Research Article

An Efficient and Secure Malicious User Detection Scheme Based on Reputation Mechanism for Mobile Crowdsensing VANET

Zhihua Wang ^{1,2}, Jiahao Liu ¹, Chaoqi Guo ¹, Shuailiang Hu ¹, Yongjian Wang ³,
and Xiaolong Yang ²

¹School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China

²School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

³National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

Correspondence should be addressed to Yongjian Wang; wjy@cert.org.cn

Received 29 September 2021; Accepted 29 November 2021; Published 21 December 2021

Academic Editor: Xiaojie Wang

Copyright © 2021 Zhihua Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasing development of wireless communication technology and Vehicular Ad hoc Network (VANET), as well as the continuous popularization of various sensors, Mobile Crowdsensing (MCS) paradigm has been widely concerned in the field of transportation. As a currently popular data sensing way, it mainly relies on wireless sensing devices to complete large-scale and complex sensing tasks. However, since vehicles are highly mobile in this scenario and the sensing system is open, that is, any vehicle equipped with sensing device can join the system, the credibility of all participating vehicles cannot be guaranteed. In addition, malicious users will upload false data in the sensing system, which makes the sensing data not meet the needs of the sensing tasks and will threaten traffic safety in some serious cases. There are many solutions to the above problems, such as cryptography, incentive mechanism, and reputation mechanisms. Unfortunately, although these schemes guaranteed the credibility of users, they did not give much thought to the reliability of data. In addition, some schemes brought a lot of overhead, some used a centralized server management architecture, and some were not suitable for the scenario of VANET. Therefore, this paper firstly proposes the MCS-VANET architecture-based blockchain, which consists of participating vehicles (PVs), road side units (RSUs), cloud server (CS), and the blockchain (BC), and then designs a malicious user detection scheme composed of three phases. In the data collecting phase, to reduce the data uploading overhead, data aggregation and machine learning technologies are combined by fully considering the historical reputation value of PVs, and the proportion of data uploading is determined based on the historical data quality evaluation result of PVs. In the data quality evaluation phase, a new reputation computational model is proposed to effectively evaluate the sensing data, which contains four indicators: the reputation history of PVs, the data unbiasedness, the leadership of PVs, and the spatial force of PVs. In the reputation updating phase, to achieve the effective change of reputation values, the *logistic* model function curve is introduced and the result of the reputation updating is stored in the blockchain for security publicity. Finally, on real datasets, the feasibility and effectiveness of our proposed scheme are demonstrated through the experimental simulation and security analysis. Compared with existing schemes, the proposed scheme not only reduces the cost of data uploading but also has better performance.

1. Introduction

In recent years, with the continuous development of mobile communication technology and wireless communication technology, especially the continuous maturity and wide application of 5G technology [1], various smart terminals and wireless sensing devices have sprung up in an endless stream, such as smart phones, smart wearable devices, and

GPS. Usually, they have certain computing, storage, communication, and sensing capabilities, so their explosive popularity and increasing development have greatly promoted the emergence of Mobile Crowdsensing (MCS) [2]. As a popular sensing mode in Internet of Things (IoT) system [3], MCS mainly completes large-scale, high-cost, and complex sensing tasks through the fusion of sensing technology and wireless communication technology [4]. Specifically, it takes lots

of ordinary individuals or mobile terminal devices as basic sensing nodes, forms a MCS network through the mutual collaboration between these nodes, and then completes data collecting according to task requirements. Compared with traditional wireless sensor networks, MCS network has lower deployment cost and faster data collecting and has been widely used in environmental monitoring [5], city management [6], social services [7], and other fields. At present, with the continuous advancement of Intelligent Transportation Systems (ITS) [8] and the growing rise of Vehicular Ad hoc Network (VANET) [9], MCS-VANET has attracted the attention of academia and industry. In MCS-VANET, as the basic sensing unit, vehicles are usually equipped with lots of wireless sensing devices and sensors, such as wireless Bluetooth, speedometer, global navigation positioning device, and gyroscope. Vehicles will rely on these devices to connect with servers, RSUs, on-board units (OBUs), etc. and complete the secure and efficient data collecting and task sharing between vehicles to vehicles, vehicles to RSUs, and vehicles to cloud [10]. However, as MCS-VANET is highly dynamic and self-organized, usually vehicles with certain sensing capabilities can be used as sensing users to participate in sensing task, the system cannot guarantee the credibility of all users and the reliability of data, nor can it avoid malicious users uploading false data to disrupt the sensing network. For example, in the road traffic, the sensors in vehicles are used to detect whether the road surface is bumpy. Once lots of malicious users upload false data, it will cause vehicles to get into trouble which causes road congestion, and in severe cases, traffic accidents may occur. Therefore, in terms of ensuring the security of MCS-VANET, it is very important to consider the credibility of all users and the reliability of data.

In order to effectively solve the problem of malicious users uploading false data, many literatures have carried out a large number of studies. Hamida and Javed [11] signed to verify channel aware information by combining the ECDSA technology with k mean clustering. The scheme ensured the confidentiality of sensing data through encryption and authentication and guaranteed the user credibility to a certain extent. However, the scheme had caused complex encryption and decryption overhead; on the other hand, the scheme was not evaluated from the sensing data quality itself, and there still existed false data. In MCS environment, Jiang et al. [12] put forward the incentive mechanism based on sensing data quality, while constructing the model into the reverse auction framework to promote the sensing data quality within the system. However, this scheme involved a lot of economic theories, and its practicability is not high in MCS-VANET.

The reputation mechanism mainly evaluates the credit degree of entities in a multidimensional manner according to their behaviours to achieve the dynamic reputation updating, which is more suitable for MCS-VANET. Michiardi and Molva [13] proposed a reputation mechanism based on "Core," which integrated the direct reputation, indirect reputation, and functional reputation of nodes to calculate the global reputation of nodes. Although the mechanism realizes the detection of malicious users, its reputation evaluation

scheme lacked accuracy, because the historical reputation values of nodes are important in reputation evaluation. The higher the historical reputation value of the node, the higher the probability of uploading real data in the next round.

Blockchain is a decentralized database and a shared distributed ledger that is immutable, traceable, open, and transparent. It is a new application mode that integrates distributed data storage [14], point-to-point transmission [15], consensus mechanism [16], encryption algorithm [17], and other computer technologies. It can be widely applied in MCS-VANET to provide a fair, secure, and transparent reputation evaluation for vehicles.

Li et al. [18] proposed a blockchain-based decentralized crowdsourcing architecture, where workers verified the quality of sensing data based on the established smart contracts, in order to detect malicious users. Zhao et al. [19] combined edge computing with the blockchain and proposed a reputation management scheme. However, the above schemes lacked appropriate indicators to measure the quality of sensing data.

Therefore, in MCS-VANET, in order to effectively detect the malicious users uploading false data in the sensing system, ensure system security, reduce the overhead of data uploading, and achieve the better performance, this paper proposes an efficient and secure malicious user detection scheme with reputation mechanism-based blockchain in MCS-VANET. In general, our main contributions are summarized as follows.

