LDV: A Lightweight DAG-Based Blockchain for Vehicular Social Networks

Wenhui Yang, *Student Member, IEEE*, Xiaohai Dai, *Student Member, IEEE*, Jiang Xiao, *Member, IEEE*, and Hai Jin, *Fellow, IEEE*

***Abstract*—As social networks are integrated into the *Vehicular Ad Hoc Networks* (VANETs), the emerging *Vehicular Social Networks* (VSNs) have gained massive interests. However, the security and privacy of data generated by various applications in VSNs is a great challenge, which blocks the further development of VSNs. The emerging blockchain technology seems to be a good catalyst for the development of VSN with its high security and irreversible features, which can be also a data management tool for rapidly generated data of VSNs with tamper proof. However, the full duplicates of blockchain data need to be stored in each node to ensure security, which is unacceptable for vehicles with limited resource. In this paper, to address the above storage challenge, a lightweight *Directed Acyclic Graph* (DAG) based blockchain (LDV) is proposed for resource-constrained VSNs. Specifically, based on the in-depth analysis of VSNs, we propose the social-based data reduction approach. In detail, each node only stores the interested data within the topic groups of interest and ignores the irrelevant data. To avoid the huge storage cost within large-scale groups with large amounts of data, we further present the historical data pruning method within a group, which meets the storage requirement by reducing the number of duplicates stored in each node. Experimental results show that LDV can save 97.13% storage space and has good scalability.**

***Index Terms*—Vehicular social networks, blockchain, data reduction.**

I. Introduction

TODAY *Vehicular Social Networks* (VSNs) have attracted massive interests from both academia and industry thanks to the promise of advancing the *Vehicular Ad Hoc Networks* (VANETs) with social networks. In particular, the distributed commuters (e.g., drivers, passengers, *Road Side Units* (RSUs), and vehicles) in VSNs of similar routine or social behaviours, can group into virtual communities and transmit the socially-aware data on roadways. By aggregating the social characteristics among the commuters of mutual interests, VSNs have fostered a myriad of prospective applications. For example, drivers can incorporate the passenger's utility to develop real-time demand-supply recommendation systems [1], passengers can sociailize with physically closed users via music, photos, and video [2], RSUs can broadcast the shared traffic conditions with vehicles in range for road safety and emergency warning [3], and the mobility patterns of vehicles along the same road segments can facilitate intelligent traffic control [4]. In return, these applications will generate huge amounts of data, including traffic information, social information, and privacy information such as routine locations or user preferences. The ever-growing volume and high variety of data require novel data storage method for VSNs with limited capacity by nature [5].

Furthermore, there exists malicious commuters who disseminate false information to others. These attacks will manipulate and violate the VSNs data in holistic environment. The lack of secure data storage will result in misbehaving and discrepancy of vehicular commuters. For instance, the selfish drivers will post false parking information in order to win a parking space for him/her [6], multiple indentities can be forged by malicious commuters to post false information misleading others into congested routes, namely Sybil attack [7].

As a result, a critical design aspect of VSNs is to provide a scalable data storage scheme without compromising the security. Unfortunately, recent work in VSNs primarily attempts to process the data and investigate the social characteristics, e.g., the small-world features investigated in [8] and the user behavior studied in [9]. All the aforementioned literatures we examined have ignored the fundamental storage issue, thus impeding them to meet the desired data storage requirements in VSNs.

In this paper, we remedy these deficiencies by empowering VSNs with blockchain technology as the basis of data storage. Blockchain has shown its merits of distributed consensus-enabled irreversibility and cryptographic hashing algorithms, when originated from the well-known digital currency Bitcoin [10] in the financial industry. Inspired by this, the built-in tamper-resistant traits of blockchain can enable secure data storage in distributed holistic VSNs environment.

