Developing a Recommender Model for Blockchain-Based IoT with Deep Reinforcement Learning

Elnaz Rabieinejad Shahriar Mohammadi Mahdi Yadegari

Information Technology Department, K.N. Toosi University of Technology, Tehran, Iran, elnazrabinezhad@email.kntu.ac.ir
Information Technology Department, K.N. Toosi University of Technology, Tehran, Iran, mohammadi@kntu.ac.ir
Information Technology Department, K.N. Toosi University of Technology, Tehran, Iran, myadegari@mail.kntu.ac.ir

Abstract—**With developments in human societies and the information and communication technology, the Internet of Things (IoT) has penetrated various aspects of daily life and different industries. The newly emerging blockchain technology has become a viable solution to the IoT security due to its inherent characteristics such as distribution, security, immutability, and traceability. However, integrating the IoT with the blockchain technology faces certain challenges such as latency, throughput, device power limitation, and scalability. Recent studies have focused on the role of artificial intelligence methods in improving the IoT performance in a blockchain. According to their results, there are only a few effects on the improvement of IoT-based performance with limited power. This study proposes a conceptual model to improve the blockchain throughput in IoT-based devices with limited power through deep reinforcement learning. This model benefits from a recommender agent based on deep reinforcement learning in the mobile edge computing layer to improve the throughput and select the right mining method.**

Keywords: IoT, blockchain, deep reinforcement learning

I. INTRODUCTION

The Internet of things (IoT) has penetrated, influenced, and facilitated various aspects of daily life. In other words, the IoT refers to the connection of sensors and devices to a network in which they can interact with each other and with users. The IoT has facilitated many tasks for humans by providing the foundation for exchange and communication between different devices without human intervention. At the same time, since this technology encompasses various features and applications, there are great concerns about the security of the IoT-based devices and potential threats [1]. The IoT security challenges include fraudulent node detection, authentication, trust management, data confidentiality, network security, and access control [2]. The blockchain technology is the missing link in the process of dealing with the IoT security and privacy issues. This technology might be exactly what the IoT needs to track billions of connected devices, process transactions, and coordinate equipment [3]. A blockchain is a distributed cryptographic ledger that provides a platform for authentication with no need for a third party by uses a peer-to-peer (p2p) network to measure and monitor the accuracy of the ongoing operations. In fact, a blockchain is a viable technique for enhancing the IoT security and privacy due to its inherent features such as decentralization, security, immutability, and traceability [4]. At the same time, it should be noted that the integration of a blockchain with the IoT faces such challenges as limited capacity and memory of the IoT-based devices, scalability, latency, and low throughput [5, 6].

Artificial intelligence (AI) is a solution to the blockchain–IoT integration challenges [7]. AI can make better decisions, a task which is difficult for humans. A blockchain includes various parameters such as security, efficiency, and distribution that must be dealt with. AI can facilitate decision-making, make the blockchain more automated and efficient, and improve the blockchain security by anticipating attacks and detecting abnormal behavior [7-9]. Reinforcement learning is a branch of artificial intelligence in which an agent adjusts its policy with sequential learning through the rewards received from the result of his action. Unlike supervised and semi-supervised learning techniques, reinforcement learning usually requires no prior knowledge of the environment, a feature which makes reinforcement learning suitable for use in the blockchain [7]. In addition to all the benefits of using artificial intelligence to solve the blockchain–IoT integration challenges, there are certain problems such as the limited power and memory of the IoT-based devices, the scalability paradox, data privacy, and the cost of large-scale blockchain applications [7, 10].

This paper proposes a novel model based on deep reinforcement learning (DRL) for the blockchain-based IoT to deal with challenges. The proposed model considers the low throughput problem in the blockchain-based IoT and the use of a blockchain in resource-constrained IoT-based devices. Throughput is the maximum number of transactions that a system handles per second [11]. A solution to the throughput challenge is to increase the block size and decrease the block construction time; however, this solution increases the transfer time as every miner must check the validity of all transactions in the block [12]. Therefore, a goal of the proposed model is to increase throughput by considering the balance between throughput and transmission delay. The model employs a recommender agent based on

deep reinforcement learning and located in the mobile edge computing (MEC) layer to optimally adjust the size and construction time of the block. The model also determines how to perform the mining tasks (with the miner's device or by offloading to the MEC) in order to minimize the mining delay and energy consumption. Deep reinforcement learning benefits from integrating the two methods of Q-learning and bidirectional long short-term memory (BiLSTM), also known as Q-learning + BiLSTM. In fact, BiLSTM is employed to maintain long-term dependencies and improve Q-learning performance in high-dimensional problems [13].

