Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5031112, 18 pages
https://doi.org/10.1155/2022/5031112

*Research Article*

# An Efficient Data Sharing Scheme for Privacy Protection Based on Blockchain and Edge Intelligence in 6G-VANET

**Zhihua Wang**,[1] **Yingheng Xu**,[1] **Jiahao Liu**,[1] **ZhenYu Li**,[1] **Zeminghui Li**,[1] **Hongyong Jia**,[1] **and Dinghua Wang**[2]

[1]*School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China*
[2]*National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China*

Correspondence should be addressed to Hongyong Jia; hyjia@zzu.edu.cn and Dinghua Wang; wangdinghua666@163.com

Received 1 May 2022; Revised 13 June 2022; Accepted 23 June 2022; Published 19 July 2022

Academic Editor: Zhaolong Ning



With the substantial increase in the number of smart cars, vehicular ad hoc network (VANET), where data can be shared between vehicles to enrich existing vehicle services and improve driving safety, is gaining more and more attention, thus creating a more efficient intelligent transportation system. Moreover, the in-depth research and development of 6G and AI technology further strengthen the interconnection of various entities in VANET and can realize edge intelligence, which fundamentally enhances the efficiency of data sharing. However, reliable transmission and secure storage of data have always been a great challenge in data sharing. Although some schemes store shared data in the blockchain, most of the consensus mechanisms they use employ full nodes to verify signature information and timestamps, which cannot effectively judge the reliability of the shared data itself. Some other schemes use scoring mechanisms to evaluate data uploaded by vehicles, but these methods can be affected by network hardware failures and cannot effectively detect duplicate data. In addition, participants' privacy may also be disclosed in the process of data sharing, such as participants' location and identity information. Therefore, to address the above problems, this paper proposes a data sharing scheme in 6G-VANET, which can not only ensure the reliability and security of shared data but also protect the privacy of participants. Firstly, a consortium chain is adopted to realize the secure storage of shared data in 6G-VANET, which meets the requirements of tamper-proof and traceability of data. Secondly, a voting consensus mechanism is designed in combination with smart contract to ensure the reliability of data. Thirdly, the trained word2vec natural language processing model is deployed to edge nodes to realize edge intelligence, effectively eliminate the duplicate shared data, and enhance storage efficiency. Finally, a participant privacy protection mechanism is designed using the Private Set Intersection (PSI) protocol, and a secure and efficient data sharing scheme is finally realized. The effectiveness of the proposed scheme is demonstrated by security analysis and experimental evaluation. The experimental results show that the time and space overhead of blockchain can meet the practical requirements, and the proposed PSI protocol of large-scale vehicles can be completed in a short time.

## 1. Introduction

Currently, vehicles are equipped with various wireless communication modules and an increasing number of sensors, which enable vehicles to generate a large amount of data during driving; for example, self-driving vehicles can collect 1 GB of data per second from in-vehicle cameras, radars, GPS, and other sensors [1]. Vehicles can share road information through these sensors to nearby vehicles or to the cloud, which has promoted to the rapid development of VANET. In general, the data shared between vehicles can be divided into two types, which are subjective type data and objective type data. Objective information is mainly related to road, traffic, and environment information, such as road congestion or damage, weather conditions, and parking lot location occupancy information. Subjective information is mainly related to the rating of service quality, such as the rating of roadside hotels or car repair shops.

In the future, with the increase of infrastructure and the number of vehicles, as well as larger network coverage,

limited wireless resources and even wired connections greatly limit the development of Internet of Things (IOT) [2], and 5G network will not meet the requirements of VANET [3]. In contrast, 6G can seamlessly integrate various wireless networks spread over ground, underwater, air, and space [4], enabling complete automation. Therefore, in a oriented 6G-VANET, this big data can be collected and shared collaboratively among vehicles for higher quality of service [5].

In the process of sharing data, if the accuracy of data and the privacy of participants cannot be guaranteed, it will endanger people's life and property safety on the one hand and reduce the user's participation on the other hand. Therefore, security and efficient vehicle data sharing can improve the safety of vehicle driving process and improve road, traffic, and environment management; so, it is gradually becoming a current research hotspot.

In order to create a secure and efficient VANET data sharing environment, reliable collection and secure storage of data is always a great challenge. In a large-scale VANET, malicious vehicles spreading false information will greatly harm the traffic system and cause some losses to the vehicle users [6]. A number of scholars have studied the reliability of shared data, and Kang et al. [7] and Xie et al. [8] proposed a sharing scheme based on shared data and videos, respectively, with a reputation mechanism that can effectively prevent the proliferation of malicious data. However, network or hardware failures or viruses may cause vehicles to be rated too low or too high in a given event. Next, the use of reputation value may suffer from cooperative attacks, where some malicious vehicles intentionally give a high score to a certain vehicle, and then this vehicle can send some false messages. Again, false data may also be sent when the identity information of a vehicle with a high historical reputation value is stolen. Finally, most of the current data sharing articles do not involve data deduplication mechanisms, and uploading duplicate data can put pressure on computing storage resources. At the same time, secure storage of data is also important for data sharing. Due to heavy data traffic load, traditional centralized data storage may face many problems and challenges. Attackers may tamper with false data stored in data center nodes or evade punishment by launching various attacks (e.g., DDoS attacks and single point of failure attacks) on data center nodes to compromise data integrity, which leads to the requirement for distributed storage in VANET. In recent years, blockchain technology has received more and more attention and research in VANET because of its decentralization, anonymity, traceability, and tamper-evident features. Smart contracts in blockchain have the characteristics of being tamper-proof, distributed, and automatically triggered, which can ensure trusted transactions without the third party. Therefore, data sharing based on blockchain is a feasible solution to ensure reliable collection and secure data storage.

In order to achieve large-scale data sharing in VANET, it will be a challenge whether the communication resources as well as the computation and storage resources can meet the practical requirements. Due to the limitation of communication resources, it is difficult for vehicles to transmit massive amounts of data on a large scale in a short period of time, and the data generated by vehicles becomes increasingly fine-grained and complex, which increases the burden of data transmission [9]. Since its early days, VANET has used communication standards such as dedicated short range communication (DSRC), long-term evolution (LTE), and 802.11P. Due to the complex road environment, these technologies present challenges in terms of speed, delay, reliability, and connectivity during data propagation [10]. Although the communication resources of the Internet of things can be expanded through 5G, with the rapid growth of IOT terminals and big data services, the shortage of communication resources is still not effectively solved [11]. In order to solve this problem, a new VANET communication standard, cellular vehicular network (C-V2X), has been proposed internationally. The standard realizes assisted driving through vehicle cooperation, improves driving safety and traffic efficiency, and reduces costs. On November 18, 2020, Federal Communications Commission (FCC) decided to allocate the 5.9 GHz frequency band (5.850-5.925 GHz) originally allocated to DSRC (ieee802.11p) to WiFi and C-V2X, of which 30 MHz bandwidth (5.895-5.925 GHz) was allocated to C-V2X, which marked that the United States officially abandoned DSRC and turned to C-V2X. C-V2X will become a new international standard for VANET communication. The decoupling of data control in Software-Defined Network (SDN) enables more flexible control and optimization of the network and improves network performance [12]. The deployment of blockchain smart contracts in the SDN control plane can effectively prevent the SDN controller from being hijacked and is therefore suitable for the management of large-scale vehicular networks. In addition, Network Functions Virtualization (NFV) can decouple software and hardware and can provide network functions on standard servers. Therefore, by further combining SDN with NFV, a more efficient network system can be built for the network requirements of 6G-VANET.

For computational and storage resources, cloud computing can provide a large amount of computational resources and massive storage space, but the cloud cannot meet the latency requirements of in-vehicle terminals when they provide applications and services for transportation and entertainment [13]. To relieve the pressure on the cloud and meet the demand for computational and storage resources in VANET, MEC (mobile edge computing, MEC) becomes a natural candidate. MEC will complete computation at the local edge, which will undoubtedly greatly improve processing efficiency and reduce the load on the cloud. However, if the data uploaded by users is not filtered, MEC will also face great pressure on computing and storage. Traditional data similarity judgment methods include Levenshtein distance, Jaccard similarity coefficient, and cosine similarity, but these methods have low accuracy. In order to improve the data processing efficiency of MEC, it needs to be combined with edge intelligence. The sentence similarity methods based on machine learning include doc2vec and word2vec. The difference between them is that doc2vec is suitable for judging long multiparagraph text, while word2vec is suitable for judging short single paragraph text. Although there are a lot

of shared data in VANET, the data shared by each vehicle during driving is not large; so, word2vec has become a perfect candidate. word2vec is a natural language processing toolkit based on deep learning launched by Google in 2013. It is a model for learning semantic knowledge from a large number of text messages in an unsupervised way. It can obtain word vectors in sentences, which is simple and efficient; so, it has attracted extensive attention. word2vec uses a standard three-layer neural network model, which is initially used to predict the relationship between word vectors. word2vec can be combined with cosine similarity to further judge the similarity of sentences. Deploying the model in MEC combined with the timestamp of shared data can effectively remove duplicate data in a short time and save the overhead of computing and storage resources. However, the traditional MEC is deployed in RSU (Road Side Unit) or BS (Base Station). Because the location of these facilities is difficult to move, the coverage will be limited and cannot meet the actual needs of large-scale VANET. In order to meet these challenges, with the characteristics of simple deployment and convenient movement, UAV (Unmanned Aerial Vehicle) can be combined with MEC to provide users with ubiquitous edge computing services through computing offload [14].

Currently, there is a lack of incentives in most data sharing schemes, and vehicle users may be reluctant to participate in sharing data unless they can benefit from such participation. Therefore, users who contribute to sharing accurate data should receive some rewards. Zhang et al. [15] proposed a reward and punishment mechanism (VCoin) to encourage user participation, but the paper only mentions an idea and does not incorporate a practical application.

