Incentive Mechanism for Edge Computing-Based Blockchain: A Sequential Game Approach

Wenlong Guo, Zheng Chang, *Senior Member, IEEE*, Xijuan Guo, Peiliang Wu, and Zhu Han, *Fellow, IEEE*

***Abstract*—Due to its distributed characteristics, the development and deployment of the blockchain framework are able to provide feasible solutions for a wide range of Internet of Things (IoT) applications. While the IoT devices are usually resource-limited, how to make sure the acquisition of computational resources and participation of the devices will be the driving force to realize blockchain at the network edge. In this article, an edge computing-based blockchain framework is considered, where multiple edge service providers (ESPs) can provide computational resources to the devices for mining. We mainly focus on investigating the trading between the devices and ESPs in the computational resource market, where ESPs act as the sellers and devices act as the buyers. Accordingly, a sequential game model is formulated and by exploring the sequential Nash equilibrium (SE), the existence of the optimal solutions of selling and buying strategies can be proved. Then, a deep Q-network-based algorithm with modified experience replay update method is applied to find the optimal strategies. Through theoretical analysis and simulations, we demonstrate the effectiveness of the proposed incentive mechanism on forming the blockchain via the assistance of edge computing.**

***Index Terms*—Blockchain, edge computing, incentive mechanism, mining.**

Manuscript received 30 September 2021; revised 17 February 2022; accepted 25 March 2022. Date of publication 6 April 2022; date of current version 9 September 2022. This work was supported in part by the NSFC under Grant 62071105, in part by the China Postdoctoral Science Foundation under Grant 2018M631620, in part by the Beijing Natural Science Foundation under Grant 4202026, in part by the NSFunder Grant CNS-2107216 and Grant CNS-2128368. Paper no. TII-21-4284. *(Corresponding author: Zheng Chang.)*

Wenlong Guo, Xijuan Guo, and Peiliang Wu are with the College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China (e-mail: guowenlong2106@163.com; xjguo@ysu.edu.cn; peiliangwu@ysu.edu.cn).

Zheng Chang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China, and also with the Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland (e-mail: zheng.chang@jyu.fi).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: zhan2@uh.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3163550.

Digital Object Identifier 10.1109/TII.2022.3163550

I. Introduction

THE emergence of Internet of Things (IoT) will be the driving force of the development of the future information and communication technology (ICT) [1]. However, due to distributed and resource-constrain natures of IoT, the security mechanism design is critical for its wide deployment. Recently, the blockchain has evolved from the original digital currency to extensive IoT applications due to its distributed, tamper-resistant, retrospective, and transparent features [2]. As a well-known decentralized ledger-based framework, blockchain is able to provide secure transactions and trust in a trustless network environment. The node (or so called miner) in blockchain executes some computation tasks to obtain an unverified block. When one miner successfully addresses the consensus protocol, it could report the result to blockchain for verification. The miners will reach consensus when the verification is correct and then obtain rewards caused by the computing for consensus process (or so called mining). As we can see, the blockchain has its great potential to provide a secure IoT platform, especially when facing large-scale accesses.

Although blockchain has been widely adopted in many applications, its application in mobile services is still limited. Before adding or publishing to the blockchain, some complex computation problems, e.g., PoW puzzle, are solved to secure the integrity and validity of transactions. In this context, to facilitate blockchain applications in future mobile IoT systems, mobile edge computing can play a significant role [3]. Leveraging the computing capabilities of edge computing system, the miners with insufficient hash power can rent computational resources from edge service providers (ESPs) [4]. Thus, how to incentivize the miners to participate the blockchain process and obtain the computational resources from ESP or perform computation offloading is of profound significance [3]–[6]. Meanwhile, how to encourage multiple ESPs to provide computational resources to the miners is also crucial. Such observations motivate us to seek for game theoretic approaches to explore the interactions between multiple ESPs and multiple miners.

Recently, there are increasing interests on utilizing blockchain incentive to design the blockchain system. There are several works utilizing the mathematical methodology on designing the incentive schemes for multiple players [7]–[14]. Jiao *et al.* [7] designed an approximation algorithm and study how to maximize the social welfare of blockchain network. Xiong *et al.* [8] proposed to investigate the optimal profits of the ESPs and/or miners under different pricing strategies via game theoretic approaches. Houy [9] suggested a two-miner model to find the strategy of utilizing computation resource and find the Nash Equilibrium (NE) in the blockchain. In [10], Lewenberg *et al.* presented a cooperative game model to study the dynamic equilibrium problem that when the miners choose to participate in the mining pool. Combining the blockchain reputation and incentive mechanism, Avyukt *et al.* [11] adopted the game theoretical methods to formulate the multibuyer and multiseller data marketplace, and realize the credible evaluation of a higher-quality ecosystem. Liu *et al.* [12] proposed a blockchain-based double auction protocol, in which multiple buyers and sellers could quickly optimize a balance market cleaning price, to ensure integrity, efficiency, and incentive. For P2P energy trading, Kumari *et al.* [13] formulated a blockchain-based scheme and Q-learning algorithm to optimize the decision-making process to improve system security, transaction latency, and participant rewards. Huang *et al.* [14] proposed to achieve social welfare maximization by a truthful double auction mechanism, which the incentive and fairness of the buyers and sellers could be guarantee.

However, the current works rarely analyze the competitive relationship like seller–seller, buyer–buyer, and the dynamic competition between multiple buyers and sellers in the edge computing-based blockchain system. Motivated by the aforementioned observations, in this article, we aim at proposing a novel incentive mechanism for an edge computing-based blockchain, in order to find the optimal purchase and pricing strategies for all the involved ESPs and miners. The main contribution can be summarized as follows.

1) We consider a multi-ESPs and multiminers scenario. In the considered system, to encourage the devices to participate the mining process and ESPs to provide the computational resources, we aim to explore the relations and interactions between these two parties.
2) We mainly focus on investigating the trading process between the devices and ESPs in the computational resource market, where ESPs and miners can act as the sellers and buyers, respectively. Accordingly, a sequential game model is formulated. Then, we have proved the existence and uniqueness of the NE, and applied backward induction to find the global optimal solution.
3) To optimize strategies, a deep Q-network-based algorithm with modified experience replay update is applied to find the optimal strategies. The proposed mechanisms can help both parties obtain the best utilities in a dynamic manner and essentially stimulate the development of blockchain system. Numerical results demonstrate the effectiveness of the proposed incentive mechanism.

The rest of this article is organized as follows. The designed system model is introduced in Section II. In Section III, we formulate the resource trading market. Then, Section IV formulates the sequential equilibrium problem. Section V proposes a deep reinforcement learning (RL)-based algorithm to obtain optimal solution. In Section VI, simulation study is conducted with detailed discussions. Finally, Section VII concludes this article.

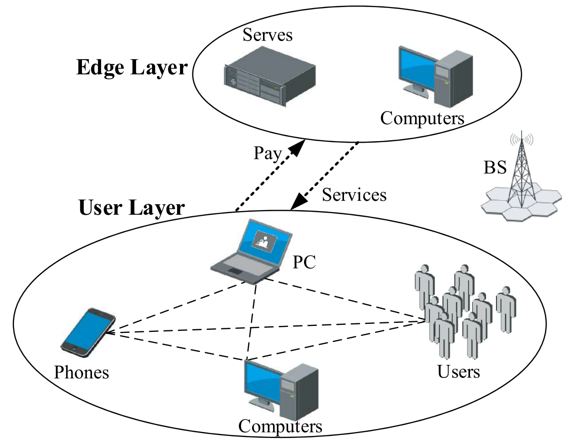


Fig. 1. System model.

