using (23).

$$ED_{secure}^{SC} = Storage_{cost}^W USD. \tag{24}$$

The proposed Q -MSEM scheme comprises IPFS-based EB, which involves only hashkey storage cost in place of complete energy data storage. Next, IPFS is an open-source, distributed storage system that is immutable and free. IPFS receives the energy data $\{C-ID, \xi_{C,t}, g_{C,t}, Cost_{C,t,\xi}, Cost_{C,t,g}\}$ from US and generate the hashkey of it. IPFS splits the ED_{secure} file into chunks that are encrypted with a random encryption key and satisfies the ESC conditions. Then, these encrypted chunks are stored into the EB . The hashkey size is 256bit for SHA-256 or 160 bits for SHA-1 in EB , hence a single tuple involves 1word-size in case of proposed scheme. The off-chain storage of ED_{secure} improves security and scalability of the scheme by adding more number of energy data at low cost. All transactions in Q -MSEM are authenticated by stakeholders due to EB characteristics like immutability, distributed, etc. that ensures no forgery with energy data.

Algorithm 2 Secure access to REM data using Blockchain.**Input:** Consumer C_ID for C , $g_{C,t}$, $\xi_{C,t}$, maximum timeslot EOD **Output:** $Cost_{C,t,\xi}$, $Cost_{C,t,g}$ for timeslot t **Initialize:** $Pay_Option=1$ for payment using EB , $EOD = 24$

```

procedure Secure_Energy_Data_Proc( $C$ ,  $\xi_{C,t}$ ,  $\xi_{C,t}$ ,  $g_{C,t}$ )
  for  $C\_ID$  consumer do
    if ( $C\_ID \notin R_{SG}$ ) then
       $R_{SG} \leftarrow$  REQUEST_ACCESS ( $R$ ,  $EOA$ ) by paying Ether
      Grant_Permission for  $\{C_{EOA}, SG_{EOA}, UP_{EOA}\}$ 
    else
      for  $t$  timeslot do
        Call Procedure REM_Agent_Proc
        Select  $g_{C,t}^{\Omega_{Ncrit}}$ ,  $\xi_{C,t}^{\Omega_{Ncrit}}$  for timeslot  $t$ 
        calculate  $Cost_{C,t,\xi} \leftarrow \xi_{Price^t} * (\xi_{C,t}^{crit} + \xi_{C,t}^{\Omega_{Ncrit}})$ 
        calculate  $Cost_{C,t,g} \leftarrow g_{Price^t} * (g_{C,t}^{crit} + g_{C,t}^{\Omega_{Ncrit}})$ 
         $ED_{secure} \leftarrow \{C\_ID, \xi_{C,t}, Cost_{C,t,\xi}, g_{C,t}, Cost_{C,t,g}\}$ 
        Generate hashkey for  $ED_{secure} \leftarrow ED_{secure}^{ipfs \leftarrow hkey}$ 

        Move  $ED_{secure}$  data to IPFS
        Store hash on  $EB$ , i.e.,  $ED_{secure}^{ipfs \leftarrow hkey}$ 
        Notify_All_Stakeholders  $\{C_{EOA}, SG_{EOA}, UP_{EOA}\}$ 
        if ( $t == EOD$ ) then
           $EVENT \leftarrow$  Generate_Energy_Cost( $C\_ID$ )
          Execute  $E\_Cost\_Payment\_Method(C\_ID)$ 
          Break
        else
          Continue
        end if
      end for
    end if
  end for
end procedure

```

6 Experimental results

In this section, we highlight results obtained from experiments and evaluate the performance of the proposed Q -MSEM scheme.

6.1 Experimental setup and dataset

The proposed Q -MSEM scheme is implemented using python programming language on Windows operating system configured as Intel(R) Core(TM) CPU @ 2.60GHz, 8GB RAM. The Open Energy Information (openEI) datasets is taken [16], which contains electricity and gas consumption data by various HAs and pre-processed using a sci-kit-learn library. The critical load data (for instance refrigerator and other) is extracted from Pecanstreet [20]. Then, hourly electricity prices is taken from PJM Data Miner as of 12th January, 2020 [22] and residential gas prices is taken from openEI [4] and hourly gas price dataset is created from it.

