A Privacy-Preserving Vehicular Data Sharing Framework atop Multi-Sharding Blockchain

Jingwei Wang*, Junqin Huang*, Linghe Kong*, Guihai Chen*, Dianle Zhou†, Joel J. P. C. Rodrigues‡§
*Shanghai Jiao Tong University, Shanghai, China
†National University of Defense Technology, Changsha, China
‡Federal University of Piauí (UFPI), Teresina, PI, Brazil
§Instituto de Telecomunicações, Portugal
Email: {wang159357, junqin.huang, linghe.kong, chen-gh}@sjtu.edu.cn; zhoudianle@nudt.edu.cn; joeljr@ieee.org

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685366

***Abstract*—Internet of Vehicles (IoV) has become an indispensable technology to bridge vehicles, persons and infrastructures, and is promising to make our cities smarter and more connected. It enables vehicles to exchange vehicular data (*e.g.*, GPS, sensors, and brakes) with different entities nearby. However, sharing these vehicular data over the air raises concerns about identity privacy leakage. Besides, the centralized architecture adopted in existing IoV systems is fragile to single point of failure and malicious attacks. With the emergence of blockchain technology, it has the chance to solve these problems due to its features of tamper-proof, traceability and decentralization. In this paper, we propose a privacy-preserving vehicular data sharing framework based on blockchain. In particular, we design an anonymous and auditable data sharing scheme using Zero-Knowledge Proof (ZKP) technology so as to protect the identity privacy of vehicles while preserving the vehicular data auditability for Trusted Authorities (TAs). In response to high mobility of vehicles, we design an efficient multi-sharding protocol to decrease blockchain communication costs without compromising the blockchain security. We implement a prototype of our framework and conduct extensive experiments and simulations on it. Evaluation and analysis results indicate that our framework can not only strengthen system security and data privacy, but also increase the data authenticity verification efficiency by 5x comparing to existing privacy-preserving schemes.**

***Index Terms*—Internet of Vehicles, blockchain, multi-sharding, scalable, privacy-preserving, zero-knowledge proof.**

I. Introduction

Internet of Vehicles (IoVs) can provide real-time communication among different entities, *e.g.*, vehicles, RoadSide Units (RSUs), pedestrians handheld devices, and aggregate vehicular data from them for safer and smarter transportation management. Due to the superiority of IoV, there are many promising explorations for IoV applications in academia [1], such as autonomous driving, vehicle management, High-Definition (HD) map, big data awareness [2], [3]. Obviously, IoV applications are driven by massive vehicular data, so that securing data privacy, authenticity and integrity during sharing is a non-negligible part in IoV systems.

However, there are some vulnerabilities in existing IoV systems [4], which will break down the safety of the vehicular data sharing paradigm: *1) System and data security*. Consider that most IoV systems are built on the centralized architecture, *i.e.*, the Client-Server (CS) model, which may suffer from single point of failure and malicious attacks [5], such as Distributed Denial of Service (DDoS) attacks, Sybil attacks, thereby disabling the functionalities of the whole IoV systems. Furthermore, by tampering vehicular data stored in the centralized database, vehicles and RSUs can be manipulated by attackers, which could cause traffic chaos. *2) Identity privacy*. Sharing vehicular data over the air can be eavesdropped and tracked by attackers, who could obtain the identity of vehicles by analyzing vehicular data patterns [6], such as driving track data. The risk of identity privacy disclosure could wear down people's enthusiasm for sharing vehicular data, which hinders the deployment of IoV systems in real world.

The emergence of blockchain technology has gained considerable attentions in recent years. Due to its beneficial characteristics, *e.g.*, decentralization, trusted execution, and tamper resistance, it is promising to solve these problems via the blockchain technology [5], [7], [8]. For example, Chen *et al.* [9] proposed a quality-driven incentive mechanism based on consortium blockchains for secure data sharing in IoV systems; Zhou *et al.* [10] designed a lightweight vehicular blockchain, namely LVBS, for secure data sharing. However, they did not consider the identity privacy disclosure of vehicles when applying blockchains. In addition, the limited performance of incumbent blockchains mismatches the demand of high throughput and mobility of IoV systems. Thus, new challenges are also emerging when introducing IoV into the blockchain-based facilities.