II. System Model

A. System Assumptions

We consider an edge computing-assisted blockchain system with M ESPs and N miners. Each ESP can provide homogeneous computational resource services to all the miners. Miners pay for the computational resources, in form of offloading the computing tasks of PoW puzzle to the ESPs. On blockchain, once the clients publish the verified requests, miners can offload the computing tasks to all the ESPs through dedicated channels via the wireless connection, then ESPs can obtain returns by providing recourse. Each ESP can provide computational services to multiple miners at the same time, and so do miners. As shown in Fig. 1, consensus mechanism makes it necessary for the miners to immediately handle the PoW puzzles.

We assume that prices set of computational resource of ESP j is $\mathbf{p}^j = [p_1^j, \ldots, p_i^j, \ldots, p_N^j]^T$, where p_i^j is the price of ESP j for miner i. We assume $p_i^j \in [p^{\min}, p^{\max}]$, where $p^{\min}$ and $p^{\max}$ are the minimum and maximum prices, respectively. The set of prices of computational resource of all the ESPs is $\mathcal{P} = \{\mathbf{p}^1, \ldots, \mathbf{p}^j, \ldots \mathbf{p}^M\}$. The set of strategies (the amounts of purchasing) of all the miners is $\mathcal{S} = \{s_1, \ldots, s_i, \ldots, s_N\}$, where s_i is the purchase strategy of miner i. We assume $s_i \in [s_{\min}, s_{\max}]$, where $s_{\min}$ and $s_{\max}$ are the minimum and maximum purchase quantity, respectively. Meanwhile, The computational capability or hash power proportion of miner i is α_i, which is expressed as

$$\alpha_i = \frac{s_i}{\sum\limits_{i \in N} s_{-i}}. \tag{1}$$

In this article, we assume the communications between the ESP and miners are perfect, and we did not consider the problem during the transmission process, caused by channel variation or spectrum usage. In practice, when the size of offloaded task is small, the transmission will not be the bottleneck due to the relatively sufficient communication and computing resources owned by the ESP. Thus, we mainly focus on the trading interaction between sellers (i.e., ESPs) and buyers (i.e., miners), to study the optimal strategy to enable MEC-based blockchain.

B. Mining Process and Consensus Mechanism

The solution of PoW puzzle is considered as a stochastic process following the Poisson distribution [19] with parameter λ. Then, the probability of successfully solving the problem μ_i is defined as

$$\mu_i = \alpha_i e^{-\lambda t_i} \tag{2}$$

where the computing delay t_i is related to the transactions or block size of each block π_b. Then, we have $t_i = \varsigma_i \pi_b$ where ς_i is a constant parameter for miner i, where $-i$ means all the remaining miners except i.

C. Degree of Satisfaction

For miner i, the amount of purchased computational resources depends on its cost and the degree of satisfaction (DoS) obtained through the mining process. For miner i, we have following definition of DoS $\mathcal{D}(s_i)$:

$$\mathcal{D}(s_i) = \log_2(1 + \mu_i). \tag{3}$$

For (3), the logarithmic function satisfies the constraint of $0 \le \mathcal{D}(s_i) \le 1$, where it is the possibility of obtaining profits. In addition. the logarithmic function of the DoS is convex and it can indicate that DoS of the miner increases as the proportion of computational resources increases, and there must exist a maximum value that optimizes the satisfaction of miners [20].

III. Resource Trading Market

First, we formulate the resource trading between the ESPs and miners as a sequential game in Section III-A and then present the corresponding game model among the ESPs and miners.

A. Sequential Game Formulation

For service provisioning, a two-stage game model is formulated. Once the strategies in the previous stages were determined, the players (either ESPs or miners) in the later stage can select the corresponding strategies. Then, the trades between multiple ESPs and miners are considered as sequential decision-making problem, where the players can make successive observations before the final decision is made. In this article, we mainly focus on the trading interaction between sellers/ESPs and buyers/miners.

Correspondingly, we formulate the trading between the ESPs and miners as a sequential game $\mathcal{G}$ with incomplete dynamic information. The competition among the miners is modeled by noncooperative game $\mathcal{G}^M$ and competition among the ESPs is defined as multioligopoly Cournot game $\mathcal{G}^E$. Sequential game is a model for making decisions along time slots based on the sequential rationality assumption. That is, all players dynamically adjust their strategies to find optimized decisions based on the current observations.

B. Mining Competition Among Multiple Miners

Blockchain players can be rewarded by participating the mining. The computational resource acquisition process of the miners is modeled as a noncooperative game, where the players are miners and the strategies are the purchasing amount from ESPs for solving the PoW puzzle. The utility of the miner consists of profit and cost functions. The profit $E(s_i)$ can be defined as follows:

1) The profit function $E(s_i)$ is a combination of fixed reward R_f and performance reward R_p, which is

$$E(s_i) = \mu_i (R_f + R_p). \tag{4}$$

2) The fixed reward R_f is the constant reward for computing a newly generated block. The fixed reward of blockchain can be regarded as an attenuation function of which the half-life is T. That is

$$R_f = R_f^{\max} \left(\frac{1}{2}\right)^{\frac{t_c}{T}} \tag{5}$$

where $R_f^{\max}$ is the constant reward from genesis block and t_c is the time point when miners start mining.

3) The performance reward R_p is related to the volume of transactions contained within the generated block, e.g., the size of each block. We have following definition:

$$R_p = r\pi_b, \tag{6}$$

where r is an evaluation factor and π_b is the size of block.

4) The participant reward $R_{\varepsilon,i}$ depends on the degree of participation in the computing process while the new block is generated, i.e.,

$$R_{\varepsilon,i} = \varepsilon \alpha_i \tag{7}$$

where ε is an evaluation factor.

5) The purchase expenditure c_i^E is paid to the ESP for computational resource, i.e.,

$$c_i^E = p_i^j s_i. \tag{8}$$

The computational cost $c_{i,M}$ is the consumption generated during the calculation process.

For miner i, the total profit comes from mining process and the cost is related to the purchase of computational resources. With the above definitions, the utility of miner i is given as

$$U_i^M(\mathcal{S}, \mathcal{P}) = E(s_i) + R_{\varepsilon,i} - c_i^E - c_{i,M}. \tag{9}$$

C. Market Competition Among Multiple ESPs

As mentioned above, the competition via multiple ESPs is formulated as a multioligopoly Cournot game. Multiple ESPs acting as the sellers select the pricing strategies at the same time, and they cannot observe each others' strategies and utilities. Based on the DoS of miners and cost c_E^j, the utility function of the ESP j can be defined as follows:

$$U_j^E(\mathcal{S}, \mathcal{P}) = \sum_{i=1}^{N} p_i^j s_i \mathcal{D}(s_i) - c_E^j. \tag{10}$$

IV. Sequential Game Analysis

In this part, the objectives of both stages are formulated in Section IV-A. The sequential game is then transformed into a static game by Harsanyi transformation in Section IV-B, and the existence and uniqueness of the Stackelberg equilibrium (SE) of two-stage static game are discussed in two separated cases in Sections IV-C and IV-D.

