DIM-DS: Dynamic Incentive Model for Data Sharing in Federated Learning Based on Smart Contracts and Evolutionary Game Theory

Yanru Chen, Yuanyuan Zhang, *Graduate Student Member, IEEE*, Shengwei Wang, Fan Wang, Yang Li, Yuming Jiang, Liangyin Chen, and Bing Guo

***Abstract*—With the development of big data, data sharing has become a hot topic. According to the previous research on data sharing, there is a problem with regard to how to design an effective incentive mechanism to make users willing to share data. First, we integrate the incentives based on reputation and payment and introduce "credibility coins" as a cryptocurrency for data-sharing transactions, to encourage users to participate honestly in the data-sharing process based on federated learning. Second, we propose a dynamic incentive model based on the evolutionary game theory to model the game process of users in data sharing and analyze the stability of their strategies. Finally, based on the results of this analysis, we use the blockchain-based smart contract technology to dynamically adjust the participation benefits of users under different conditions in order to promote users to join consortium blockchains more often and steadily to participate in model training for federated learning and obtain better model accuracy. Our work is the first to apply the evolutionary game theory to the study of incentives in federated learning, and plays a leading role in the study of incentives in federated learning. Experimental simulation validation shows that our DIM-DS model can adequately motivate users to participate in the collaborative task of data sharing and maintain stability. The model can maximize the effectiveness of the federated learning model.**

***Index Terms*—Data sharing, dynamic incentive model, evolutionary game theory, federated learning, smart contracts.**

Manuscript received 4 January 2022; revised 29 May 2022 and 2 July 2022; accepted 11 July 2022. Date of publication 18 July 2022; date of current version 21 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62072319, Grant 61772352, and Grant 62172061; in part by the Science and Technology on Communication Security Laboratory under Grant 6142103190415; in part by the Sichuan Science and Technology Program under Grant 2022YFG0041, Grant 2019YFG0184, Grant 2019JDTD0001, Grant 2022YFG0155, Grant 2022YFG0157, Grant 2021GFW019, Grant 2021YFG0152, Grant 2021YFG0025, and Grant 2020YFG0322; and in part by the Luzhou Science and Technology Program under Grant 2021CDLZ-11; and in part by the National Key Research and Development Project under Grant 2020YFB1711800 and Grant 2020YFB1707900. *(Corresponding authors: Liangyin Chen; Bing Guo.)*

Yanru Chen, Yuanyuan Zhang, Shengwei Wang, Fan Wang, Yuming Jiang, and Bing Guo are with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: yanruchen@stu.scu.edu.cn; yuanyuanzhang@stu.scu.edu.cn; wangshengwei1217@126.com; 2713940331@qq.com; jiangym@scu.edu.cn; guobing@scu.edu.cn).

Yang Li is with the Science and Technology on Security Communication Laboratory, Institute of Southwestern Communication, Chengdu 610041, China (e-mail: yishuihanly@pku.edu.cn).

Liangyin Chen is with the College of Computer Science and the Institute for Industrial Internet Research, Sichuan University, Chengdu 610065, China (e-mail: chenliangyin@scu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2022.3191671

I. Introduction

With the development of industrial big data technology, the application of big data is becoming more and more popular [1]. The key to unleashing the potential of big data is data sharing [2]. Data sharing not only plays an important role in improving the efficiency of existing industrial scenarios but also has a huge impact on the intelligence [24] and the Quality of Experience (QoE) of users in the metaverse [25], the virtual scenario of the future. Thanks to the decentralization and tamper-proof features of blockchain, more and more scholars have started to study blockchain-based data sharing schemes. However, problems such as insufficient financial incentives have led to the fact that most users are still reluctant to actively share data [21]. In order to ensure that the interests of users involved in data sharing are maximized, a suitable incentive mechanism needs to be incorporated in the blockchain.

The existing blockchain-based incentives are mainly divided into reputation-based incentives, payment-based incentives, and noncooperative game-based incentives. Reputation-based incentive methods cannot significantly increase the proportion of user participation. For example, Kang *et al.* [10] designed a data-sharing scheme based on reputation-based incentives. The scheme ensures high-quality data sharing among vehicles, but user participation is not high and it is only applicable to Internet of Vehicles. The payment-based incentive approach not only increases user expenditures but also fails to motivate users to continuously participate. For example, Li *et al.* [11] designed CreditCoin, a payment-based incentive announcement network, where the percentage of users participating in data sharing in the network was not stable and users could not verify the value of data shared by other users. Incentive methods based on noncooperative games cannot achieve win–win situation for multiple parties. For example, a knowledge pricing strategy based on a noncooperative game with market incentives proposed by Lin *et al.* [13] is not effective in motivating users to share data. Data sharing is critical in federated learning, but the federated learning approach [17] proposes to protect users' data privacy that makes the participation ratio and stability of users reduced because it greatly increases the participation cost of users and cannot effectively motivate users. The analysis yields the following problems in the current study. In summary, the shortcomings with the current research are as follows.

1) In existing federated learning-based data sharing, incentives are insufficient and the percentage of participating users is generally low.
2) The stability of user participation strategies is generally inadequate.
3) The migration capability of existing incentives is severely lacking.

To solve the above shortcomings in a general scenario and considering that the user participation strategy needs to change continuously according to the current state of the system, we propose a dynamic incentive model based on smart contracts and evolutionary game theory, called DIM-DS. This model greatly facilitates users' honest participation in the joint learning-based data sharing process by integrating reputation and payment-based incentives-based data sharing process. The model innovatively introduces evolutionary game theory to construct a dynamic incentive model that maximizes the proportion of user participation. We simulate the game process of users in data sharing, in which the system obtains high user participation rate and high strategy stability.

In summary, our contributions include the following.

1) In order to solve the problem of scarce incentives in federated learning and insufficient user participation ratio, we integrate "reputation + payment + dynamic incentives," propose the dynamic incentive model based on evolutionary game theory for the first time, and introduce "credibility coins" to encourage users to honestly participate in the federated learning-based data-sharing process. Eventually, our system maximizes user participation rates.
2) To address the stability of user participation strategies, we model the game process of users in data sharing, conduct a complete mathematical analysis and simulation experiments on the stability of user strategies, and find the conditions to satisfy the stability.
3) To solve the problem of poor migration of existing incentive mechanisms, this work focuses on generic scenarios, and the proposed incentive mechanism is highly scalable. Subsequent researchers can adapt the present incentive mechanism to its specific application scenarios by adding and removing conditions on the basis of this incentive mechanism.

The subsequent sections are organized as follows. Section II briefly describes the related work. Section III presents the DIM-DS model design. Section IV conducts the analysis of experimental results. Section V concludes our work.

II. Related Work

A. Federated Learning

Federated learning belongs to a machine learning framework, proposed by Google in 2016 for implementing local model updates for mobile end users. Its goal is to perform efficient machine learning across multiple participants or computing nodes during the exchange of data, ensuring information security and terminal and personal data privacy within the scope of legal operation [19], effectively solving the problem of data silos.

Taking users A and B as examples, when they want to train a machine learning model together, A and B are reluctant to exchange data directly due to data privacy protection and security concerns. However, if the users each build their own local models, there is a risk that their models will not be built or implemented well due to having incomplete or insufficient data. Based on a federated learning mechanism, users do not have to send their local data but instead train a model together in the form of a parameter exchange. The model is equivalent to an optimal model built by aggregating data together, but the user's local data are not moved, and therefore, there is no risk of having a data privacy breach.

Through a review of incentive design for federated learning, Zhan *et al.* [16] found that it is difficult to model the utility function of each participant (parameter server and client) in federated learning. It is impossible to apply directly in making existing incentive design work. When the total amount of data from the participating parties is insufficient, the learned models will be less effective [20]. How to motivate users to participate in collaboration is therefore a problem that must be addressed by federated learning. However, current research on incentives for federated learning is still in its early stages.

By integrating federated learning into the blockchain and building a decentralized and trusted network, data sharing in industrial intelligence analysis scenarios can be achieved. The DS2PM model [17] addresses the privacy issue during data sharing by not sharing the original data directly but rather sharing the trained federation learning model. However, existing studies have focused on incentives in the case of direct data sharing, and little attention has been given to incentives under collaborative tasks such as federated learning. In federated learning, users need to consume greater computational consumption and communication consumption to complete collaborative tasks while participating in data sharing. When there is a lack of sufficient incentive, most users are reluctant to participate in collaboration, which results in a lack of sufficient training data for federation learning and makes the final trained model less accurate.