In the process of data sharing, identity privacy and data privacy must also be considered. For shared data, data sharing can be achieved by matching similar profile attributes for socialization purposes [16]. However, during this information matching process, vehicles cannot reveal their respective sensitive information (e.g., driver's location and identity information) to each other. If these sensitive information is disclosed, the attacker can infer the user's behavior patterns, preferences, habits, and interests through the user's sensitive information, which will threaten the user's property and security [17]. PSI protocol can perfectly solve this problem by completing the information matching of users on the one hand and protecting the privacy of each participating party on the other hand. Since the privacy of the announcement type data is not as high as that of the private data, the tedious data encryption and decryption operations can be eliminated in the process of verifying the accuracy of the data, and only the vehicle users with the same characteristics need to judge the data.

Therefore, to address the above problems, a secure and efficient data sharing solution that meets privacy requirements is proposed. Firstly, considering privacy requirements, a cloud-assisted PSI protocol is designed as a blockchain voting consensus mechanism, which allows vehicles with the same characteristics to make subjective judgments about the accuracy of data to improve the accuracy of data evaluation, thus protecting user privacy and resisting complicity attacks. Secondly, to prevent data tampering and ensure data integrity, this paper uses blockchain technology to design a data storage scheme and a smart contract for voting. The vehicle sharing data as well as the voting process will generate a transaction and pack into a block, and the identity of the vehicle can be verified through the transaction address, which replaces the traditional signature verification and further enhances the trust of the whole system. Thirdly, to facilitate blockchain maintenance and slow down the storage overhead of the blockchain, this paper uses word2vec and cosine similarity comparison to de duplicate data suspected to be duplicates in a short period of time. To mitigate the communication, this paper uses MEC, SDN, and other technologies to build a novel C-V2X communication architecture to meet the communication requirements in VANET. Finally, to address the incentive requirements, this paper gives a method of reward issuance in the constructed blockchain platform, which is done by the blockchain administrator node invoking a smart contract or directly sending a transfer transaction.

Overall, three contributions are as follows.

(1) By combining SDN, MEC, blockchain, and other technologies, a novel 6G-VANET network architecture is proposed to logically ensure the data sharing process is secure and efficient

(2) A PSI protocol combining hash and random number is designed and given to edge nodes for execution, which ensures data and vehicle privacy in the data sharing process while reducing the overhead in the matching process

(3) A blockchain system for 6G-VANET is constructed, and the reliability, integrity, and tamper-evidence of data in the data sharing process are ensured by designing smart contracts and consensus mechanisms

The rest of the paper is as follows: in Section 2, the current related research work is analyzed. In Section 3, the system architecture and problem description are given. Section 4 describes the data sharing scheme process in detail. Section 5 gives the security proof and experimental results. Section 6 concludes, and a future plan is made.

## 2. Related Work

*2.1. Data Sharing.* In traditional data sharing approaches, mostly centralized servers are used to collect and store the data shared by all entities. Ali et al. [18] proposed a cloud secure data sharing approach (SeDaSC) to ensure data confidentiality and integrity, and Dong et al. [19] proposed a framework for securely sharing sensitive data on a big data platform that enabled secure data delivery, storage, usage, and destruction. However, centralized storage approaches still faced the threat of security and privacy risks, where attackers could easily forge or tamper with data in open wireless communication environments, and may cause

single points of failure. In recent years, blockchain technology has attracted a lot of attention in the field of IoT and security, and many scholars have introduced blockchain into data sharing. Kouicem et al. [20] proposed a decentralized and anonymous data sharing scheme based on the construction of blockchain, smart contracts, and zero-knowledge proof. Feng et al. [21] applied blockchain and attribute-based cryptography to propose a secure efficient data sharing model and used smart contracts for authentication and access control. Zhang and Chen [22] proposed a data sharing framework using a federated blockchain to maintain the data storage and sharing system by applying smart contracts through preselected nodes and used digital signature techniques to ensure the integrity and security of data during data sharing. The consensus process in most articles is more similar to [16], which only verified the digital signature and timestamp, which did not guarantee the correctness of the data itself and could not judge the repeatability of the data. The scheme proposed in this paper replaces the digital signature with the transaction information, relying on the tamper-evident nature of the transaction to be verified only once, which reduced the resource overhead and enables data deduplication by the word2vec method.

Data sharing is also widely used in the field of VANET, which can improve traffic management and promote the prosperity of in-vehicle application services. Luo [23] et al. proposed an efficient collaborative data sharing method in VANET assisted by edge computing, which could effectively solve the data sharing problem between vehicles by using RSUs and vehicles as a springboard for data transmission and storage, but in the process of data transmission that could not guarantee the integrity of data. To solve the above problems, it is a good choice to introduce blockchain into data sharing in vehicular networks. Khalid et al. [24] deployed blockchain into RSUs and used IPFS to store data related to traffic events. Fan et al. [25] and Horng et al. [26] proposed a data sharing scheme in vehicular social networks with authentication and revocation mechanisms, respectively. Ma et al. [27] proposed an attribute-based encryption algorithm and maintained by RSU, which could prevent unauthorized access to plaintext data as well as malicious tampering using blockchain, and showed that it was very feasible to use blockchain to share data in the field of vehicular networking, but it still lacked the judgment of the true accuracy and repeatability of the data.

*2.2. Privacy Protection.* With the advent of the era of big data, privacy preservation has always been a focus of attention in academia and industry. There are many privacy-preserving techniques, including differential privacy, federated learning, and secure multiparty computation. Wei et al. [28] and Zhao et al. [29] combined differential privacy and federated learning to propose a privacy-preserving framework to protect users from privacy leakage, respectively, the former by varying the artificial noise variance to satisfy differential privacy under different protection levels and the latter by scrambling the entity-generated gradient to avoid privacy threats and reduce communication costs. Zhao et al. [30] provided a comprehensive review of secure multiparty computing, and the article mentioned secure multiparty computing including OT protocols, secret sharing, obfuscated circuits, zero-knowledge proofs, homomorphic encryption, and other methods that could solve the problem of collaborative computation of private data from multiple participants in a distributed computing scenario in a secure manner. However, due to the consideration of personal privacy, strangers in the network are often unwilling to cooperate, and users may not intend to expose their personal data to the server, even if their privacy is safely protected. In order to promote this cooperative relationship, PSI protocol has gradually become a research hotspot [31].

Wang et al. [32] proposed a tag-based verifiable outsourced PSI protocol (TVDPSI) for the multisubset case, where each subset was associated with a separate tag for data classification. Liu and Zhang [11] proposed a fine-grained profile matching scheme with multiple tags attached to a subset of users for classification to achieve fine-grained matching. Wang et al. [33] proposed an application scenario for feature matching in in-vehicle social networks using techniques such as OT protocols and PSI protocols. However, most of the current PSI protocols are based on both users and have excessive overhead. The proposed scheme in this paper allows vehicles with the same features to be classified through PSI protocols, satisfying multiple users to jointly participate in set matching and reducing the overhead. Specifically, when a vehicle sends shared data to the cloud, vehicles with the same characteristics are enabled to judge this data and are combined it with the blockchain to reach a consensus mechanism. Each step of the operation generates a block, which ensures the integrity and reliability of the shared data and mitigates conspiracy attacks by using the pooled matching mechanism.

*2.3. Application of Blockchain in VANET.* In recent years, with the emergence of new technologies and applications, many academic institutions and standardization organizations have conducted research on 6G-based in-vehicle communication. Guo et al. [34] analyzed the structure of future 6G in-vehicle networks and proposed some application scenarios, one of which was that by deploying blockchain at the edge nodes at the edge nodes could provide data sharing, secure authentication, and privacy protection for users. Su et al. [5] designed a secure data storage sharing scheme in a vehicular edge network using federation chain and smart contract technology. Wang et al. [6] proposed a cloud-based trust management scheme for real-time video reporting and in-vehicle messages. Ahmed et al. [35] proposed a VANET emergency messaging protocol, specifically by storing the authentication information of vehicles and the messages sent through two chains, respectively. Gao et al. [36] combined SDN and blockchain to propose a trust-based model for controlling malicious behavior in vehicular networks. Zhang et al. [37] combined blockchain with deep reinforcement learning to propose a trust model to secure the communication links between vehicles. Ortega et al. [38] proposed a content-centric network (CCN) based on blockchain and network slicing to replace the traditional client-server architecture, thus eliminating the risk of data

Table 1: Comparison of research literatures.

| Research literatures | Advantages | Disadvantages |
|---|---|---|
| Data sharing [12–21] | Ensure the security of data storage and improve storage efficiency through various technologies. | The correctness and repeatability of the data itself cannot be guaranteed. |
| Privacy protection [22–26] | Through various technologies, the privacy of each participant is protected, and the information security of users is guaranteed. | The actual implementation process is too expensive and is not suitable for large-scale VANET. |
| Application of blockchain in VANET [5, 6, 27–31] | The sensitive information in VANET is stored in the blockchain, which improves the safety of users during driving. | The consensus process of the blockchain used is inappropriate and cumbersome. |

tampering and improving the robustness of VANETs. In this paper, a high-performance VANET architecture by combining blockchain with 6G is designed, SDN and MEC technologies, where blockchain is mainly deployed into the SDN control plane, and a data chain as well as a contractual link is designed to ensure the security of the whole system, respectively.

By comparing the above literatures (Table 1), we summarize their advantages and disadvantages.

## 3. Preliminaries

This section focuses on our solution architecture and process as well as the final security requirements and design goals to be achieved.

*3.1. System Model.* This subsection gives our proposed architectural model, namely, the blockchain-based 6G-VANET secure and trusted data sharing management model.