Nevertheless, the security of blockchain relies on the highly redundant distributed ledger feature of blockchain, i.e., each node ensures secure storage at the cost of maintaining a complete history of transactions linked by blocks. Taking Bitcoin as an example, the current ledger of blockchain data has exceeded 210 GB, where each full node in Bitcoin network is required to store a full copy. The storage cost of the entire Bitcoin network

Manuscript received September 1, 2019; revised November 24, 2019; accepted December 18, 2019. Date of publication January 8, 2020; date of current version June 18, 2020. This work was supported by the Technology Innovation Project of Hubei Province of China under Grant 2019AEA171, in part by the National Science Foundation of China under Grants 2018YFB1004805 and 61702203, and in part by Hubei Provincial Natural Science Foundations under Grant 2018CFB133. The review of this article was coordinated by Prof. H. Li. *(Corresponding author: Jiang Xiao.)*

The authors are with the National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Laboratory and the Cluster and Grid Computing Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: ywh@hust.edu.cn; daixh@hust.edu.cn; jiangxiao@hust.edu.cn; hjin@hust.edu.cn).

Digital Object Identifier 10.1109/TVT.2020.2963906

becomes exorbitant with large amounts of data. The similar situation can be witnessed in VSNs. According to the statics,[1] the global vehicles in used have come to around 1.2 billion already in 2015, and the scale is likely to reach 2 billion or more by 2035.[2] The total amount of data generated by billions of vehicles is exceedingly enormous, which become worse when the full copies of data need to be stored on each vehicle. More seriously, the total amount of data will become larger when the full copies of data (i.e., the data generated by various of devices in VSNs) grows rapidly. Such rapid increasing data will lead to significant storage overhead for resource-constrained commuters in VSNs. It is non-trivial to apply the conventional blockchain to store socially-aware VSNs data with low storage overhead guarantee. Therefore, to design a lightweight blockchain system for VSNs with the strict restriction of security is of great importance and urgency.

To this end, we present LDV, a *Directed Acyclic Graph* (DAG) based blockchain system to enable lightweight and secure data storage for resource-constrained VSNs. LDV introduces DAG to be the underlying VSNs data structure. Specifically, DAG provides promising properties of higher efficiency and scalability than conventional block-based by organizing the data in the format of transactions directly [11]. The key insight of LDV lies on the fact the storage burden can be reliefed by only storing the data in grouped commuters of common interests. To further reduce the storage overhead in large-scale groups, we decrease the number of duplicates, and prune the historical data with little usefulness to make room for storage of useful real-time information. In more detail, the design of LDV is based on the in-depth analysis of social characteristics of VSNs:

- In VSNs, the commuters are usually care about the information of interest, and pay little attention to the irrelevant information that is useless to them.
- The expired historical data are of little value to real-time decision-making in the rapidly changing transportation scenario.

Therefore, only the relevant data and the recent information need to be stored for normal vehicular nodes, which reduce the storage requirement largely. To evaluate the effect of data reduction approach, we have implemented a prototype system and conduct some experiments. The experimental results show that 97.13% storage space can be saved.

In summary, this paper makes three contributions:

- We conducted an in-depth analysis of the social relationships in VSNs. To the best of our knowledge, this is the first attempt deeply combining the social relationship to design a lightweight blockchain system for VSNs.
- We design a social-based data reduction approach to reduce the storage cost of blockchain based on the social relationship of VSNs. To further reduce the storage cost within a single group, we propose the pruning method of historical data utilizing the feature of real-time in VSNs.
- We have implemented the data reduction approaches in our prototype system, and some experiments are conducted to evaluate the effect of data reduction.

The remainder of this paper is structured as follows: we introduce the background and related works of this paper in Section II. Then, the design of the lightweight DAG-based blockchain system is described in Section III. The evaluations and discussions of LDV are described in Section IV and Section V respectively. Finally, the paper is concluded in Section VI.

II. Background and Related Work

A. Social Relationship in VSNs

With the help of VANETs, data transformation between mobile vehicles is feasible. As shown in the lower layer of Fig. 1, VANETs are comprised of vehicles, RSUs and communication links between them. The data exchange in VANETs relies on multiple hops between the above components. Although data transmission is convenient through VANETs, the value and semantic of transmitted data is extremely limited in VANETs, which bring little improvement to transportation network.

Furthermore, social network enable the information transformation with rich semantics in VSNs rather than simple data transmission. In *Online Social Networks* (OSNs), people with common interest share information with each other. Fortunately, vehicular network can be provided with the similar characteristic when integrated with OSNs. Similarly, vehicles can also share information that is relevant to their interests with others through VSNs, such as social information or traffic information about a particular road.