B. Current Incentives

Current incentive mechanisms can be divided into three types: 1) reputation-based [4], [10]; 2) payment-based [5], [6], [11], and 3) noncooperative game-based approaches (also called "tit-for-tat" approaches) [7], [8], [13]. With the development of blockchain technology, the incentive mechanism, one of its underlying core technologies, is also receiving attention. Based on this technology, each node can share distributed ledgers without resorting to a third-party trusted organization. The open and transparent nature of the ledger effectively guarantees the fairness of transactions. By designing a reasonable incentive mechanism, the blockchain nodes that participate in the consensus can maximize their personal benefits while ensuring the security and reliability of the blockchain system [9], [14], [15].

A reputation-based incentive mechanism [4], [10] is applied in a blockchain system. In this system, nodes decide whether to transact with a target node based on its reputation value.

Payment-based incentives [5], [6], [11] can be understood as being essentially a reward-based approach. In this approach, users are rewarded for completing transactions. For example, in blockchain applications, the most typical example is the bitcoin system. To maintain the extension of the system, nodes in the system are incentivized to create new blocks and publish them across the network to reach consensus. At the same time, they are rewarded with bitcoins and a fee for completing the transaction. The noncooperative game-based approaches [7], [8], [13] allow for direct auditing of user behavior and the calculation of benefits. However, user decisions in this approach depend entirely on the decisions of the remaining users. This finding occurs because of the high fluctuation of user strategies and the inconsistency of contributions among users.

In addition, all of the above studies lack rational analysis of users. The studies judge participation in the data-sharing process by only the fixed benefits of the data-sharing transactions posted. In fact, there is no clear definition of the value of the data. Users are unable to judge whether the policy is the best one. The user's own decisions are influenced by the policies of the remainder of the group to which they belong. Users can choose to participate in sharing in conjunction with the information they know or not. None of the data-sharing studies based on current incentives have analyzed the stability of user strategies, which is detrimental to the federated learning process.

C. Evolutionary Game Theory

Game theory is the study of different individuals or teams in the game process, based on the opponent's strategy, to adjust the corresponding strategy under agreed upon conditions. By finding the Nash equilibrium point, the user can be kept in a certain fixed strategy. Once the strategy changes, the user's income will decrease. For example, Deng and Huang [12] proposed an incentive mechanism based on mixed strategy games. Zhan *et al.* [22] proposed a game-based incentive mechanism for a federated learning platform, which combines distributed deep learning and crowdsensing together for big data analytics on mobile clients. Due to the use of the noncooperative game model [16], users are relatively constrained in their choices and are unable to change their strategies depending on the situation. Lim *et al.* [26] used an evolutionary game approach to address the problem of reward allocation and resource allocation for data owners without discussing the participation rate of data owners in the whole system.

Traditional game theory studies the behavior of a single, static game in a state of perfect rationality, whereas evolutionary game theory adds repetition and dynamics to the traditional game theory in a state of finite rationality, thus extending the static traditional game analysis [16]. Evolutionary game theory is a dynamic game theory. It no longer assumes that the users of a game are perfectly rational. Rather, it combines biological evolution with the idea that users in a game will continuously adjust their strategies in the process of making decisions to achieve game equilibrium. Therefore, our work introduces evolutionary game theory into the data-sharing process and investigates the stability of user participation strategies.

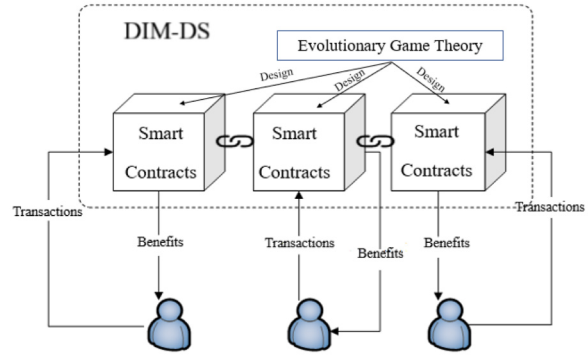

Fig. 1. Overall framework of DIM-DS's incentive mechanism diagram.

Although all of the above studies have achieved some results, the fact that the existing studies on data-sharing incentives do not maximize the proportion of participating users in the system and do not analyze the stability of user participation strategies have not been addressed. Besides, existing incentives are designed based on specific scenarios and are poorly migratory, so they cannot be widely applied. Therefore, we propose a dynamic incentive model for data sharing in federated learning based on smart contracts and evolutionary game theory (DIM-DS).

III. Model Design

The DIM-DS model consists mainly of incentives based on evolutionary game theory and smart contracts based on incentives. The two are combined together to form the DIM-DS model. The DIM-DS model introduces the concept of "credibility coins" by combining a payment reward mechanism with a credibility incentive. Users are trained in a federated learning model to earn "credibility coins." Combined with the deductive game theory, the model simulates the game behavior of users participating in data sharing and allows them to adjust their strategy in relation to the information they currently have. Thus, the model facilitates the users' participation in collaboration. Combined with smart contract technology, users can join the consortium chain sooner to participate in the federated learning model training by dynamically adjusting their reward income. Thus, the model gains higher accuracy. The model consists of two parts: 1) an evolutionary game theory-based incentive mechanism and 2) a smart contract design based on evolutionary game theory incentives, as shown in Fig. 1. The federated chain is used for reasons of the superior computational performance of the blockchain [18]. We use the form of storing user data summaries on-chain and data off-chain. After the data requester sends a data-sharing request, the system retrieves the data owner and makes the data owner respond to the data-sharing request. In the process, we propose an incentive mechanism to encourage more data requestors and owners to participate in sharing.

The main stages involved in credibility coins are as follows.

1) At the user registration stage, the system allocates a certain number of credibility coins to new users, and protects the privacy of the number of reputation coins held by the user by means of encryption.

TABLE I
MEANING OF DIM-DS MODEL SYMBOLS

Symbol	Meaning
P	The totality of users participating in data sharing
N	Collection of related data providers
S	Strategy space available for data
U	Payout matrix of both sides of the game

2) At the data request stage, to avoid users posting multiple requests, the data requestors needs to pay a certain number of credibility coins as a deposit for posting data request announcements.
3) At the data-sharing stage, the degree of contribution from the data provider is assessed based on the similarity of the data and the quality of the trained model. The data provider who contributes more to the model will share more credibility coins. The number of credibility coins available to the data provider is the product of the deposit paid by the requesting party and the contribution value of the data provider.
4) After the stage of the federated learning is completed, the data requester pays a credibility coin to obtain the trained model, and other users can also pay a corresponding credibility coin to obtain the trained model.

A. Evolutionary Game Theory Part

The content of this section is mainly to research incentives based on the evolutionary game theory. We build the DIM-DS model. Our work analyzes the stability of our model to find a stable point where all system users participate in data sharing and collaboration tasks. In response to a data-sharing request, the data providers can choose whether to participate in data sharing. In this model, it is assumed that the data providers participating in data sharing have bounded rationality. Data providers make corresponding decisions based on the profit situation, with the maximization of personal benefits as the strategic goal. It should be noted that participating in data sharing mentioned in this section is a federated learning collaboration task in the process of participating in sharing.

1) Model Building: The DIM-DS model is a symmetrical user-participating data-sharing evolution model composed of quaternion array $G = (P, N, S, U)$. The users in this model are related data holders in the data request node selection stage. The meaning of each symbol is in Table I.

In the data-sharing game model, relevant data providers make decisions about whether to participate in data sharing by calculating the sharing benefits and sharing costs. User data-sharing benefits can be divided into two categories: 1) direct benefits and 2) collaboration benefits.

Direct benefits refer to the benefits that data providers can obtain from participating in data sharing. For the ith data provider, direct benefits are mainly related to the similarity ρ_i of data sharing request and the degree of credibility it own. The credibility is measured by the number m_i of credibility coins owned by the data holder. We record the direct benefit of this component for participating in data sharing as $k_1\rho_i + k_2 m_i (k_1 > 0, k_2 > 0)$. This formula indicates that the benefits of user participation in data sharing are positively correlated with the similarity ρ_i of the requested data and the credibility coins held by the user. Users with higher data similarity or higher credibility will receive more benefits as an incentive to participate honestly in the collaboration.