As shown in Figure 1, the proposed architecture in this paper mainly consists of the following entities:

*Vehicles*: vehicles, as the main component of the SDN data plane, are equipped with a large number of sensors as well as storage and wireless communication modules, and the communication between vehicles is carried out through wireless networks. Vehicles can generate road condition related announcement messages while driving, and these announcement messages can be passed to higher levels for judgment with the help of RSU, which can be shared to all vehicles in the area. The communication technology mentioned in the architecture diagram is C-V2X, which is a cellular network-based wireless communication technology for vehicles

*UAVs*: UAVs can act as edge nodes to provide edge computing services to remote areas

*RSUs*: RSUs are equipped with processing, storage, and wireless communication modules, and the RSU acts as a bridge between the SDN controller and the vehicle

*Base stations*: the base stations are deployed to multiple areas based on geographic location and are equipped with MEC functions as the main component of the SDN control plane. Each base station is also equipped with a data chain and a contract chain. The function of the data chain is to store the data shared by vehicles, and the function of the contract chain is to add vehicle information and vehicle voting. The function of the MEC is to perform the corresponding set matching and block out calculation operations and provide storage resources to store the information on the blockchain

*Certification authority* (CA): the CA is assumed to be a fully trusted authority that distributes key information for each vehicle as a way to generate the vehicle user's address in the blockchain. If any vehicle user sends a false message during data transmission, the CA can find the real identity of the malicious user and add him/her to the blacklist

*Cloud*: after the edge nodes store the shared messages from vehicles, they are also sent to the cloud. Some remote areas are not covered by the SDN control area and cannot receive the messages in time; so, the final data sharing also needs to broadcast the messages to the rest of the vehicles through the cloud. Thus, vehicles can be more aware of traffic conditions and help improve the security and effectiveness of system transmission

This architecture is mainly built based on the three SDN planes, at the bottom layer is the SDN data plane, which mainly consists of vehicles and RSUs. In this layer, the vehicle is responsible for collecting data during the driving process, and the RSU acts as a bridge between the SDN controller and the vehicle. The middle layer is the SDN control plane, which mainly consists of 6G base stations and UAVs with MEC functions, and the UAV can provide data collection, edge computing, and other services for vehicle users to expand the wireless network cache [39]. This plane deploys smart contracts and blockchain and can also deploy different contracts according to different situations, reflecting the programmability of the SDN control plane. Since the smart contracts cannot be changed after they are deployed, the SDN control plane can be prevented from being hijacked. To relieve the computational pressure of a single SDN controller, this paper proposes a distributed SDN controller system, which also reflects the decentralized feature of blockchain and can effectively prevent single-point attacks. The top layer is the SDN application plane, which mainly consists of CA and the cloud. The CA can issue keys to vehicles for authentication, and the cloud can provide various services to vehicles. Our scheme uses NFV to coordinate the devices in each SDN area and break the limitation of incompatible hardware and software. Also, in order to optimize network resource allocation, network slicing techniques can be used to assign priorities to different shared data requirements, giving priority to services with high network requirements.

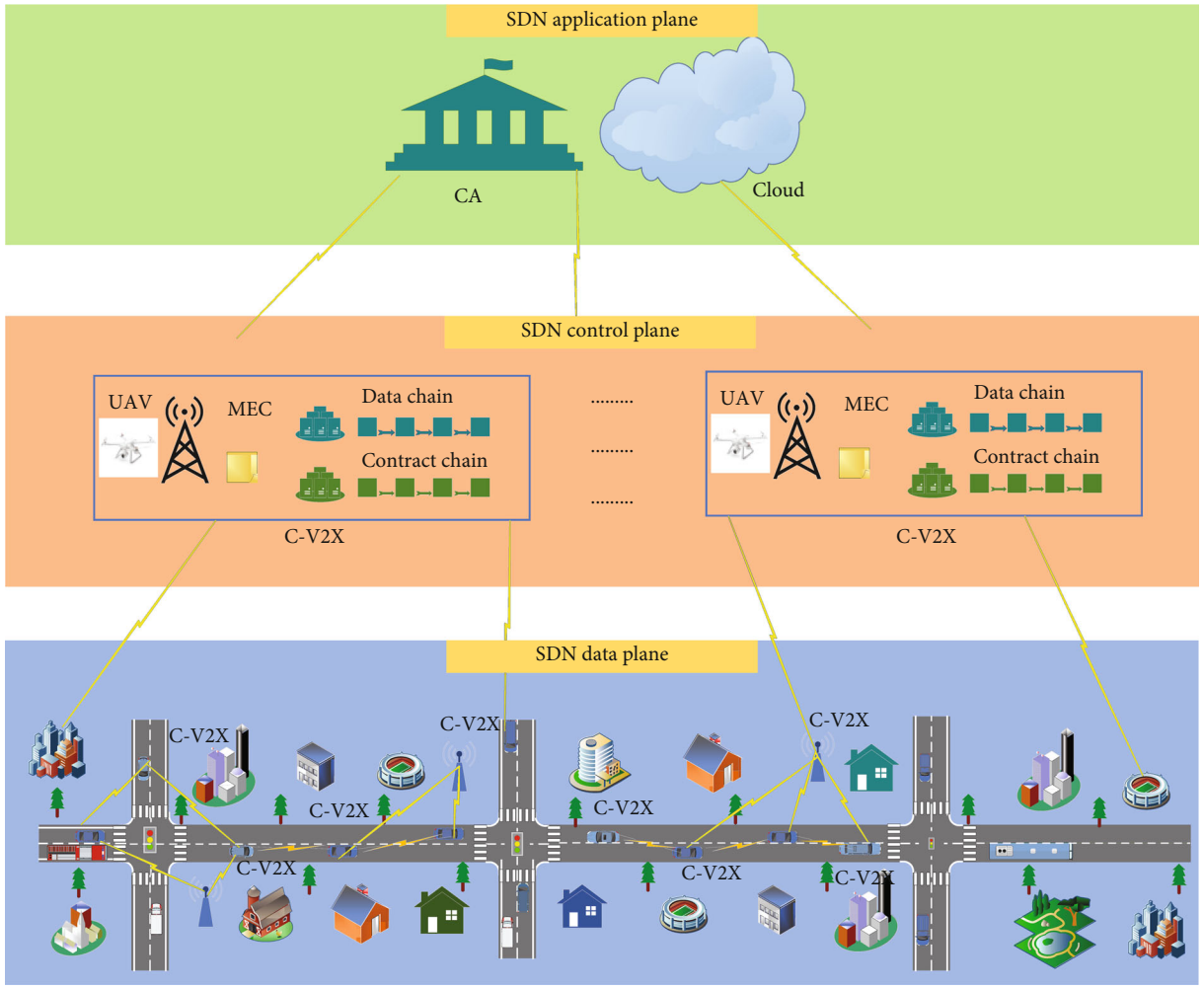


Figure 1: System architecture.

Table 2: Summary of notations.

| Symbols | Description |
|---|---|
| $V_i$ | Legitimate vehicle in the VANET |
| $PK_{vi}$, $SK_{vi}$ | Vehicle's public key and private key |
| CA | Certificate authority for managing vehicle enrollment |
| $V_p$ | Vehicle with providing data |
| $V_q$ | Vehicles with voting rights |
| $M_p$ | Data submitted by vehicle |
| $T_x$ | Transactions arising from vehicles |
| $B_s$ | Base station with mobile edge computing |

*3.2. Problem Description.* In our data sharing scenario, the main entities are the legal vehicle $V_i$, the trusted authority CA for issuing keys, the data provider $V_p$, the set of voters $V_q$ with the same set properties as $V_i$, and the base station $B_s$ with MEC function, where $V_p$ and $V_q$ belong to $V_i$. It is assumed that the CA is fully trusted, the MEC can correctly match the aggregated information of the vehicle, and the CA can trace the real identity information of the user. Simply put, after $V_p$ shares data, MEC selects $V_q$ through PSI protocol, then $V_q$ votes and judges the data, and finally stores the data in the blockchain for other users to access. The notations used in this paper are shown in Table 2, and the detailed process will be described in Chapter 5.

*3.3. Secure Data Sharing Process with Privacy Protections and Incentives.* The data sharing task process is shown in Figure 2.

(1) Vehicle registration: vehicles join the network for the first time to be issued a pair of public and private keys by the CA in the SDN application plane as the vehicle's identity

(2) Vehicle data sending: vehicles located in the SDN data plane can collect traffic-related data and send them to the SDN control plane in the form of transactions. After receiving the data shared by vehicles, the SDN control plane performs data deduplication according to the time stamp

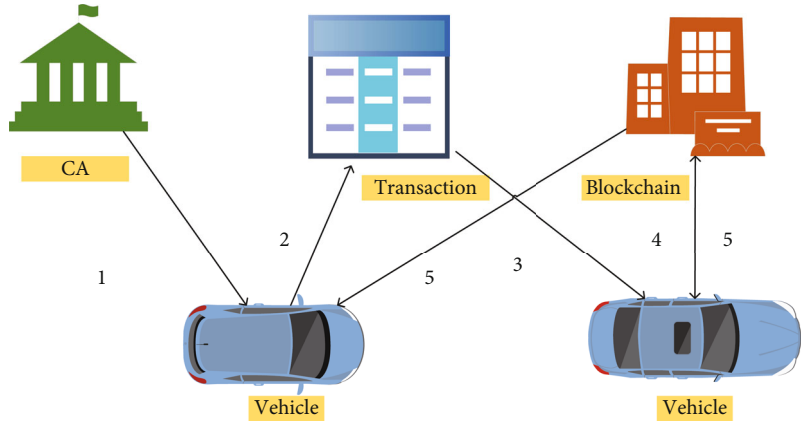


Figure 2: Task process.

(3) Voting consensus: the SDN control plane performs set matching based on the vehicle information in the current area, vehicles with the same characteristics can obtain voting rights and judge the accuracy of the data, and saves the voting records in the contract chain

(4) Packing the transaction into a block: after passing the PBFT consensus protocol, the transaction can be saved into the data chain, and then the cloud and edge nodes can broadcast the data out

(5) Issuing rewards: certain rewards should be given to those vehicles that share useful data and those that participate in voting normally

*3.4. Security Requirements and Design Goals.* The data sharing scheme proposed in this paper mainly includes the following security requirements and design goals.

*3.4.1. Security Requirements*

*(1) Protecting the Privacy of Each Participant.* When a user sends information to an edge node, an attacker may steal the user's information through eavesdropping attacks. Secure multiparty computing protects the sensitive information of vehicle users (e.g., driver's location and identity information) from disclosure, in which individual participants cannot access each other's information. In addition, to process these information, even if the attacker steals some information, he cannot get any useful content from it.