- (1) In order to ensure the security of sensing system, a blockchain-based architecture is proposed in the MCS-VANET, which includes the four entities of PVs, RSUs, CS, and BC. And the blockchain is structured into three parts and integrated into the proposed three-layer architecture. Therefore, the proposed architecture can effectively avoid the single-point attack and provide a secure and fair sensing environment for the PVs and RSUs
- (2) In order to effectively realize the malicious user detection, an efficient and secure malicious user detection scheme is proposed in the MCS-VANET, which fully considers the historical reputation value of PVs and puts forward the reputation model from four aspects: the reputation history of the PVs, the data unbiasedness, the leadership of the PVs, and the spatial force of the PVs. Therefore, the proposed scheme can effectively evaluate the data quality and ensure the fairness of PVs
- (3) In order to reduce the overhead of data uploading, a data uploading method is designed by combining data aggregation technology with machine learning technology. On the one hand, this method reduces the overhead of storage and computing and, on the other hand, improves the quality of sensing data

The rest of this article is organized as follows. Section 2 presents the related work of malicious user detection, Section 3 presents the preliminaries of this paper, Section 4

TABLE 1: Comparison of research literatures.

Research categories	Advantages	Disadvantages
Cryptography [20–22]	The credibility of user is guaranteed.	The sensing data quality cannot be guaranteed; the overhead is too large.
Incentive mechanism [23–26]	The credibility of user and effective evaluation of data are guaranteed.	The user fairness cannot be guaranteed; the practicality of the scheme is poor.
Reputation mechanism [27–31]	The reputation of user is evaluated from multiple aspects.	The historical reputation is not fully considered; the user fairness and the accuracy of detection methods cannot be guaranteed.

proposes the malicious user detection scheme, Section 5 verifies the effectiveness of the proposed scheme through security and performance analysis, and finally, Section 6 is the conclusion.

2. Related Work

As is well known, malicious user detection is very important in MCS-VANET; it is often used to ensure security and reliability of sensing services. Therefore, there are lots of research literatures in this aspect, which can be classified into the following three categories based on different research perspectives, as shown in Table 1.

2.1. Cryptography. Tzeng et al. [20] proposed an identity authentication scheme based on the Random Oracle Model, which realized the detection of malicious users. However, the scheme relied on a complex certificate and secret key management system, which increased the system overhead. In order to solve the problem of high overhead of certificate management, Basudan et al. [21] proposed a certificateless aggregation signature scheme based on fog computing in the road condition monitoring. Kong et al. [22] proposed a secure query mechanism for the vehicle fog data publishing based on homomorphic encryption and batch verification technology in the architecture of fog computing in VANET, in which RSUs verified the correctness of recovered data requests. All the schemes above are based on cryptography to achieve user identity authentication. Although the credibility of participating users is guaranteed, the reliability of sensing data cannot be guaranteed and the computational overhead is large.

2.2. Incentive Mechanism. Yang et al. [23] proposed an incentive mechanism based on user behaviour, in which different incentive measures were set according to the information uploaded by users. Similarly, Peng et al. [24] constructed a MCS incentive mechanism, which evaluated the quality of sensing data based on the EM algorithm, quantified the effective contribution of users by using the principle of information entropy, and provided incentives for each participant based on its effective contribution. Wen et al. [25] proposed a quality-driven auction incentive mechanism, QDA, in which participating users received corresponding reward incentives based on the quality of the data rather than working time, while obtaining sensing data with higher reliability by constructing a probabilistic model. Park and Beak [26] proposed an incentive mechanism based

on one-way hash chain, which guaranteed the detection of malicious nodes through PKI technology and promoted the mutual collaboration of nodes, but was not evaluated from the sensing data quality itself to detect users with selfish and malicious nodes in the VANET. Most of the above schemes adopt economic theories such as auction mechanism, but they are not suitable for the highly dynamic and self-organizing characteristics of VANET. Furthermore, it is unfair that users who fail the auction cannot participate in sensing tasks.

2.3. Reputation Mechanism. Reputation mechanism has been widely used in wireless networks [32] and IoT systems to evaluate user reliability and detect malicious users. Sun et al. [27] put forward the two-layer fog architecture in MCS-VANET, integrated the zero-knowledge proof technology into it, and evaluated the vehicle reputation according to the offset value of sensing data and the historical reputation value of PVs. Nan et al. [28] proposed a CSII model based on cross-spatial multiple interaction to evaluate user reputation from four aspects: the data collecting time, the cover rate, the correlation, and historical reputation value. Song and Ma [29] proposed a method to evaluate the comprehensive reputation of users in VANET by using self-observation data and second-hand information data and taking historical reputation value into account. The above schemes only consider the user's latest historical reputation value and cannot fully reflect the user's real reputation, which seriously affects the user fairness. Due to the poor quality of sensing data caused by uncontrollable factors such as sensing intensity and damage degree, the user's reputation value becomes smaller. Yang et al. [30] proposed a decentralized reputation management system for VANET based on blockchain, in which reputation values were aggregated in RSU according to ratings generated by the message receiver. However, in this scheme, each PV needed to store and maintain the reputation value of adjacent vehicles, which brought a lot of communication overhead. Cai et al. [31] combined zero-knowledge proof and precommitted homomorphic attributes to identify the improper behaviour of users. However, this scheme did not evaluate the data quality, and the accuracy of malicious user detection was not high.

Therefore, in general, although these schemes guarantee the credibility of users, the quality of sensing data is not well evaluated. Therefore, a reputation mechanism is proposed based on blockchain, which fully considers the historical reputation value of sensing users and puts forward a comprehensive evaluation method of data quality by integrating

FIGURE 1: The system architecture of MCS-VANET.

different indicators. At the same time, machine learning is integrated with data aggregation technology, which effectively realizes the detection of malicious users and reduces the system overhead.

3. Preliminaries

In this section, some preliminaries are introduced. Among these, the system architecture of our paper is firstly proposed, which contains four entities. Then, the malicious user detection process is proposed, and finally, security requirements and design goals of this paper are given in detail.

3.1. System Architecture. In this system architecture, there are mainly four entities; the details are shown in Figure 1.

PVs: participating vehicles are usually equipped with various types of sensing devices and have certain sensing capabilities. They mainly collect sensing data in the system.

RSUs: RSUs have certain storage, computing, and sensing capabilities and play a vital role in the system, mainly for data quality evaluation and updating.

CS: cloud server mainly performs system initialization, interacts with external requesters, obtains sensing task requirements, and deploys a database based on secure media such as disks.

BC: blockchain mainly provides a reliable sensing environment for PVs and RSUs and updates and stores the reputation value of PVs. Specifically, the block is deployed on the periphery of the CS, and after completing the reputation

information collection, it is imported into the database of the CS for security publicity.

3.2. Malicious User Detection Process. The notations used in this paper are listed in Table 2.

In MCS-VANET, vehicles are defined as participants, who have certain mobility and self-organization. At this point, the sensing service has relatively high data requirements on quality, such as road condition detection service. If the PV_{ij} collects false data feedback to the server, it will lead to inaccuracy of road condition detection, and in serious cases, it will cause personal safety. Therefore, a new malicious user detection scheme is proposed, which evaluates the degree of maliciousness based on the reputation mechanism. The whole scheme consists of three phases: data collecting, data quality evaluation, and reputation updating. At the same time, the corresponding three-part structure of BC is embedded in the architecture.

Firstly, PVs of the sensing layer form a small Local Area Network (LAN) within a specific time region according to environment, region, and other factors. According to the LAN strategy, PVs publish their historical reputation values and then sign the historical reputation values and broadcast them to the LAN for the public election of the subleaders of the LAN. Once the subleaders have been successfully elected, the group is formed and then connected to a nearby RSU in the edge layer that matches the strategy. The RSU starts collecting the sensing data from PVs in the group after it was connected to the group network. Moreover, in order to

TABLE 2: Notations.