In order to address the aforementioned challenges, in this paper, we propose a privacy-preserving vehicular data sharing framework atop multi-sharding blockchain. In order to protect the identity privacy of vehicles while retaining the ability of revealing the identity of malicious vehicles for Trusted Authorities (TAs), we design an anonymous and auditable data sharing scheme taking Zero-Knowledge Proof (ZKP) technology as primitives. So as to bridge the gap between the low performance of blockchains and the high mobility of IoV systems, we design an efficient multi-sharding blockchain protocol for IoV to decrease blockchain communication costs without compromising the blockchain security. Our contributions are summarized as follows:

- We propose a privacy-preserving vehicular data sharing framework based on blockchain, where we design an anonymous and auditable data sharing scheme for protecting the

978-1-7281-8104-2/21/$31.00 ©2021 IEEE

identity privacy of vehicles while retaining the identity auditability.

- In order to achieve high scalability, low communication complexity in IoV systems, we propose an efficient multi-sharding blockchain protocol. In multi-sharding protocol, consensus nodes process multiple shards rather than one shard comparing to existing sharding protocols.
- We have implemented a proof-of-concept system for the proposed framework and conducted thorough analysis and extensive experiments. The experimental results show that the proposed data sharing framework is secure, privacy-preserving, and efficient for IoV systems.

II. Related Work

Recent advances [9], [10] have been devoted to designing a secure data sharing framework based on blockchain for IoV systems. However, there are still two challenges to be solved: *privacy disclosure* and *performance bottleneck*. We briefly review related works from these two aspects.

Privacy disclosure. There are many works aimed at protecting privacy in IoV systems. For example, Horng *et al.* [11] proposed an identity-based scheme that achieves secure data sharing in Vehicular Ad Hoc Networks (VANETs). However, this scheme relies on trusted cloud computing nodes and cannot guarantee system security. Wei *et al.* [6] designed a privacy-preserving vehicular communication scheme based on BBS04 group signature, where the group manager acts as a trusted arbiter, but the frequent updating of group members could bring a huge computing burden to the group manager. Yadav *et al.* [12] proposed a linkable location-based services scheme based on a modified Linkable Spontaneous Anonymous Group (LSAG) ring signature scheme, which also needs the trusted parties, *i.e.*, RSUs, as the signature proxies.

Performance bottleneck. Blockchain sharding [13] is one of the most popular solutions to improve the performance and scalability of blockchains, and there are many researchers focusing on design an efficient sharding protocol. For example, Zamani *et al.* [14] proposed RapidChain, the first sharding-based public blockchain protocol that achieves complete sharding of the communication, computation, and storage overhead; Kokoris *et al.* [15] designed an efficient cross-shard commit protocol that atomically handles transactions affecting multiple shards; Zhang *et al.* [16] presented CycLedger, a scalable and secure parallel protocol for distributed ledger via sharding. However, these works did not consider the high mobility of vehicles, so they are not suitable for IoV systems where cross-shard transactions happen frequently.

III. Attack Model

We consider three types of attacks/threats in the proposed system: *1) Attacks from vehicles*. We assume vehicles are not trusted. Malicious vehicles could report bogus data to the system (*bogus data attack*) or disguise as other honest vehicles (*impersonation attack*). *2) Attacks from RSUs*. RSUs and other infrastructures are assumed to be semi-trusted. Attackers may manipulate a small fraction of RSUs to perform Sybil attack, thereby controlling the network to obtain revenue. Attackers could also manipulate RSUs to broadcast false information. *3) Privacy disclosure*. Attackers can access all transactions (*i.e.*, vehicular data) due to the transparency of blockchains. Thus, attackers may infer the identity of vehicles by tracking a certain vehicle account (*i.e.*, public key) and analyzing its vehicular data, which cause identity privacy disclosure.

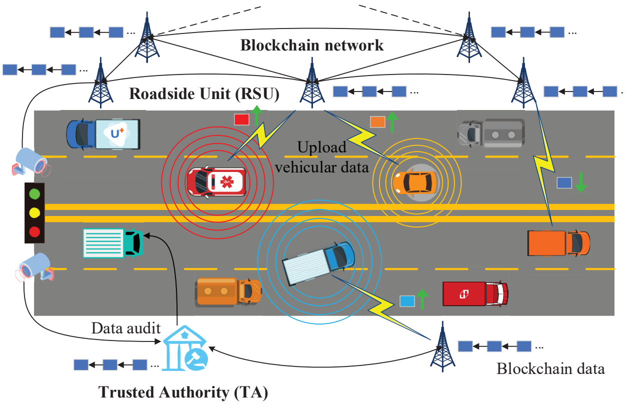

Fig. 1: The overall design of vehicular data sharing framework.