*(2) Preventing Malicious Behavior of Edge Nodes.* As an intermediary for set matching, edge nodes may obtain user information from them. Therefore, the malicious behavior of the edge node must be prevented, and the correctness of the calculation of the edge node can be verified, so as to prevent this type of man in the middle attack.

*(3) Ensuring the Security of Data Storage.* In the process of uploading and storing data, an attacker may attack the data storage system and tamper with the data. Therefore, the data shared by vehicle users must be traceable and tamper proof. Since the data needs to be voted on by other vehicle users in the process of storage, the voting behavior of participating users need to be constrained.

*3.4.2. Design Goals*

*(1) Reducing the Overhead Generated in the Process of Data Storage.* A suitable blockchain storage architecture should be designed a blockchain consensus block-out mechanism so that the communication overhead and storage overhead are sufficient to meet the actual requirement.

*(2) Designing a PSI Protocol That Can Quickly Match the Set of Large-Scale Vehicles.* The traditional PSI protocol is only applicable to two parties, but there are many vehicle users in VANET. In order to make the proposed scheme work smoothly, it is essential to design a lightweight multiparty PSI protocol.

*(3) Reducing System Low Cost.* The cost overhead of the whole data sharing system is simulated according to the current blockchain system, so that the system can be reasonably applied to the reality.

## 4. Data Sharing Scheme

In order to make the data sharing scheme run smoothly and ensure the security of data storage, we deploy the blockchain into MEC. Blockchain is a distributed database that backs up all transactions to each node. It is decentralized, tamper proof, irreversible, and traceable. It consists of a series of block links. Each block contains the current block number, the hash value of the current block, the hash value of the parent block, the hash value of the transaction, the timestamp, and other information. We have designed two chains based on Ethereum to realize the reliability of scheme storage. One is the data chain, which stores the hash value of the sent data transaction. Through this hash value, we can query the details of the vehicle shared data. The other is the contract chain, which stores the transaction information of contract deployment and contract call, from which we can query the specific situation of vehicle voting. A PBFT-based


Figure 3: Data transaction information.

consensus mechanism is designed by combining the PSI protocol and smart contracts. The PSI protocol is used to select voters, and smart contracts are used to implement the voting process. The specific process includes system initialization, transaction generation, voting right vehicle selection, transaction authentication, data feedback, and incentive mechanism.

*4.1. System Initialization.* A new vehicle $V_i$ joins the system for the first time to obtain a key pair $PK_{vi}$ and $SK_{vi}$ to prove its legitimate identity, which can be distributed by the CA located in the SDN application plane. Among them, in order to protect the security of the key, the public and private keys can be set as 64-bit hexadecimal strings by the secp256k1 elliptic curve with reference to the Ethernet ECDSA algorithm. In order to ensure the privacy of the user while not exposing the public key, the public key should be subjected to Keccak-256 hashing operation to take the last 40 bits to generate the account address of each vehicle $V_i$. In the process of data sharing, it is necessary to ensure that malicious vehicles can be identified and CA can reveal the identity of malicious vehicles, impose some penalties, and add them to the blacklist. But before that, a sufficient number of vehicles $V_i$ need to authenticate the message $M_p$ to determine if it is a malicious message.

Next is the deployment of smart contracts, which are deployed to the edge nodes $B_s$ of each distributed SDN control plane, which are also blockchain administrator nodes. The role of the smart contract is mainly to add the information of the vehicle sending the message for other vehicles with voting rights to vote on the message and return the final number of votes to judge the accuracy of the message. When a vehicle $V_p$ sends a message, the blockchain administrator node first loads $V_p$'s address into the smart contract, followed by the rest of the vehicles $V_q$ for voting judgment.

To prevent malicious users from continuously uploading data and thus wasting resources, the upload frequency of each user is also limited. For some similar announcement messages, the MEC can filter the uploaded duplicate data based on the timestamp.

*4.2. Transaction Generation*

*4.2.1. Transactions Sent.* If a vehicle $V_p$ wants to share data $M_p$, $M_p$ is first packed into a transaction $T_x$, and the details of $T_x$ are shown in Figure 3.

*hash* indicates the current hash value of $T_x$; if $T_x$ has been packed into blocks, *blockHash* and *blockNumber* will show the corresponding block information. *from* indicates who initiated the $T_x$, and since the transaction sender needs to unlock the account before sending the transaction, the identity of the sender can be proved here. *to* is the recipient of the $T_x$, which here can refer to the administrator node $B_s$ in the SDN control plane that is responsible for broadcasting the message. *input* is the data information contained in this $T_x$, referring to the data information shared by the vehicle $V_p$, or the data information generated by the invocation of the smart contract. For a data transaction $T_x$ the details of *input* are the hexadecimal strings of the data sent, and after the vehicle $V_p$ sends the data $V_q$, it then authenticates the content of the message.

*4.2.2. Data Deduplication.* When a user uploads data, similar recent data may be uploaded. In order to save computational storage resources, users should compare the data with recent data by timestamps before uploading, and similar data should be discarded. Commonly used sentence similarity measures are vector space cosine similarity, which uses the cosine of the angle between two vectors in vector space as a measure of the size of the difference between two individuals. The closer the cosine value is to 1, the closer the angle is to 0 degrees, that is, the more similar the two vectors are, as follows.

Suppose there are two similar data sets: $A$, there was a car accident here and $B$, there was a traffic accident here. Take the intersection of which $C$: (there was a car traffic accident here). After that, $A$ and $B$ are compared with $C$ and expressed in the form of a vector, with the same character as 1 and different characters as 0.At this time, $A = [1, 1, 1, 1, 0, 1, 1]$, and $B = [1, 1, 1, 0, 1, 1, 1]$. After that, the cosine formula is used to calculate the cosine values of $A$ and $B$. The closer the cosine value is to 1, the more similar the two vectors are, and the two sets of data are more similar.

$$\cos\theta = \frac{\sum_{i=1}^{n}(X_i \times Y_i)}{\sqrt{\sum_{i=1}^{n}(X_i)2} \times \sqrt{\sum_{i=1}^{n}(Y_i)2}}. \quad (1)$$

The similarity between the two data sets $A$ and $B$ can be calculated by equation (1) as 0.83. However, the above method is a statistical-based method, and the statistical-based method cannot satisfy the semantic similarity matching. The following method is a method-based deep learning, which solves the semantic similarity matching to a certain extent.

The implementation process of judging sentence similarity through word2vec is as follows. First, the sentences are divided into words, the word2vec word vectors are trained using the Gensim library to obtain the word vectors corresponding to each word, then all the word vectors are summed and averaged to obtain the sentence vectors, and finally, the cosine values of the two sentence vectors are calculated. The similarity between $A$ and $B$ is 0.91 calculated by the Google open source word2vec model; so, the judgment

by word2vec is obviously much better. The neural network model of word2vec is shown in Figure 4.

Among them, the input layer is a single hot code of multidimensional vector. For example, the input layer corresponding to incident in $A$ is $A' = [0, 0, 0, 0, 1, 0]$, but in the actual training process, the dimension of the input layer will be very large. The hidden layer is a multidimensional vector weight matrix; that is, each word can be expressed as a multidimensional vector, and the output layer is the probability of similar single correspondence.

In this paper, the word2vec model is put into each edge node of the SDN control plane. Whenever a vehicle is to send data, the edge node compares this data with all the recently uploaded data and defines a threshold of size 0.8 to make reasonable trade-offs.

*4.3. Voting Vehicle Selection.* Once $V_p$ has authenticated his identity, he can collect the information of $V_i$ to start matching, and the $V_i$ with voting rights eventually forms $V_q$. This process is a PSI protocol process with many vehicle users in VANET; so, a cloud-assisted PSI protocol is to be proposed.

The PSI protocol is a privacy-preserving protocol, assuming that Alice and Bob have their respective sets of attributes $X$ and $Y$. Through this protocol, Alice and Bob can compute the intersection of $X \cap Y$. At the same time, this process does not reveal any information that is not in $X \cap Y$. The specific implementation of this protocol is left to the computationally powerful MEC. The specific implementation of this protocol is left to the computationally powerful MEC to perform the matching. Alice and Bob first divide their respective sets into t parts based on different attributes, where $t$ parts are not necessarily the same, as shown in (2) and (3).

$$X_A = \chi_1 \| \chi_2 \| \cdots \| \chi_{t1}, \tag{2}$$

$$X_B = \chi_1 \| \chi_2 \| \cdots \| \chi_{t2}. \tag{3}$$

After that, MEC performs the set intersection operation between set $A$ and set $B$ as $C$, where $\chi_1 \chi_2 \chi_{t3} \in X_A / X_B$

$$X_C = \chi_1 \| \chi_2 \| \cdots \| \chi_{t3}. \tag{4}$$

A number of scholars have also studied cloud-assisted schemes for this protocol, Li et al. [40] proposed a PSI protocol based on homomorphic encryption, and Kerschbaum [41] proposed a one-way function PSI protocol, but the use of homomorphic encryption in vehicular networking brought pressure on computation and storage and hash functions alone may receive brute force cracking. The set is matched to the edge nodes for processing, but the edge nodes may be dishonest; so, to protect the user's privacy, the set attributes are to be fuzzed, to reduce the overhead of the whole system, in this paper, by using encryption, random number, and hash value matching for this protocol. Before each set matching, the same encryption once for each attribute in the set is performed, after which a random number is added to the ciphertext and finally, the corresponding hash value is calculated and given to the edge node for matching. Since the voters are only obtained based on the matching results, the user does not need to decrypt the set attributes, and the user only needs to know whether he or she has the right to vote, thus further protecting the privacy of each participant. The specific form of ensemble matching is as follows.