A. Problem Formulation

In the formulated game, multiple miners need to compete for the resources and the optimal utility in a noncooperative manner. The miners cannot observe the each others' information (e.g., purchasing demand and the probability of successfully mining) and the strategies of all miners are executed at the same time. Thus, the subgame of miners is considered as a static game with incomplete information. In this stage, the optimization problem (P1) is formulated as

$$\begin{aligned} \text{P1}: \ & \max_{\mathcal{S}} U_i^M(\mathcal{S},\mathcal{P}) \\ & s.t.\ \mathcal{D}(s_i) \in [0,1] \end{aligned} \tag{11}$$

The competition among the ESPs is modeled as a multi-oligopoly Cournot game. The ESPs cannot obtain each other's information and the subgame of ESPs is also a static game with incomplete information. In this stage, the subgame of the ESP aims at addressing the following problem, i.e.

$$\begin{aligned} \text{P2}: \ & \max_{\mathcal{P}} U_j^E(\mathcal{S},\mathcal{P}) \\ & s.t.\ \mathcal{D}(s_i) \in [0,1] \end{aligned} \tag{12}$$

Based on the presented sequential game model, a two-stage iterative method is required to reach a SE. This two-stage update will iterate until the conditions in *Definition 1* are satisfied.

Definition 1: Let $\mathcal{S}^*$ be a solution for P1 and $\mathcal{P}^*$ denotes a solution for P2, Then, the point $(\mathcal{S}^*,\mathcal{P}^*)$ is a Sequential equilibrium for the game if for any $(\mathcal{S},\mathcal{P})$ the following conditions are fulfilled:

$$\begin{aligned} U_j^E(\mathcal{S}^*,\mathcal{P}^*) &\geq U_j^E(\mathcal{S},\mathcal{P}) \\ U_i^M(\mathcal{S}^*,\mathcal{P}^*) &\geq U_i^M(\mathcal{S},\mathcal{P}). \end{aligned}$$

In the considered two-stage game, the miners choose their purchasing strategies after observing the information of all the ESPs. The miners inevitably purchase the resources that can optimize their own utilities, and the ESPs will accordingly adjust their pricing strategies to reach equilibrium in a dynamic manner. The overall game is considered as a sequential game with incomplete information. As mentioned, ESPs may tend to obtain the miners' private information to complete the formulated game. Based on the historical interaction records (e.g., incentive, consumption, and probability) in the computing resource trading market, ESPs can predict the private information and formulate the next-step strategies. In addition, the communication between the ESPs and miners can also contains some of the information, which could accelerate the decision-making process.

B. Harsanyi Transformation

For the formulated sequential game with incomplete information, we choose to add a virtual player Ω to transform the dynamic game into a two-step static game [15]. After Ω chooses the participants, who will formulate the strategies in the next step, the dynamic game is transformed into a two-stage static game. Backward induction method is then used to gradually reverse from the later stage of decision-making to the previous stage. That is, the study of subgame NE of the previous stage will have to add the later equilibrium as the basis. When the subgame in each stage reaches NE, the game will turn into the SE, i.e., the global optimal solution of formulated problem.

In the following, two decision-making scenarios are discussed separately: ESP-first-select case and miner-first-select case, based on which set of players act first in the second step of Harsanyi Transformation. Next, we turn each of cases into a two-stage decision-making problem, and study the SE of these two cases, respectively.

C. ESP-First-Select (EFS) Case

In this part, we study the case that ESPs first set the price. Thus, in the first stage, the ESPs first set the price, and the miners purchase the resources from the ESPs in the second stage. In the following, backward induction method is used to solve the optimization problem through reverse deduction.

1) Game of Miners in EFS: The utility function of miner i obtaining computational resources from ESP j can be expressed as follows:

$$U_i^M = e^{-\lambda t_i}\alpha_i(R_f+R_p) + \varepsilon\alpha_i - p_i^j s_i - c_{i,M}. \tag{13}$$

Take the first and second derivatives of U_i^M with respect to s_i, respectively, we obtain that

$$\begin{cases} \frac{\partial U_i^M}{\partial s_i} = e^{-\lambda t_i}\frac{\partial \alpha_i}{\partial s_i}(R_f+R_p) + \varepsilon\frac{\partial \alpha_i}{\partial s_i} - p_i^j \\ \frac{\partial^2 U_i^M}{\partial s_i^2} = \left(e^{-\lambda t_i}(R_f+R_p)+\varepsilon\right)\frac{\partial^2 \alpha_i}{\partial s_i^2}. \end{cases} \tag{14}$$

The first and second derivatives of α_i with respect to s_i is given as

$$\begin{cases} \frac{\partial \alpha_i}{\partial s_i} = \frac{\sum\limits_{i\in N} s_{-i}}{\left(\sum\limits_{i\in N} s_i\right)^2}, i \in N \\ \frac{\partial^2 \alpha_i}{\partial {s_i}^2} = -2\frac{\sum\limits_{i\in N} s_{-i}}{\left(\sum\limits_{i\in N} s_i\right)^3}. \end{cases} \tag{15}$$

Then, we are able to find U_i^M is convex with respect to s_i^*. Accordingly, there must be at least one s_i^*, which enables to optimize the utility of miner i while the condition of $\frac{\partial U_i^M}{\partial s_i} = 0$ can be satisfied.

Next, the fixed point method is used to explore the existence of NE and we can obtain the following two theorems.

Theorem 1: In the formulated sequential game $\mathcal{G}$, there exists fixed point(s).

Proof: Obviously, $\mathcal{S}$ and $\mathcal{P}$ are all nonempty sets. Because the domain of $\mathcal{S}$ and $\mathcal{P}$ all contain upper bounds, so $\mathcal{S}$ and $\mathcal{P}$

belong to the subnonempty compact spaces of Euclidean space $\Re$. In addition, the utility function is a strictly convex function. We can also see that the solution set are convex set. Moreover, it can be easily obtained that the function is continuous. Above all, the strategy sets of this game are nonempty convex and compact sets, and the utility functions are continuous.

Then, the utility function of miners could be a continuous mapping in the total sets of strategy and utility. According to the definition of Brower's fixed point theorem [16], the utility function must have a fixed point, i.e., there is a s_0 in $\mathcal{S}$, which enables $s_0 = U(s_0)$. The proof is now completed. □

Theorem 2: The defined utility functions have the fixed points.

Proof: The strategy set of the game $\mathcal{G}^M$ is an nonempty convex and compact set, and the utility function is continuous. Therefore, the defined utility function must have the fixed points. Due to the limitation of the space and detailed proof can be found in [16], we omit here. □

Assume that $e^{-\lambda t_i}(R_f + R_p) = \Phi$, we can obtain

$$\Phi \frac{\partial \alpha_i}{\partial s_i} + \varepsilon \frac{\partial \alpha_i}{\partial s_i} - p_i^j = 0. \tag{16}$$

Due to the fact

$$s_i = \sum_{i \in N} s_i - \sum_{i \in N} s_{-i} \tag{17}$$

and we have

$$s_i^* = \sqrt{\frac{(\Phi + \varepsilon) \cdot \sum_{i \in N} s_{-i}}{p_i^j}} - \sum_{i \in N} s_{-i}. \tag{18}$$

To this end, the optimal purchasing strategy for miner i, which maximizes the utility is expressed as follows:

$$s_i^* = \frac{(N-1)}{\sum_{i \in N} \frac{p_i^j}{\Phi + \varepsilon}} - \frac{p_i^j}{\Phi + \varepsilon} \left(\frac{(N-1)}{\sum_{i \in N} \frac{p_i^j}{\Phi + \varepsilon}} \right)^2. \tag{19}$$

Then, we will study the uniqueness of the NE. In noncooperative game problems, the sequential equilibrium solution problem (SEP) and the variational inequality (VI) problem have some common similarities. Thus, the problem of refining NE can be transformed into the VI problem [17]. Based on the uniqueness of the NE in the formulated miners' subgame with nonempty convex compact set of strategy, there are mainly two methods to construct the VI problem.