Contributions: The significant research contributions are listed as below:

● Developing a novel model that consists of six layers including perception, data, network, consensus, mobile edge computing, and application. This model can generally be used for the IoT-based devices with limited resources and rich resources. Considering the power of a device and its required power, it is possible to determine how to perform mining tasks.

● This model proposes a method for adjusting the block size and the block construction time by using the recommender agent, which benefits from Q-learning + BiLSTM, to improve the blockchain throughput.

● It determines the optimal mining method for every miner by using the recommender agent, which employs the Q-learning + BiLSTM. The optimal method is an operation that minimizes the mining delay and energy consumption.

Organization: Section II presents a review of the literature, whereas Section III describes the blockchain-based IoT mode and explains how the recommender agent operates. Finally, Section IV concludes the paper.

II. Related Work

The combination of AI and the blockchain technology has been drawing attention in recent years. It has introduced two sections of artificial intelligence to improve the blockchain technology and its performance in order to enhance the AI performance. Since this study seeks to enhance the use of a blockchain in the IoT, the AI–blockchain literature was reviewed to improve blockchain performance in the field of the Internet of Things. Table 1 compares and reviews these studies. These studies analyze artificial intelligence approaches adopted (machine learning (ML), deep learning (DL), reinforcement learning (RL), deep reinforcement learning (DRL), and their contributions.

Table 1.Related works

Reference	Artificial intelligence approach				Contribution
	RL	*DRL*	*DL*	*ML*	
[5]				✓	Analyzing and categorizing machine learning adoption for a blockchain from different goal-based, layer-based, interaction-based, and smart app-based perspectives
[6]			✓		Providing a decentralized network architecture based on software-defined networking, fog, and blockchain computing as well as a deep learning-based IoT attack detection mechanism
[7]		✓			Improving scalability in a blockchain by considering latency, distribution, and the IoT security through deep reinforcement learning
[8]			✓		Providing a collaborative intrusion detection system based on the blockchain framework and deep learning to for the IoT security and privacy
[9]			✓		Designing a blockchain-based intelligent architecture for the IoT to maintain security and decentralized processing of big data from the IoT-based devices
[10]		✓			Providing an optimal privacy model to offload mining and other computing mobile tasks to the mobile edge server through deep reinforcement learning
[11]	✓				Surveying on challenges and advantages of reinforcement learning in the blockchain-based IoT
[12]			✓		Improving the fork problem in a blockchain by managing and decreasing the block transmission delay through deep learning
[13]		✓			Adopting deep reinforcement learning to increase the successful transfer rate of the IoT requests to a blockchain

III. Proposed Model

The research environment includes a space that is geographically limited but consists units or adjacent buildings. The area intended for the model can be thought of a smart building consisting of several separate departments or a university campus including adjacent buildings under the same management. In this environment, objects are different smart devices, a limited number of which are available. These smart objects are composed of industrial devices with high power as well as smart devices with low power and memory. For instance, the environment includes smartphones that have problems with a blockchain running. Resources-constrained IoT-based devices can offload mining tasks to the mobile edge computing layer. In [1], a model was developed to improve the blockchain throughput for the IoT-based devices through the deep Q-learning method. In this paper, the feedforward neural network was used for deep reinforcement learning. Unlike the feedforward neural network, the recursive neural network can retain and use information regarding past states to better predict better the value of Q [2, 3]. At the same time, the reviewed paper considered industrial IoT (IIoT)-based devices. Often designed for larger objects than smartphones or wireless devices, the IIoT aims to

connect industrial appliances [4]. Therefore, the current paper proposes a model to improve throughput for resource-constrained IoT-based devices by using a Q-learning method equipped with a recursive neural network.