TABLE II
MEANING OF THE DIM-DS MODEL EARNINGS COMPONENT SYMBOLS
(NOTE: IF ρ , m, AND C ARE NOT SUBSCRIPTED, THEY REFER TO GENERAL USERS)

Symbol	Meaning
ρ_i	Similarity of user data for the i_{th} data provider
k_1	Scaling parameters between data similarity and benefit
k_2	Scaling parameters between the number of credited coins and benefit
m_i	Number of credibility coins held by users for the i_{th} data provider
λ	Collaboration benefit correlation coefficient
C_i	User data sharing participation cost for the i_{th} data provider
I_i	Dynamic incentive participation benefits for the i_{th} data provider

When both users i and j involved in the decision choose to participate in data sharing, the fusion of data and information between the two parties promotes the process of federated learning and reduces the difficulty of solving data request problems. Both parties involved in data sharing can reap the benefits of collaboration in the federated learning process. Based on the above analysis, the synergistic benefits between users can be recorded as formula $\lambda(k_1\rho_i + k_2 m_i)$ and $\lambda(k_1\rho_j + k_2 m_j)(\lambda > 1)$. According to the rationality analysis in the real environment, the benefits of collaboration need to be met at $1 + 1 > 2$ to promote the sharing of data between the two parties.

In the data-sharing stage, users need to spend a certain amount of data-sharing costs. These costs consist mainly of the time and computational costs to be consumed during the training phase of the federated learning model. The system first calculates the average training cost of a particular model in federated learning. This is used as a standard to calculate the confidence level of the training cost reported by the data provider of that model, which is calculated in [23]. This confidence level is used to calculate the user cost that is recognized by the system, denoted as C. Users who choose not to participate in data sharing do not consume these costs, and they also do not receive the benefits of data sharing.

In addition, the goal of this incentive mechanism is to encourage more users to participate in data sharing. When the initial number of users participating in data sharing is too small, users can be more inclined not to participate in decision making under fixed income conditions. Therefore, to encourage system members to participate more, the system is set up to give a portion of the reward to users who participate in the sharing phase when the initial number of participating users is small. The dynamic income of the reward is recorded as $\boldsymbol{I}$.

In conjunction with the above analysis, the symbols involved in the building of the DIM-DS model in this section are shown in Table II.

For the similarity $\boldsymbol{\rho}$ of user data, we still follow the method defined in our previous work [17].

2) Payoff Matrix: Assuming that the users participating in the data-sharing evolutionary game are A and B, the

TABLE III
PAYOFF MATRIX FOR CALCULATING THE BENEFITS OF A'S PARTICIPATION

User A's decision	User B's decision: Participation	User B's decision: Nonparticipation
Participation	$\lambda(k_1\rho_A+k_2m_A)+I_A-C_A$	$(k_1\rho_A+k_2m_A)+I_A-C_A$
Nonparticipation	0	0

TABLE IV
PAYOFF MATRIX FOR CALCULATING THE BENEFITS OF B'S PARTICIPATION

User A's decision	User B's decision: Participation	User B's decision: Nonparticipation
Participation	$\lambda(k_1\rho_B+k_2m_B)+I_B-C_B$	0
Nonparticipation	$(k_1\rho_B+k_2m_B)+I_B-C_B$	0

decision-making situation of the two users can be divided into the following three categories.

First, both A and B participate in data sharing. In this case, the relevant data providers ($P = P_1, P_2, \ldots, P_n$) that participate in the data-sharing federated learning all participate in the federated learning task. The benefits obtained by the users from participating in the data sharing belong to the collaborative benefits. The benefits of both parties P_A and P_B in this case are formula $\lambda(k_1\rho_A+k_2m_A)+I_A-C_A$ and $\lambda(k_1\rho_B+k_2m_B)+I_B-C_B$. Second, neither A nor B participates in data sharing. In this case, none of the data providers will participate in the federated learning task. Therefore, under this type of decision, the user cannot obtain benefits, and the user does not need participation costs. In other words, the benefits of both parties are zero. Third, only one user between A and B participates in data sharing. In this case, the benefits of users who participate in data sharing are direct benefits, in other words, formula $(k_1\rho_A+k_2m_A)+I_A-C_A$ for P_A or $(k_1\rho_B+k_2m_B)+I_B-C_B$ for P_B. However, the benefits of users who do not participate in data sharing are zero.

3) Reproduction of Dynamic Equations: Since users of this system can only choose to participate or not to participate, the game process of user A and user B conforms to the "general two-person symmetric game."

The payment matrix for calculating the benefits of A's participation in the decision is shown in Table III.

The payment matrix for calculating the benefits of B's participation in the decision is shown in Table IV.

Assume that in a group N composed of all relevant data providers in this system, and the probability that each user chooses to participate in the data sharing strategy is x. Therefore, the probability that users choose not to participate in data sharing is $1-x$, where x is a function of time t, $x \in [0, 1]$. User A must participate if he wants to receive the benefit. Then, A's benefit is $x[\lambda(k_1\rho_A+k_2m_A)+I_A-C_A]+(1-x)[(k_1\rho_A+k_2m_A)+I_A-C_A]$, where $x[\lambda(k_1\rho_A+k_2m_A)+I_A-C_A]$ is the payoff to user A if B participates with probability x and $(1-x)[(k_1\rho_A+k_2m_A)+I_A-C_A$ is the payoff to user A if B does not participate with probability $(1-x)$. The situation of user B is analyzed in the same way as user A.

According to the above analysis, we can obtain the benefits of user P_i who choose to participate in the strategy during the data-sharing decision-making stage, as shown in

$$u(s_1,x) = x[\lambda(k_1\rho_i+k_2m_i)+I_i-C_i] + (1-x)[(k_1\rho_i+k_2m_i)+I_i-C_i]. \quad (1)$$

The benefits of users who choose not to participate in the strategy are shown in

$$u(s_2,x) = 0. \quad (2)$$

The average revenue for the users is given in

$$\begin{aligned}\overline{u} &= x\cdot u(s_1,x)-(1-x)\cdot u(s_2,x)\\ &= x[(\lambda(k_1\rho_i+k_2m_i)+I_i-C_i)\\ &\quad + (1-x)((k_1\rho_i+k_2m_i)+I_i-C_i)]. \end{aligned}\quad (3)$$

According to the Malthusian equation, the user's replication dynamic equation in the DIM-DS model can be expressed by the growth rate of the sharing strategy, as in

$$\begin{aligned}F(x) &= \frac{\mathrm{dx}}{\mathrm{dt}} = x\cdot[u(s_1,x)-\overline{u}]\\ &= x(1-x)\cdot[u(s_1,x)-u(s_2,x)]\\ &= x(1-x)\cdot[x(\lambda(k_1\rho_i+k_2m_i)+I_i-C_i)\\ &\quad + (1-x)((k_1\rho_i+k_2m_i)+I_i-C_i)]. \end{aligned}\quad (4)$$

4) Stability Analysis: The stability of the above DIM-DS model is analyzed using the local stability analysis method proposed by Friedman. According to the stability theorem of the replicated dynamic equation, the stability strategy of the user can be judged and analyzed by bringing the equilibrium point x into (4).

We set $F(x) = 0$, and then, we can obtain three stable states of the user's replication dynamic equation $F(x)$, which are $x_1 = 0$, $x_2 = 1$, and $x_3 = ([C-k_1\rho_i-k_2m_i-I]/[(\lambda-1)(k_1\rho_i+k_2m_i)])$, respectively.

In analysis based on the evolutionary stability strategy (ESS) of the game and the stability theorem, a steady state should be stable when there is a small disturbance in the dynamic system. In other words, it should satisfy

$$\begin{aligned}F'(x) &= [(1-2x)(x(\lambda(k_1\rho_i+k_2m_i+I_i-C_i)\\ &\quad + (1-x)((k_1\rho_i+k_2m_i)+I_i-C_i))\\ &\quad + x(1-x)(\lambda-1)(k_1\rho_i+k_2m_i)] < 0. \end{aligned}\quad (5)$$

Substituting the above three stable (x_1, x_2, x_3) points into (5), we can get

$$\begin{cases} F'(x_1) = (k_1\rho_i+k_2m_i)+I_i-C_i) \\ F'(x_2) = -[\lambda(k_1\rho_i+k_2m_i)+I_i-C_i] \\ F'(x_3) = \frac{(C-k_1\rho_i-k_2m_i-I_i)[\lambda(k_1\rho_i+k_2m_i)+I_i-C_i]}{(\lambda-1)(k_1\rho_i+k_2m_i)}. \end{cases} \quad (6)$$

Combining (6), we analyze the player's ESS from the following situations.