The original set $X$ is as follows:

$$X = \chi_1 \| \chi_2 \| \cdots \| \chi_{t1}, \tag{5}$$

where $\chi_1, \chi_2 \cdots \chi_t$ is the value of the attribute in the set $X$. After performing an encryption operation, the set is as follows:

$$X' = \mathrm{Enc}(\chi_1) \| \mathrm{Enc}(\chi_2) \| \cdots \| \mathrm{Enc}(\chi_t). \tag{6}$$

The set the set is as follows after adding the random number $\lambda$

$$X'' = \mathrm{Enc}(\chi_1)\lambda \| \mathrm{Enc}(\chi_2)\lambda \| \cdots \| \mathrm{Enc}(\chi_t)\lambda. \tag{7}$$

Then, a hash value for each attribute is computing in the set

$$X'h = \mathrm{Hash}(\mathrm{Enc}(\chi_1)\lambda \| \mathrm{Hash}(\mathrm{Enc}(\chi_2)\lambda \cdots \| \mathrm{Hash}(\mathrm{Enc}(\chi_t)\lambda). \tag{8}$$

The random hash value of each set attribute can be obtained from (8), which can prevent violent inference attacks on set elements by untrustworthy edge nodes.

A set $V_q$ of vehicle users that intersects with $V_p$ is returned after set matching, and the same set elements as $V_p$ are returned for each vehicle user in $V_q$. The set matching process is shown in Algorithm 1,assuming that the number of vehicles is $m$ and the number of set elements of each vehicle is $n$, and the time complexity of the algorithm is $O(m^* n)$.

To avoid malicious edge nodes returning wrong set information, the final set intersection result should be verifiable. Zheng and Xu [42] and Qian et al. [43] proposed a digital signature-based set verification method, since it is not considered that the explicit text of the set intersection, but only the computational correctness here, using the OT protocol for verification. Suppose Alice and Bob first divide the intersection returned by MEC into $t$ parts according to different attributes and arrange them in some order, when t1 = t2, then the sum is the hash value of the set attributes.

$$\begin{aligned} X_A &= \omega_1 \| \omega_2 \| \cdots \| \omega_{t1}, \\ X_B &= \varphi_1 \| \varphi_2 \| \cdots \| \varphi_{t2}. \end{aligned} \tag{9}$$

Alice then sends Bob a set $S$ of arbitrary attribute number strings ($S$ is the combination of any of the values 1, 2, 3,....., $p$ any combination of values, here assume $S = 1$, 3, 5). Subsequently, Bob calculates the heterogeneous value

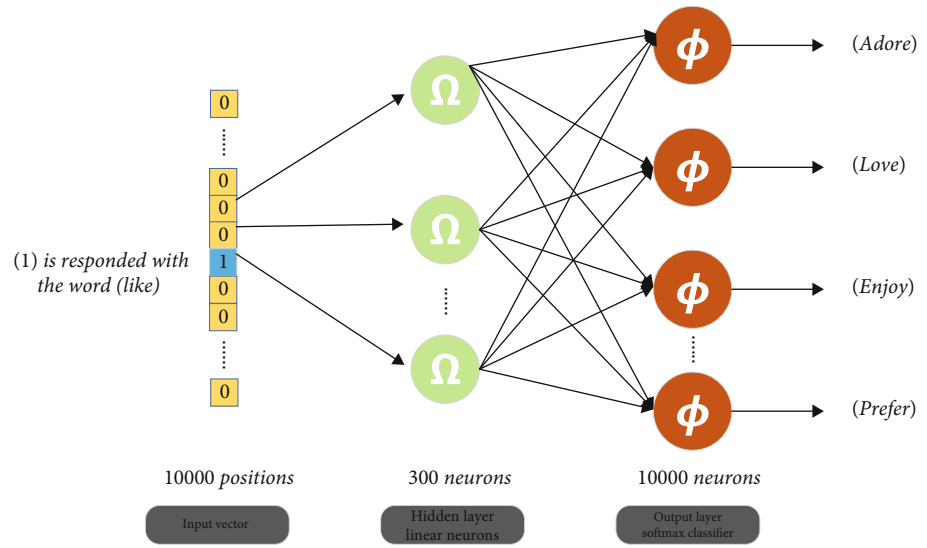


Figure 4: word2vec neural network model.

Input:
The collection content of $\mathbf{V_p}$;
Set B of the remaining vehicles in $\mathbf{V_i}$;
Output:
$\mathbf{V_q}$ with voting rights;
1: for each $i \in$ Bdo
2: if $\mathbf{V_p} \cap i! = \emptyset$ then
3: Add vehicle $i$ to set $\boldsymbol{V_q}$;
4: else
5: Vehicle $i$ has no voting rights;
6: end if
7: end for
8: if $\boldsymbol{V_q} == \emptyset$ then
9: Check the historical interactive information of $\mathbf{V_p}$ to decide whether the transaction is valid;
10: else
11: return $\boldsymbol{V_q}$;
12: end if

Algorithm 1: Calculating intersection to get the vehicles with voting rights.

of the corresponding attribute number $S$ in his own intersection as $B$.

$$B = \varphi_1 \bigoplus \varphi_3 \bigoplus \varphi_5. \tag{10}$$

Alice also has to perform the corresponding calculation noted as $A$.

$$A = \omega_1 \bigoplus \omega_3 \bigoplus \omega_5. \tag{11}$$

From (10) and (11), the set of attributes numbered as $S$ with different or values can be derived, if $A = B$, and then $X_A = X_B$.

In this, the SDN control plane collects all $V_i$ messages and $V_p$ in the current area for ensemble matching, vehicles with the same ensemble attributes are considered to the same characteristics, and these vehicles have the power to initially vote on the messages. To further mitigate the conspiracy attack, the vehicles with the final voting power are selected randomly by 50% from $V_q$. Since the voting process should follow the PBFT protocol, which is a protocol in which the minority obeys the majority, it is necessary to ensure that more than 2/3 of the nodes vote before subsequent operations. In some special cases it may encounter that the set matching is an empty set or the number of vehicles of $V_q$ is less than 3. In this case, it is referred that the past interaction information of $V_p$ to determine whether

the message can be shared, and the interaction information refers to whether $V_p$ has a record of multiple malicious votes. The interaction information records are determined as follows.

$$R = \frac{N}{M}, \tag{12}$$

where $R$ denotes the confidence level of the current message, $N$ denotes the record of correct $V_p$ history voting, and $M$ denotes the total number of $V_p$ voting. If ≥0.8, and the message of the current vehicle is considered correct.

If $V_p$ joins the vehicle network for the first time and fails to match the set of vehicles, it is initially recognized that the message sent by $V_p$ is correct. After each vote is completed, the edge node will record the voting information, and the vehicles that vote incorrectly for many times should be added to the blacklist.

*4.4. Transaction Certification.* After the vehicle sends a message for the first time, the transactions generated are still pending and are not packed into blocks and need to go through the consensus mechanism and the mining operation performed by miners before the transactions can be packed into blocks. The introduction of blockchain will bring challenges to the delay overhead of the system [44]. Cryptocurrencies like Bitcoin put some restrictions on the generation of each block in order to ensure the security of the blockchain and prevent attackers from having a lot of arithmetic power to attack the various nodes of the blockchain. Most cryptocurrencies do this by finding a random number and then calculating a special form of hash value by a hashing algorithm, such as the first $X$ bits are all zeros. However, generating a block by this form will waste a lot of computational resources, and some articles use an incentive mechanism to make cars become miners to mine but still consume a lot of time for computation in practical applications. In order to generate a block in a short time, the scheme proposed in this paper does not impose restrictions on the form of generating blocks, but a lot of verification of PBFT consensus mechanism is needed during the process of block generation; so, the security of the block chain can also be protected to some extent.

Once the set is matched, the content of the message can be voted on, and the voting process is represented as a smart contract. Before that, the information of $V_p$ has to be loaded into the smart contract, and this process is a smart contract invocation. In order to rationalize the use of resources, the voting process is restricted in the smart contract so that each vehicle can only cast one vote for a given vehicle address at a time. Vehicle voting modifies the vote value defined in the smart contract, and this process generates a transaction packaged into a block as a record of each vehicle's voting process, further enhancing the trust of the system and alleviating the phenomenon of false voting. The smart contract voting algorithm is shown in Algorithm 2; assuming a total of $n$ users vote, the time complexity of the algorithm is $O(n)$.

Input:
Transaction sent by $\mathbf{V_p}$;
$\mathbf{V_q}$ with voting rights;
Number of $\mathbf{V_q}$ as Q;
Output:
The number of votes P;
Whether the transaction is valid or not;
1: for each $i \in \boldsymbol{V_q}$ do
2: ifVehicle $i$ verified the transactionthen
3: P+=1;
4: end if
5: end for
6: if $P \geq \mathbf{2/3} *$ Q then
7: Package transactions into blocks;
8: else
9: Ignore this vote;
10: end if

Algorithm 2: Voting process.

$V_q$ judges the $M_p$ in $T_x$ and adds one to the number of votes if the message is correct and finally counts the total number of votes received. After that, the PBFT protocol is followed, and the transaction is considered correct when the number of votes exceeds 2/3 of the total.

The miners mainly consist of BS with MEC function in each distributed SDN controller, which reflects the decentralized feature of blockchain. Secondly, MEC can calculate the hash value of a block in a short period of time with its powerful computing power, so that transactions can be packed into blocks and data can be shared out faster. Trusted data is propagated and replicated into the blockchain, which is maintained by decentralized edge nodes, making it accessible from anywhere. The associated computation is offloaded to the edge nodes, thus making the fully use of storage and computational resources. Once a transaction is successfully packed into a block, the blockchain administrator node can broadcast the messages in the transaction.

The transaction validation process is shown in Figure 5. When $V_p$ sends a data transaction $T_x$, the blockchain administrator broadcasts the transaction to the rest of the vehicles in the SDN control domain for validation and finally obtains the validation result. When a data transaction $T_x$ is successfully packaged into blocks, the whole blockchain system will synchronize according to the longest chain principle. This principle not only compares the length of the blockchain but also the content of the previous blockchain. If an attacker wants to change the direction of the entire blockchain, he needs to control more than 50% of the nodes in the system, which is extremely difficult to achieve in practice.