Parameters	Description
m	The number of subregions
n	The number of PVs in the subregion
PV_{ij}	The j -th participating vehicle in the i -th subregion
whole_region	The whole sensing region
sub_region $_i$	The i -th sensing subregion
n_i	The number of PVs in sub_region $_i$
sub_leader $_i$	The subleader in sub_region $_i$
uid $_{ij}$	The identity information of PV_{ij}
info $_{ij}$	The package information uploaded by PV_{ij}
pk $_{ij}$	The public key of PV_{ij}
sk $_{ij}$	The secret key of PV_{ij}
t_{ij}^k	The k -th historical reputation value of PV_{ij}
t_{ij}	The mean of t -time historical reputation values of PV_{ij}
$t_{new_{ij}}$	The latest reputation value of PV_{ij}
raw_data $_{ij}$	The raw sensing data of PV_{ij}
value_data $_{ij}$	The valuable sensing data of PV_{ij}
PER $_{ij}$	The proportion of data uploading by PV_{ij}
EVA_old $_{ij}$	The historical data quality evaluation result of PV_{ij}
EVA_new $_{ij}$	The new data quality evaluation result of PV_{ij}
history $_{ij}$	The reputation history of PV_{ij}
quality $_{ij}$	The data unbiasedness of PV_{ij}
quality_old $_{ij}$	The old data unbiasedness of PV_{ij}
data_bias $_{ij}$	The deviation between value_data $_{ij}$ and the mean value of all data of PV_{ij}
leader $_{ij}$	The leadership of PV_{ij}
leader_old $_{ij}$	The old leadership of PV_{ij}
leader_times $_{ij}$	The number of times of PV_{ij} as subleader
space $_{ij}$	The space force of PV_{ij}
space_old $_{ij}$	The old space force of PV_{ij}

reduce the data uploading overhead, the proposed scheme uses machine learning and data aggregation technology to upload different proportions of sensing data according to historical data quality evaluation result EVA_old $_{ij}$ and carries out corresponding reputation evaluation by defining the four indicators of history $_{ij}$, quality $_{ij}$, leader $_{ij}$, and space $_{ij}$. From the perspective of fairness, these four indicators are from the perspective of the past and the present, respectively, and provide sufficient sensing opportunities to PVs. Finally, the evaluation results and other information such as the vehicle's identity information uid $_{ij}$ are uploaded. Once the monitoring interface of the blockchain server listens to this information, it starts timing and at the same time

starts the block generation program of the server to process data packets. At this point, the user's latest reputation value can be obtained through the *logistic* model function curve of the reputation updating phase, that is, if the evaluation result is low, the reputation value will be reduced; if the evaluation result is large, the reputation value will be increased. The latest reputation value and other information of PVs are recorded in the block. When each block completes a phase of information collection, the next block is automatically generated and so on. The blocks are then imported into a database stored in the CS. This process makes use of the nonrepudiation and traceability of blockchain to process the information collected and uploaded by users and the information updated by reputation computing, in order to ensure the openness and transparency of the reputation evaluation process of MCS-VANET. It is noted that, in this paper, the range of reputation value is within $[0, 1]$, where 0 and 1 represent two extreme cases, that is, special malice and special honesty. And it can be expressed by the following formula.

$$t_{ij}^k, t_{new_{ij}} = \begin{cases} 0, & PV_{ij} \text{ is always malicious;} \\ 1, & PV_{ij} \text{ is always honest;} \\ 0 \sim 1, & \text{other situations;} \end{cases} \quad 1 \leq i \leq m, 1 \leq j \leq n. \quad (1)$$

Through this process, the detection of malicious users becomes obvious through the reputation value. The overall process of malicious user detection is shown in Figure 2.

3.3. Security Requirements. Based on the above system architecture and detection process in the MCS-VANET, the security requirements of the system are as follows.

- (a) *Decentralization.* As is known to all, centralized management is easy to cause single point of failure in the process of malicious user detection. Therefore, system security requirements need to be decentralized.
- (b) *Immutability and Tamper-proofing.* In the process of malicious user detection, a large amount of sensing data needs to be generated, transmitted, and processed continuously. Therefore, in order to accomplish the task accurately, it is necessary to ensure that the data is not tampered with and cannot be forged.
- (c) *Resistance to Attack.* To prevent malicious users from threatening the sensing system, the system must be able to effectively resist malicious users' attacks.

3.4. Design Goals. Based on the above system architecture and detection process in the MCS-VANET, the design goals are as follows.

- (a) *Low Overhead.* A lot of data is generated in the process of sensing task completion. Therefore, in order

FIGURE 2: The overall process of our proposed scheme.

to better complete the task and reduce the burden of the system, the overhead from two aspects of storage and time must be reduced.

- (b) *High Performances.* In the process of malicious user detection, different indicators should be used to evaluate the quality of sensing data. Therefore, it is necessary to ensure the availability and efficiency of the defined malicious user detection indicators.
- (c) *Robustness.* In MCS-VANET, the system architecture should be stable and normal, but there are malicious users. Therefore, the system should be robust when facing malicious user attacks.

4. The Proposed Malicious User Detection Scheme

In this section, the malicious user detection scheme is proposed, which mainly includes three phases, namely, data collecting phase, data quality evaluation phase, and reputation updating phase. The specific details are as follows.

4.1. Data Collecting. As the first phase of the malicious user detection scheme, data collecting is mainly involved in the process of sensing data collecting through various types of sensors assembled by vehicle. On the one hand, the data is huge from the perspective of computing or storage, and there are some redundant or irrelevant data to the sensing service; on the other hand, from the perspective of data quality, it is expected to collect high data quality and less false data. Therefore, data aggregation and machine learning techniques are introduced in the data collecting phase, and

this phase is divided into two processes: the subleader election and data uploading.

4.1.1. The Subleader Election. In this process, there are multiple RSUs in the whole sensing region $whole_region$, and the vehicle is constantly moving. Therefore, in order to facilitate data collecting and reduce unnecessary overhead, the whole sensing region is divided into several subregions, and the region covered by each RSU is a region sub_region_i , in which there are n participating users. Then, a subleader is chosen from each sub_region_i . In practice, the identity information of uid_{ij} is used to query its historical reputation value t_{ij}^k , and the vehicle with high mean historical reputation value t_{ij} is selected as the subleader sub_leader_i .

In order to realize the election of sub_leader_i , the following three aspects should be considered. First, from the perspective of trustworthiness, PVs with higher historical reputation values have higher trustworthiness in the whole subregion and will upload real data with a high probability. Second, in terms of fairness, it is not enough to query the historical reputation value of a vehicle only once, because there are some special situations. For example, in the past, a PV has uploaded real data for several times in succession, but in a certain time, the quality of the sensing data collected by the vehicle was poor due to the sensing device being damaged, stolen, or offline, so the vehicle was judged as a malicious participant, which is unfair. Therefore, more reputation value of the vehicle should be considered. Third, in terms of precision, it can effectively reflect the reputation of the participants over the past period by querying the historical reputation values of multiple times in the past then calculating t_{ij} .