1) *Condition I:* $(C-k_1\rho_i-k_2m_i-I_i) > (\lambda-1)(k_1\rho_i+k_2m_i)$. This condition means that the cost of users participating in data sharing is higher than the benefits that can be obtained. At this time, $F'(0) < 0$, $F'(1) > 0$, and x_3 does not exist. Therefore, x_1 is the only evolution equilibrium point of the system. We conclude that the final evolution

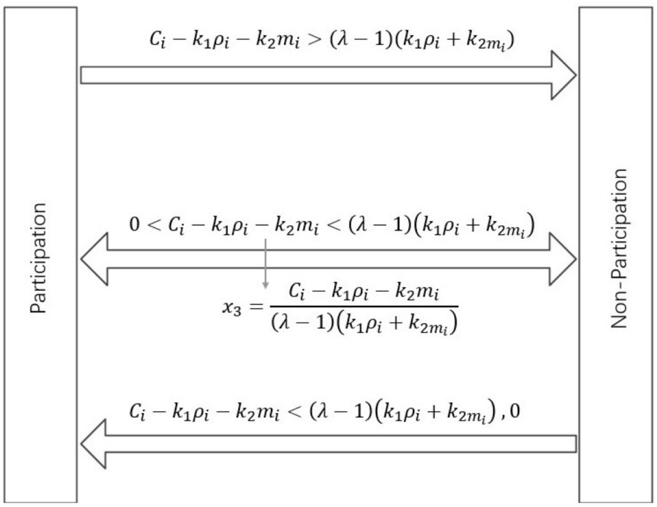

Fig. 2. Summary of the ESS trend of the DIM-DS model.

strategy ESS of the system users does not participate in data sharing.

2) *Condition II:* $(C - k_1\rho_i - k_2m_i - I_i) < (\lambda - 1)(k_1\rho_i + k_2m_i)$. This condition means that the collaboration benefits that users can obtain from participating in data sharing are higher than the participation costs. At this point, $F'(1) < 0$, and x_2 is explained as the stable equilibrium point of the evolution of the system. It is also explained that one of the evolution strategies of the users is to take part in data sharing. Under this condition, we can deduce the formula $\lambda(k_1\rho_i + k_2m_i) + I_i - C_i > 0$, but $\lambda > 1, k_1\rho_i + k_2m_i > 0$. Thus, the value of $F'(x_3)$ depends on $C - k_1\rho_i - k_2m_i - I_i$. Because of the formula $F'(X_1) = k_1\rho_i + k_2m_i) + I_i - C_i$, it is possible to analyze the different value ranges of formula $C - k_1\rho_i - k_2m_i - I_i$ according to the current conditions. The specific analysis process is as follows.

Condition (1): $C - k_1\rho_i - k_2m_i - I_i > 0$. Under this condition, $F'(x_3) > 0$, $F'(x_1) < 0$. Therefore, $x_1 = 0$ is also the evolutionarily stable equilibrium point of the model. The final decision of the system users will depend on the proportion x^* of users in the model who initially participate in data sharing. When $x^* < x_3$, the strategy tends to be the nonparticipation strategy; when $x^* > x_3$, the strategy tends to be the participation strategy.

Condition (2): $C - k_1\rho_i - k_2m_i - I_i < 0$. Under this condition, $F'(x_3) < 0$, $F'(x_1) > 0$. $x_3 < 0$ does not exist. Therefore, users will all eventually choose to participate in data sharing.

The above content analysis the evolutionary and stable trend of data sharing. The evolutionary trend under different conditions is summarized in Fig. 2. Whether users choose to participate in data sharing or not to participate in data sharing affects the evolutionary stability, depending on which conditions are met. The ESS depends on the benefits obtained from the sharing of user data and the proportion of users involved in the initial state of the system.

The participant benefits are related to the user data similarity ρ, the number of tokens held by users m, the participation cost C, and the dynamic incentive income I. Therefore, the above-mentioned relevant parameters of the participation income can be modeled and analyzed according to different conditions. Among them, in the case in which users can allocate fixed income, dynamic incentive income is a key factor in increasing the proportion of system users who participate.

Based on the above stability analysis, for the data provider P_i we can obtain the range of values of I_i that motivate users to participate in data sharing, i.e.,

$$\frac{C_i - k_1 - k_2 - I_i}{(\lambda - 1)(k_1\rho_i + k_2m_i)} < x^*. \tag{7}$$

An equivalent transformation of (7) yields an interval lower bound e.t. $I_i > C_i - [(\lambda - 1)x^* + 1][k_1\rho_i + k_2m_i]$ for I. Considering the realistic scenario where the giver of I always wants the minimum sum of I used to motivate the system users to participate in data sharing, and because satisfying (7) the conclusion that data providers always tend to participate in data sharing is obtained according to the evolutionary game theory, so we set I of each data provider P_i to $I_i = C_i - [(\lambda - 1)x^* + 1][k_1\rho_i + k_2m_i] + 1$. It can be seen that the size of I is related to the order of participation of data providers and is inversely proportional to the quality of data providers (i.e., the number of credibility coins m and data similarity ρ_i), and the initial ratio x^*. Therefore, before determining I for each data provider, they are sorted in ascending order by $k_1\rho_i + k_2m_i$.

Therefore, to encourage users to participate in data sharing, the following strategy can be used: when the initial proportion of users participating in data sharing is lower than x_3, the incentive income I is continuously increased to increase the participation proportion. The magnitude of the increase is determined by the product of data similarity and reputation value. The increase is greater as the data similarity and reputation values of the users increase. When the participation ratio exceeds that number, there is no need to increase the profit at this time. Therefore, the system can achieve the maximum participation ratio.

B. Smart Contract Design Part

To encourage users to participate in data sharing and to minimize the cost for incentives, the participation ratio of users in the current system can be queried and the user's decision can be evolved into a participation strategy by dynamically adjusting the incentive benefit. Therefore, a smart contract based on the DIM-DS model can be designed. When a user chooses to make a decision, the system executes the code to calculate the user participation rate; then, the system dynamically adjusts the incentive benefit based on the participation rate.

1) Design of Variables: A UML class diagram of smart contracts based on the DIM-DS model is shown in Fig. 3.

The relevant variables in this smart contract are collated as in Table V. benefit is fixed benefit and dynamic incentive benefit that can be earned by users while participating in data sharing. Four main variables are maintained: k_1, k_2, cost, and cooperation, which correspond to k_1, k_2, C, and λ

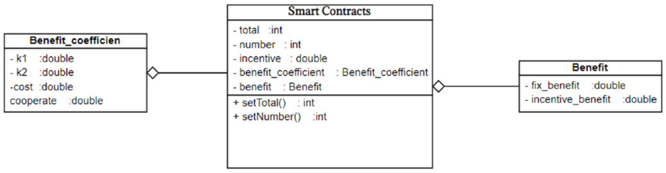

Fig. 3. UML class diagram of smart contracts based on the DIM-DS model.

TABLE V
MEANING OF THE VARIABLES RELATED TO SMART CONTRACTS BASED ON THE DIM-DS MODEL

Symbol	Meaning
$total$	The total number of data providers related to the data request
$number$	Total number of users currently participating
$benefit$	Fixed income and incentive income received by users
$incentive$	Dynamic incentive income
k_1	Similarity correlation coefficient
k_2	Credit currency correlation coefficient
$cost$	Cost of participation
$cooperation$	Coefficient of income from collaborative tasks

Algorithm 1: Smart Contracts

Input: the set of data providers $\boldsymbol{P}, \boldsymbol{\rho}, \boldsymbol{m}, \boldsymbol{C}$, ***cooperation***
Output: ***expected benefit***

```
total ← setTotal(P)
Sort by k1ρi + k2mi ascending order
for i ← 1 to total do
    x' = (cost_i − k1ρi − k2mi − incentive_i) / ((cooperation − 1)(k1ρi + k2mi))
    x* = number / total
    if x* < x' then
        incentive_i = cost − [(cooperation − 1)x* + 1][k1ρ + k2m] + 1
    end
    user join in sharing data setNumber()
end
return
```

in Table II of the DIM-DS model building, respectively. The variable incentive corresponds to I.

2) Algorithm Function Design: The above smart contract contains three functions. The functions's relevant meanings are as follows.

setTotal(): This function sets the total number of users, defined as total, in the evolutionary game model DIM-DS based on the total number in the set $P = \{P_1, P_2, \ldots, P_n\}$, which consists of relevant data providers queried by the data request.

setNumber(): When users participate in data sharing, this function updates the total number of users currently participating in the system.

The algorithm details of smart contracts are shown in Algorithm 1.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

To synthesize the impact of the factors in the DIM-DS model on the system user policy, we simulate different factors under different conditions. The ode45 function is used to solve the replication dynamic (4). Setting different values of k_1, k_2, C, λ, and I based on previous studies and the analysis of condition A and condition B in the stability analysis, we verify the effect of the parameters on the ESS of the system and the dynamic incentive income on user data sharing under various conditions. We accomplish this goal by fixing the values of the other incentive-related parameters and varying linearly only the values of the incentive-related parameters to be analyzed. We later simulate the evolution to show how the proportion of user participation varies with time t. Here, t varies continuously.