2) Game of ESPs in EFS: Based on the above definitions, the utility of ESP j obtained from serving miner i is

$$U_j^E = p_i^j s_i \mathcal{D}(s_i) - c_E^j. \tag{20}$$

After we get the optimal purchasing strategy, which maximizes the utility of miners, the best pricing strategy for ESP can be applied in a similar manner in the first stage. After substituting (3) and (19) into (20), one can arrive

$$U_j^E = p^j s_i^* \mathcal{D}(s_i^*) - c_E^j. \tag{21}$$

Then, with some calculations, conclusions can be easily obtained that the U_j^E is also a convex function with respect to p^j while $\frac{\partial^2 U_j^E}{\partial (p_i^j)^2} < 0$. The optimal pricing strategy $(p^j)^*$ that makes ESP maximize the profits while the condition $\frac{\partial U_j^E}{\partial p^j} = 0$ can be satisfy.

To this end, the following observations can be made: there must exist an unique α_i^*, which makes the Hessian matrix to be negative definite. That is, the optimization of EFS is true. Through backward induction method, we can also find the existence of an unique strategy combination (i.e., SE) $(\mathcal{S}^*, \mathcal{P}^*)$ which optimizes the problem P1 and P2.

D. Miner-First-Select (MFS) Case

In this case, rational miners first select the purchasing strategy based on the observations. Then, ESPs select the pricing strategy accordingly. Here, we also utilize backward induction and first study the second stage of this game.

1) Game of ESPs in MFS: In the second stage of MFS, the ESP select a pricing strategy. The considered homogeneous ESPs (with the same strategy space) simultaneously choose their strategies in the Cournot game. To maximize utilities, miners inevitably expect much more computational resources with lower prices. Obviously, the increase in the amount of computational resources provided by multiple ESPs will have an negative impact on selling prices.

Let q^j be the strategy of ESP j for selecting the provided quantity, where the set of quantity strategy is $\Theta = (q^1, .., q^j, \ldots, q^M)$ and $q^j \in [q^{\min}, q^{\max}]$, where $q^{\min}$ and $q^{\max}$ are the minimum and maximum quantities, respectively. We assume that the seller decides the resource quantity strategy at time $t+1$ based on the profits of moment t, that is

$$q^j(t+1) = q^j(t) + \vartheta q^j(t) \frac{\partial U_j^{E,\Theta}}{\partial q^j} \tag{22}$$

where ϑ is a positive value of the relative adjusting speed of q^j, and the presentation of U_j^E is given in (24).

Let's consider a Cournot duopoly game, where a twice differentiable and quadratic nonlinearity inverse demand function [21] can be denoted as follows:

$$p^j = p\left(Q^E\right) = a_B - b_B \left(Q^E\right)^2 \tag{23}$$

where a_B and b_B are positive constants of demand function, $Q^E = \sum_{j=1}^M q^j$ is the total quantity ESPs provided. Also we have $Q^E = \sum_{i=1}^N s_i$.

For simplification, we now redefine the utility function of ESP through the economic method, and ignore the impact of $\mathcal{D}(s_i)$ from miner i. Thus, the utility of ESP j would be

$$U_j^E = \left(p^j - c_E^j\right) q^j = \left(a_B - b_B \left(Q^E\right)^2 - c_E^j\right) q^j \tag{24}$$

Take the first derivative of U_j^E with respect to q^j and we have

$$\frac{\partial U_j^{E,\Theta}}{\partial q^j} = -2q^j Q^E \frac{\partial Q^E}{\partial q^j} + \left(a_B - b_B \left(Q^E\right)^2 - c_E^j\right). \tag{25}$$

Once the condition $\frac{\partial U_j^E}{\partial q^j} = 0$ was true, the optimal selling strategy for ESP j can be denoted as $(q^j)^* = f(p^j)$.

Thus, the optimal quantity of ESP j was obtained by

$$\left(q^j\right)^* = \arg\max_{q^{(j)}} \left[\left(a_B - b_B\left(Q^E\right)^2 - c_E^j\right) q^j\right]. \tag{26}$$

Then, the optimization problem P2 can be transformed into P2' under the condition of ignoring DoS for ESP, that is

$$\begin{aligned} &\mathbf{P2'} : \max_{\mathcal{S},\boldsymbol{\Theta}} \mathbf{U_j^{E,\Theta}} \\ &s.t.\ \mathbf{C1} : \mathcal{D}\left(s_i\right) \in [0,1]. \end{aligned} \tag{27}$$

Therefore, we define the Lagrange function of P2' as

$$\begin{aligned} L\left(q^j, \zeta_i\right) = &\left(a_B - b_B\left(Q^E\right)^2 - c_E^j\right) q^j \\ &- \sum_{i=1}^{N} \zeta_i \left(\mathcal{D}\left(s_i\right) - 1\right) \end{aligned} \tag{28}$$

where $\zeta_i > 0$ is Lagrange multiplier corresponding to constraint C1. After the KKT condition can be obtained,the Lagrange method to solve the optimal problem to find the optimal $(q^j)^*$ and then $(p^j)^*$. Due to the space limitation, we omit it here.

Based on the mentioned fixed points theorem and the strategy $q^j(t+1)$ at slot $t+1$, the dynamic equation of the ESPs could be expressed by mapping function from the previous time slot

$$\begin{cases} q^1\left(t+1\right) = q^1(t) + \vartheta q^1(t) \frac{\partial U_1^E}{\partial q^1} \\ \cdots \\ q^M\left(t+1\right) = q^M(t) + \vartheta q^M(t) \frac{\partial U_M^E}{\partial q^M}. \end{cases} \tag{29}$$

We consider study the eigenvalues of the Jacobian matrix of the aforementioned mapping function to research the stability of the NE

$$\mathrm{J} = \begin{bmatrix} \frac{\partial q^1(t+1)}{\partial q^1(t)} & \frac{\partial q^1(t+1)}{\partial q^2} & \cdots & \frac{\partial q^1(t+1)}{\partial q^M} \\ \cdots & \cdots & & \cdots \\ \frac{\partial q^M(t+1)}{\partial q^1(t)} & \frac{\partial q^M(t+1)}{\partial q^2(t)} & \cdots & \frac{\partial q^M(t+1)}{\partial q^M(t)} \end{bmatrix}. \tag{30}$$

Through two typical cases, i.e., the selection of adaptive strategies adjustment strategies, the stability of the NE could be verified and the necessary conditions to be satisfied.