The past period of time is set as the past two months [33]. On the one hand, if the statistics are less than two months, it is not enough to reflect the historical reputation of the PVs; on the other hand, if the statistics are more than two months, it will bring large computing and storage overhead. In addition, if PV_{ij} is an initial vehicle, which is newly added to the sensing system, its t_{ij} is initialized to 0.5 [34]. It can be expressed by the following formula.

$$t_{ij} = \frac{1}{t} \sum_{k=1}^t t_{ij}^k, \quad t > 0. \quad (2)$$

In a specific time interval, the PVs form a small temporary LAN. Then, in sub_region_i covered by each RSU, according to environmental and other factors, the PVs package their own t_{ij} and uid_{ij} as $info_{ij}$, sign $info_{ij}$ with their own key pair (sk_{ij}, pk_{ij}) assuming that the PVs have already obtained the key pair, and broadcast the signature information $S_{ij} = \text{SIGN}(sk_{ij}, info_{ij})$ on the temporary LAN. Then, by comparing t_{ij} of the PVs and verifying their signature information $Ver_{ij} = \text{VERIFY}(pk_{ij}, info_{ij})$, the vehicle with the largest t_{ij} is elected. When more than a certain number of PVs in the network agree that the vehicle becomes the sub-leader of the network sub_leader_i , the election ends. Then, sub_leader_i will complete the subsequent organization and supervision work and carry out the next stage of sensing data collection, aggregation and distribution. If there is a problem in the election, the group network fails to be established and this RSU cannot be accessed, then the above process can be repeated. On the one hand, this RSU will record the election process, issue a notice broadcast $RSU_Broadcast$ within the region of each coverage, and inform sub_leader_i corresponding to PVs according to each sub_region_i to facilitate the following data uploading. On the other hand, this RSU will encrypt the other information generated by calculation in the region and upload it to the CS for later public announcement. Therefore, the election process is described by the following formula.

$$\begin{cases} info_{ij} = t_{ij} \oplus uid_{ij}, \\ S_{ij} = \text{SIGN}(sk_{ij}, info_{ij}), \\ \text{User_Broadcast} : S_{ij} = \text{SIGN}(sk_{ij}, info_{ij}), \\ Ver_{ij} = \text{VERIFY}(pk_{ij}, info_{ij}), \\ \text{RSU_Broadcast} : sub_region_i \mapsto sub_leader_i. \end{cases} \quad (3)$$

4.1.2. Data Uploading. After completing the election of sub_leader_i , each PV begins to collect data. This process mainly relies on various sensing devices and measurement devices assembled by PVs. Sensing devices can make PVs to collect data information about the surrounding environment. Measurement devices are usually composed of trusted password module devices, which can initialize the devices of PVs, reorganize the collected information, calculate and output it, and upload the corresponding identification to RSU and block-

chain. Before the data is uploaded to the edge layer, in order to reduce the overhead of data transmission and enhance the sensing data quality, data aggregation and machine learning technologies are introduced.

As a branch of machine learning [35], deep learning [36] simulates the workflow of human brain through neural network models, and its architecture generally includes input layer, hidden layer, and output layer. Here, a deep learning model based on fully connected neural networks is constructed, as shown in Figure 3.

In this model, the four indicators for data quality evaluation are transformed into one-dimensional vector $\vec{fac}_{ij}$ as the input of model training, and the model returns a parameter PER_{ij} as the output, which is in $(0, 1)$. Therefore, the system will set different proportion of data uploading according to EVA_old_{ij} . Among them, EVA_old_{ij} is calculated according to the data quality evaluation function $e()$ in Section 4.2. Moreover, all indicators used in this evaluation process are past indicators, so PV_{ij} can query them based on the identity information. In general, the larger EVA_old_{ij} is, the greater the probability that the PV_{ij} will upload real data in the next round of data collecting is. Therefore, the larger the proportion of the $value_data_{ij}$ uploading is. The process is described by the following formula.

$$\vec{fac}_{ij} = \left(history_{ij}, quality_old_{ij}, leader_old_{ij}, space_old_{ij} \right), \quad (4)$$

$$EVA_old_{ij} = e\left(history_{ij}, quality_old_{ij}, leader_old_{ij}, space_old_{ij} \right), \quad (5)$$

$$value_data_{ij} = raw_data_{ij} * PER_{ij}, \quad PER_{ij} \in [0, 1]. \quad (6)$$

In this paper, the data quality is a very critical information and determines the degree of final task completion. The reliable parameter PER_{ij} can be obtained by training the deep learning model, which is very effective for obtaining high-quality sensing data. Finally, uid_{ij} is matched with $value_data_{ij}$. The sensing data uploading information formula is as follows.

$$\text{Data_Upload} : uid_{ij} \mapsto value_data_{ij}. \quad (7)$$

4.2. Data Quality Evaluation. Once the PVs of the sensing layer complete the data collecting, the RSUs of the edge layer begin the data quality evaluation. In order to evaluate the quality of sensing data from multiple aspects and accurately reflect the reputation of the PVs, the following four indicators are proposed to evaluate the quality of sensing data: $history_{ij}$, $quality_{ij}$, $leader_{ij}$, and $space_{ij}$.

4.2.1. $history_{ij}$. Firstly, the reputation history of PV_{ij} $history_{ij}$ is considered. This indicator is mainly affected by t_{ij} obtained by PVs participating in t -time sensing tasks in the past two months, where t_{ij} is $[0, 1]$. k_0 is used as an

FIGURE 3: A deep learning model based on a fully connected neural network.

influencing factor, which mainly affects $history_{ij}$. Generally, the larger t_{ij} , the greater the probability the PVs will upload real data, so the higher the quality of the sensing data and vice versa. The relationship is expressed by the following formula.

$$history_{ij} = k_0 * t_{ij}, (k_0 \neq 0). \quad (8)$$

4.2.2. $quality_{ij}$. Next, the data unbiasedness of PV_{ij} $quality_{ij}$ is proposed. This indicator evaluates the quality of the sensing data from the current perspective, and it can most intuitively evaluate the reliability of PV_{ij} from the data itself. The deviation $data_bias_{ij}$ between $value_data_{ij}$ and the mean value of all data $(1/mn) \sum_{i=1}^m \sum_{j=1}^n value_data_{ij}$ is used as the influencing factor of unbiasedness, and k_1 is also set as the influencing factor so that $quality_{ij}$ can be defined within a certain numerical range. Generally speaking, the greater $data_bias_{ij}$, the smaller $quality_{ij}$. The smaller $data_bias_{ij}$, the greater $quality_{ij}$. It can be expressed by the following formula.

$$data_bias_{ij} = value_data_{ij} - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n value_data_{ij}, \quad (9)$$

$$quality_{ij} = k_1 * data_bias_{ij}, (k_1 \neq 0). \quad (10)$$

4.2.3. $leader_{ij}$. Then, in order to reflect the fairness of PVs, the leadership of PV_{ij} $leader_{ij}$ is defined, which is mainly related to the times $leader_times_{ij}$ of PV_{ij} as sub_leader_i . In order to define its value within a certain range, the influence factor k_2 is set. Moreover, whether PV_{ij} servers as a sub_leader_i is related to t_{ij} . Generally, the more times PV_{ij} serves as a subleader, the greater its historical reputation value is, the stronger its leadership is, and the greater the probability that it will upload real data and vice versa. Therefore, it can

be expressed by the following formula.

$$leader_{ij} = k_2 * \frac{\ln(leader_times_{ij} + 1)}{e}, (k_2 \neq 0). \quad (11)$$