Note that, we assume TAs are trusted and secure. Attackers cannot manipulate TAs or steal TAs' secret keys. And we assume that attackers cannot break the cryptographic primitives, including hash inversion attack, digital signature forgery, etc. We conduct security analysis under the above assumptions.

IV. Vehicular Data Sharing Framework atop Blockchain

A. Overview

Fig. 1 shows the overall design of the proposed blockchain-based vehicular data sharing framework. In this framework, we have three types of roles: vehicles, RSUs, and TAs. All of these roles have their own blockchain accounts, which are unique identities used for making transactions in the blockchain.

RSUs and TAs are the static infrastructures of IoV systems which have higher computing power, so that they act as consensus nodes, *i.e.*, full blockchain nodes. The RSUs are the road side infrastructures (*e.g.*, traffic lights, cameras, street lamps) and responsible for interacting with vehicles. More specific, RSUs collect, synchronize vehicular data from vehicles and other RSUs, and transmit vehicular data to other vehicles nearby, in the form of blockchain transactions. TAs are also full blockchain nodes and responsible for vehicular data audit, as shown in Fig. 1. If there are some malicious vehicles upload bogus data, or disguise as other vehicles, TAs have the ability to reveal the real identity of the vehicles and punish them. In comparison, vehicles with high mobility and limited computing resources act as light blockchain nodes. In order to reduce the storage and computation overhead, the light blockchain nodes do not store blockchain data and participate in the process of consensus. They obtain or send vehicular data through RSUs nearby in the form of blockchain transactions.

Owing to the decentralization architecture of blockchain, we do not need a trusted centralized server to store, process ve-

TABLE I: Description of symbols in the anonymous and auditable data sharing scheme.

Symbols	Description
(msk,mpk)	Public-private key pair of TA
(sk,pk)	Public-private key pair of vehicle user
$CertGen$	Algorithm to generate certificate of a vehicle pk
$cert_{pk}$	Certificate for a legal vehicle pk
$Auth$	Algorithm to generate proof
m	Transaction that user wants to send
ek	User's public key pk encrypted with mpk
u	Proof generated by $Auth$
$Verify$	Algorithm to verify the validity of proof
$Reveal$	Algorithm to reveal the identity of user

hicular data, which strengthen the system reliability. Moreover, vehicular data stored in the blockchain are tamper-proof, which can ensure the integrity of on-chain data. However, on-chain vehicular data may leak the identity privacy of vehicles due to the transparency feature of blockchain. In addition, the limited performance of incumbent blockchains cannot satisfy the high mobility and throughput demand of IoV systems.

To solve the above two challenges, we firstly design an anonymous and auditable data sharing scheme for protecting the identity privacy of vehicles while preserving the data auditability for TAs. And then we propose an efficient multi-sharding blockchain protocol, which can improve the performance of blockchains for IoV systems.

B. Anonymous and Auditable Data Sharing Scheme

In order to protect the identity privacy of vehicles, ensure the authenticity of vehicular data, and reveal the identities of malicious vehicles, we design an anonymous and auditable privacy preserving scheme based on ZKP technology.

This scheme should satisfy the following objectives: *1)* Vehicles can know about the authenticity of obtained vehicular data, but there is no way to reveal the real identities of vehicle through tracking and analyzing its vehicular data. *2)* TAs can audit the real identities of malicious vehicles if necessary. *3)* The scheme should be non-interactive, as the communication bandwidth of mobile vehicles is very limited. *4)* The computation of vehicular data authenticity verification should be efficient even negligible.

For the above objectives, we design the primitives of this scheme based on non-interactive zero-knowledge proof technology, namely zk-SNARK [17]. The basic idea of this scheme is that each vehicle has a public-private key pair to represent the identity of the vehicle. In order to join the IoV system, the vehicle needs to register its identity information to TAs and get a certificate. When generating a transaction, the vehicle prove its legality through its public-private key pair and certificate, while other vehicles cannot obtain any identity information of the vehicle. At the same time, the vehicle identity encrypted with the TA's public key is also included in the transaction, so TA can directly decrypt it to obtain identity if necessary.