A. Logical Layers of the Proposed Model

Figure 1 demonstrates the logic layers of the proposed model. These layers are described as below:

- Perception layer: This layer includes all the IoT-based devices. Because of a private blockchain, the IoT-based devices need to be registered first to interact with the blockchain.
- Data layer: The data layer consists mainly of transactions and data blocks. Containing a few transactions, every block is chained to the previous block and forms an orderly list of blocks. The block consists of two sections, *i.e.* the block header and the main data. The header determines the block metadata including the block version, the hash of previous and current blocks, the timestamp, the Merkel root [14], and other information. In every block, all transactions are hashed separately through the hash algorithm. These hash values are then combined in pairs and re-hashed until a single hash value is obtained. This value is known as the Merkel tree hash value. The Merkel tree hash value can be used to easily and quickly check all transactions recorded in a block [15]. In every block, the main data contain all the transactions in it—the data of this layer are encrypted by the hash function and the asymmetric encryption algorithm [16].
- Network layer: This layer includes the communication structure, a p2p network, and an authentication mechanism. The IoT-based devices use different communication mechanisms to transmit information. A communication infrastructure is the basis for the data block transfer and communication of members in the p2p network. In a p2p network, nodes are connected via wired and wireless connections, allowing block propagation between blockchain nodes. Every node is responsible for checking the accuracy of the received block and distributing it to the neighboring nodes [17]. A member of the p2p network was selected as the registrar responsible for authenticating and registering the new IoT-based devices that want to participate in the blockchain.
- Consensus layer: Consensus is the foundation of a blockchain and is a mechanism for trusting the data used. This layer is responsible for reaching a distribution agreement for the accuracy of a data block. The consensus can be achieved through various algorithms of proof of work (PoW), proof of stack (PoS), practical byzantine fault tolerance (PBFT), etc. Every blockchain must have a consensus mechanism, according to which it rewards the nodes that participated in implementing the consensus algorithm. Without the reward, the nodes have no motivation for approving the blocks that do not belong to them [18].The proposed model provides no rewards for the authentication process because all of these processes are part of its workflow. All of these devices are internal members of the company [16].
- Mobile edge computing (MEC) layer: This layer includes computing and storage resources. If none of the IoT-based devices have the computing power and memory required to perform blockchain operations, they can perform tasks by using the resources within this layer [10]. This layer also includes a recommender agent that, by receiving data and using deep reinforcement learning, can adjust the block size and the block construction time to improve throughput and decision-making on local mining or offloading to minimize latency and energy consumption.
- Application layer: In this layer, the proposed model can be used by the lower layers in various industries. These services include smart homes, smart universities, and smart health.