4.2.4. $space_{ij}$. Finally, the space force of PV_{ij} $space_{ij}$ is defined based on the degree of congestion in sub_region_i . The degree of congestion in sub_region_i will affect the efficiency of data transmission. In serious cases, if sub_region_i is too crowded, part of the sensing data will be lost. The indicator is closely related to n_i in sub_region_i . The larger n_i is, the more crowded sub_region_i will be, which will lead to congestion and noise in the transmission process, and the deviation or delayed uploading of the sensing data will eventually lead to the reduction of its reputation value and vice versa. Therefore, this relationship can be expressed by the following formula.

$$space_{ij} = \frac{n_i}{\sum_{i=1}^m n_i}. \quad (12)$$

To sum up, based on the above four factors, the evaluation function $e()$: $EVA_new_{ij} = e(history_{ij}, quality_{ij}, leader_{ij}, space_{ij})$ is defined to describe the quality of the sensing data, which is mainly evaluated by the above four indicators. The quality of the sensing data is positively correlated with $history_{ij}$ and $leader_{ij}$, while negatively correlated with $quality_{ij}$ and $space_{ij}$. Therefore the above relationship can be expressed as the following formula.

$$\begin{aligned} e(history_{ij}, quality_{ij}, leader_{ij}, space_{ij}) \\ = history_{ij} + quality_{ij} + leader_{ij} + space_{ij}, \end{aligned} \quad (13)$$

$$\left\{ \begin{array}{l} \text{history}_{ij} = k_0 * t_{ij}, (k_0 \neq 0), \\ \text{quality}_{ij} = k_1 * \text{data_bias}_{ij}, (k_1 \neq 0), \\ \text{leader}_{ij} = k_2 * \frac{\ln(\text{leader_times}_{ij} + 1)}{e}, (k_2 \neq 0), \\ \text{space}_{ij} = \frac{n_i}{\sum_{i=1}^m n_i}, \\ \text{data_bias}_{ij} = \text{value_data}_{ij} - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \text{value_data}_{ij}. \end{array} \right. \quad (14)$$

Therefore, the formula is expressed as follows.

$$\begin{aligned} \text{EVA_new}_{ij} &= k_0 * t_{ij} + k_1 \\ &\quad * \left(\text{value_data}_{ij} - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \text{value_data}_{ij} \right) \\ &\quad + k_2 * \frac{\ln(\text{leader_times}_{ij} + 1)}{e} + \frac{n_i}{\sum_{i=1}^m n_i}. \end{aligned} \quad (15)$$

4.3. Reputation Updating. After RSUs complete the above sensing data quality evaluation and upload the data information to the upper CS, the BC server is triggered to transfer information. Then, the task of comparing uid_{ij} is started, and the tasks such as reputation updating and broadcasting are automatically performed on the BC server. In this paper, the *logistic* function model [37] is introduced. The model is applicable to the MCS-VANET, especially when reputation is updated. In view of the changes of the reputation value, in the early stage, due to the lack of collection of sensing data and relatively small number of PVs, the growth of reputation value is slow. With the continuous growth of PVs, honest PVs are also increasing. After the evaluation of data quality, the quality of the data is relatively high, and the reputation value increases rapidly at this time. Then, the reputation value will gradually converge with the saturation of the PVs in the sensing region and the completion of the sensing task. Therefore, a *logistic* function model curve is used to effectively describe the changes of the reputation value of PVs, in order to better identify the malicious PVs and reduce interference to the MCS-VANET. This function curve is marked as the L curve, which is defined as follows.

$$L(x) = \frac{1}{1 + e^{-x}}. \quad (16)$$

In order to reflect the reputation updating, the new data quality evaluation result EVA_new_{ij} is taken as the input of the *logistic* function, and the output result of $L(x)$ is taken as $t_{\text{new}_{ij}}$. Once the reputation updating process is completed, RSUs will send an updating broadcast to the CS at the server layer, match uid_{ij} of PVs with $t_{\text{new}_{ij}}$, and record the updated result in the CS. The broadcast format is as follows.

$$\text{RSU_Broadcast} : \text{uid}_{ij} \mapsto t_{\text{new}_{ij}}. \quad (17)$$

5. Security and Performance Analysis

5.1. Experimental Environment. Our experiment is set on the Lenovo computer with a CPU of AMD Ryzen 7 4800H with Radeon Graphics, a basic frequency of 2.90 GHz, and a memory of 16 GB. The algorithm is implemented using the PyCharm tool and the Air Quality Online Monitoring and Analysis Platform from China (<https://http://www.aqistudy.cn/>) and Beijing Municipal Ecological and Environment Monitoring Center (<http://www.bjmemc.com.cn/>) datasets. The experiment sets the scene of urban air quality monitoring. The 2015-2018 PM2.5 statistics of Beijing in the three years are used as a simulated sensing dataset. The data range is between 0 and 500 (unit: mg/m^3). Each PV collects 10 pieces of data at a time, and its final sensing data is the mean value of the 10 pieces of data.

5.2. Security Analysis. Firstly, the MCS-VANET architecture based the blockchain meets the characteristic of decentralization. The blockchain is essentially a distributed data system. In the blockchain system with lots of nodes, each node is highly autonomous and can also be verified. Moreover, all nodes store all transaction records of the blockchain without the participation of any third party, which ensures the security of the system. Centralized systems are vulnerable to a single point of attack, and most third parties are semitrusted. In the proposed architecture, the decentralized characteristic of blockchain can well solve security problems caused by untrusted third party.

Then, all data transmitted in the proposed scheme meet the requirements of immutability and tamperproofing. In this paper, the blockchain is structured as three parts embedded into the system architecture. The PVs at the sensing layer serve as input to the blockchain for data collecting, the blockchain provides a secure computing environment for RSUs at the edge layer, and blocks are deployed outside the CS and eventually imported into the server's database. In the malicious user detection process, when the identity information, sensing data, and reputation information of the PV are calculated to trigger the smart contract of the blockchain, they are packaged together for verification and stored in the trusted chain as the unique identification of the PV. The sensing data will be uploaded to the blockchain server through the edge nodes, and the data generated in the process will be reorganized and reproduced through the special Translation Control Register (TCR) equipment in the blockchain for traceability and will be publicized in the blockchain. All data is signed by the corresponding entity, and no node can forge the message signature without the private key. These keys are also bound to the blockchain and are usually stored in tamperproof devices. If they were to be successfully attacked, it would require the malicious attacker to have more than 51% of the computing power in a blockchain system, something that cannot be done in polynomial time.

Finally, the malicious user detection algorithm of this scheme meets the effectiveness and can resist malicious

attacks. It is assumed that there are $p(p > 0)$ PVs in the whole MCS environment, and there are $q(0 < q < p)$ malicious PVs. The security analysis of the scheme algorithm is carried out by defining the malicious occupancy rate $\theta = q/p$. The specific analysis process is as follows.

When $0 < \theta < 1/2$, the number of malicious PVs in the system is less than half of all PVs. At this moment, it can be considered that there are fewer malicious PVs in the system. In this case, the honest PVs are dominant, and the maliciousness generated by malicious PVs has little impact on the system. After the election of subleader and the collection of sensing data, the data quality is relatively high, and the latest reputation value calculated through the quality evaluation phase and the *logistic* model is also correspondingly large. Even if there are a small number of malicious PVs who upload false data many times, the reputation value evaluated by the algorithm will also be reduced, and the corresponding reputation results will also be recorded and made public in the blockchain, so the detection of malicious users is obvious.