Specifically, the primitives and workflow of the anonymous and auditable data sharing scheme are described as follows, Table I explains the meaning of the symbols:

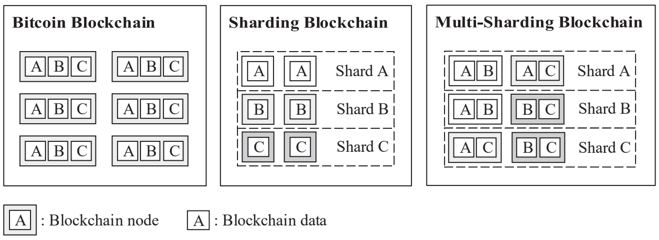

Fig. 2: The comparison of consensus approach in different blockchain protocols.

1) Setup. During the system initialization process, the TA generates a public-private key pair $(mpk,\ msk)$. mpk is known to all system participants as built-in information in the system.

2) $CertGen$. When a new vehicle user joins the system, a public-private key pair $(pk,\ sk)$ will be generated. The vehicle user will give his identity and pk to the TA, and the TA will use its private key msk to generate a certificate $cert_{pk}$ to prove the legality of the vehicle.

3) $Auth(m, sk, pk, cert_{pk}, mpk, ek) \rightarrow (m, ek, u)$. When a vehicle wants to launch a new transaction in the system, it needs to use the sk,pk,$cert_{pk}$,mpk fields to run the $Auth$ algorithm. In particular, the vehicle first calculates $ek = enc(mpk, pk||m)$, which encrypts the vehicle's public key pk concatenating message m with mpk. Let $x = (sk, pk, cert_{pk})$ is the private input, $y = (m, mpk, ek)$ is the public input. Then $Auth$ algorithm is built on the zk-SNARK proving algorithm $Prover(x, y, PP)$ [17]. This algorithm outputs m, ek and u, which are broadcast to the blockchain network.

4) $Verify(m, mpk, ek, u) \rightarrow 0/1$. Consensus nodes (*i.e.*, RSUs) can run the $Verify$ algorithm and output $0/1$ to verify whether a signature is legal. The algorithm is built on the zk-SNARK verifying algorithm $Verifier(y, u, PP)$ [17].

5) $Reveal(ek, msk) \rightarrow pk$. When a malicious situation occurs, the TA can decrypt the ek to obtain the vehicle's public key pk, thereby revealing the identity of the vehicle.

C. Multi-Sharding Protocol

Multi-sharding is a blockchain sharding protocol that focuses on improving the performance of cross-shard transactions. Generally, the basic idea of multi-sharding is that by maintaining multiple shards, consensus nodes can directly process cross-shard transactions. In the existing blockchain sharding protocols, the consensus nodes would only store the data of one shard, and only process the transactions of the corresponding shard. Therefore, when a transaction involves the data of multiple shards, the consensus node needs to communicate with other shards. Unfortunately, cross-shard communication may need $O(c^2)$ communications in two-phase commit (2PC) scheme [15], [16], [18] to ensure safety (c is the node number in one shard).

The point of multi-sharding is that by maintaining multiple shards, consensus nodes can directly process cross-shard transactions between these shards. Note that maintaining multiple shards by a consensus node is not the same as maintaining a larger shard. Because different consensus nodes will choose to

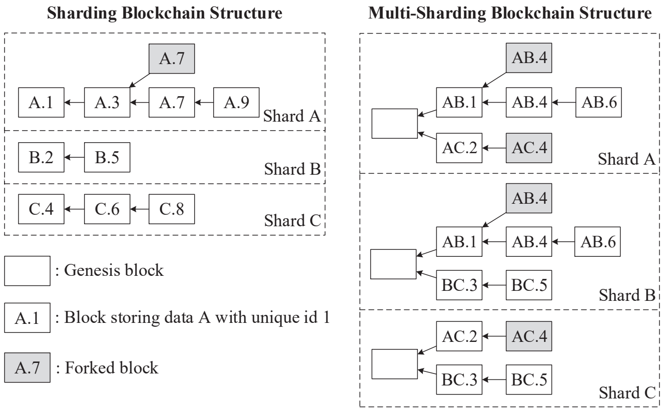

Fig. 3: The comparison of blockchain structure between multi-sharding and existing sharding protocols.

maintain different shard sets. Another feature of multi-sharding is detaching the concepts of shard and consensus group. In the existing sharding protocols, nodes that maintain a shard will form a consensus group and only reach consensus within the shard. In multi-sharding, consensus group is the nodes which store the same multiple shards. And the nodes in a same consensus group will reach consensus, decide execution order and result of the transactions in these shards. As shown in Fig. 2, nodes storing “BC” in shard B will form a consensus group with nodes storing “BC” in shard C.