Firstly, the collected energy (electricity and gas) data is pre-processed for cleaning. Then, unwanted observations, for instance, irrelevant information and duplicate data values are removed from data. Next, we filter out the unwanted outliers, which are observed in

Table 3 Simulation parameter settings

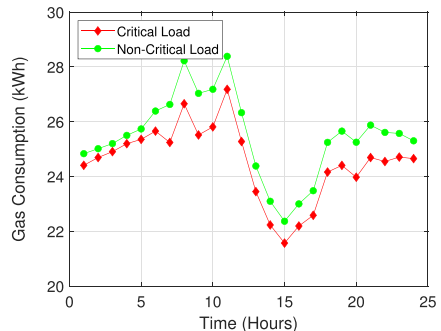
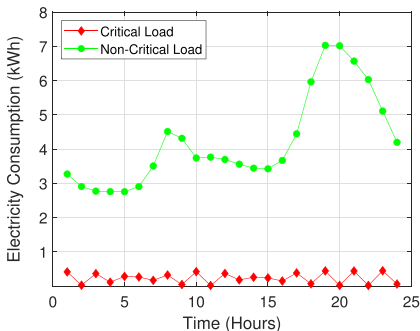
Parameter	Values
REM agent	2
γ_C	0.8
δ	0.1
ϵ	0.6
ς	0.001
ζ	0.95
χ	0.01
EG_{price}	10.928

both the electricity and gas consumption data. Then, we address the issue of zero and null value occurrence by employing a linear interpolation mechanism. Next, zero and null values are replaced by the interpolated values and then, cleaned energy data is prepared. Here, the major aim for cleaning is to remove errors, handle redundancy, increase energy data reliability for optimal analysis, and to make efficient use of memory resources.

6.2 Results and comparative analysis of Q-MSEM

For simulation, a residential house is considered with critical HA (refrigerator), and non-critical/controllable HA (AC, Basic facility, lights, and others). In case of gas consumption, room heater is considered as critical HA and rest as non-critical HA. Initially, all simulation parameters are set as illustrated in Table 3. Then, Algorithm 1 is used for energy consumption reduction and Q-value converges based on dynamic consumption of electricity and gas. The proposed scheme is effortlessly expendable with more number of HAs, which can be included due to flexible and scalable architecture of Q-MSEM scheme.

Figure 3a and b shows the accumulated critical and non-critical electricity and gas consumption of a specific consumer, which evident that non-critical HAs play major contribution in energy consumption through out the day. So, only non-critical energy load can be adjusted in DRP of REM system as critical load are those load which require energy through out a day, which is discussed in detail in Section 4.



(a) Electricity consumption of a consumer

(b) Gas consumption of a consumer

Fig. 3 Hourly energy consumption

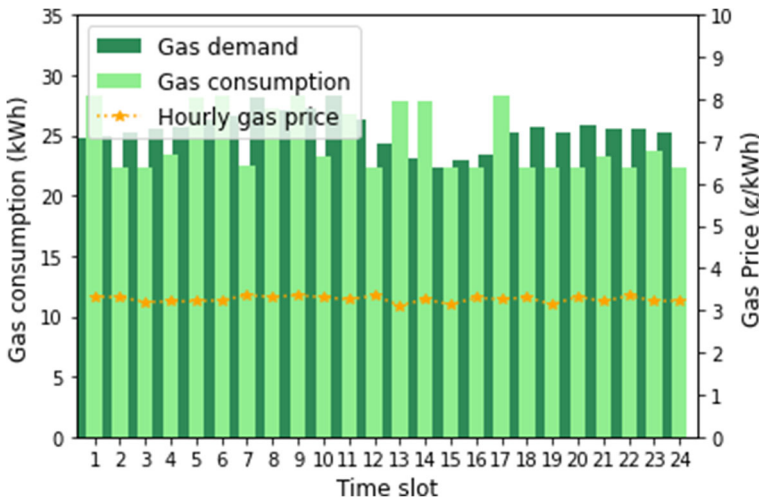


Fig. 4 Gas consumption reduction using proposed *Q-MSEM* scheme

Figures 4 and 5 describe the energy consumption reduction for electricity and gas of a consumer based on dynamic energy load associated with Q-learning. It depicts from the figure that energy consumption is reduced during peak hour (as prices increases the energy demand of consumer gets reduce). More, the non-peak hour having high consumption to balance the dissatisfaction cost to consumer in terms of energy demand. Figure 6 shows the total reduction in a day for electricity and gas. Before applying *Q-MSEM* scheme, electricity consumption is 106.8 (kWh), which reduces to 103.67 (kWh), and in case of gas consumption reduces from 1200.27 (kWh) to 1179.95 (kWh).