When $1/2 < \theta < 1$, the number of malicious PVs in the system is more than half of all PVs. At this moment, it can be considered that there are more the malicious PVs in the system. In this case, the malicious PVs are dominant, and there will be lots of users uploading false data. As the proposed algorithm takes into account the historical reputation value of PVs for many times, the data unbiasedness, the leadership of PVs, and the spatial force of PVs, in order to leverage the advantages of blockchain in the system, each block node cooperates with each other, all state information will be recorded, and malicious PVs will be found quickly. At this point, the reputation value evaluated by the algorithm decreases correspondingly rapidly, so it is difficult for malicious PVs to participate in the next round of sensing tasks. In terms of the immutability of blockchain, all nodes have formed a strong computing power. Even if there are many malicious PVs involved, their cost of counterfeiting is relatively high, so it is difficult to pose a threat to the system.

5.3. Performance Analysis. This paper will analyze the performance of the proposed malicious user detection scheme from the following four aspects and test it on real dataset.

5.3.1. The Proportion of Data Uploading. A large amount of data is used to train the model based on a fully connected neural network proposed in this paper, and a relationship model between EVA_old_{ij} and PER_{ij} is obtained, as shown in Figure 4. It can be seen that, although there will be certain fluctuations due to various factors of specific scene, the proportion of the sensing data uploading is approximately proportional to the EVA_old_{ij} . Namely, the larger EVA_old_{ij} is, the larger PER_{ij} is. This solves the problem high overhead of data uploading and storage when the PVs upload raw data.

5.3.2. Time Overhead. In order to reduce the overhead caused by data uploading, our scheme proposes the subleader election and data aggregation processes. In the process of election, the user with higher t_{ij} in each sensing region

FIGURE 4: The relationship between EVA_old_{ij} and PER_{ij} .

is selected as sub_leader_i . In the process of data aggregation, in order to further improve the quality of sensing data, the sensing data is collected according to t_{ij} of the PVs, and the final aggregated data will be used as valuable sensing data. Therefore, the data aggregation time is observed by setting different PVs in the following experiments.

In the experiment, a total of 100 PVs is simulated to participate in PM2.5 task collection. It can be seen from Figure 5 that the data aggregation time increases with the increase of PVs. This is because the more PVs, the more data, the aggregation process takes more time. The aggregation time of our scheme is 3.25 seconds when the number of users is 10; and when the number of PVs is 100, the aggregation time is 31.81 seconds, that is, the average aggregation time for each PV is about 0.3 seconds, which is acceptable and applicable for air quality monitoring. Moreover, by comparison of the proposed scheme with the B-PPRM scheme [34], it is found that although the data aggregation time also increases linearly with the increase of PVs, the overall aggregation time overhead of the B-PPRM scheme is larger than that of our scheme. Specifically, when the number of PVs is 10, the aggregation time is 5.54 seconds; and when the number is 100, the aggregation time reaches 40.24 seconds, that is, the average aggregation time of each PV is about 0.4 seconds. This is because the B-PPRM scheme combines homomorphic encryption technology, which requires complex encryption and decryption operations involving data, while our scheme focuses on the reputation evaluation and does not involve cryptography overhead.

Moreover, although the average time of data aggregation for each PV between B-PPRM scheme and our proposed scheme is only about 0.1 second, this gap will become larger and larger with the increasing number of PVs in the MCS system. The more users the system has, the greater the burden it carries, especially for air quality monitoring, only a large number of PVs can ensure the accuracy of the sensing data, so it can be seen that the data aggregation process with low overhead is very important. Meanwhile, by comparing Figure 4, it can be found that the time overhead of the proposed scheme is smaller because the machine learning model

FIGURE 5: Data aggregation time under different malicious user detection schemes.

is used for training before data uploading to get the proportion of data uploading instead of uploading all the raw data. In general, the proposed scheme is effective in supporting lots of users to participate in sensing tasks.

5.3.3. The Impact of Different Indicators on Reputation Updating. In the process of reputation updating, four indicators are firstly proposed, that is, $history_{ij}$, $quality_{ij}$, $leader_{ij}$, and $space_{ij}$, to evaluate the quality of sensing data. Then, based on the quality evaluation results, reputation updating is performed on *logistic* model. Therefore, in order to test the effectiveness of the proposed scheme, in the following experiments, different $quality_{ij}$, $leader_{ij}$, and $space_{ij}$ are set to analyze the impact on reputation updating.

(1) *The Impact of $quality_{ij}$ on Reputation Updating.* $quality_{ij}$ reflects the data deviation in the process of sensing data collecting. Generally speaking, the smaller the deviation, the greater the unbiasedness, the higher the data quality evaluation result, and the higher the reputation value calculated by the *logistic* model and vice versa. Therefore, in order to better reflect the impact of $quality_{ij}$ on reputation value, $history_{ij}$, $leader_{ij}$, and $space_{ij}$ are fixed, the initial reputation value of each PV is set as 0.5, and 10 vehicles with different data unbiasedness are simulated. It can be seen from Figure 6 that, when $quality_{ij} > 0$, it indicates that the PV is uploading the real data. The larger $quality_{ij}$, the smaller the deviation of the sensing data, the higher the evaluation result of the data quality, and the higher the reputation value calculated by the *logistic* model, the faster the growth of reputation value. Moreover, it can be found that the data quality in the sensing system is getting better and better as the task is completed. When the number of tasks is about 40, since the data is much unbiased at this moment, it can be consid-

ered that there is almost no false data in the whole sensing system, and the reputation value will converge to 1. On the contrary, when $quality_{ij} < 0$, it indicates that the PV is uploading the false data. The smaller $quality_{ij}$, the larger the deviation of the sensing data, the lower the data quality evaluation result, and the lower the reputation value calculated by the *logistic* model, the faster the reputation value declines. Moreover, it can be found that when the number of tasks is about 40, it indicates that our scheme has given enough opportunities for this PV to participate in, avoiding the unfair behaviour of reducing the reputation value due to the large deviation of data collecting in some special situations such as damaged or offline sensing equipment. In addition, as the task is gradually completed, the PV continues to upload false data, which makes the data deviation become larger and larger, and the data unbiasedness become smaller and smaller. At this moment, there is almost no real data in the sensing system, and our scheme is intolerable to this situation, so its reputation value will gradually converge to 0.

(2) *The Impact of $leader_{ij}$ on Reputation Updating.* $leader_{ij}$ is mainly related to the number of times of the PV acting as a subleader. The more times a PV serves as a subleader, the higher its historical reputation value and the greater its leadership, the greater the probability that the PV will upload high quality data in the next round of data collecting, so the higher its reputation value calculated by the *logistic* model. In order to better reflect the influence of this indicator on the reputation, the same $history_{ij}$ and $space_{ij}$ are set, its initial reputation value is set as 0.5, and five PVs with different leadership are simulated. Since the number of times of PVs as subleaders cannot be negative, then $leader_{ij} \geq 0$. As shown in Figure 7, for PVs with different $leader_{ij}$, as the tasks are gradually completed, there are more and more honest users and more and more high-quality sensing data

FIGURE 6: The impact of $quality_{ij}$ on reputation updating.