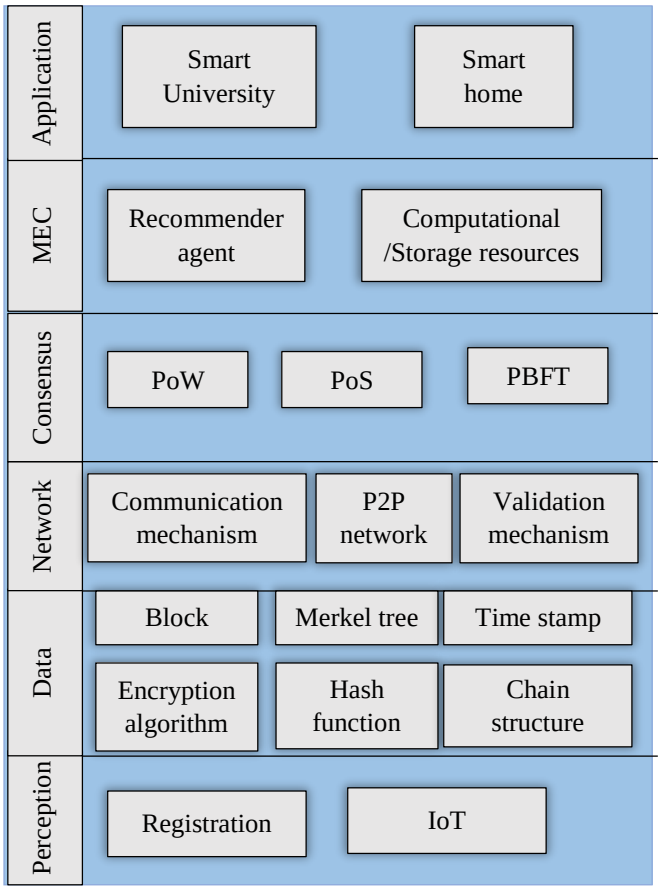

Figure 1.Logical layers of the proposed model

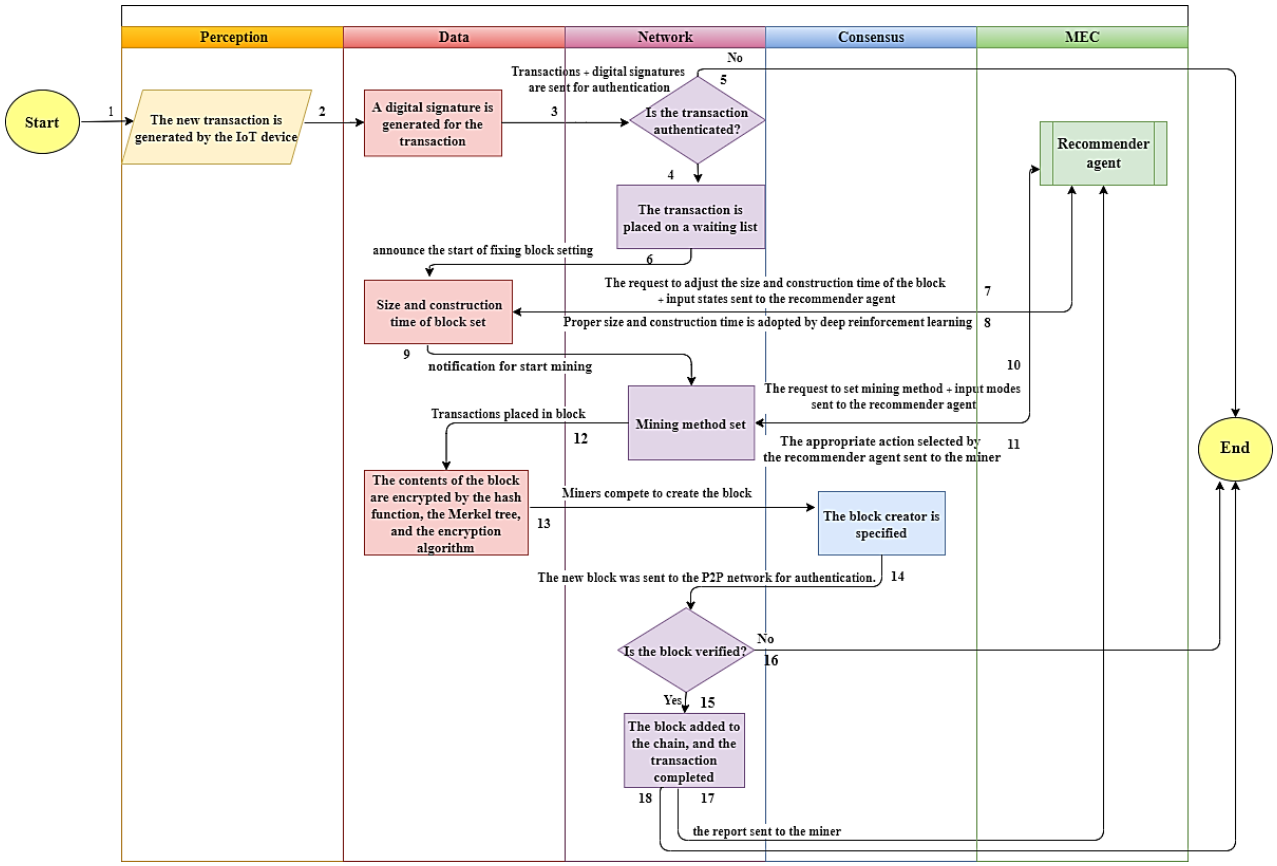

Figure 2. The proposed model workflow

B. Model Workflow

According to Figure 2, the workflow of the proposed model is as follows:

1. A new request is generated on the IoT-based device.
2. The transaction is hashed by the hash function, and then the private key of the IoT-based device is applied to it in order to generate a digital signature.
3.The transaction is broadcast along with its digital signature for authentication on the p2p network. Every member of the p2p network employs the public key to decrypt the encrypted transaction and then compare it with the received transaction hash. This ensures the integrity of the transaction and the credibility of the generating device.
4. The transaction is validated by the p2p network members and is then placed on a waiting list for completion.
5. If members do not approve the transaction of the p2p network, the process will end and will be notified to the requesting device.
6. When an approved transaction is available, a notification will be sent to set the block size and the block construction time.
7. To adjust the block size and the block construction time, a setting request is sent along with the current states of the system including the average transaction size, the computing power of the IoT-based devices, and the transfer rate between the two IoT-based devices to the recommender agent. The sent states are then input to the policy section to increase throughput.
8. The recommender agent adjusts the block size and the block construction time to improve the throughput through Q-learning + BiLSTM.
9. After setting the block size and the construction time, a notification is sent to the miners to start the mining process.
10. Miners send a request to the recommender agent to decide how to perform the mining operation (by using the device itself or offloading to the mobile edge computing layer). Along with the current system state, this request includes the system transactions, the power of the wireless channel between the miner device and the mobile edge computing server, and the amount of computing energy of the device. These are the policy-making section input states regarding how to perform tasks to minimize delay and energy consumption.
11.The recommender agent, uses the Q-learning + BiLSTM technique to select the appropriate right mining method for the miners to minimize latency and energy.
12. After the policies are set, the transaction placed in the block and the block contents are encrypted by the hash function, the encryption algorithm, and the Merkel tree.
13. The IoT-based devices compete, and the consensus process determines the block creator.

14. The blocks created for authentication are broadcast on the same network.
15. If a block is confirmed, it is then added to the chain, and the transaction is completed.
16. If a block is not approved, the process ends.
17. At the end of each round and transaction, a report is sent by the blockchain system from each miner to the recommender agent. The report is a sequence (state, action, reward, and next state) that shows the amount of reward or loss received in the s^t state with the a^t action and the next state which is reached by acting. The reward or loss received in the blockchain settings section indicates the amount of throughput earned by the block construction size and time. The reward in each miner excludes the total latency and energy consumption. As the energy and latency increase, the amount of reward decreases. The report is stored in the recommender memory and is then used to train BiLSTM.
18. The process ends.

C. Recommender Agent

Q-learning is employed in the deep reinforcement learning section, which is equipped with a recursive neural network. In addition to solving the problem of learning Q in high dimensions, the recursive neural network can also maintain long-term dependencies in it [2].Unlike the feedforward neural network, learning signals move in only one direction from the input layer to the hidden layers and then to the output layer in the recursive neural network. It contains a feedback loop that allows previously acquired information to remain in the network and not be lost. Therefore, it is able to preserve and use past data to predict Q values better. The recursive neural network has shown good performance in partially observable Markov decision processes, and problems can easily be solved by the Markov decision process [3].

A problem with the recursive neural network is the vanishing gradient; thus, long short-term memory(LSTM) uses an extended recursive neural network with additional gates to memorize long sequences of its input data. This method overcomes the problem of vanishing gradient and makes it possible to retain information over long periods. The vanishing gradient occurs when the input data disappear over time due to the large layer passage, and long-term dependencies become challenging to maintain [19].
The gate inspired the ability to make decisions about keeping or ignoring information. A long short-term memory model takes essential features from the input and stores them for a long time. Decisions to delete or retain information are made based on the weight values assigned to the training process. BiLSTM is a development of the long short-term memory in which two long short-term memories are applied to the output data. In the first round, a long short-term memory is employed in the input sequence (for example, the forward layer), whereas another long short-term memory is used in reverse (for example, the backward layer). BiLSTM improves the learning of long-term dependencies and the morel accuracy [19].