As an instance, Fig. 1 depicts a common social scenario of VSNs when integrated with OSNs. In more detail, the lower layer describes the physical communication links among vehicles through the ad hoc network. The virtual social relationships of vehicles are shown in the upper layer, in which several topic groups are formed according to the interests of different vehicles. In this scenario, people can subscribe to any topic which they are interested in and join the topic group freely. In this way, drivers can easily receive information from subscribed topics. For example, while a driver has subscribed to a topic about the traffic conditions at lane *A*, information about lane *A* will be notified to this driver in time unless he/she leaves this topic.

Although vehicular networks integrated with social characteristics can achieve rich semantic and valuable information transformation, the introduction of social features is likely to deteriorate the security and privacy of VSNs by analyzing the exposed social information, which needs to be tackled carefully.

B. Blockchain Technology

With the prosperity of Bitcoin [10] and other blockchain systems [12], [13], blockchain is believed to provide a data storage service with high security and good privacy protection in distributed environment. By adopting distributed consensus algorithm, the blockchain nodes can reach an agreement on stored data and thus each node will store the same data

[1] https://www.statista.com/

[2] https://www.greencarreports.com/

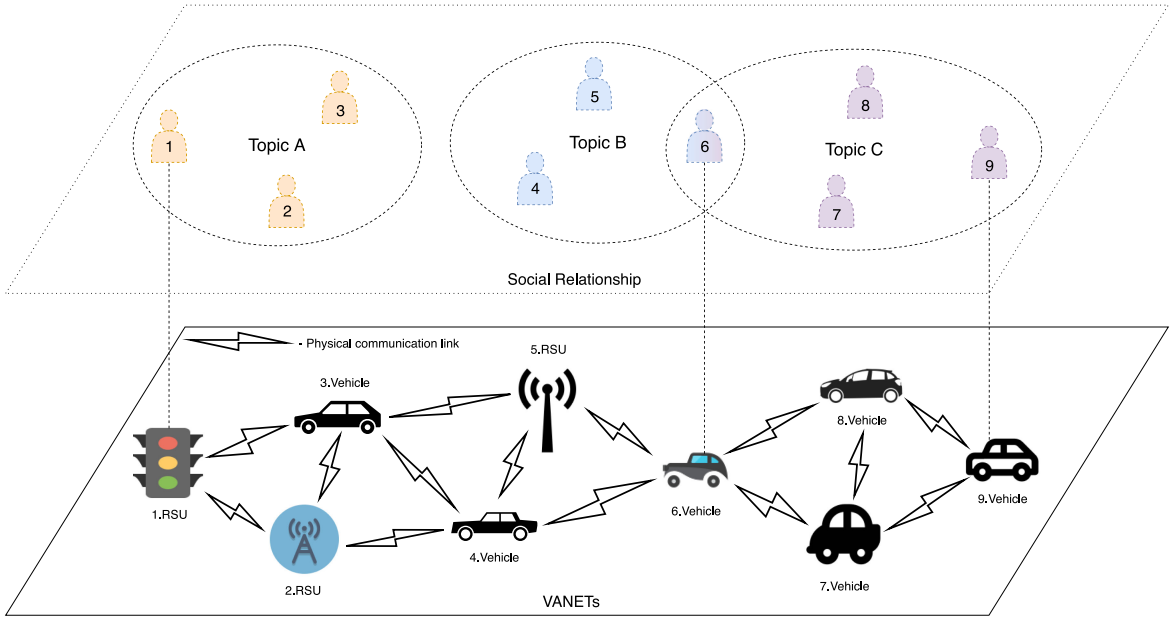

Fig. 1. Social relationship in VSNs.

eventually. Thanks to the use of consensus algorithm and secure cryptographic hash algorithm as well as Merkle tree, the computational power must exceed 50% to tamper with data in blockchain, which is able to resist attack from malicious nodes. Moreover, all nodes interact with each other through addresses composed of some alphanumeric characters, which achieves good anonymities in blockchain. Therefore, it can realize good privacy protection for data stored in blockchain system due to the data sender represented by address which is not known to others. As a result, blockchain can bring a secure data storage for VSNs to alleviate the problems of VSNs described aboved.