Figure 7a shows the Q-value convergence and initially it has high convergence then it learns with trial and error approach and result to maximum Q-values and compared with other existing approaches such as Baseline 1 (Lu et al. [15]) and Baseline 2 (Xu et al. [31]).

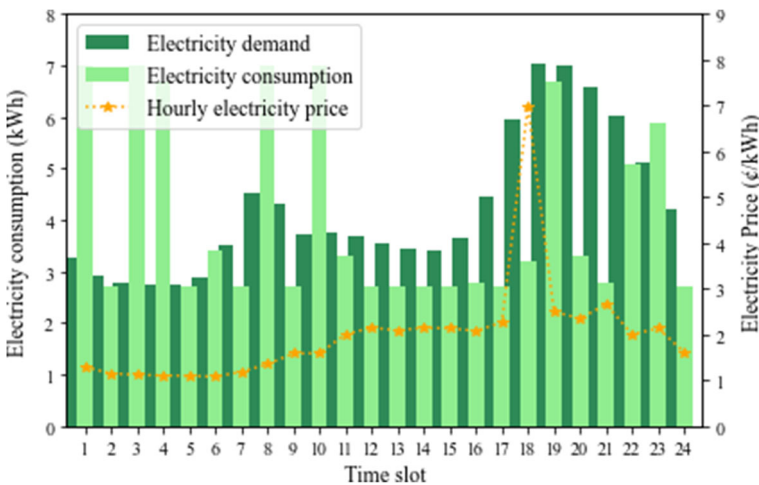


Fig. 5 Electricity reduction using proposed *Q-MSEM* scheme

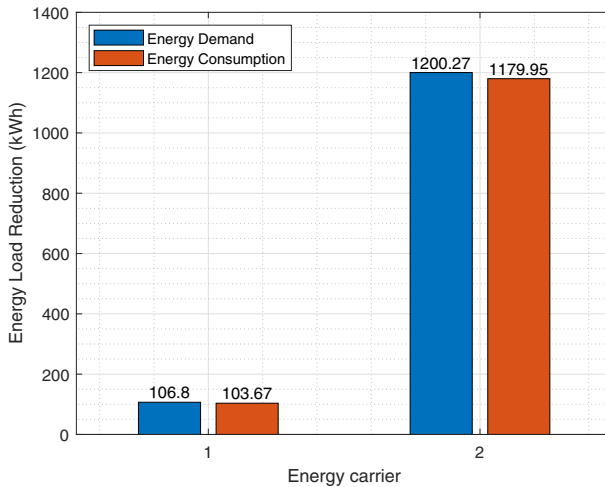


Fig. 6 Total Energy reduction using proposed *Q-MSEM* scheme

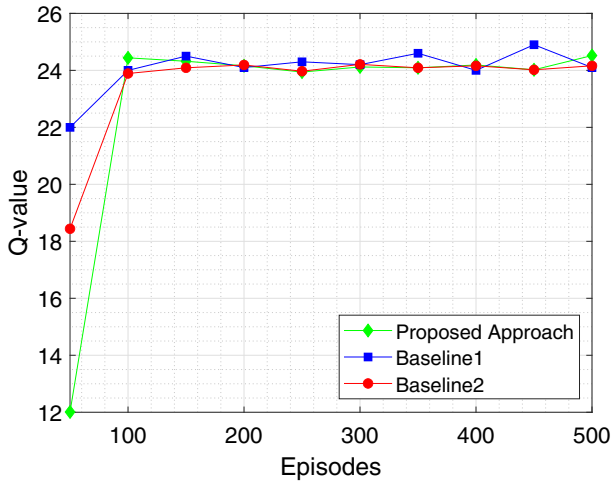
Figure 7b illustrate the hourly electricity cost reduction of the proposed approach. Graph shows the effectiveness of the proposed scheme while comparing with the other existing approaches: Baseline 1 (Lu et al. [15]), Baseline 2 (Ahrarinnouri et al. [1]), and Baseline 3 (Rastegar et al. [24]). The total cost reduction is reported as 15.82% in case of electricity in the proposed *Q-MSEM* scheme, which is moderately high compare to Baseline 1 (7.3%), Baseline 2 (12%), and Baseline 3 (10%). The total gas consumption cost reduction is achieved 1.72% in the proposed *Q-MSEM* scheme.