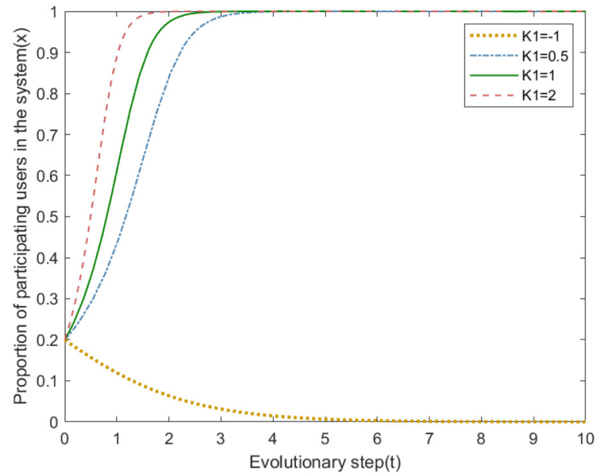

Fig. 4. Evolutionary plot of the change in data similarity coefficient k_1.

A. Influence of the Factors on ESS

1) Impact of Data Similarity on ESS: Data similarity is an important factor in the calculation of user benefits. The evolutionary process of the proposed DIM-DS model is obtained by increasing the data similarity coefficient k_1 while other factors remain unchanged. As shown in Fig. 4, in this case, $k_2 = 0.2$, $C = 5$, $\lambda = 2$, $I = 3$, $\rho = 0.8$, $m = 10$, and the initial proportion of participating users in the system is 0.2.

Fig. 4 shows that the data similarity positively affects the users' participation in data sharing in the DIM-DS model. As the data similarity coefficient increases, the likelihood of the system converging to $x = 1$ increases, and convergence is accelerated. The probability of users choosing to participate in data sharing is improved. It can be verified that the incentive mechanism provides a better incentive for users with higher data similarity.

2) Impact of Creditworthiness on the ESS: To guarantee credible sharing in the data-sharing process, the degree of credibility is incorporated into the calculation of the user's benefits. The evolutionary process of the proposed DIM-DS model is obtained by increasing the credibility degree coefficient k_2 while other factors held constant, as shown in Fig. 5. The values of the remaining relevant parameters are $k_1 = 1$, $C = 5$, and $\lambda = 2$, $I = 3$, $\rho = 0.8$, and $m = 10$. Fig. 5 shows that the degree of credibility positively affects the user participation in data sharing in the DIM-DS model and that the degree of

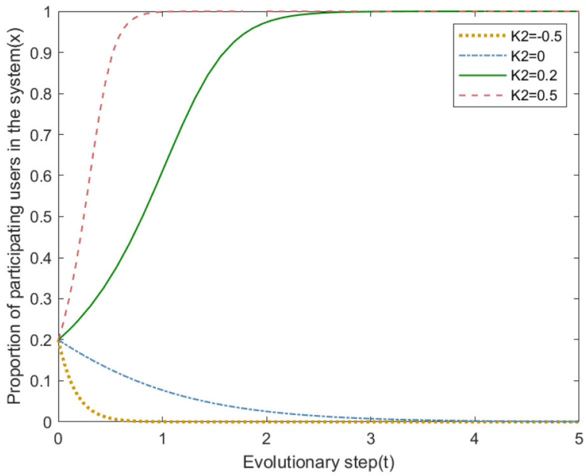

Fig. 5. Evolutionary graph of the change in the user credibility coefficient k_2.

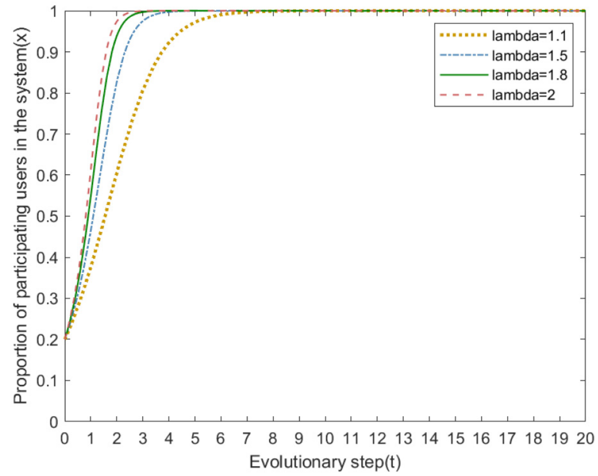

Fig. 6. Evolutionary graph of the change in the coefficient of return to collaboration λ.

credibility is an important factor in user participation in decision making. As the user reputation coefficient increases, the likelihood of the system converging to $x = 1$ increases, and the convergence rate is accelerated. The probability of users choosing to participate in data-sharing improves.

3) Impact of Collaboration on ESS: As the federated learning task in DS2PM [17] proposed in our previous work requires multiple parties to be involved in training the same machine learning, the task is a collaborative task for the user. To analyze the impact of the collaboration task on the level of user engagement, the evolutionary process of the proposed DIM-DS model was achieved by increasing the value of the collaboration gain coefficient while other factors remained constant, as shown in Fig. 6. The values of the remaining relevant parameters are $k_1 = 1$, $k_2 = 0.2$, $C = 5$, $I = 3$, $\rho = 0.8$, and $m = 10$. Fig. 6 shows that increasing the appropriate collaboration gain for the specificity of the federated learning task positively affects user engagement data sharing in the DIM-DS model. The rate of user evolution accelerates as the coefficient of collaborative gain increases. It can also be analyzed from Fig. 6 that we have designed the incentives in such a way that the ultimate strategy of the users is to participate in data sharing. This finding means that users pursue their individual interests while maximizing their collective interests, thus satisfying the requirement for incentive compatibility.

4) Impact of Participation Costs on ESS: Users consume a certain amount of computing resources and time costs when participating in the data-sharing process. To verify the influence of this part of the value on the evolutionary trend of the system, this section verifies the influence of the participation cost on the ESS by reducing the participation cost while other factors remain unchanged. The evolutionary process of the DIM-DS model obtained is shown in Fig. 7. The values of the remaining relevant parameters are $k_1 = 1$, $k_2 = 0.2$, and $\lambda = 2$, $I = 3$, $\rho = 0.8$, and $m = 10$. Fig. 7 shows that the magnitude of the participation cost negatively affects user participation in data sharing in the DIM-DS model. As the user participation cost decreases, the probability of the system converging to $x = 1$ increases and converges faster.

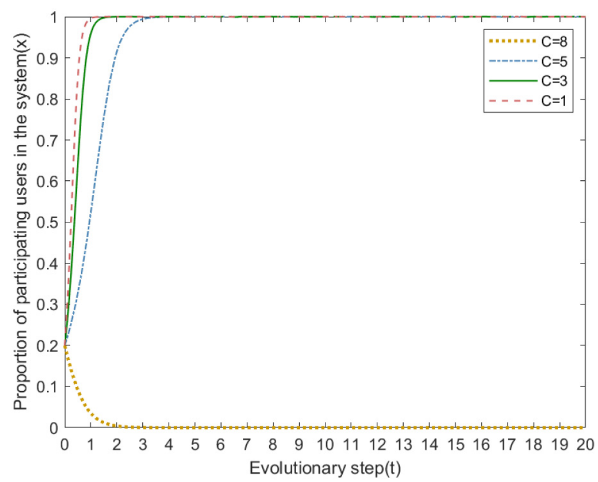

Fig. 7. Evolutionary graph of the change in user participation cost C.

The probability of users choosing to participate in data sharing is enhanced. Users are more likely to choose not to participate in the decision when the benefits they gain from data sharing do not meet the losses from the participation costs.

5) Impact of Incentive Benefits on ESS: As the participation cost, data similarity and reputation of the system's users are fixed values, the benefits calculated are also fixed. When the fixed benefit does not provide sufficient incentive for the user to participate in the strategy, users are more likely to choose not to participate. However, increasing the weight of each parameter could lead to excessive participation benefits for system users, and the data requestor might not be able to meet the shared benefits of the data provider. Therefore, it is necessary to reasonably control the weight proportion of each factor. At the same time, the system dynamically adjusts the benefits by introducing dynamic incentive benefits I to promote the users' participation in data sharing.