Then we introduce multi-sharding protocol in this framework more detailedly. For ease of description, we denote a RSU as $r \in R$, a vehicle as $v \in V$.

Shard. We divide the entire system into A areas according to location, denote as $a_i \in Area$. Each area contains several RSUs, and vehicles move in different areas. We regard an area as a shard, denoted a shard as $Shard_i \in S$. When a vehicle generates data in an area a_i, this data will be attached to the corresponding shard $shard_i$.

Multi shards. Consensus nodes in the system (including RSUs and TAs) will store k shards. In addition to the shard where the RSU is located, consensus nodes will randomly store $k-1$ shards from the remaining $A-1$ shards. That is, a consensus node $n \in N_i$ in the area a_i will store all the data in the shards $\{shard_i, C_{A-1}^{k-1}(shard_1, ..., shard_{i-1}, shard_{i+1}, ...)\}$. In the following description, in order to simplify the discussion, we temporarily assume $k = 2$. The consensus nodes only store the data in two shards $shard_i$ and $shard_j$, and the shard is randomly calculated by the blockchain address $addr$ of the consensus node and a public random number $RAND$: $j = HASH(addr||RAND)modA$. We denote the consensus nodes set storing the data in $shard_i$ and $shard_j$ as N_{ij}. Note that $N_{ij} \subset N_i \cup N_j$.

Block structure. There will also be a large number of transactions involved in multiple areas. The transaction tx_{ij} involved in the area a_i and area a_j will be processed by node set N_{ij}. Nodes in N_{ij}, as miners, will verify the authentication of tx_{ij} and add it to the block $block_{ij}$. Obviously, all transactions in $block_{ij}$ generated by N_{ij} only involve transactions between vehicles in the area a_i and area a_j.

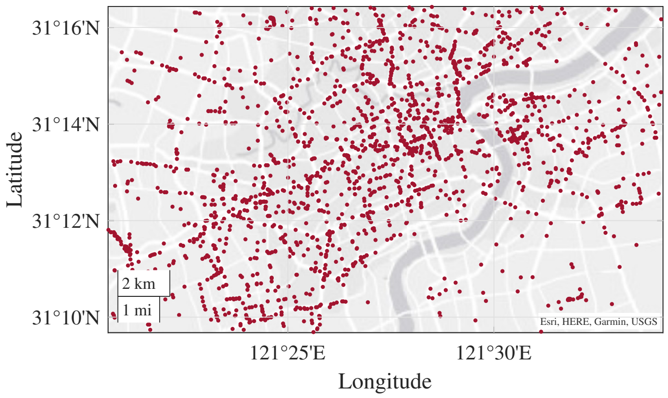

Fig. 4: Visualization of the SUVnet on March 1, 2007, at 09:00. Every dot represents a taxi.

Blocks in a shard form a tree. Under this design, the structure of data in a shard is no longer a chain, but a tree, as shown in Fig. 3. All blocks in a shard have a common ancestor, which is the genesis block that does not store transactions. All $block_{ij}$ will form a chain, which is mined by N_{ij}. Finally, a shard will contain $A-1$ chains.

Ordering blocks in a shard. Since the data structure of the shard is a tree, the ability to defense double-spending attacks has disappeared. We explain this problem through an example. Suppose the vehicle v in the $shard_i$ initiates two transactions at the same time to request the data in the $shard_j$ and $shard_k$, and the price of one data is 1 token and the balance in the account of vehicle v is also 1 token. Unfortunately, the miners in N_{ij} and N_{ik} both think that tx_{ij} and tx_{ik} are legal, and add them to the block. Since these two transactions are on different branches of a shard tree, it is impossible to directly rollback the block through transaction conflict. Therefore, an attacker can easily complete a double spending attack.

A scheme proposed in a sharding work OHIE [19] to order blocks can be effectively migrated to solve this problem. *i.e.*, by introducing a defined timestamp, the consensus nodes will locally transform the tree structure in a shard into a chain structure, so the above double-spending attack can be defended by a block rollback.