*4.5. Disinformation Feedback and Incentives.* Some malicious vehicles may form a group and vote for malicious false messages by pooling and matching, which can mislead the rest of the vehicles and endanger traffic safety, and messages such as road information may be outdated, such as road damage,

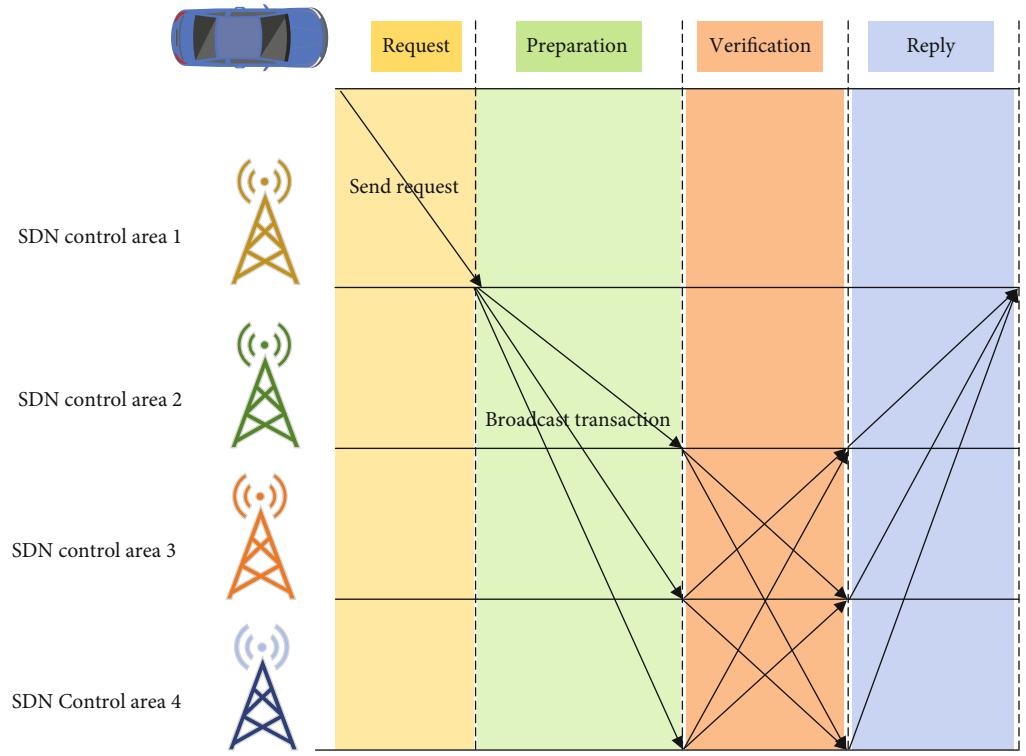


Figure 5: Transaction verification process.

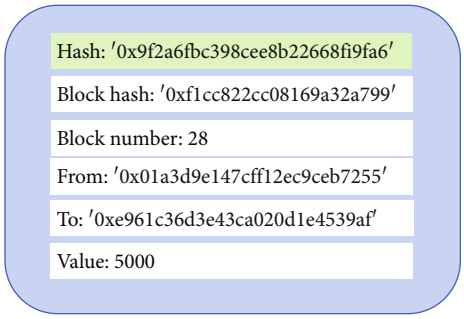


Figure 6: Currency transaction information.

traffic jams caused by traffic accidents. The above two types of messages can be feedback by the rest of the vehicles to initiate a vote again and mark a response to the corresponding message after reaching a certain number of votes.

Since the process of sharing data consumes certain resources, incentives are essential in order to encourage user participation. However, designing effective incentives to encourage secure data sharing among multiple users remains a daunting task. In this paper, an incentive mechanism in the form of cryptocurrency is designed by the Ether platform, where vehicles that share correct messages and participate in voting can receive certain rewards. Blockchain transactions can send not only data but also cryptocurrency, the incentive mechanism is issued by blockchain administrator nodes, and the specific amount can be referred to the real-time Ether price. The transaction information is shown in Figure 6, where *value* indicates the amount of this transfer, and the rest of the content is similar to the data transaction information. The reward issuance process is shown in Algorithm 3; assuming that a total of $N$ users participate in the data sharing, the time complexity of the algorithm is $O(n)$.

## 5. Security Analysis and Experimental Evaluation

*5.1. Security Analysis.* The proposed scheme in this paper uses PSI protocol to let vehicles with the same characteristics vote on messages and randomly selects a certain number of vehicles from the voting vehicles as the final vehicles with voting rights, which mitigates the conspiracy attack to a certain extent. In order to protect the privacy of each participant, the edge nodes perform set matching and then inform each vehicle separately whether it has the voting right or not, the vehicles cannot get any information from each other, and the vehicles with the voting right will get the transaction information of the shared data for the next judgment.

In addition, the scheme can prevent malicious users from performing set snapping to obtain other users' set information, because vehicle users can only perform set matching while sharing data and cannot match all the time. In order to avoid malicious edge nodes stealing users' set information or returning wrong set information, we add random numbers to the set and encrypt them, which can effectively prevent malicious edge nodes from brute force cracking. Even if the attacker steals some information, he cannot get any useful content from it and can verify the

Input:
Transaction sent by $\mathbf{V_p}$;
$\mathbf{V_q}$ with voting rights;
Output:
Whether the participating vehicles can be rewarded;
1: for each $i \in \boldsymbol{V_q}$ do
2: if Vehicle $i$ correctlyverified the transactionthen
3: The administrator sends a transfer transaction to Vehicle $i$;
4: end if
5: end for
6: if $\mathbf{V_p}$ shared the correct messagethen
7: The administrator sends a transfer transaction to $\mathbf{V_p}$;
10: end if

Algorithm 3: Reward process.

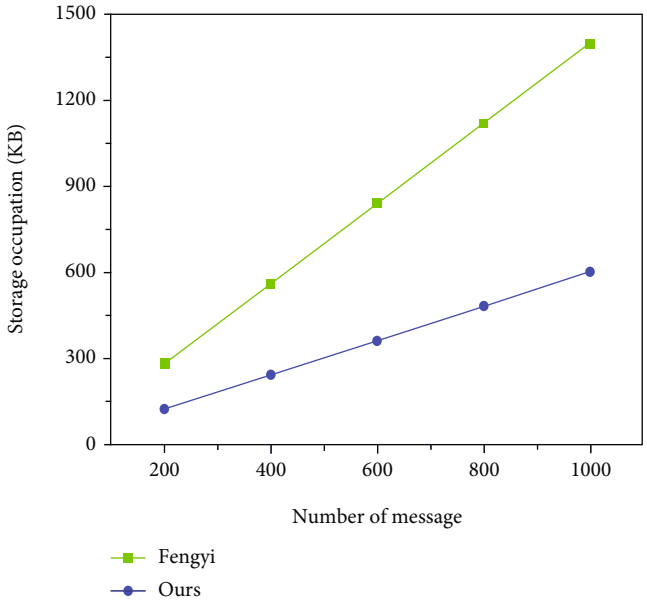


Figure 7: Storage overhead of shared data.

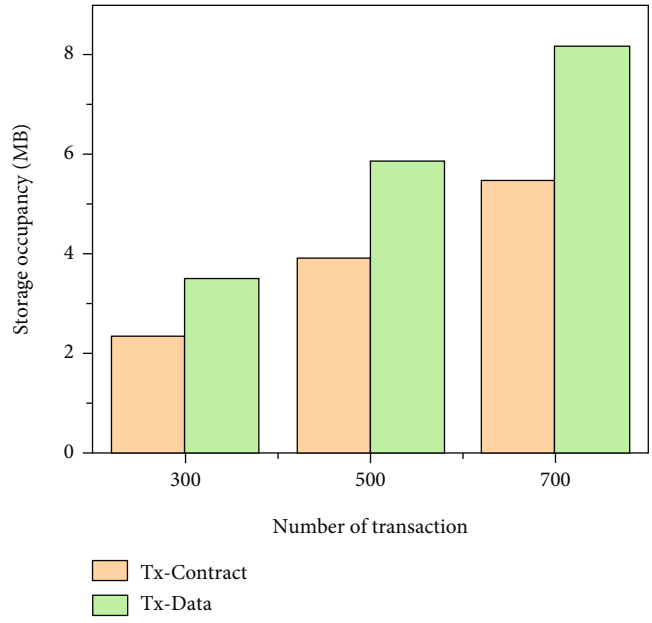


Figure 8: Storage space occupied by transactions.

results returned by the edge node with OT protocol ; so, it can effectively prevent the attacker from eavesdropping attacks and man in the middle attacks of edge nodes.

The data shared by the vehicle is loaded into a transaction, which is then packed into blocks, relying on the tamper-evident nature of the transaction and the blocks to effectively prevent the data from being maliciously tampered with, thereby protecting the storage security of the data. Each transaction contains the sending and receiving addresses of the data, which can meet the traceability of the data. In addition, the user's voting process will also generate a transaction and pack into a block, which can record the user's voting process and thus constrain the user's behavior.

### 5.2. Experimental Evaluation

*5.2.1. Experimental Settings.* The computer configuration for the experiment is Windows10 system AMD R7 4800H processor with 16G RAM. The simulation platform of Ethernet is Ganache2.13.2 version, and the execution script is Node 14.15.4 and Web3 0.20.1, the smart contract solidity language version 0.4.26. The experimental environment of PSI protocol is python pycryptodome Cryptographic library and hashlib module of sha256 hash algorithm.