This section ensures that the other factors remain unchanged and verifies the impact of the dynamic incentive benefit I on the ESS by adding I. In this case, the dynamic incentive starts I from a negative number and makes linear increases to adjust

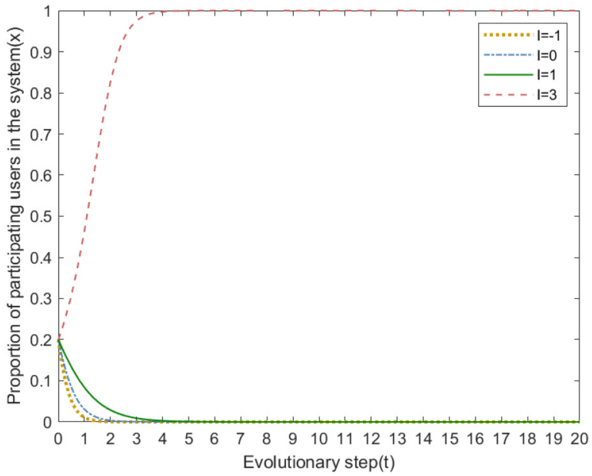

Fig. 8. Evolutionary graph of the change in incentive gain I.

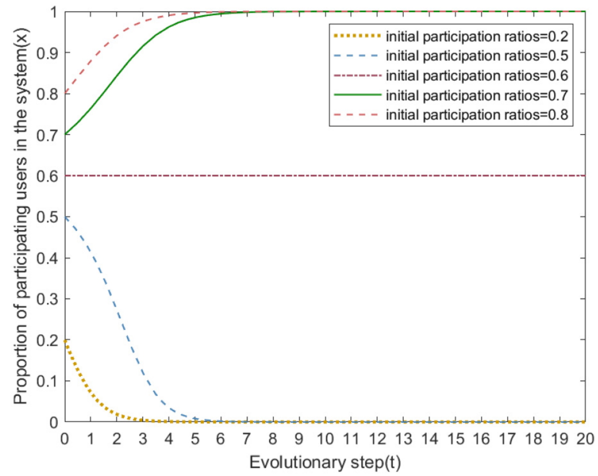

Fig. 9. Evolutionary graphs for different initial participation ratios.

it. The evolution of the DIM-DS model obtained is shown in Fig. 8. The values of the remaining relevant parameters are: $k_1 = 1$, $k_2 = 0.2$, $C = 5$, $\lambda = 1.5$, $\rho = 0.8$, and $m = 10$. From Fig. 8, it can be seen that $I = -1$ indicates that the incentive does not increase the return on top of the user's original fixed return. However, it makes the user hold less return. Therefore, the final evolutionary strategy of the system user is not to participate in data sharing. When $I = 0$, the incentive maintains the original fixed return. The user's final evolutionary strategy is to slow down the rate of user evolution, although the user still does not participate. Similarly, when $I = 1$, the incentive helps to increase the user's original fixed benefit. However, the incentive is still not sufficient to make the user's evolutionary strategy tend to participate in the decision, although the rate of evolution is somewhat slowed. Therefore, it is possible to analyze from this finding that increasing the incentive gain promotes the user's participation in data sharing, even if the final evolutionary strategy of the model tends toward participation in decision making. This conclusion can be verified by the graph of $I = 1$.

The above analysis shows that as the incentive gain increases, the likelihood of the system converging to $x = 1$ increases and the rate of convergence is accelerated. The probability of users choosing to participate in data sharing is improved.

6) Impact of the Initial User Participation Rate on ESS: In this section, analysis is conducted for the user participation ratio. Therefore, each parameter is set as follows: $k_1 = 1$, $k_2 = 0.2$, $C = 6$, $\lambda = 1.5$, $I = 2$, $\rho = 0.5$, and $m = 10$. Substituting the above parameters into the stable point $x_3 (x_3 = ([C - k_1\rho - k_2 m - I]/[(\lambda - 1)(k_1\rho + k_2 m)])$ of the replication dynamic equation $F(x)$ of (5), we can derive a participation ratio threshold for the replicated dynamic equation conditional on this parameter, which is thres $= 0.6$. According to this threshold value, the trend of the user evolutionary strategy under different initial participation proportions is studied, and Fig. 9 is obtained.

From Fig. 9, it can be determined that when the initial participation ratio of the system is below the threshold value of 0.6, users are more inclined not to participate in the current calculated return scenario. When the initial participation ratio of the system is just under thres condition, the user participation ratio of the system always remains at 0.6, where thres is the ESS under the current parameter condition. When the user participation ratio of the system increases, the evolution strategy of the system users gradually shifts to the participation strategy, and the ratio is higher. The remaining users choose the participation strategy faster.

TABLE VI
VERIFICATION OF THE PARAMETERS OF THE DYNAMIC INCENTIVE MECHANISM

Test item	parameters		Test item	parameters	
	Θ	I		Θ	I
item1	0.2	0	item9	0.7	0
item2	0.2	1	item10	0.7	1
item3	0.2	2	item11	0.7	2
item4	0.2	3	item12	0.7	3
item5	0.45	0	item13	0.95	0
item6	0.45	1	item14	0.95	1
item7	0.45	2	item15	0.95	2
item8	0.45	3	item17	0.95	3

B. Analysis of Dynamic Motivational Effects

1) User Strategy Stability Analysis: To verify the effect of the proposed dynamic incentive mechanism on the evolutionary outcome of the system user policy, the analysis is conducted in this section in conjunction with different initial participation ratios and incentive gains I. The parameters in the replication dynamic equations of the DIM-DS model, except for the incentive gain I, are held constant on the basis of subsection impact of the initial user participation rate on ESS. Let the initial participation ratio be Θ. The values of Θ and the values of I for the experimental design in this section are shown in Table VI.

Simulation analysis combined with the parameter values in the table results in the user evolution strategy graph shown in Fig. 10.

Fig. 10 shows that when the value of I is 0, the final evolutionary strategy of the system users is nonparticipation,

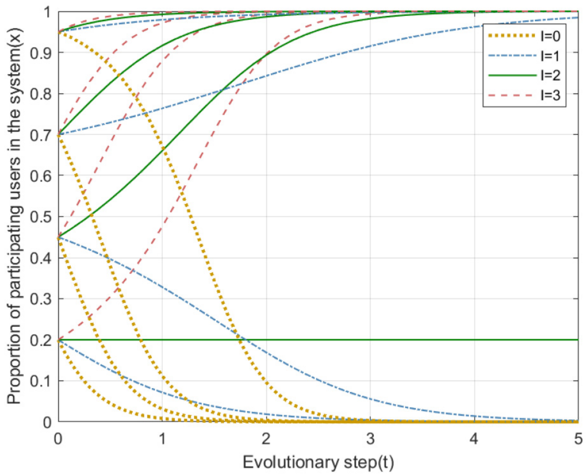

Fig. 10. Evolution curve diagram under different incentive income and initial participation ratios.

regardless of the initial participation ratio Θ. When increasing the value of I to 1, the user's evolutionary strategy is still nonparticipation when the initial participation ratio is low ($\Theta = 0.2$ and $\Theta = 0.45$). However, when the value of Θ is 0.7, the user's decision gradually evolves to a participation strategy. When the value of I is 2, the user evolution strategy stabilizes at that point except when the value of Θ is 0.2. However, by the time the participation ratio is raised, the final evolution strategy of the system users is to participate in data sharing. The users' evolutionary strategy is to participate in data sharing when the incentive gain continues to increase to a value of 3 for I.

Based on the above analysis, it can be verified that when the initial proportion of participating users in the system is low, such as $\Theta = 0.2$, the incentive gain can be increased appropriately ($I = 3$) to promote user participation in data sharing. This approach allows the number of participating users in the system to increase. The corresponding proportion of participating users is increased. The incentive gain can be gradually adjusted; for example, when the value of Θ is 0.45, the value of the incentive gain I is 2, and it still allows the evolutionary strategy of the users to tend toward participation. Similarly, when the value of I is 1, the participation ratio $\Theta = 0.7$ still promotes users' participation in data sharing, although the participation gain is relatively reduced.

2) Impact of the Participation Strategy on the Model Accuracy: To evaluate the impact of the proposed dynamic incentive mechanism on the federated learning process in this data sharing study, an experimental analysis was conducted through the accuracy of the machine learning models trained with different numbers of user participants. The model accuracy was calculated as shown in

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}). \quad (8)$$

The meanings of the symbols in the above equations are listed in Table VII.