Case 1: when there are two ESPs and both ESPs adjust the quantity according to the income at the previous time slot. That is, when (22) is satisfied, the Jacobian matrix is as follows:

$$\mathrm{J}_1 = \begin{bmatrix} J_{1,1} & J_{1,2} \\ J_{1,3} & J_{1,4} \end{bmatrix} \tag{31}$$

where

$$\begin{cases} J_{1,1} = (q^1)^2\vartheta{-}2 - b_B - 4\vartheta + (q^2)^2\vartheta{-}2 - b_B \\ \qquad + q^1q^2{-}2\vartheta b_B + \vartheta{-}2 - 2b_B + \vartheta a_B - c_E^j + 1 \\ J_{1,2} = 2{-}2 - b_B q^2 - 2b_B q^1 \\ J_{1,3} = 2{-}2 - b_B q^1 - 2b_B q^2 \\ J_{1,4} = (q^2)^2\vartheta{-}2 - b_B - 4\vartheta + (q^1)^2\vartheta{-}2 - b_B \\ \qquad + q^1q^2{-}2\vartheta b_B + \vartheta{-}2 - 2b_B + \vartheta a_B - c_E^j + 1. \end{cases} \tag{32}$$

The eigenvalues can be obtained as follows:

$$\begin{aligned} &(\lambda_{1,1}, \lambda_{1,2}) \\ &= \frac{-2 \pm \sqrt{\left(J_{1,1}{+}J_{1,4}\right)^2 - 4\left(J_{1,1}J_{1,4}{-}J_{1,2}J_{1,3}\right)\left(J_{1,1}{+}J_{1,4}\right)}}{4\left(J_{1,1}J_{1,4}{-}J_{1,2}J_{1,3}\right)\left(J_{1,1}{+}J_{1,4}\right)}. \end{aligned} \tag{33}$$

Substituting (32) into (33), for given q^1 and q^2, we can see that the stability of NE is relevant to ϑ since a_B and b_B are only coefficient constraints.

Case 2: when there are two ESPs, if one of them is a strategic choice to adjust based on the observations of the previous slot, and the other is an adaptive adjustment, that is

$$\begin{aligned} q^1\left(t+1\right) &= q^1(t) + \vartheta q^1(t) \tfrac{\partial u^1}{\partial q^1} \\ q^2\left(t+1\right) &= \left(1-\beta\right) q^2(t) + \beta q^1(t) \end{aligned} \tag{34}$$

where β is the adjustment speed.

Similarly, the Jacobian matrix can be defined as

$$\mathrm{J}_2 = \begin{bmatrix} J_{2,1} & J_{2,2} \\ J_{2,3} & J_{2,4} \end{bmatrix} \tag{35}$$

where

$$\begin{cases} J_{2,1} = J_{1,1} \\ J_{2,2} = \vartheta q^1 \left(2\left(-2 - b_B\right) q^2 - 2b_B q^1\right) \\ J_{2,3} = \beta \\ J_{2,4} = \frac{\partial q^2(t+1)}{\partial q^2(t)} = 1 - \beta. \end{cases} \tag{36}$$

Similarly, conclusion can be easily drew that he stability of NE is related to the speed of adjustment ϑ, β. When the NE is stable, the utility of the ESPs cannot be increased by altering the quantity or price.

The combined analysis of Case 1 and Case 2 shows that when ESPs select the quantity of computing resources, no matter what kind of strategy is using, as long as the appropriate adjustment speed is selected, the optimal sale strategy $\mathbf{q}^* = [(q^1)^*, (q^2)^*, \ldots, (q^M)^*]^{\mathrm{T}}$ can be achieved and the stability can be guaranteed.

Based on (23), when ESP j selects the optimal quantity $(q^j)^*$, the optimal pricing scheme could be obtained as follows:

$$\left(p^j\right)^* = a_{\mathrm{BE}} - b_{\mathrm{BE}}\left(\left(Q^E\right)^*\right)^2. \tag{37}$$

The selection of adaptive strategies adjustment strategies, the stability of the NE could be verified and the necessary conditions to be satisfied. When ESPs select the quantity of computing resources, no matter what kind of strategy is using, as long as the appropriate adjustment speed is selected, the optimal sale strategy $\mathbf{q}^* = [(q^1)^*, \ldots, (q^M)^*]^{\mathrm{T}}$ can be achieved and the stability can be guaranteed.

2) Game of the Miners in MSF: Based on the idea of backward induction, after observing the pricing strategy of ESPs, miners choose the purchasing strategy. Thus, the utility of miner i at this time slot is

$$U_i^M = e^{-\lambda t_i}\alpha_i\left(R_{f,i} + R_{v,i}\right) + \varepsilon\alpha_i - \left(a_{BE} - b_{BE}\left(\left(Q^E\right)^*\right)^2\right)s_i - c_{i,M}. \tag{38}$$

The proof of the existence and uniqueness of NE is similar to above analysis. The existence of NE can be determined by the fixed point theorem, and the uniqueness can be proved by Hessian matrix, which we omit it here due to the space limitation.

E. Summary

To study the timer-shaft-based incentive optimization problem, the incentive mechanism under two cases are separately studied. Then, the existence and uniqueness of the SE of these two cases are studied through backward induction. We are able to find there are SEs for the formulated games which enable the optimal utility.

V. Proposed Solution

As we can see, over a certain amount of time slots, the optimization problem needs to obtain the complete information about the future time slots to reach the optimal solution for the next time slot, which means that absence of prior information may degrade its achievable performance. Therefore, we intend to utilize the RL-based algorithm to obtain solution without aforementioned prior knowledge.

A. RL Framework Formulation

In our considered system, the agent is the network controller, and the environment consists of all the entities in the network. In each time slot l, the agent chooses an action a_l from the action space, which decides the resource trading. The agent obtains a reward or punishment from the environment after applying an action. This scheme aims at maximizing the cumulative received rewards within interactions. The RL problem comprises of a single or multiple agents and an environment. Based on a chosen policy, the agent can take actions to interact with the environment. Briefly, there are three elements in the RL framework: action $\mathbf{a}$, state $\mathbf{s}$, and reward $\mathbf{r}$. The state space, action space, and reward of the DRL-based framework at time slot l are defined in the following.

1) State: We define the *state space* $\Psi = (\psi_l \in \Psi, l = 1, 2, \ldots)$ is a set of the following factors: the DoS, the probability of successfully reward, computational capability, etc, which is the observation of the current environment at time slot l.

2) Action: We consider the *action space* of agent i is $\mathcal{A} = \{a_l \in \mathcal{A}, l = 1, 2, \ldots\}$, where is strategy of the blockchain players (i.e., $\mathcal{S}$ and $\mathcal{P}$) at time slot l.

3) Reward: After executing the chosen action, the agent will obtain a reward in certain state in each time slot. As the target of the RL is to obtain reward maximization, the defined reward needs to be positively related to the objective function. For the considered problem, we define reward as the utility functions (i.e., objective function of P1 or P2). In the simulations, as iterative scheme is used for finding the optimum, we use objective function of P2 as the reward.

B. Proposed DQN-Based Solution

DQN uses a neural network (NN) $Q(\boldsymbol{\psi}, \mathbf{a}; \theta)$ to represent Q-function, where θ is the weights of the NN. By updating θ at each iteration, the Q-network is trained to approximate the real Q-values. When it is applied to Q-learning, NN improve the performance of flexibility at the cost of stability [22]. In this context, DNN is proven to be a robust learning with better performance [23]. Comparing with the Q-learning, there are following major improvements in the DQN.

The hierarchical layers of convolution filters in the DNN can be used to exploit the local spatial correlations. By such, the high-level features of input data are extracted. The second one is that experience replay can store its experience tuple $e(l) = (\psi_l, a_l, r_l, \psi_{l+1})$ at time slot l into a replay memory $\mathcal{O}$. The relay can randomly sample batches $\hat{\mathcal{O}}$ from the memory to train the DNN. Such a process enables DQN to learn from different past experience rather than from the current one. In addition, while using one network for estimating the Q-values, the target Q-values that compute the loss of each action in the training process can be generated by a second network. Such a procedure is able to make the DQN stable.