*5.2.2. Storage Overhead.* Due to the introduction of blockchain in VANET, it inevitably leads to more storage usage. Zeng et al. [45] proposed a blockchain-based data sharing scheme Fengyi, but no smart contract was used; so, by comparing in terms of shared data storage, as shown in Figure 7, the memory space occupied by Fengyi for one data sharing is about 1.4 kb. As shown in Figure 7, Fengyi takes about 1.4 kb of memory space for one data sharing, while the sharing of announcement messages is targeted, eliminating part of the encryption and decryption operations, and theoretically, the storage overhead of one data

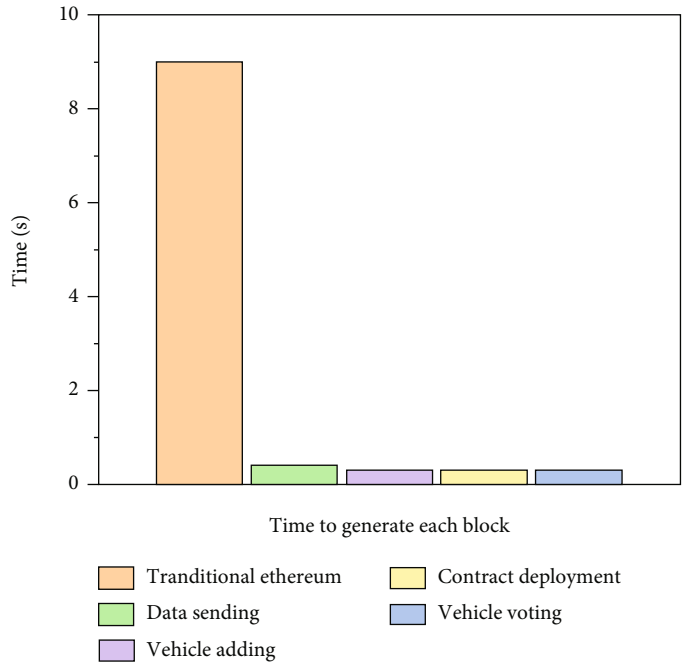


Figure 9: Block-out time comparison.

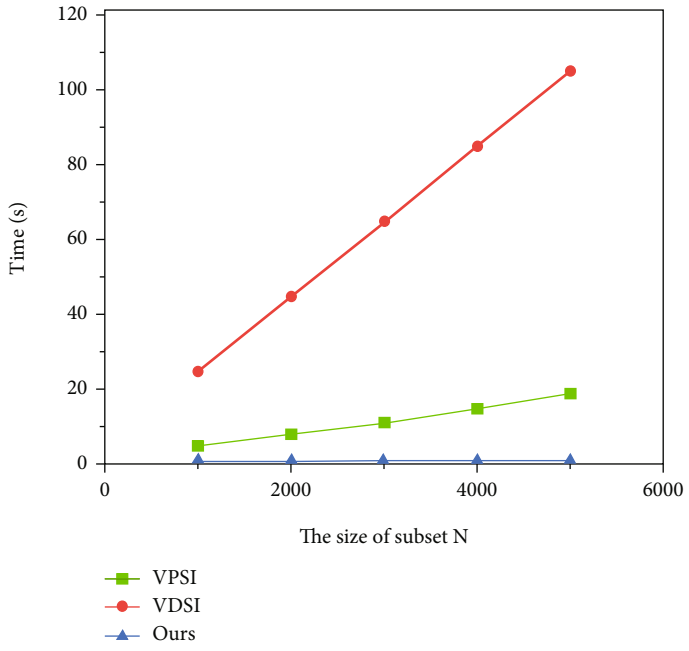


Figure 10: Time to outsource set attributes to edge nodes.

Table 3: Cost consumption.

| Process execution | Gas | Eth | $ |
|---|---|---|---|
| Contract deployment | 926414 | 0.00853 | 15.4 |
| Vehicle adding | 138482 | 0.00097 | 1.7 |
| Vehicle voting | 1146919 | 0.00114 | 2.1 |
| Data sending | 118428 | 0.00057 | 1.0 |

sharing is about 0.6 kb, which is a great improvement in the data storage performance.

To further evaluate the storage overhead of the scheme proposed in this paper, the real storage overhead with Ganache-CLI is simulated, which includes the storage of the two smart contracts of adding voting vehicles and vehicle voting, as well as the transactions generated by sending data and packing the information into blocks, and filtered the block and transaction information in the output results. As

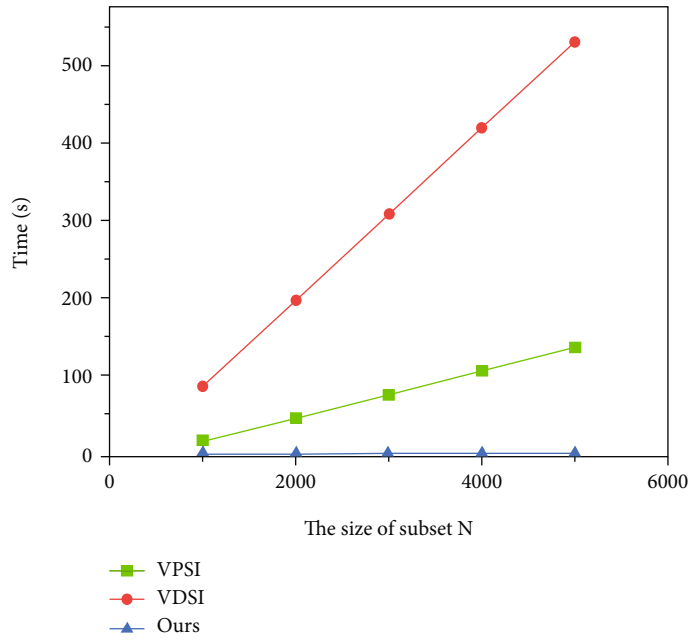


Figure 11: Set matching time.

seen in Figure 8, the overall storage occupancy increases linearly and is positively correlated with the number of transactions. Sending data transactions (Tx-data) in Figure 8 refers to transactions sent by vehicle shared data, and this type of transaction is directly initiated by Web3 objects, which require specifying the transaction initiator and transaction receiver as well as the transaction information. The storage overhead of this part is mainly related to the size of the data in the transaction information, and the data storage field in the experiment is set to empty, as in the storage simulation in Fengyi. The contract invocation transactions in Figure 8 refer to the contract transactions (Tx-contract) generated by adding voting vehicles and vehicle voting, which are generated differently from sending data transactions and are generated by invoking smart contracts. Since adding vehicle information and vehicle voting are essentially two methods in the same smart contract, the block storage overhead is actually the result of invoking the same smart contract, and eventually, the two transactions are tested to occupy the same amount of memory.

For the storage overhead of the vehicles themselves, our scheme is similar to the VANET data sharing scheme mentioned by Shen et al. [46], where the storage overhead of the vehicles is constant and is only the respective key pair is stored, and the only difference being that our key pair is the address of the blockchain account and a random AES key that is encrypted for each set of attributes.

*5.2.3. Communication Overhead.* Figure 9 shows the time taken to pack various transactions into blocks generated by traditional Ethereum, vehicle send data, add vehicle information, and vehicle voting. Unlike traditional blockchains, the block-out difficulty is not set. This is mainly because our approach does not consume as much arithmetic power as traditional blockchains, such as using the POW consensus mechanism. PBFT is used to let some nodes implement the consensus process instead of all nodes, thus greatly increasing the throughput of the whole network. As seen in Figure 9, starting from the deployment of smart contract requires about 1255 ms to complete message, which is sent to the packing into operation, but in the practical application of smart contract that needs to be deployed only once, there will be other factors (such as data transmission rate and bandwidth) and time overhead of the vehicle to subjective judgment message. So, the final time will be greater than that.

The time of outsourcing the set attributes to edge nodes is simulated by using the python pycryptodome cryptographic library and sha256 in the hashlib module. Compared with VDSI mentioned by Guo et al. [34] and VPSI mentioned by Ahmed et al. [35], digital signature is not used to verify the set matching result, but use OT protocol for verification; so, the overhead of signature generation is reduced at this stage. Since it is not needed to calculate the decryption of the set matching result, and the set matching result is only for selecting voters, a more lightweight symmetric encryption algorithm AES is used, which generates a random key for each set matching and encrypts all set attributes with this key to ensure that the result is the same after encrypting the hash of each set matching, which is convenient for the edge nodes to calculate, thus replacing the VPSI. This replaces the asymmetric encryption algorithm ElGamal used in VPSI. From Figures 10 and 11, it can be seen that the processing of the set by asymmetric encryption and digital signature methods and the matching time of the set by the cloud is in seconds, and this is only the result of two-party set operation, which does not meet the communication requirements of large-scale 6G-VANET. Our proposed method as shown in Figures 10 and 11 is millisecond in time and supports multiparty set operations,

where our matching scheme is to match the encrypted hash value directly, unlike the first two which compute the cipher text, and to facilitate the display of the set matching time, 200 vehicle messages are simulated at a time, and the number of set elements is 1000-5000. The set matching time is simulated. It can be seen that MEC can select vehicle voters in a short time, which is applicable to the dynamic changes of large-scale VANET and greatly improves the efficiency of VANET. In order to enable voters to judge the accuracy of the message more realistically, a threshold value is also set for screening the number of identical elements in the intersection of matching results according to the actual situation of pooled matching.

*5.2.4. Cost.* The cost overhead of the system is calculated based on the units of gas consumed by transactions and smart contracts. Gas is a unit of measurement used to measure the computational power of executing transactions in the Ethereum platform. For cost analysis, several parameters are considered: (1) gas consumed for contract deployment, (2) gas consumed for adding voting vehicles, (3) gas consumed for vehicle voting, and (4) gas consumed for sending data transactions. All the above four parameters are the total amount of gas consumed from transaction generation to packing into blocks in a single time. Both gasPrice and gasLimit in ganache-CLI are default values.

Table 3 shows the cost of each implementation process. The dollar consumption in the last column refers to the Ethereum price on May 30, 2022, where $1\,\text{eth} \approx 1800$ dollars.