V. Analysis and Evaluation

In order to truly reflect the performance of our framework, we used a real-time traffic dataset, SUVnet [20], which was collected from over 4000 taxis in Shanghai, China, as shown in Fig. 4. We implemented a prototype of our framework. We used jsnark[1] to implement zk-SNARK and chose *Groth16* [17] as our underlying zk-SNARK scheme. We implemented the multi-sharding blockchain prototype in C++ and conducted it on 10 nodes, in which one node simulates the transaction generating and other 9 nodes work as distributed consensus nodes. The configuration of every node is Intel Core i7 with 16GB of RAM.

A. Anonymous and Auditable Data Sharing Scheme

Analysis. For explaining the availability of our anonymous and auditable data sharing scheme, we give a brief analysis

[1] [Online]. Available: https://github.com/akosba/jsnark

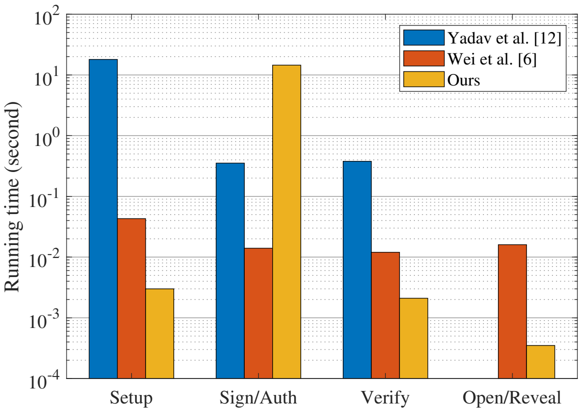

Fig. 5: Running time comparison of privacy-preserving schemes between ours and other works (y-axis is log-scale).

on three attack types in Section III. *1) Against attacks from vehicles*. The data audit and penalty feature of TAs can punish bogus data attacks and impersonation attacks from vehicles. The premise is the unforgeability of our anonymous and auditable scheme. Unforgeability means if attackers do not use their own public-private key pair or legal certificate, attackers cannot generate legal proofs to be verified by other trusted nodes. Unforgeability corrects as this attack clearly violates the ZKP primitives. *2) Against attacks from RSUs*. Manipulated RSUs can conduct Sybil attacks and possible double-spending attacks. These attacks can be defended by the blockchain design. *3) Privacy disclosure*. Attackers would extract privacy information from the authentication, which is straightforward for the ZKP primitives. In addition, attackers can attempt to link the multiple data from single vehicle to obtain some sensitive information. *Unlinkability* means the attackers cannot link two data from a vehicle. Unlinkability holds because for attackers, the outputs of $Auth$ algorithm are all random values which cannot be distinguished.

Evaluation. In Section IV-B, we mentioned the design objectives of our anonymous and auditable scheme are mainly four points, and the last two objectives are in response to the performance. Obviously, the third objective of reducing communication interactions have been achieved by zk-SNARK primitives. For illustrating the performance of our scheme, we evaluated the running time comparison of our scheme and two related works, which are respectively the ring signature based [12] and the group signature based [6] schemes.

We used this experimental environment to test the anonymous and auditable scheme we proposed, and the experimental results are shown in the Fig. 5. Note that, the step of *Setup* (corresponding to $CertGen$ algorithm in our scheme) will be executed only once. For one data in the network, generating authentication process *Sign/Auth* will be executed only once by the data owner. While verifying authentication process *Verify* will be executed by many consensus nodes. It can be seen from Fig. 5 that the *Sign/Auth* step takes a bit longer time in our scheme comparing to the other two schemes. But this does not significantly affect the performance of the framework. On the contrary, for each data, *Verify* step may be executed for hundreds times, which greatly affect the scheme performance. Obviously, our scheme performs better in *Verify* process as the running time is 5x smaller than the related works.

TABLE II: The comparison of multi-sharding protocol and existing blockchain protocols.

	Bitcoin	Sharding	Multi-sharding*
Storage redundancy	O(n)	O($\frac{n}{m}$)	O($\frac{n}{\sqrt{m}}$)
Bandwidth consumption	O(n)	O($\frac{n^2}{m^2}$)	O($\frac{n}{\sqrt{m}}$)
Throughput	O(1)	O(m)	O(m)
Security	n	$\frac{n}{m}$	$\frac{n}{m}$

[1] n: total number of consensus nodes, m: shard number, $\frac{n}{m} = c$: number of consensus nodes in a single shard.

[2] * when set the shard number $m' = \sqrt{m}$ in multi-sharding protocol, multi-sharding can gain the same security parameter and throughput improvement. Meanwhile, multi-sharding performs better in bandwidth consumption, with the cost of more storage consumption.