6.3 Ethereum smart contract

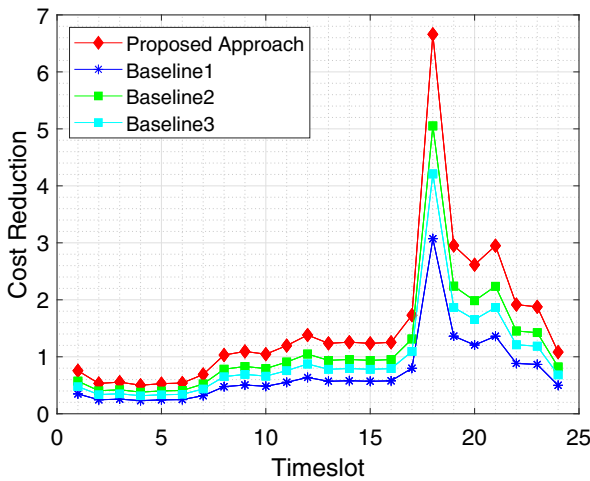
In the proposed *Q-MSEM* scheme, after performing the DRP activity using optimal energy load, the details of electricity and gas consumption, hourly energy prices are required to be shared in real-time with all stakeholders using Algorithm 2. The ESC is developed in Remix IDE and deployed over Truffle suits to perform block verification on EB.

Figure 8 shows the *Q-MSEM* interface uses ESC for real-time electricity and gas data access and decision making. It supports several real-time functionalities related to energy data publishing on EB with high energy data security. It facilitates several security features, which are time-stamps dependent such as tx.origin, re-entrancy, execution cost, and others. It is one of the critical tasks to design ESC as it can not be altered after deployment. Hence, we verified the security vulnerabilities of the proposed *Q-MSEM* scheme on open-source tool Mythril [3]. Figure 9 shows the successful verification of security vulnerabilities of ESC on Mythril.

Figure 10a shows comparative analysis between conventional EB-based system and the proposed *Q-MSEM* scheme in terms of data storage cost. In a conventional EB-based system, the energy data storage cost is calculated using (24), whereas in the proposed scheme IPFS mechanism is incorporated, which is distributed in nature and is a low-cost storage system compared to conventional EB such as Li et al. [13] & Jindal et al. [9]. Figure 10a



(a) Convergence of Q-value



(b) Cost comparison

Fig. 7 Comparative analysis of the proposed *Q-MSEM* scheme and existing approaches: (a) Convergence of Q-value and (b) Cost comparison

evident that as *Ether* price upsurge (11th January, 2020 to 20th January, 2020), then energy storage cost gets impacted slightly due to the off-chain storage facility compare to the conventional-EB system, which has a huge impact on price change.

Figure 10b illustrate scalability comparison of the proposed *Q-MSEM* scheme based on the transactions time and blocks mined during execution of ESC. In the proposed scheme, only the hash-key of energy data is sent to the EB for storage as complete multimedia data

consumptionAler...	address consumer, address UP, address gri
costAlertMessage	address consumer, address UP, address gri
electricityPriceH...	uint256 time, uint256 hPrice
energyCost_Pay...	address consumer
energyCostGene...	address consumer
energyDetailsSet	uint256 consumerID, uint256 bill, uint256 t
gasPriceHourlyGet	uint256 time, uint256 hGPrice
Grant_Permission	address Consumer
registrationRequ...	bytes32[] hashKey

Fig. 8 Smart contract interface

stores on IPFS off-chain database. Here, hash-key size is 256bit (SHA-256) /160 bits (SHA-1). Hence, *Q-MSEM* scheme facilitates to accommodate more transactions on the EB and service more consumers at the smart grid. With the increase of the end-consumers, more blocks need to be mined with the increase request, which resulted to upsurge the transaction time. In the proposed *Q-MSEM* scheme, 256 bits haskey is generated for energy data that is comparatively less than the actual size of a transaction. So, *Q-MSEM* scheme facilitates adding more transactions on the EB to service more end-consumers.