FIGURE 7: The impact of $leader_{ij}$ on reputation updating.

in the system, so the reputation value is rising after the evaluation of the proposed scheme. The larger $leader_{ij}$ is, the more convincing the PV is in the subregion, and the more reliable PVs are, the more probability they will upload reli-

able data, so their reputation value increases faster and faster. It can be found that with the continuous completion of tasks, when the number of tasks is about 35, there is little false data in the system. After the data quality evaluation,

FIGURE 8: The impact of $space_{ij}$ on reputation updating.

the reputation value will converge to 1. It can also be found that, compared with Figure 6, the reputation value of PVs converges to 1 when the number of tasks is about 40, which indicates that the impact of the $leader_{ij}$ of PVs is greater than that of the $quality_{ij}$. This is because with the continuous completion of the task, the number of subleaders of PV is increasing, and the change of its $leader_{ij}$ is an increasing process, while the $quality_{ij}$ is contingent to a certain extent. Meanwhile, when $leader_{ij} = 0.4$ or 0.5 , its reputation value quickly converges to 1, and the number of tasks is less than 20, which does not conform to the actual sensing situation. Therefore, $leader_{ij}$ should not be too large and is recommended to be between 0.1 and 0.3.

(3) *The Impact of $space_{ij}$ on Reputation Updating.* Similarly, in order to better reflect the impact of this indicator on reputation, the same $history_{ij}$ and $leader_{ij}$ are set, with an initial reputation value of 0.5, and five PVs with different $space_{ij}$ are simulated. $space_{ij}$ mainly reflects the degree of congestion in the sensing subregion of the PVs. For PVs with a certain degree of congestion, their spatial force $space_{ij}$ represents the degree of chaos and noise of the sensing environment. Therefore, in the process of PM2.5 data collected by PVs, there will be deviations or uploading delays in the data collected by sensing devices. However, the quality of the data uploaded is real data, so the reputation value calculated by the *logistic* model will also increase. The larger $space_{ij}$ is, the more PVs in the subregion are, the more crowded the subregion will be, and the slower the growth

rate of its reputation value will be after the data quality evaluation. As shown in Figure 8, with the increasing number of tasks, when the number of tasks is about 50, there are more and more high-quality sensing data in the system. It can be considered that there is almost no false sensing data. So the reputation value will continue to converge to 1 after evaluation. However, by comparing it with Figures 6 and 7, it can be found that when the number of tasks is about 35 or 40, its reputation value will converge to 1, which shows that $space_{ij}$ of PVs has a much smaller impact on reputation than its $leader_{ij}$ and $quality_{ij}$. Moreover, when the number of tasks is less than 5, the reputation value has a slight downward trend, which is because the number of tasks completed in the early stage is less, the sensing data collected by PVs is less, and the reliable data is relatively less. Therefore, the impact of its $space_{ij}$ is relatively large, and its reputation value will be reduced after the data quality evaluation.

5.3.4. *Robustness against Malicious PVs.* Robustness can be used to reflect the effectiveness of the proposed scheme in resisting malicious PVs by setting different degrees of maliciousness. The experiment simulates four PVs with different degrees of maliciousness; their degrees of maliciousness are 10%, 30%, 50%, and 90%, respectively. The initial reputation value of PVs is set to 0.63. When a PV shows a 10% probability of malicious behaviour, the impact on its $quality_{ij}$, $leader_{ij}$ and $space_{ij}$ is minimal. After the data quality evaluation, its data quality is still high, so the reputation value calculated by the *logistic* model will continue increasing. When a PV shows malicious degree with a probability of 30%, it

FIGURE 9: Robustness to resist different malicious PVs.

will have a great impact on its $quality_{ij}$, $leader_{ij}$, and $space_{ij}$. For example, if the deviation of data quality becomes larger, $quality_{ij}$ becomes smaller, and its leadership is weakened, its reputation value will be reduced after the data quality evaluation. Therefore, as shown in Figure 9, when the number of tasks is between 20 and 40, the reputation value will be reduced. However, as the proposed scheme is based on fairness and fully considers the historical reputation of PVs and other factors, there is still a certain degree of tolerance for PVs. Therefore, with the continuous completion of tasks, there are more and more honest PVs more and more high-quality data in the system, and the reputation value obtained by updating is also increasing. When the number of tasks is about 50, the reputation value will eventually converge to 1. When the maliciousness of a PV is 50%, it indicates that this PV alternately uploads real and false data. After data quality evaluation, its reputation value will increase and decrease. This situation is unbearable. This is because, on the one hand, as the quality of the sensing data collected by the system is getting higher and higher, there are fewer and fewer false sensing data; on the other hand, if the PVs collect false data, then it will cause a waste of computing and storage resources for the sensing device. Therefore, with the continuous completion of tasks, when the number of tasks is about 60, there will be a large number false or low-quality sensing data in the system. After the evaluation of the proposed scheme, its reputation value keeps decreasing and will eventually converge to 0. When the maliciousness of the PV is 90%, it indicates that the quality of the sensing data is extremely poor. The situation is extremely intolerable. With the continuous completion of the task, there is almost no real sensing data in the system. After the evaluation of the proposed scheme, its reputation value will decrease at a fas-

ter speed and eventually converge to 0. Therefore, through this experiment, in order to ensure the quality of sensing data, it is recommended that the probability of maliciousness of PV should not exceed 50%. Meanwhile, it can also be found that the proposed scheme is effective in resisting malicious PVs.

6. Conclusion

With the continuous development of wireless communication and sensing technology, MCS now has been widely applied due to its low maintenance cost and large amount of data collecting. The rapid development of the VANET further promotes the in-depth research of MCS. The vehicle is no longer just as a common vehicle, but as an important sensing node to complete the corresponding data collecting task. However, in MCS-VANET, there are inevitably malicious nodes who upload false data, which will seriously affect network performance and even cause immeasurable consequences.

Therefore, this paper proposes an architecture of MCS-VANET based on blockchain, which includes four entities, that is, PVs, RSUs, CS, and BC. Based on this architecture, a malicious user detection scheme is designed based on a reputation mechanism to solve the problem of malicious users uploading false data. Specifically, the proposed scheme consists of three phases. Firstly, in the data collecting phase, in order to reduce the overhead of data uploading, the sub-leader of the sensing region is elected to complete data aggregation, and the proportion of data uploading is determined according to the evaluation results of historical data quality of PVs by machine learning technology. Secondly, in the data quality evaluation phase, in order to effectively

evaluate the sensing data quality, the four indicators of history_{ij}, quality_{ij}, leader_{ij}, and space_{ij} are introduced for evaluation. Then, in the reputation updating phase, in order to effectively reflect the changes of the reputation value, the *logistic* model function curve is integrated into it, and the whole transmission and computational process is performed in the blockchain which is structured in three parts. Finally, the security analysis is carried out, and a large number of simulation experiments are carried out on the real dataset to evaluate the proposed scheme. Compared with the existing schemes, it can be seen that the proposed scheme has less overhead and higher security.

In the future work, our scheme will be further improved with the current popular technology to solve the problem of malicious user detection in different scenarios. Moreover, a lightweight incentive mechanism will be designed to further motivate sensing user to upload real sensing data, in order to reduce the threat caused by false data to the system and improve the effectiveness of malicious user detection.