According to Figure 3, the state is first received from the environment in the reinforcement learning process in the proposed model. The neural network selects the action that maximizes the Q function value (in the absence of an appropriate action for the state, a random action is taken). Available transactions for processing, the amount of computational energy of the miner device, and the wireless channel power between the miners and the mobile computing edge are received while adopting reinforcement learning to find the optimal policy in performing mining tasks. Policies should be set to select an action at each stage to minimize the amount of energy and time spent performing the mining task. The time required to offload tasks to the mobile edge computational layer includes the delay in loading a task into the mobile computing edge, the queue delay indicating the time that the task has to wait for processing, and the delay in processing a task through the mobile computing edge. The energy spent offloading tasks to the mobile computing edge includes the energy to load tasks to the mobile computing edge, the energy required to process the task, and the energy needed to execute the mobile computing edge. If a device wants to perform its tasks locally, the consumed time and energy include the time spent processing each bit of data as well as the amount of energy spent processing each bit of data [10]. Algorithm 1 shows the DRL-based task offloading algorithm.

Using deep reinforcement learning to increase throughput, the current state of the environment includes the transaction size, the computing power of IoT-based devices, and the transfer rate between different IoT-based devices. The block size can be increased, whereas the block construction time can be decreased in order to improve the throughput. Generally, the time required to process data and add a new block to the chain includes the time it takes to generate a new block and the time it takes to broadcast the block and authenticate it. As the block size increases due to increased transactions per block, the block transfer takes more time, whereas the total latency increases [10]. To achieve the desired latency, it is assumed that blocks must be issued and validated at regular intervals. The time interval between the final approval of a block in the blockchain must be less than the block generation time interval. The earlier generated block must complete processing before producing another block. If this condition is not met, zero rewards will be given. Algorithm 2 shows the DRL-based throughput optimization algorithm.