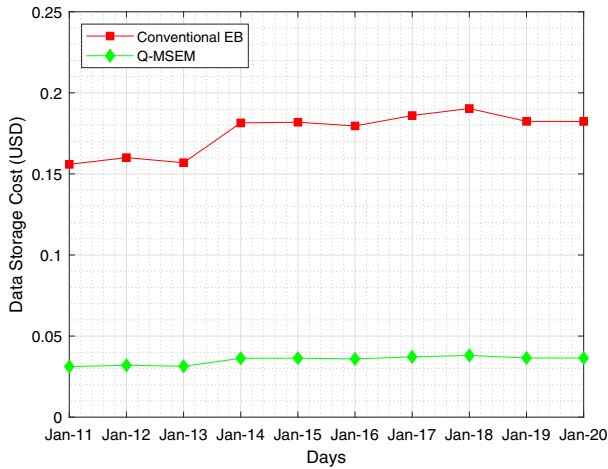
The proposed *Q-MSEM* scheme benefited multicarrier REM systems with data security though it only works in an energy consumption environment with finite & discrete action spaces and state. To extend the proposed *Q-MSEM* scheme to richer environments, function approximators are needed to be applied instead of storing the full state-action table, which is often infeasible. Therefore, an extension of this research work will be employing deep RL to handle uncertain load patterns.

```

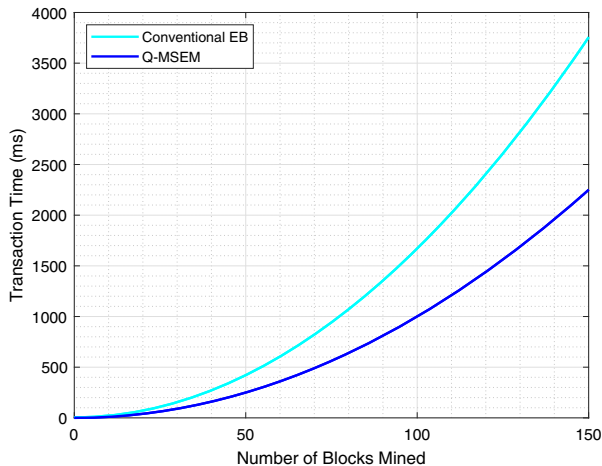
aparna@LAPTOP-MI4D38R6: ~
aparna@LAPTOP-MI4D38R6:~$ /home/aparna/.local/bin/myth analyze SecureQmsem.sol --solc 0.4.22
The analysis was completed successfully. No issues were detected.

```

Fig. 9 Security verification on Mythril



(a) Storage cost comparison



(b) Scalability comparison

Fig. 10 Comparative analysis of conventional EB and the proposed *Q-MSEM* scheme: (a) Storage cost comparison and (b) Scalability comparison

7 Conclusion

In this paper, we proposed *Q-MSEM* scheme, a multiagent-based secure REM approach for multimedia grid communication using Q-learning with the aim to reduce energy consumption with a notch of discomfort in DRP. Motivated by the fact that electricity prices are impacted by natural gas prices, so, DRP requires energy consumption reduction for gas and electricity both in multicarrier REM systems, which improve smart grid reliability and reduce the burden of energy generation and supply on the grid with QoS. So, we have adapted an RL-based approach using Q-learning, to incorporate dynamic energy consumption to reduce load. The empirical results validated the effectiveness of the proposed approach compared to state-of-the-art approaches.

In the future, real-time energy prices, working time of HAs, and fault detection with the effect of Current Transformer (CT) saturation will be an extension of this research work using deep RL to handle uncertain load patterns.