DQNs are optimized by minimizing

$$\mathcal{L}(\theta) = \mathbb{E}[y_l - Q(\psi_l, a_l; \theta))^2] \tag{39}$$

where y_l is the target Q-value, and it can be expressed as

$$y_l = r(\psi_l, a_l) + \xi \max_{a_{l+1}} Q^*(\psi_{l+1}, a_{l+1}; \theta^-). \tag{40}$$

θ^- is a target network parameter that is frozen for some iterations when the online network $-Q(\boldsymbol{\psi}, \mathbf{a}; \theta)$ is updated by gradient descent. Specially, the network controller chooses a_l at time slot l, obtains reward r_l and goes to the next state ψ_{l+1}. Accordingly, an experience replay memory $\mathcal{O}$ is used to store the vector $(\psi_l, a_l, r_l, s_{l+1})$.

C. Modified Experience Replay Update Method

We consider to randomly select two historical sequences in the experience pool and remove the empirical value with a larger number of Niche (i.e., the distance is greater than a predetermined value, similar to the concept of variance), and then put the current sampling result into the experience replay. Accordingly, we propose a new experience replay update algorithm, which is mainly used for secondary update of the weights after initialization.

Algorithm 1: Modified Experience Replay Update Method for DQN.

Input: $\mathcal{A}$, Ψ, θ
Output: Optimal strategy $a^* = \arg\max_{a' \in \mathcal{A}} Q(\psi_{l+1}, a_{l+1}; \theta)$
Initialize $Q(\psi_l, a_l)$ in the prioritized replay memory D with the size of N^D
Initialize the main Q-network with input pairs $(\psi_l, a_l, r_l, \psi_{l+1})$ and the target $Q(\psi_{l+1}, a_{l+1})$
Initialize the parameters of online Q-network to measure the loss value
Repeat
while $\forall \psi$ and a, s.t. $Q(\psi, a)$ not converge **do**
  Step 1: At the beginning of decision episode l, randomly select the action a_l with probability ε according to ε-greedy policy.
  Step 2: Execute action a_l, received instant reward signal r_l and the new state s'.
  Step 3: Generate the sequence $(\psi_l, a_l, r_l, \psi')$.
  Step 4:
  if $(\psi, a, r, \psi')_{c_l}^{Niche} < (\psi, a, r, \psi')_{d_l}^{Niche}$
  $(\psi, a, r, \psi')_{d_l} \leftarrow (\psi_l, a_l, r_l, \psi_{l+1})$
  else $(\psi, a, r, \psi')_{c_l} \leftarrow (\psi_l, a_l, r_l, \psi_{l+1})$
  Step 5: Randomly sample a mini-batch of the state sequence $(\psi_{l'}, a_{l'}, r_{l'}, \psi')$ from D
  Step 6: Calculate the target Q-value by (40) and loss function
  Step 7: Randomly generate new weight ω'
end while

As shown in *Algorithm 1*, action a_l is first selected randomly based on the probability $1 - \varepsilon$ of the ε-greedy strategy at time slot l. Then, based on state ψ_l and immediate reward r_l, a new sequence combination $(\psi_l, a_l, r_l, s_{l+1})$ will be generated when the latest state s_{l+1} is obtained. Next, randomly select two sequence combinations and replace the greater number of niche by the new sequence. Then, randomly sample a mini-batch in the experience replay, gradually approach the target *Q*-value through (40), and update the key parameters of the current *Q*-network according to the loss function (39). Once it gradually converges, the iterative process will be interrupted and the optimal action $a^* = \arg\max_{a_{l+1} \in \mathcal{A}} Q(\psi_{l+1}, a_{l+1}; \theta)$ will be the output.

Based on the improved experience replay update method, the presented DQN algorithm can achieve a rapid convergence with the help of experience replay sampling. Thus, the players in edge computing-based blockchain system can quickly find the optimal participation strategy based on the information set to achieve the optimal utility.

VI. Performance Evaluation

In this section, we conduct numerical simulation of the designed edge computing-based blockchain incentive mechanism. For the value of some key parameters, we refer to [7]. In detail, we set the following parameters to: $R_f^{\max} = 50$, $T = 10^7$, $r = 10^5$, $\varepsilon = 10^{-6}$, $\lambda = \frac{1}{600}$, $\eta = 10^{-2}$, $c_{i,M} = 10^{-3}$, $c_E^j = 10^{-4}$, $\xi = e - 1$, and $\varepsilon = 5 \times 10^{-2}$.

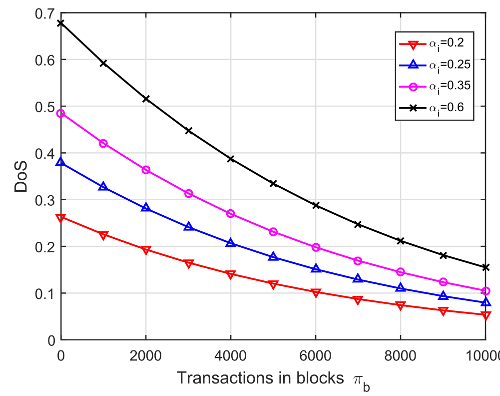


Fig. 2. Size of block versus DoS.

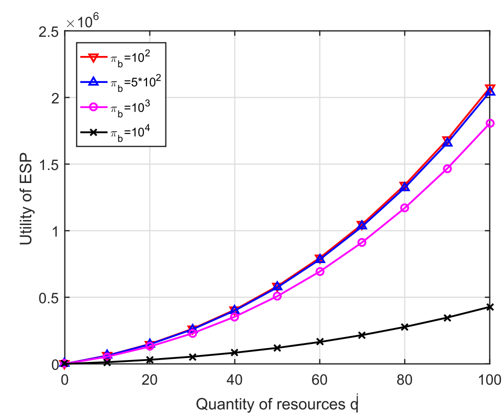


Fig. 3. Quantity versus utility of ESP.

Fig. 2 shows the relations between the number of transactions in block π_b and the formulated DoS. It can be found that, when miners own fixed computational capability or hash power proportion α_i, the DoS increases as the size of block increases. It is mainly because that the increase of number of transactions in the block (i.e., the block size) directly makes it more difficult to solve a new PoW puzzle. Then more computational resource is needed for mining. For the case of fixed number of transactions π_b in the block, it is obvious that the increase in α_i directly leads to an increment of DoS. Based on the definition, the DoS is mainly determined by the probability of successfully mining, which is further affected by the factors such as the number of transactions, time delay and the proportion of hash power.

Fig. 3 illustrates the relations between the quantity of computational resource q^j provided by ESP and the utility j U_j^E. It can be found that when the block size π_b is constant, the increase of the quantity q^j leads to the increase of the utility of ESP j. Based on the definition of the utility, U_j^E is nonlinear w.r.t. q^j. Then, the price of computational resource would decrease, which is determined by the inverse price-demand function. Meanwhile, the utility of miners and the DoS increase, which leads to the incremental of the utility. Also, it can be easily found that the

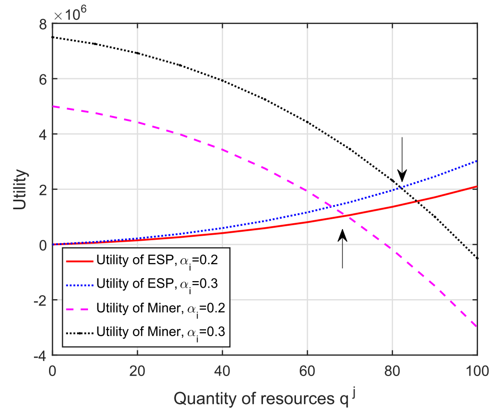


Fig. 4. Quantity versus utility of players.