## 6. Conclusion

In this paper, we design a new data sharing scheme in 6G-VANET. Specifically, we design a new VANET architecture combining 6G, SDN, edge computing, and other technologies and then realize the effective deduplication of shared data by word2vec. Finally, two different forms of blockchain are designed to store various information needed in the process of sharing data. Among them, the vehicle layer acts as the SDN data plane to collect all kinds of shared data. MEC deployed in the SDN control plane is mainly used to calculate vehicle set matching and block producing operation of blockchain. In addition, we deployed word2vec and two blockchains into MEC, taking full advantage of MEC's computational storage resources and embodiment edge intelligence. In order to reflect the decentralized characteristics of blockchain and relieve the pressure on the SDN control plane, this paper designs the SDN control plane as distributed according to different regions, and the certification center and other service coproviders are deployed in the SDN application plane.

For judging the accuracy of the data, most current data sharing schemes are by calculating the reputation value of the data itself or the data provider, but this may be subject to some problems such as collusion attacks, network, or hardware failures. There are also some articles that use all-node voting; however, for certain announcement types of shared messages, those vehicles in different characteristics may not be able to accurately determine the authenticity of the messages. This paper designs a lightweight PSI protocol for vehicle-specific matching, in which vehicles with the same characteristics obtain voting rights. The voting process is carried out in the form of smart contracts, and each vehicle leaves a block transaction record when voting is completed, which constrains users to participate normally, and false announcement messages can be stopped from the source.

In our future work, the performance of the solution by other simulation platforms will be evaluated. In addition, more data security protection features will be considered, such as state access, and extend more application services on the blockchain platform. Finally, we will consider combining the data sharing mechanism and reputation value in this paper to create a more secure and efficient data sharing system.

## Data Availability

The proposed scheme and its analysis need only theoretical and experimental support. There is no additional data set to be provided in this paper.

## 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 Henan Provincial Department of Science and Technology Project (No. 212102210408) and by the Key Scientific Research Project of Henan Province (No. 22A550036).

## References

[1] W. Xu, H. Zhou, N. Cheng et al., "Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2018.

[2] X. Liu and X. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5379–5388, 2019.

[3] M. S. Omar, S. A. Hassan, H. Pervaiz et al., "Multiobjective optimization in 5G hybrid networks," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1588–1597, 2017.

[4] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[5] Z. Su, Y. Hui, and Q. Yang, "The next generation vehicular networks: a content-centric framework," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 60–66, 2017.

[6] Z. Wang, J. Liu, C. Guo, S. Hu, Y. Wang, and X. Yang, "An efficient and secure malicious user detection scheme based on reputation mechanism for mobile crowdsensing VANET," *Wireless Communications and Mobile Computing, vol.*, vol. 2021, article 5302257, pp. 1–16, 2021.

[7] J. Kang, R. Yu, X. Huang et al., "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4660–4670, 2018.

[8] L. Xie, Y. Ding, H. Yang, and X. Wang, "Blockchain-based secure and trustworthy internet of things in SDN-enabled 5G-VANETs," *IEEE Access*, vol. 7, pp. 56656–56666, 2019.

[9] J. He, Y. Ni, L. Cai, J. Pan, and C. Chen, "Optimal dropbox deployment algorithm for data dissemination in vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 632–645, 2018.

[10] Z. Wang, H. Wang, Y. Wang, and X. Yang, "CLASRM: a lightweight and secure certificateless aggregate signature scheme with revocation mechanism for 5G-enabled vehicular networks," *Wireless Communications and Mobile Computing, vol.*, vol. 2022, article 3646960, pp. 1–20, 2022.

[11] X. Liu and X. Zhang, "Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 5971–5980, 2018.

[12] S. Garg, K. Kaur, S. H. Ahmed, A. Bradai, G. Kaddoum, and M. Atiquzzaman, "MobQoS: mobility-aware and QoS-driven SDN framework for autonomous vehicles," *IEEE Wireless Communications*, vol. 26, no. 4, pp. 12–20, 2019.

[13] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—a key technology towards 5G," *ETSI White Paper*, vol. 11, no. 11, pp. 1–16, 2015.

[14] Z. Ning, P. Dong, M. Wen et al., "5G-enabled UAV-to-community offloading: joint trajectory design and task scheduling," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 11, pp. 3306–3320, 2021.

[15] L. Zhang, M. Luo, J. Li et al., "Blockchain based secure data sharing system for internet of vehicles: a position paper," *Vehicular Communications*, vol. 16, pp. 85–93, 2019.

[16] Y. Qian, X. Xia, and J. Shen, "A profile matching scheme based on private set intersection for cyber-physical-social systems," in *2021 IEEE Conference on Dependable and Secure Computing (DSC)*, pp. 1–5, Aizuwakamatsu, Fukushima, Japan, 2021.

[17] 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*, vol. 9, no. 1, pp. 120–133, 2022.

[18] M. Ali, R. Dhamotharan, E. Khan et al., "SeDaSC: secure data sharing in clouds," *IEEE Systems Journal*, vol. 11, no. 2, pp. 395–404, 2017.

[19] X. Dong, R. Li, H. He, W. Zhou, Z. Xue, and H. Wu, "Secure sensitive data sharing on a big data platform," *Tsinghua Science and Technology*, vol. 20, no. 1, pp. 72–80, 2015.

[20] D.-E. Kouicem, A. Bouabdallah, and H. Lakhlef, "An efficient and anonymous blockchain-based data sharing scheme for vehicular networks," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6, Rennes, France, 2020.

[21] C. Feng, K. Yu, A. K. Bashir et al., "Efficient and secure data sharing for 5G flying drones: a blockchain-enabled approach," *IEEE Network*, vol. 35, no. 1, pp. 130–137, 2021.

[22] X. Zhang and X. Chen, "Data security sharing and storage based on a consortium blockchain in a vehicular ad-hoc network," *IEEE Access*, vol. 7, pp. 58241–58254, 2019.

[23] G. Luo, H. Zhou, N. Cheng et al., "Software-defined cooperative data sharing in edge computing assisted 5g-vanet," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 1212–1229, 2019.

[24] A. Khalid, M. S. Iftikhar, A. Almogren, R. Khalid, M. K. Afzal, and N. Javaid, "A blockchain based incentive provisioning scheme for traffic event validation and information storage in VANETs," *Information Processing & Management*, vol. 58, no. 2, article 102464, 2021.

[25] K. Fan, Q. Pan, K. Zhang et al., "A secure and verifiable data sharing scheme based on blockchain in vehicular social networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5826–5835, 2020.

[26] S.-J. Horng, C.-C. Lu, and W. Zhou, "An identity-based and revocable data-sharing scheme in VANETs," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15933–15946, 2020.

[27] J. Ma, T. Li, J. Cui, Z. Ying, and J. Cheng, "Attribute-based secure announcement sharing among vehicles using blockchain," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10873–10883, 2021.

[28] K. Wei, J. Li, M. Ding et al., "Federated learning with differential privacy: algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[29] Y. Zhao, J. Zhao, M. Yang et al., "Local differential privacy-based federated learning for internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8836–8853, 2020.

[30] C. Zhao, S. Zhao, M. Zhao et al., "Secure multi-party computation: theory, practice and applications," *Information Sciences*, vol. 476, pp. 357–372, 2019.

[31] 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*, p. 1, 2021.

[32] Q. Wang, F. Zhou, J. Xu, and S. Peng, "Tag-based verifiable delegated set intersection over outsourced private datasets," *IEEE Transactions on Cloud Computing*, vol. 10, no. 2, pp. 1201–1214, 2020.

[33] X. Wang, X. Kuang, J. Li, J. Li, X. Chen, and Z. Liu, "Oblivious transfer for privacy-preserving in VANET's feature matching," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4359–4366, 2020.

[34] H. Guo, X. Zhou, J. Liu, and Y. Zhang, "Vehicular intelligence in 6G: networking, communications, and computing," *Vehicular Communications*, vol. 33, p. 100399, 2022.

[35] M. Ahmed, N. Moustafa, A. F. M. S. Akhter et al., "A blockchain-based emergency message transmission protocol for cooperative VANET," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[36] J. Gao, K. O.-B. O. Agyekum, E. B. Sifah et al., "A blockchain-SDN-enabled internet of vehicles environment for fog computing and 5G networks," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4278–4291, 2019.

[37] D. Zhang, F. R. Yu, and R. Yang, "Blockchain-based multi-access edge computing for future vehicular networks: a deep compressed neural network approach," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2021.

[38] V. Ortega, F. Bouchmal, and J. F. Monserrat, "Trusted 5G vehicular networks: blockchains and content-centric networking," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 121–127, 2018.

[39] X. Wang, Z. Ning, S. Guo, M. Wen, L. Guo, and V. Poor, "Dynamic UAV deployment for differentiated services: a multi-agent imitation learning based approach," *IEEE Transactions on Mobile Computing*, p. 1, 2021.

[40] S. Li, S. Zhou, Y. Guo, J. Dou, and D. Wang, "Secure set computing in cloud environment," *Journal of Software*, vol. 27, no. 6, pp. 1549–1565, 2016.

[41] F. Kerschbaum, "Collusion-resistant outsourcing of private set intersection," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pp. 1451–1456, Trento, Italy, 2012.

[42] Q. Zheng and S. Xu, "Verifiable delegated set intersection operations on outsourced encrypted data," in *2015 IEEE International Conference on Cloud Engineering*, pp. 175–184, Tempe, AZ, USA, 2015.

[43] Y. Qian, J. Shen, P. Vijayakumar, and P. K. Sharma, "Profile matching for IoMT: a verifiable private set intersection scheme," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3794–3803, 2021.

[44] Z. Ning, S. Sun, X. Wang et al., "Blockchain-enabled intelligent transportation systems: a distributed crowdsensing framework," *IEEE Transactions on Mobile Computing*, p. 1, 2021.

[45] C. Zeng, Y. Wang, F. Liang, and X. Peng, "Fengyi: trusted data sharing in VANETs with blockchain," in *2020 IEEE 25th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pp. 11–20, Perth, WA, Australia, 2020.

[46] J. Shen, T. Zhou, J. Lai, P. Li, and S. Moh, "Secure and efficient data sharing in dynamic vehicular networks," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8208–8217, 2020.