References

- Ahrarinouri M, Rastegar M, Seifi AR (2021) Multiagent reinforcement learning for energy management in residential buildings. *IEEE Trans Indust Inf* 17(1):659–666
- Al Kawasmi E, Arnautovic E, Svetinovic D (2015) Bitcoin-based decentralized carbon emissions trading infrastructure model. *Syst Eng* 18(2):115–130
- ConsenSys, Mythril security analysis tool, <https://github.com/ConsenSys/mythril>, online; Accessed: 20-December-2020
- Eia: Natural gas prices, https://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_m.htm, online; Accessed: 20-December-2020
- Fredrik B, Hossein F (2018) A feasibility study of blockchain technology as local energy market infrastructure. Master Thesis: Norwegian University of Science and Technology, pp 1–73. [Online]. Available: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2502356?show=full>
- Gazafroudi AS, Shafie-khah M, Heydarian-Forushani E, Hajizadeh A, Heidari A, Corchado JM, Catalão JP (2019) Two-stage stochastic model for the price-based domestic energy management problem. *Int J Elect Power Energy Sys* 112:404–416
- Gutierrez-Martinez VJ, Moreno-Bautista CA, Lozano-Garcia JM, Pizano-Martinez A, Zamora-Cardenas EA, Gomez-Martinez MA (2019) A heuristic home electric energy management system considering renewable energy availability. *Energies* 12(4):671
- Huang Y, Wang L, Guo W, Kang Q, Wu Q (2018) Chance constrained optimization in a home energy management system. *IEEE Trans Smart Grid* 9(1):252–260
- Jindal A, Aujla GSS, Kumar N, Villari M (2019) Guardian: Blockchain-based secure demand response management in smart grid system. *IEEE Trans Serv Comput*, pp 1–1
- Kumari A, Tanwar S (2020) Secure data analytics for smart grid systems in a sustainable smart city: Challenges, solutions, and future directions. *Sustain Comput Inf Sys* 28:100427
- Kumari A, Gupta R, Tanwar S, Tyagi S, Kumar N (2020) When blockchain meets smart grid: Secure energy trading in demand response management. *IEEE Netw* 34(5):299–305
- Kumari A, Shukla A, Gupta R, Tanwar S, Tyagi S, Kumar N (2020) Et-deal: A p2p smart contract-based secure energy trading scheme for smart grid systems. In: *IEEE INFOCOM 2020-IEEE conference on computer communications workshops (INFOCOM WKSHPs)*. IEEE, pp 1051–1056
- Li Z, Kang J, Yu R, Ye D, Deng Q, Zhang Y (2018) Consortium blockchain for secure energy trading in industrial internet of things. *IEEE Trans Indust Inf* 14(8):3690–3700
- Lu R, Hong SH (2019) Incentive-based demand response for smart grid with reinforcement learning and deep neural network. *Appl Energy* 236:937–949
- Lu R, Hong SH, Yu M (2019) Demand response for home energy management using reinforcement learning and artificial neural network. *IEEE Trans Smart Grid* 10(6):6629–6639
- Open energy information: Smart meters data from houses, <https://openei.org/datasets/files/961/pub>, online; Accessed: 02-December-2020
- Paterakis NG, Erdinc O, Bakirtzis AG, Catalão JP (2015) Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. *IEEE Trans Indust Inf* 11(6):1509–1519
- Patyn C, Reymond M, Rădulescu R, Deconinck G, Nowé A (2018) Reinforcement learning for demand response of domestic household appliances. *Adapt Learn Agents 2018 Proc*, pp 1–7
- Paudyal P, Ni Z (2019) Smart home energy optimization with incentives compensation from inconvenience for shifting electric appliances. *Int J Elect Power Energy Sys* 109:652–660
- Pecan street dataport, <https://www.pecanstreet.org/dataport/>, accessed: 21-December-2020
- Pezeshki Z, Mazinani SM (2019) Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: a survey. *Artif Intell Rev* 52(1):495–525
- pjm data miner <https://www.pjm.com/markets-and-operations/etools/data-miner-2.aspx>, accessed: 20-December-2020
- Rajalakshmi A, Lakshmy K, Sindhu M, Amritha P (2018) A blockchain and ipfs based framework for secure research record keeping. *Int J Pure Appl Math* 119:1437–1442

24. Rastegar M, Fotuhi-Firuzabad M, Lehtonen M (2015) Home load management in a residential energy hub. *Electr Power Syst Res* 119:322–328
25. Remani T, Jasmin E, Ahamed TI (2018) Residential load scheduling with renewable generation in the smart grid: a reinforcement learning approach. *IEEE Syst J* 13(3):3283–3294
26. Reynolds J, Rezgui Y, Kwan A, Piriou S (2018) A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control. *Energy* 151:729–739
27. Roustai M, Rayati M, Sheikhi A, Ranjbar A (2018) A scenario-based optimization of smart energy hub operation in a stochastic environment using conditional-value-at-risk. *Sustainable Cities Soc* 39:309–316
28. Sharifi AH, Maghouli P (2019) Energy management of smart homes equipped with energy storage systems considering the par index based on real-time pricing. *Sustainable Cities Soc* 45:579–587
29. Wang J, Zhong H, Ma Z, Xia Q, Kang C (2017) Review and prospect of integrated demand response in the multi-energy system. *Appl Energy* 202:772–782
30. Wood GG (2014) Ethereum: a secure decentralised generalised transaction ledger. *Ethereum Proj Yellow Paper* 151:1–32
31. Xu X, Jia Y, Xu Y, Xu Z, Chai S, Lai CS (2020) A multi-agent reinforcement learning-based data-driven method for home energy management. *IEEE Trans Smart Grid* 11(4):3201–3211
32. Yu M, Hong SH (2017) Incentive-based demand response considering hierarchical electricity market: a stackelberg game approach. *Appl Energy* 203:267–279

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.