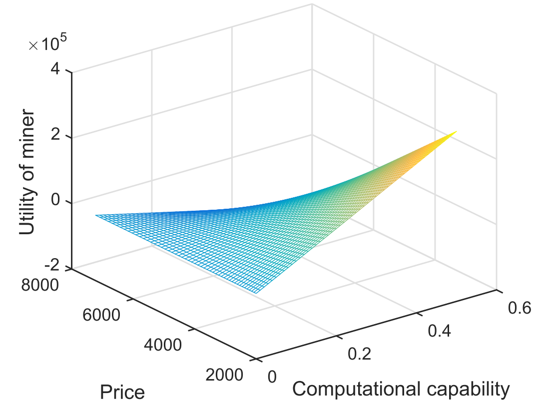


Fig. 6. Computational capability and price versus utility of miner.

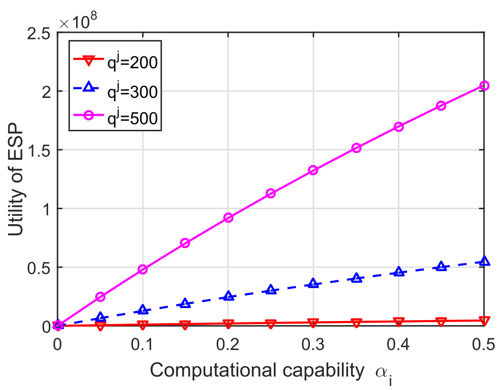


Fig. 5. Computational capability versus utility of ESP.

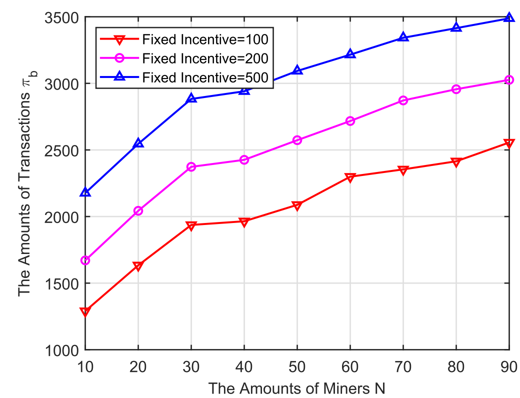


Fig. 7. The amounts of miners versus transactions.

bigger π_b would make a smaller utility U_{ESP}^j for the case of fixed quantity q^j. That is due to fact that a larger number of transactions in block makes the process of solving PoW puzzle harder, which decreases the DoS of miners as well as the utility of ESP.

In Fig. 4, we plot the relationship between the resource quantity q^j and the utilities of ESP j and miner i. We can see that the increase of q^j makes U_i^E become larger, but decrease U_i^M. The utilities of players have an intersection point, which is the SE point. At the intersection point, the utilities of players in blockchain reach a balance where neither of players can change the strategy without performance loss. We can also find from this figure that the utility must be optimized. For the case of fixed quantity q^j, the bigger hash power proportion α_i is, the greater utility of player will be. This is mainly because the increase of α_i would lead to the increased probability of successfully mining μ_i and a better DoS. Furthermore, based on the defined inverse price-demand function, the fixed quantity q^j makes a constant price of computational resource, so the increase of α_i shows the fact that miners (except miner i) decrease the demand of service, then it will also increase the DoS of miner i.

In Fig. 5, we plot the relation between the proportion of computational resource and the utility of ESP. It can be found that the increase of hash power proportion α_i leads to the increase of U_j^E. For the case of fixed α_i, the bigger the quantity q^j, the higher utility will be. Moreover, U_j^E would be negative if the quantity q^j are blow 200, where should be the minimum amount provided by ESP j in this case.

As shown in Fig. 6, we can see that some factors, i.e., the proportion of computational capability α_i and the price of the computational resources p^j can affect the utility of miners U_i^M. In the figure, when the price remains the same, the utility U_i^M increases with the increment of α_i. U_i^M gradually decreases with the increase of p^j for the case of fixed α_i. The main reason is that the increase of α_i and the decrease of p^j would lead to the increase of the probability of successfully mining and the DoS, so as to bring a positive effect on the expected reward for miner i. Further, with the change in α_i and p^j, it can be easily observed that there must exist strategic point(s), which enables to optimize the utility of miners.

Fig. 7 implies the relations between the amounts of miners participating in mining and system efficiency. Specifically, the system efficiency is characterized by the number of transactions contained in a single block in this paper. As for the newly generated block, while the award of the certain block and the average time of solving the PoW puzzle keep constant, the larger amounts of miners will cause the amount of transactions/data contained in the block to increase. In other words, the efficiency of the

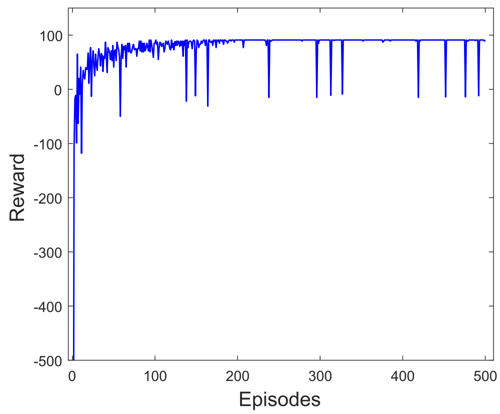


Fig. 8. Convergence performance.

blockchain network increases as the number of miners increases. Likewise, when the number of miners remains unchanged, the efficiency of the system improves as the award increases.

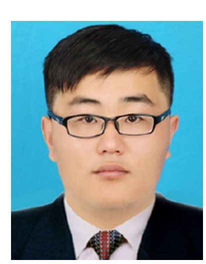
We evaluate the proposed experience replay update method for DQN algorithm in Fig. 8. Through numerical simulation, we characterize the reward performance for players of each episode. With the increase of number of the episode, the algorithm shows positive effects. Then, the result obtained by all blockchain participants converges to a relatively stable value, which represents a good convergence performance. For the modeled sequential decision problem, the proposed algorithm can enable participants to obtain the optimal reward. In addition, Fig. 8 also shows that the proposed scheme can significantly reduce the number of target steps to achieve the optimal strategy and utility.

VII. Conclusion

In this article, an edge computing-based blockchain framework is considered, where multiple ESPs can offer computational resources to the devices for mining. We mainly focus on investigating the trades between rational devices and ESPs in the computational resource market, where ESPs can act as the sellers and devices as the buyers. Accordingly, a sequential game model is formulated and by exploring the sequential NE, the existence of optimal incentive solutions can be proved. Then, a Deep Q-Network-based algorithm with modified experience replay update method is applied to find the optimal strategies. Through theoretical analysis and simulations, we demonstrate the effectiveness of the proposed incentive mechanism on forming the blockchain. In the future, we will take transmission-related metric into the consideration when designing the interactions between blockchain players.