B. Multi-Sharding Protocol

Analysis. To claim the advancement of multi-sharding protocol, we provide an informal theoretical comparison of CycLedger with existing sharding blockchain protocols, seen in Table II. The existing sharding protocols we compared are based on 2PC cross-shard communication protocol like Omniledger [15], Cycledger [16], Chainspace [18], which have similar settings on consensus approach, cross-shard communication, thereby achieving similar performance. Other cross-shard communication protocols in previous works have their own application limitations, *e.g.*, RapidChain [14] only supports simple currency transfers in the system.

We defined 4 evaluation indicators to compare the performance of our solution and previous works. The *storage redundancy* represents the storage resources consumed by a transaction in the entire system. *Bandwidth consumption* represents the amount of communication need to transmit for a transaction, which characterizes the performance of cross-shard communication. For simplicity, we assume that all transactions in the system are cross-shard transactions, and these transactions only involve two shards. *Throughput* represents the theoretical throughput level under different schemes. The *Security* parameter represents the number of nodes in a consensus region, which can qualitatively describe the possibility of malicious nodes successfully manipulating a consensus group.

For the previous sharding protocol which using 2PC cross-shard communication, the storage can be $O(\frac{n}{m})$ and the security can be $\frac{n}{m}$ (n: total number of consensus nodes, m: shard number). As one shard can be seen as a relatively independent consensus group, m shards in sharding protocol can theoretically gain m times throughput than Bitcoin protocol. Most importantly, 2PC cross-shard communication involves in $O(c^2)$ interactions, which is a bottleneck of sharding protocol. In multi-sharding protocol, we can get a similar result: the storage redundancy is $O(\frac{n}{m})$, the bandwidth consumption can be $O(\frac{n}{m})$, $O(m^2)$ throughput and $\frac{n}{m^2}$ security. Security is a problem that cannot be despised in public blockchains. In multi-shard, we can simply reduce the number of shards to achieve the same level of security as the previous sharding protocol, *i.e.*, reduce the

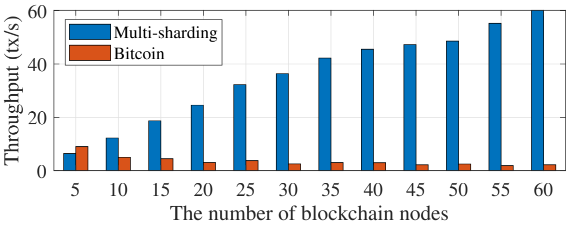

Fig. 6: Throughput comparison between multi-sharding blockchain and Bitcoin blockchain.

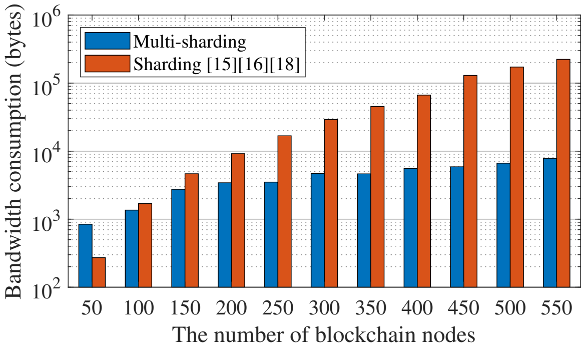

Fig. 7: Bandwidth consumption comparison between sharding blockchain and multi-sharding blockchain (y-axis is log-scale).

number of shards in the multi-shard to $m' = \sqrt{m}$, and results are shown in fourth column in Table II. From the table, we can find that multi-sharding protocol have the same throughput and security compared with previous sharding protocol. While the bandwidth consumption of multi-sharding is better than the existing sharding protocol at the cost of higher storage consumption.

Evaluation. We conducted simulation experiments on Bitcoin, sharding and multi-sharding protocols to evaluate the correction of the above analysis. Fig. 6 shows the throughput comparison between multi-sharding and Bitcoin protocol when setting each shard contains 5 blockchain nodes. We can find that with the increasing number of the consensus nodes, the throughput of multi-sharding will near-linearly increase and then realize scalability. Fig. 7 shows the bandwidth consumption comparison between sharding and multi-sharding protocol when setting the number of shards to 5. When fixing the number of shards and increasing the total number of nodes in the network, the bandwidth consumption of one hundred transactions would be unaffordable in existing sharding protocols, while multi-sharding performs outstanding.