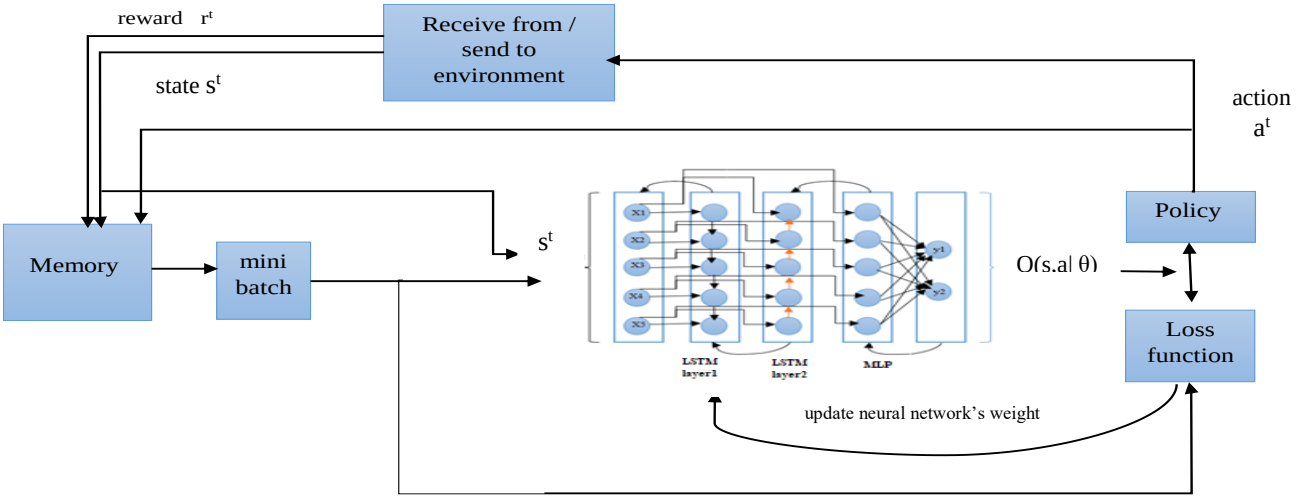

Figure 3.Q learning + BiLSTM recommender agent

Algorithm1. DRL-based task offloading algorithm

1. A memory D is defined by weight N.
2. We define a BiLSTM with pair input (state, action) and an estimated value-function function with random weight θ.
3. **For** t = 1, 2,... **do the following step**s:
4. The device observes the transactions in the blockchain. These transactions include newly arrived transactions (D0) and transactions in the buffer (D1). (Processing priority assigns to transactions in the buffer.)
5. Estimate the wireless channel's power (Gn) between the nth miner and the mobile computing edge.
6. The computational power of the miner (En) is received.
7. the input state as $S^t=\{D_1^t,D_0^t,G^t,E_n\}$ defines.
8. The action (a^t)with ε probability is selected randomly; otherwise, we choose the action that maximizes the value of the q value function in this case. (a^t= argmax Q (s^t, a, θ))
9. The selected action or the miner does its task or offloads it to the mobile computing edge layer. Rewards and the next state are observed. Rewards each stage describe as $R^n=-C^t(s,a)$ .In this formula C^t (s, a) represents the total cost and energy spent.
10. The performed transaction is stored in memory as a sequence (s^t,a^t,r^t,s^{t+1}).
11. A small batch of transactions in memory is randomly sampled.
12. The value of the Q function is calculated by the equation $y^k = r^k+\Upsilon * \max Q(s^{k+1},a' , \theta)$. This value indicates the amount of reward we expect to achieve in the future.
13. We use the gradient descending with the formula $(y^k – Q (s^k, a^k, \theta))^2$ to reduce the loss.
14. **End**

IV. Discussion and Conclusions

The blockchain technology is an efficient strategy to improve the IoT security vulnerabilities due to its inherent properties such as cryptography, transparency, immutability, and decentralization. However, the blockchain–IoT integration faces certain challenges such as latency, scalability, throughput, and the use of resource-constrained IoT-based devices. Known as a subfield of artificial intelligence, deep reinforcement learning is an excellent strategy to solve the blockchain–IoT integration problems. This study proposed a novel model through deep reinforcement learning to increase the throughput with usability in the IoT-based devices with limited power. Using a recommender agent, this model adjusts the block size and construction time to increase the throughput and determine the optimal way to perform mining for the miner. Therefore, this paper proposes a novel model for the blockchain-based IoT. This model can be used for IoT-based devices with low or rich resources.

It employs Q-learning + BiLSTM to adjust an optimal policy to perform the mining task and improve the throughput. However, collecting all the information in the mobile edge computing layer and using a recommender agent to assign policies in the event of a hacker attack would entail certain risks such as information leakage and privacy breaches. At the same time, deployment of the recommending agent in the mobile computing layer would make it possible for devices with limited resources to use the agent simply. However, sending and receiving requests can cause delays, especially when there are many requests. In future works, we are going to improve the proposed model to solve latency and security problems and present the results through simulation.

Algorithm 2. The DRL-based throughput optimization algorithm

1. A memory D is defined by weight N.
2. A BiLSTM is defined with a pair input (state, action) and an estimated value function with a random weight θ.
3. **For** t = 1, 2, ... **take the following steps:**
4. The average transaction size (χ) in the blockchain is calculated.
5. The computing power (c) of IoT-based devices that are members of the blockchain network is calculated.
6. The transfer rate between nodes *i* and *j* of the IoT is calculated as Rij.
7. The input state is defined as (s^t = (χ, c, R).
8. The action (a^t) with a probability of ε is selected randomly; otherwise, the action that maximizes the Q value function is selected in this case (a^t= argmax Q (s^t, a, θ)).
9. According to the selected action, the block size and the block construction time are adjusted, and the next reward and mode are observed. If the condition for finalizing block *i* before generating block *i+1* is true, the $\frac{\lfloor S^B/\chi \rfloor}{T^I}$ reward is given; otherwise, no reward will be given. In the mentioned formula, S^B indicates the block size, whereas χ indicates the average transaction size, and T^I indicates the block construction time.
10. The performed transaction is stored as a sequence (s^t,a^t,r^t,s^{t+1}) in memory.
11. A small transaction batch is randomly sampled in memory.
12. The value of the Q function is calculated by the equation $y^k = r^k + \Upsilon * \max Q(s^{k+1}, a', \theta)$. This value indicates the amount of the expected reward for the future.
13. The descending gradient is used through the formula $(y^k - Q(s^k, a^k, \theta))^2$ to reduce the loss.
14. **End.**