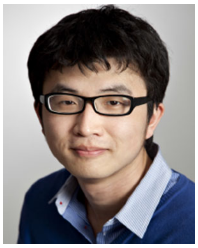
References

[1] L. Zhao, J. Wang, J. Liu, and N. Kato, "Optimal edge resource allocation in IoT-based smart cities," *IEEE Netw.*, vol. 33, no. 2, pp. 30–35, Mar./Apr. 2019.
[2] P. K. Sharma, S. Singh, Y. Jeong, and J. H. Park, "DistBlockNet: A distributed blockchains-based secure SDN architecture for IoT networks," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 78–85, Sep. 2017.
[3] Z. Chang, W. Guo, X. Guo, Z. Zhou, and T. Ristaniemi, "Incentive mechanism for EDGE computing-based blockchain," *IEEE Trans. Ind. Informat.*, vol. 16, no. 11, pp. 7105–7114, Nov. 2020.
[4] S. Guo, Y. Dai, S. Guo, X. Qiu, and F. Qi, "Blockchain meets edge computing: Stackelberg game and double auction based task offloading for mobile blockchain," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5549–5561, May 2020.
[5] N. C. Luong, Z. Xiong, P. Wang, and D. Niyato, "Optimal auction for edge computing resource management in mobile blockchain networks: A deep learning approach," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.
[6] Z. Zhang, Z. Hong, W. Chen, Z. Zheng, and X. Chen, "Joint computation offloading and coin loaning for blockchain-empowered mobile-edge computing," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9934–9950, Dec. 2019.
[7] Y. Jiao, P. Wang, D. Niyato, and Z. Xiong, "Social welfare maximization auction in edge computing resource allocation for mobile blockchain," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, 2018, pp. 1–6.
[8] Z. Xiong, S. Feng, W. Wang, D. Niyato, P. Wang, and Z. Han, "Cloud/Fog computing resource management and pricing for blockchain networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4585–4600, Jun. 2019.
[9] N. Houy, "The bitcoin mining game," *Ledger*, vol. 1, pp. 53–68, 2016.
[10] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, "Bitcoin mining pools: A cooperative game theoretic analysis," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2015, pp. 919–927.
[11] A. Avyukt, G. Ramachandran, and B. Krishnamachari, "A decentralized review system for data marketplaces," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency*, 2021, pp. 1–9.
[12] L. Liu, M. Du, and X. Ma, "Blockchain-based fair and secure electronic double auction protocol," *IEEE Intell. Syst.*,vol. 35, no. 3, pp. 31–40, May 2020.
[13] A. Kumari, R. Gupta, and S. Tanwar, "PRS-P2P: A prosumer recommender system for secure P2P energy trading using Q-learning towards 6G," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2021, pp. 1–6.
[14] J. Huang, Y. Xu, B. An, and M. Xiao, " Blockchain-based double auction for edge cloud resource trading with differential privacy," in *Proc. IEEE 18th Int. Conf. Mobile Ad Hoc Smart Syst.*, 2021, pp. 615–620.
[15] M. Gairing, B. Monien, and K. Tiemann, "Selfish routing with incomplete information," *Theory Comput. Syst.*, vol. 42, no. 1, pp. 91–130, 2008.
[16] D. Gale, "The game of hex and the brouwer fixed-point theorem," *Amer. Math. Monthly*, vol. 86, no. 10, pp. 818–827, 1979.
[17] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.
[18] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjarungnes, "Game Theory in Wireless and Communication Networks: Theory, Models, and Applications," Cambridge, U.K.: Cambridge Univ. Press, 2012.
[19] A. Shoker, "Sustainable blockchain through proof of exercise," in *Proc. IEEE 16th Int. Symp. Netw. Comput. Appl.*, Cambridge, MA, USA, 2017, pp. 1–9.
[20] D. Niyato, E. Hossain, and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach," *IEEE Trans. Mobile Comput.*, vol. 8, no. 8, pp. 1009–1022, Aug. 2009.
[21] H. N. Agiza and A. A. Elsadany, "Chaotic dynamics in nonlinear duopoly game with heterogeneous players," *Appl. Math. Comput.*, vol. 149, pp. 843–860, Feb. 2004.
[22] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction." Cambridge, MA, USA: MIT Press, 1998.
[23] V. Minh *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.



Wenlong Guo received the B.Eng. degree in electronics engineering from Yantai University, Yantai, China, in 2015. He is currently working toward the Ph.D. degree in computer science and technology with the College of Information Science and Engineering, Yanshan University, Qinhuangdao, China.

He is currently an visiting Ph.D. with the University of Jyväskylä, Jyväskylä, Finland. His research interests include blockchain, UAV, resource allocation, and mobile computing.



Zheng Chang (Senior Member, IEEE) received the B.Eng. degree from Jilin University, Changchun, China, in 2007, the M.Sc. (Tech.) degree from Helsinki University of Technology (Now Aalto University), Espoo, Finland, in 2009, and the Ph.D degree from the University of Jyväskylä, Jyväskylä, Finland, in 2013.

Since 2008, he has held various research positions with Helsinki University of Technology, University of Jyväskylä and Magister Solutions Ltd., Finland. He has been awarded by the Ulla Tuominen Foundation, the Nokia Foundation and the Riitta and Jorma J. Takanen Foundation for his research excellence. He has been awarded as 2018 IEEE Communications Society best young Researcher for Europe, Middle East, and Africa region. He has authored or coauthored more than 140 papers in journals and conferences, and received best paper awards from IEEE TCGCC and APCC in 2017. His research interests include IoT, cloud/edge computing, security and privacy, vehicular networks, and green communications.

Dr. Chang serves as an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS, *Springer Wireless Networks and International Journal of Distributed Sensor Networks*, and a Guest Editor for the IEEE NETWORK, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINE, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, etc. He was the exemplary reviewer of IEEE WIRELESS COMMUNICATION LETTERS in 2018. He has participated in organizing workshops and special sessions in Globecom'19, WCNC'18-22, SPAWC'19, and ISWCS'18. He also serves as Symposium Chair for ICC'20 and Publicity Chair for INFOCOM'22.

Xijuan Guo received the Ph.D. degree from Yanshan University, Qinhuangdao, China.

She is currently a Professor with the College of Information Science and Engineering, Yanshan University. Her current research interests include high performance computing, cloud computing, image processing, and wireless communications.

Peiliang Wu received both the B.Sc. degree in computer science and technology from Yanshan University, Qinhuangdao, China, in 2004, and the Ph.D. degree in computer application technology from Yanshan University, in 2010.

He is currently a Professor of artificial intelligence and intelligent robotics with the Department of Computer Science and Engineering, School of Information Science and Engineering, Yanshan University. His research interests include IoT robot system, multiagent system, artificial intelligence, robot manipulation skill learning, computer vision, and pattern recognition.

Dr. Wu was a Postdoctor with The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation Chinese Academy of Sciences, China, and an academic visitor in the School of Engineering and the Built Environment, Edinburgh Napier University, U.K. He has served as a Local Chair for both Artificial intelligence and Internet of Things Forum, International Future Intelligent Computing Conference 2020 and China Conference on Knowledge Graph and Semantic Computing 2022.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively.

From 2000 to 2002, he was a R&D Engineer of JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as in the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, Big Data analysis, security, and smart grid.

Dr. Han was a recipient of NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018, AAAS fellow since 2019, and ACM distinguished Member since 2019. He is a 1% highly cited Researcher since 2017 according to Web of Science. He is also the winner of the 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."