The machine learning experiment environment is built in Python (version 3.7.4). The TensorFlow library (version 2.3.2) and Keras Library (version 2.4.3) are introduced to build

TABLE VII
MEANING OF SYMBOLS IN THE MODEL ACCURACY EQUATIONS

Symbol	Meaning
TP	Number of predicted positive classes as positive classes
FP	Number of negative classes predicted to be positive
TN	Predicted number of negative classes as negative classes
FN	Number of predicted positive classes to negative classes

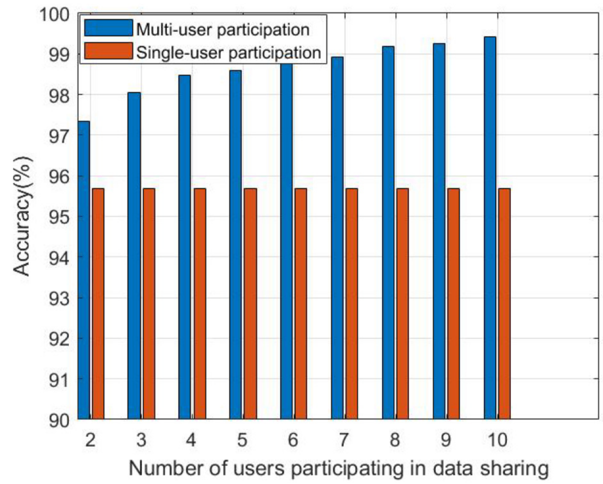

Fig. 11. Model accuracy rate under different user participation numbers.

machine learning models. The dataset used in the experiment is the MNIST dataset. Each data point is a picture of handwritten digits containing 0 to 9. MNIST contains 60 000 training datasets and 10 000 test datasets. The convolutional neural network (CNN) model is used as an experimental training model. During the experiment, the training dataset was randomly divided into ten datasets of equal size, and each group contained 6000 data points, which were assigned to ten users. Each user can choose whether to participate in the model training of federated learning based on his local dataset. Then, the experiment simulates the model training situation under the participation of different numbers of users, with only one user participating as the comparison object. The experimental results obtained are shown in Fig. 11.

Fig. 11 shows that the larger the number of collaborative users who participate in the training of the model, the larger the total training data of the model. Thus, the accuracy rate is higher.

C. Comparative Analysis

Our work proposes for the first time to solve the incentive problem in the process of federated learning based on evolutionary game theory. Through the above simulation and experimental analysis, the proposed DIM-DS model can effectively encourage users to participate in data sharing. In the system simulation experiments, our work sets the total number of users in the system to 3000. To correspond to Section IV-A, each coefficient $k_1 = 1$, $k_2 = 0.2$, and $\lambda = 2$ is set. We consider the participation cost of users to be normally distributed with a mean of 10 and the data correlation to be randomly distributed within [0, 1]. The simulation comparison results are shown in Fig. 12. For the same set

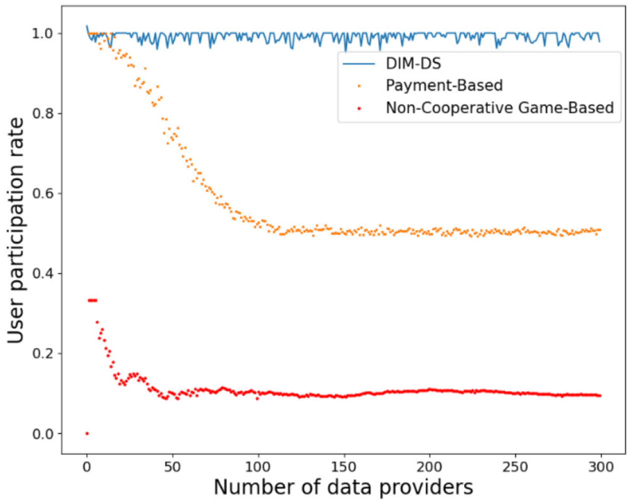

Fig. 12. Participation rates of users in different sizes under the condition of satisfying the evolutionary game.

of data providers (set size from 1 to 300), we compare the user participation rate of the DIM-DS incentive mechanism with the payment-based incentive mechanism [11] and the noncooperative game-based incentive mechanism [13] under the condition that the evolutionary game is satisfied. When the evolutionary game condition is satisfied, users tend to adopt the "participation" strategy, and the user strategy is more stable. At the same time, the payment-based incentives are not stable in terms of user participation rate, and the noncooperative game-based incentives have limited incentive effect, as shown in Fig. 12. Under the condition of satisfying the evolutionary game, the incentive benefit received by the data provider always makes it inclined to adopt the "participation" strategy, so that all data participants in the system can participate in the data sharing under the condition of sufficient time. Payment-based incentives, such as [11], do not take into account the relationship between the amount of payment and users' willingness to participate, so when the total incentive amount is certain, the strategy adopted by users is not stable and the incentive effect is mediocre. Meanwhile, the incentive mechanism based on noncooperative game [13] is not very effective because the benefits between users are tit-for-tat, so only those users with high final benefits tend to participate in data sharing.

1) *Fairness:* Usually, the value of data is evaluated by a third-party central agency, and a corresponding pricing plan is given. However, the third-party central agency could cooperate with dishonest users to conduct common evils and act unfairly against users. In the established DIM-DS model, users obtain expected benefits by invoking smart contracts. The expected benefits are automatically calculated based on the user's data similarity and credibility. The expected benefits do not require review by a third-party central agency to ensure the participants' fairness. In collaborative tasks, data similarity and contribution value are positively correlated, i.e., the higher the contribution value, the higher the fixed benefit, and therefore fair.
2) *Bounded Rationality:* The dynamic incentive model proposed in our work combines the ideas of evolutionary game theory. Compared with the current game theory, it no longer assumes that users are completely rational. However, it can continuously adjust their own decisions based on the limited information they have obtained. It conforms to the characteristics of the user's limited rationality in the real world.
3) *Stability:* The DIM-DS model analyzes the equilibrium points of the system user strategies under different conditions based on the "Nash equilibrium" in game theory. The model combines smart contract technology and dynamically adjusts the user's participation benefits in such a way that the user's final evolution strategy tends to participate to make decisions and to maintain stability. The stability of user participation includes two aspects, one is that the user participation rate is subject to small changes in the size of data providers in the system, and the other is that the user participation rate can fluctuate up and down at a high level. As the size of data providers in the system increases, the change of user participation rate is influenced by different incentives. Under the condition of satisfying the evolutionary game, the participation rate of data providers is always maintained above 95%, and it also has high theoretical stability (analyzed in Section III). Ceteris paribus, under the payment-based incentive mechanism, the user participation rate decreases rapidly as the data provider size increases when the data provider size is small, so it is less stable in small and medium-sized data-sharing systems (data provider size is less than 100), while it can have better stability in large-scale systems (data provider size is generally above 100), and the user participation rate is maintained at around 50%. Under the incentive mechanism based on noncooperative game, the user participation rate shows more stable results, but the participation level is insufficient and can only incentivize the providers with high data competitiveness.
4) *Incentive Compatibility:* The criterion of incentive compatibility is to achieve the consistency of the ultimate goals of the mechanism designer and the demander. In other words, while individuals pursue the maximization of personal interests, they also maximize collective interests. In the incentive mechanism designed in our work, the data provider related to the data request also belongs to the mechanism demander. As a mechanism designer, the goal is to enable each data provider to participate in federated learning to obtain better model effects. In the mechanism designed in our work, each participant can choose whether to participate in the collaborative task of federated learning. However, it can gain benefits only by participating in the strategy. Under the dynamic incentive mechanism, each user is allowed to participate in the collaboration. The dynamic incentive mechanism maximizes the effect of the final model trained by federated learning and meets the setting of incentive compatibility.

TABLE VIII
SECURITY COMPARISON OF VARIOUS DATA-SHARING SCHEMES

Indicator	CreditCoin [11]	Kang [10]	Lin [13]	DIM-DS
No third party audit required			√	√
Fairness		√	√	√
Bounded rationality	√	√		√
Stability				√
Incentive compatibility	√	√	√	√

5) *Security analysis:* We discuss the security of the system for possible collusion attacks as follows. The key to collusion attacks is that multiple malicious users join together to gain benefits, but in our system, even if multiple malicious users join together, their own reputation coin amounts and data similarity do not change, so they gain the same benefits as they would have without joining together. Therefore, our model is resistant to collusion attacks. In our system, the attack behavior of malicious users is generally to participate in the collaborative task by low quality or even wrong data, but low quality data cannot give malicious users enough benefit to cover the cost, so malicious users have to participate in the collaborative task honestly if they want to get the benefit.

Based on the above analysis, combined with the representative incentive schemes introduced in the overview part of our work, the incentive effects are compared. The comparison results are shown in Table VIII.