Data Availability

The experiment uses the datasets of the Air Quality Online Monitoring and Analysis Platform from China (<https://http://www.aqistudy.cn/>) and Beijing Municipal Ecological and Environment Monitoring Center (<http://www.bjmecm.com.cn/>). The experiment sets the scene of urban air quality monitoring at the same time. The 2015–2018 PM2.5 statistics of Beijing data used to support the fundings of this study are included within the article. The data range is between 0 and 500 (unit: mg/m³). All data included in the proposed scheme are available.

Conflicts of Interest

The authors declare that they have no conflicts of interest in this paper.

Acknowledgments

This work was sponsored in part by the National Natural Science Foundation of China under Grant 61941113, Grant 61971033, and Grant 61671057; by the Henan Provincial Department of Science and Technology Project (No. 212102210408); and by the Henan Provincial Key Scientific Research Project (No. 22A520041).

References

- [1] Z. Ning, P. Dong, X. Wang et al., "Partial computation offloading and adaptive task scheduling for 5G-enabled vehicular networks," *IEEE Transactions on Mobile Computing*, 2020.
- [2] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.
- [3] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: challenges, solutions, and opportunities," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2419–2465, 2019.
- [4] Z. Wang, C. Guo, J. Liu et al., "Accurate and privacy-preserving task allocation for edge computing assisted mobile crowdsensing," *IEEE Transactions on Computational Social Systems*, 2021.
- [5] P. Dutta, P. M. Aoki, N. Kumar et al., "Common sense: participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems - SenSys '09*, pp. 349–350, New York, 2009.
- [6] J. Zhang, B. Guo, H. Chen, Z. Yu, J. Tian, and A. Chin, "Public sense: refined urban sensing and public facility management with crowdsourced data," in *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, Beijing, China, 2015.
- [7] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G. S. Ahn, and A. T. Campbell, "BikeNet: a mobile sensing system for cyclist experience mapping," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 1, pp. 1–39, 2010.
- [8] Z. Ning, S. Sun, X. Wang et al., "Blockchain-enabled intelligent transportation systems: a distributed crowdsensing framework," *IEEE Transactions on Mobile Computing*, 2021.
- [9] M. N. Mejri, J. Ben-Othman, and M. Hamdi, "Survey on VANET security challenges and possible cryptographic solutions," *Vehicular Communications*, vol. 1, no. 2, pp. 53–66, 2014.
- [10] A. Alnasser, H. Sun, and J. Jiang, "Cyber security challenges and solutions for V2X communications: a survey," *Computer Networks*, vol. 151, pp. 52–67, 2019.
- [11] E. B. Hamida and M. A. Javed, "Channel-aware ECDSA signature verification of basic safety messages with *k*-means clustering in VANETs," in *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, pp. 603–610, Crans-Montana, Switzerland, 2016.
- [12] L.-Y. Jiang, F. He, Y. Wang, L.-J. Sun, and H.-p. Huang, "Quality-aware incentive mechanism for mobile crowd sensing," *Journal of Sensors*, vol. 2017, Article ID 5757125, 14 pages, 2017.
- [13] P. Michiardi and R. Molva, "Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks," in *Advanced communications and multimedia security*, pp. 107–121, Boston, MA, 2002.
- [14] R. K. Raman and L. R. Varshney, "Distributed storage meets secret sharing on the blockchain," in *2018 Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, 2018.
- [15] Z. Li, A. V. Barenji, and G. Q. Huang, "Toward a blockchain cloud manufacturing system as a peer to peer distributed network platform," *Robotics and Computer-Integrated Manufacturing*, vol. 54, pp. 133–144, 2018.
- [16] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun, "A review on consensus algorithm of blockchain," in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 2567–2572, Banff, AB, Canada, 2017.
- [17] F. Gao, "Data encryption algorithm for e-commerce platform based on blockchain technology," *Discrete & Continuous Dynamical Systems-S*, vol. 12, no. 4&5, p. 1457, 2019.
- [18] M. Li, J. Weng, A. Yang et al., "Crowdbc: a blockchain-based decentralized framework for crowdsourcing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 6, pp. 1251–1266, 2018.

- [19] K. Zhao, S. Tang, B. Zhao, and Y. Wu, "Dynamic and privacy-preserving reputation management for blockchain-based mobile crowdsensing," *IEEE Access*, vol. 7, pp. 74694–74710, 2019.
- [20] S. F. Tzeng, S. J. Horng, T. Li, X. Wang, P. H. Huang, and M. K. Khan, "Enhancing security and privacy for identity-based batch verification scheme in VANETs," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3235–3248, 2015.
- [21] S. Basudan, X. Lin, and K. Sankaranarayanan, "A privacy-preserving vehicular crowdsensing-based road surface condition monitoring system using fog computing," *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 772–782, 2017.
- [22] Q. Kong, R. Lu, M. Ma, and H. Bao, "A privacy-preserving and verifiable querying scheme in vehicular fog data dissemination," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1877–1887, 2018.
- [23] G. Yang, S. He, and Z. Shi, "Leveraging crowdsourcing for efficient malicious users detection in large-scale social networks," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 330–339, 2016.
- [24] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: a quality based incentive mechanism for crowdsensing," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 177–186, New York, 2015.
- [25] Y. Wen, J. Shi, Q. Zhang et al., "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4203–4214, 2015.
- [26] J. S. Park and S. J. Beak, "Securing one-way hash chain based incentive mechanism for vehicular ad hoc networks," *Peer-to-Peer Networking and Applications*, vol. 7, no. 4, pp. 737–742, 2014.
- [27] G. Sun, S. Sun, H. Yu, and M. Guizani, "Toward incentivizing fog-based privacy-preserving mobile crowdsensing in the Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4128–4142, 2019.
- [28] W. Nan, B. Guo, S. Huangfu, Z. Yu, H. Chen, and X. Zhou, "A cross-space, multi-interaction-based dynamic incentive mechanism for mobile crowd sensing," in *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, Bali, Indonesia, 2014.
- [29] J. H. Song and C. X. Ma, "A reputation-based scheme against malicious packet dropping for mobile ad hoc networks," *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*, vol. 3, pp. 113–117, 2009.
- [30] Z. Yang, K. Yang, L. Lei, K. Zheng, and V. C. Leung, "Blockchain-based decentralized trust management in vehicular networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1495–1505, 2018.
- [31] C. Cai, Y. Zheng, Y. Du, Z. Qin, and C. Wang, "Towards private, robust, and verifiable crowdsensing systems via public blockchains," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 4, pp. 1893–1907, 2021.
- [32] X. Wang, Z. Ning, S. Guo, M. Wen, and V. Poor, "Minimizing the age-of-critical-information: an imitation learning-based scheduling approach under partial observations," *IEEE Transactions on Mobile Computing*, 2021.
- [33] C. Perger, S. Fritz, L. See et al., "A campaign to collect volunteered geographic information on land cover and human impact," *GI_Forum*, pp. 83–91, 2012.
- [34] L. Ma, X. Liu, Q. Pei, and Y. Xiang, "Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 786–799, 2018.
- [35] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [36] Z. Ning, S. Sun, X. Wang et al., "Intelligent resource allocation in mobile blockchain for privacy and security transactions: a deep reinforcement learning based approach," *SCIENCE CHINA Information Sciences*, vol. 64, no. 6, pp. 1–16, 2021.
- [37] P. J. Verhulst, "Notice sur la loi Que la Population Suit Dans Son Accroissement," *Correspondance Mathematique Et Physique*, vol. 10, pp. 113–126, 1838.