REFERENCES

[1] S. Tanwar, et al., "Machine learning adoption in blockchain-based smart applications: the challenges, and a way forward". IEEE Access, vol.8, pp. 474-488,2019

[2] S. Rathore, B.W. Kwon and J.H. Park, "BlockSecIoTNet: blockchain-based decentralized security architecture for iot network". Network Computer Applications,vol.143,pp. 167-177, 2019

[3] Y. Liu, F.R. Yu, X. Li, H. Ji, and V.C. Leung, "Blockchain and machine learning for communications and networking systems". IEEE Communications Surveys, vol.22(2), pp. 1392-1431, 2020

[4] O. Alkadi, N. Moustafa, B. Turnbull and K.-K.R. Choo, "A deep blockchain framework-enabled collaborative intrusion detection for protecting iot and cloud networks". IEEE Internet of Things Journal, 2020.

[5] S.K. Singh, S. Rathore and J.H. Park, "Blockiotintelligence: A blockchain-enabled intelligent iot architecture with artificial intelligence". Future Generation Computer Systems,vol. 110, pp. 721-743, 2020

[6] D.C. Nguyen, P.N. Pathirana, M. Ding, A. Seneviratne, and S. Management, "Privacy-preserved task offloading in mobile blockchain with deep reinforcement learning". IEEE Transactions on Network, 2020.

[7] F. Jameel, et al., "Reinforcement learning in blockchain-enabled iiot networks: a survey of recent advances and open challenges". Sustainability, vol.12(12), pp. 5161,2020.

[8] F. Jameel, M. Nabeel, M.A. Jamshed and R. Jäntti. "Minimizing forking in blockchain-based iot networks". in 2020 IEEE International Conference on Communications Workshops (ICC Workshops), 2020.

[9] Z. Xiong, et al., "The best of both worlds: a general architecture for data management in blockchain-enabled internet-of-things". IEEE Network,vol. 34(1), pp. 166-173, 2020.

[10] M. Liu, F.R. Yu, Y. Teng, V.C. Leung, and M. Song, "Performance optimization for blockchain-enabled industrial Internet of Things (iiot) systems: a deep reinforcement learning approach". IEEE Transactions on Industrial Informatics, vol.15(6), pp. 3559-3570,2019

[11] C. Chen, V. Ying and D. Laird, "Deep q-learning with recurrent neural networks". Stanford Cs229 Course Report,vol. 4,pp. 3. 2016.

[12] D. Han, K. Doya and J. Tani, "Self-organization of action hierarchy and compositionality by reinforcement learning with recurrent neural networks". Neural Networks,vol.129,pp. 149-162,2020.

[13] P. Helmiö. "Open source in industrial internet of things: a systematic literature review", 2017.

[14] R.C. Merkle. "Protocols for public key cryptosystems. in 1980 IEEE Symposium on Security and Privacy", 1980.

[15] Y. Yu, Y. Li, J. Tian and J. Liu, "Blockchain-based solutions to security and privacy issues in the internet of things". IEEE Wireless Communications, vol.25(6), pp. 12-18, 2018.

[16] W. Villegas-Ch, X. Palacios-Pacheco and M. Román-Cañizares, "Integration of iot and blockchain to in the processes of a university campus". Sustainability, vol.12(12), pp. 4970, 2020

[17] L. Ismail, H. Materwala and S. Zeadally, "Lightweight blockchain for healthcare". IEEE Access, vol.7, pp. 149935-149951, 2019.

[18] A.D. Dwivedi, G. Srivastava, S. Dhar and R. Singh, "A decentralized privacy-preserving healthcare blockchain for iot". Sensors, vol.19(2). pp. 326, 2019

[19] S. Siami-Namini, N. Tavakoli and A.S. Namin. "The performance of lstm and bilstm in forecasting time series". in 2019 IEEE International Conference on Big Data (Big Data), 2019.