V. CONCLUSION

To address the existing studies on data-sharing incentives that do not maximize the proportion of participating users in the system or analyze the stability of user participation strategies, our work proposes a dynamic incentive model for data sharing in federated learning (DIM-DS) based on smart contracts and evolutionary game theory. We compared the impact of different numbers of users participating in model training on model accuracy and verified that the DIM-DS proposed in our work maximizes the proportion of system users participating. Our approach can effectively encourage users to participate in federated learning collaboration and maintain stability, which is significantly better than other solutions. Meanwhile, the proposed DIM-DS model is based on the general data-sharing scenario in federated learning, so it can be used in multiple scenarios of federated learning. Our next work will consider its application to specific scenarios.

REFERENCES

[1] "China institute of electronic technology standardization," Ind. Big Data, White Paper, Apr. 2019, p. 17. [Online]. Available: http://www.cesi.cn/images/editor/20190401/20190401145953698.pdf

[2] J. Wang, "Analysis of the current situation of big data applications in China," *Softw. Integr. Circuits*, no. 11, pp. 72–87, 2020.

[3] "Guiding Opinions on the Development of Industrial Big Data." Ministry of Industry and Information Technology Publish. 2020. [Online]. Available: http://www.cac.gov.cn/2020-05/16/c_1591178516420666.htm

[4] R. Jurca and B. Faltings, "An incentive compatible reputation mechanism," in *Proc. IEEE Int. Conf. E-Commerce*, 2003, pp. 285–292.

[5] S. Zhong, J. Chen, and Y. R. Yang, "Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks," in *Proc. IEEE INFOCOM 22nd Annu. Joint Conf. IEEE Comput. Commun. Soc.*, 2003, pp. 1987–1997.

[6] B. B. Chen and M. C. Chan, "MobiCent: A credit-based incentive system for disruption tolerant network," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[7] U. Shevade, H. H. Song, L. Qiu, and Y. Zhang, "Incentive-aware routing in DTNs," in *Proc. IEEE Int. Conf. Netw. Protocols*, 2008, pp. 238–247.

[8] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 5, pp. 2438–2455, Sep./Oct. 2021.

[9] Y. Yuan and F. Wang, "Blockchain technology development status and outlook," *J. Autom.*, vol. 42, no. 4, pp. 481–494, 2016.

[10] J. Kang *et al.*, "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4660–4670, Jun. 2018.

[11] L. Li *et al.*, "CreditCoin: A privacy-preserving blockchain-based incentive announcement network for communications of smart vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2204–2220, Jul. 2018.

[12] X. Deng and M. Huang, "P2P incentive mechanism based on hybrid strategic game," *Electron. Technol. Appl.*, vol. 36, no. 6, pp. 127–129, 2010.

[13] X. Lin, J. Li, J. Wu, H. Liang, and W. Yang, "Making knowledge tradable in edge-AI enabled IoT: A consortium blockchain-based efficient and incentive approach," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6367–6378, Dec. 2019.

[14] H.-N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for Internet of Things: A survey," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8076–8094, Oct. 2019.

[15] M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7702–7712, Oct. 2019.

[16] Y. Zhan, J. Zhang, Z. Hong, L. Wu, P. Li, and S. Guo, "A survey of incentive mechanism design for federated learning," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 2, pp. 1035–1044, Apr.–Jun. 2022, doi: 10.1109/TETC.2021.3063517.

[17] Y. Chen *et al.*, "DS2PM:A data sharing privacy protection model based on blockchain and federated learning," *IEEE Internet Things J.*, early access, Dec. 13, 2021, doi: 10.1109/JIOT.2021.3134755.

[18] K. Yue *et al.*, "A survey of decentralizing applications via blockchain: The 5G and beyond perspective," *IEEE Commun. Surveys Tuts*, vol. 23, no. 4, pp. 2191–2217, 4th Quart., 2021.

[19] W. Li, H. Qi, R. Xu, X. Zhou, and K. Li, "Research progress and trends of data centre network traffic scheduling," *J. Comput. Sci.*, vol. 43, no. 4, pp. 600–617, 2020.

[20] C. Wang and J. Feng, "Research on the application of federated learning in the insurance industry," *J. Acad. Insur. Prof.*, vol. 34, no. 1, pp. 13–17, 2020.

[21] W. Chen, Y. Mu, X. Liang, and Y. Gao, "Medical data sharing model based on blockchain," *J. Phys. Conf. Ser.*, vol. 1267, no. 1, 2019, Art. no. 12014.

[22] Y. Zhan, P. Li, K. Wang, S. Guo, and Y. Xia, "Big data analytics by CrowdLearning: Architecture and mechanism design," *IEEE Netw.*, vol. 34, no. 3, pp. 143–147, May/Jun. 2020.

[23] J. Kang, Z. Xiong, D. Niyato, H. Yu, Y.-C. Liang, and D. I. Kim, "Incentive design for efficient federated learning in mobile networks: A contract theory approach," in *Proc. IEEE VTS Asia–Pacific Wireless Commun. Symp. (APWCS)*, 2019, pp. 1–5.

[24] M. Xu *et al.*, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies,and challenges," 2022, *arXiv:2203.05471*.

[25] W. Y. B. Lim *et al.*, "Realizing the metaverse with edge intelligence: A match made in heaven," 2022, *arXiv:2201.01634*.

[26] W. Y. B. Lim *et al.*, "Decentralized edge intelligence: A dynamic resource allocation framework for hierarchical federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 3, pp. 536–550, Mar. 2022.

Yanru Chen received the master's degree in management (finance) from the University of Melbourne, Melbourne, VIC, Australia, in 2018. She is currently pursuing the Ph.D. degree with the Embedded Systems and Big Data Management Laboratory and the Internet of Things Laboratory, College of Computer Science, Sichuan University, Chengdu, China, under the supervision of Prof. B. Guo.

Her research interests include blockchains, wireless sensor networks, cognitive psychology, personal financial big data, computational finance, and financial data space.

Yuanyuan Zhang (Graduate Student Member, IEEE) received the master's degree from the College of Computer Science, Sichuan University, Chengdu, China, in 2020, where she is currently pursuing the Ph.D. degree with the College of Computer Science, under the guidance of Professor Liangyin Chen.

She is interested in blockchains, wireless sensor network, machine learning, and big data.

Shengwei Wang is currently pursuing the master's degree in electronic information with the College of Computer Science, Sichuan University, Chengdu, China.

His research interests include blockchains, Internet of Things, big data analytics, and computer networks.

Fan Wang received the master's degree from the College of Computer Science, Sichuan University, Chendu, China, in 2021.

Her research interests include blockchains, Industrial Internet, and edge computing.

Yang Li received the B.S. degree in physics and the Ph.D. degree in radio physics from Peking university, Beijing, China, in 2007 and 2012, respectively.

He is currently a Senior Engineer with the Department of Science and Technology, Communication Security Laboratory, Institute of Southwestern Communication, Chengdu, China. His research interests include blockchains, quantum cryptography, quantum secure communication, and quantum information.

Yuming Jiang received the M.S. and Ph.D. degrees in mechanical manufacture and automation from Sichuan University, Chengdu, China, in 1990 and 2001, respectively.

He is a Professor with the College of Computer Science, Sichuan University. His current research interests include blockchains and big data.

Liangyin Chen (Member, IEEE) received the Ph.D. degree from the College of Computer Science, Sichuan University, Chengdu, China, in 2008.

He joined the School of Computer Science, Sichuan University, where he has been a Professor since 2014. He works as the Director of Sichuan Big Data Analysis and Fusion Application Technology Engineering Laboratory. From 2009 to 2010, he was a Visiting Researcher with the University of Minnesota, Minneapolis, MN, USA, under the supervision of Prof. T. He has authored or coauthored more than 100 papers, many of which were published in premier network journals and conferences, such as IEEE Communications Surveys & Tutorials, IEEE Internet of Things Journal, IEEE Transactions on Mobile Computing, IEEE Transactions on Information Forensics and Security, IEEE Transactions on Industrial Informatics, *Knowledge-Based Systems*, IEEE International Conference on Computer Communications, ACM International Conference on Multimedia, IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, and ACM Conference on Embedded Networked Sensor Systems. His research interests include wireless sensor networks, Intelligent Internet of Things and IoT security, Industrial Internet, blockchains, and big data.

Bing Guo received the M.S. and Ph.D. degrees in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 1999 and 2002, respectively.

He is currently a Professor with Sichuan University. His current research interests include blockchains, green computing, and big personal data.