VI. Conclusion

In this paper, we propose a privacy-preserving vehicular data sharing framework atop multi-sharding blockchain. First, we design an anonymous and auditable data sharing scheme based on ZKP for protecting the identity privacy of vehicles while retaining conditional auditability. Second, we propose an efficient multi-sharding blockchain protocol, which can achieve lower communication complexity compared to the existing sharding protocols and is more practical for IoV systems. Evaluation and analysis results indicate that our framework can efficiently strengthen the system security and protect the identity privacy.

Acknowledgment

This work was supported in part by NSFC grant 61972253, U1908212, 72061127001, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. Linghe Kong is the corresponding author.

This work is partially funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the Project UIDB/50008/2020; and by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 313036/2020-9.

References

[1] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of vehicles in big data era," *IEEE CAA Journal of Automatica Sinica*, 2018.

[2] Z. Jiao, H. Ding, M. Dang, R. Tian, and B. Zhang, "Predictive big data collection in vehicular networks: A software defined networking based approach," in *IEEE GLOBECOM*, 2016.

[3] A. Ouya, B. M. De Aragon, C. Bouette, G. Habault, N. Montavont, and G. Z. Papadopoulos, "An efficient electric vehicle charging architecture based on lora communication," in *IEEE SmartGridComm*, 2017.

[4] Z. Lu, G. Qu, and Z. Liu, "A survey on recent advances in vehicular network security, trust, and privacy," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[5] S. Cao, S. Dang, X. Du, M. Guizani, X. Zhang, and X. Huang, "An electric vehicle charging reservation approach based on blockchain," in *IEEE GLOBECOM*, 2020.

[6] L. Wei, J. Liu, and T. Zhu, "On a group signature scheme supporting batch verification for vehicular networks," in *International Conference on Multimedia Information Networking and Security*, 2011.

[7] T. Jiang, H. Fang, and H. Wang, "Blockchain-based internet of vehicles: Distributed network architecture and performance analysis," *IEEE Internet of Things Journal*, 2019.

[8] W. Dong, Y. Li, R. Hou, X. Lv, H. Li, and B. Sun, "A blockchain-based hierarchical reputation management scheme in vehicular network," in *IEEE GLOBECOM*, 2019.

[9] W. Chen, Y. Chen, X. Chen, and Z. Zheng, "Toward secure data sharing for the iov: A quality-driven incentive mechanism with on-chain and off-chain guarantees," *IEEE Internet of Things Journal*, 2020.

[10] Z. Su, Y. Wang, Q. Xu, and N. Zhang, "Lvbs: Lightweight vehicular blockchain for secure data sharing in disaster rescue," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[11] S.-J. Horng, C.-C. Lu, and W. Zhou, "An identity-based and revocable data-sharing scheme in vanets," *IEEE Transactions on Vehicular Technology*, 2020.

[12] V. K. Yadav, S. Verma, and S. Venkatesan, "Linkable privacy-preserving scheme for location-based services," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[13] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. G. Sirer, *et al.*, "On scaling decentralized blockchains," in *FC*, Springer, 2016.

[14] M. Zamani, M. Movahedi, and M. Raykova, "Rapidchain: Scaling blockchain via full sharding," in *ACM CCS*, 2018.

[15] E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, E. Syta, and B. Ford, "Omniledger: A secure, scale-out, decentralized ledger via sharding," in *IEEE S&P*, 2018.

[16] M. Zhang, J. Li, Z. Chen, H. Chen, and X. Deng, "Cycledger: A scalable and secure parallel protocol for distributed ledger via sharding," in *IEEE IPDPS*, 2020.

[17] J. Groth, "On the size of pairing-based non-interactive arguments," in *EUROCRYPT*, Springer, 2016.

[18] M. Al-Bassam, A. Sonnino, S. Bano, D. Hrycyszyn, and G. Danezis, "Chainspace: A sharded smart contracts platform," in *NDSS*, ISOC, 2018.

[19] H. Yu, I. Nikolic, R. Hou, and P. Saxena, "Ohie: Blockchain scaling made simple," in *IEEE S&P*, 2020.

[20] H.-Y. Huang, P.-E. Luo, M. Li, D. Li, X. Li, W. Shu, and M.-Y. Wu, "Performance evaluation of suvnet with real-time traffic data," *IEEE Transactions on Vehicular Technology*, 2007.