It seems that blockchain can solve aforementioned problems of VSNs very well. However, current blockchain systems are extremely resource intensive, the power used in mining each year is about tens of terawatt-hours [14]–[16], which is unacceptable for resource-constrained VSNs. In addition, popular blockchain systems mainly adopt chain-based structure like Bitcoin and Ethereum. As shown in Fig. 2(a), the chain-based design processes transactions and blocks in a sequential approach, which results in poor performance in terms of throughput. Both Bitcoin and Ethereum hava a low throughput compared to Visa. For example, the average throughput of Bitcoin is estimated to 7 *Transactions Per Second* (TPS) [17]. Obviously, the low throughput and resource consumption of chain-based blockchain is not suitable for VSNs with rapid data generation and constrained resource.

Recently, a novel DAG-based blockchain is proposed to improve the scalability of blockchain, such as IOTA [18], Byteball [19] and Nano [20]. Due to the adoption of graph structure, the processing of transactions can be done in a parallel manner, which is different from the sequential way in chain-based blockchain, as shown in Fig. 2(b). In other words, chain-based blockchain processes only one block at a time while DAG-based

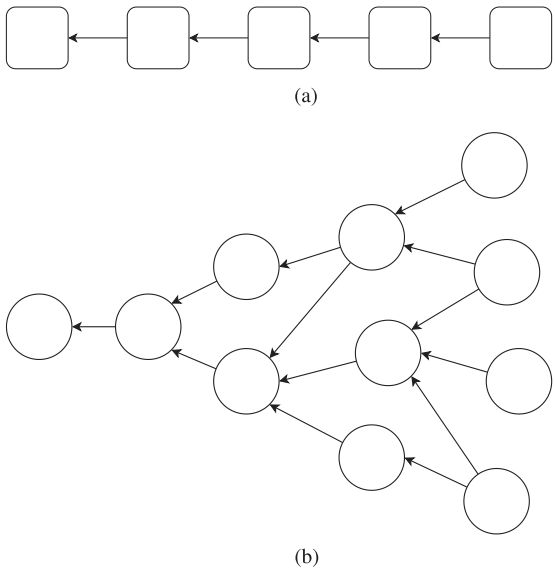

Fig. 2. The comparison of chain-based blockchain and DAG-based blockchain. (a) The structure of chain-based blockchain. (b) The structure of DAG-based blockchain.

blockchain deals with multiple blocks at the same time. Besides, because of the low resource requirements of DAG-based blockchain by avoid the massive useless computation in mining, it is more suitable for resource-constrained VSNs. Therefore, we adopt DAG-based blockchain in the next design to compensate for the shortcomings of VSNs. Unfortunately, the high throughput of blockchain will further increase the storage cost [21], which is conflict with the resource-constrained devices in VSNs. As a result, a lightweight and high throughput blockchain sytem need to be designed for VSNs.

C. Related Work

There are various of researches focusing on vehicular social networks and blockchain. We introduce these works from the perspective of data management in VSNs and data reduction in blockchain respectively.

1) Data Management in VSNs: Most of researches about data management of VSNs are focusing on data analysis, data processing, and data security including privacy protection in VSNs.

Wang *et al.* [1] proposes an real-time recommendation system for drivers and passengers to try to satisfy the their requirement and profit at the same time by analyzing the data generated by taxi. Meanwhile, some of the studies try to analyze the social characteristic in VSNs. Concretely, the small-world features are studied in [8]. The user behavior of publishing information influenced by external environment in VSNs is investigated in [9].

Data processing is also studied by many previous literatures. Yang *et al.* [22] propose a keyword extraction metric to improve the query performance of information in VSNs. Kong *et al.* [4] propose a data generation approach of private cars through the dataset of taxis to make up for the lack of private cat data. Meanwhile, efficient range query on encrypted data and secure query with privacy protection are studied in [23] and [24] respectively.

In terms of data privacy protection in VSNs, a dynamic group division algorithm [25] is presented to protect privacy of location and trajectory generated by vehicles for the scenario of 5G-based VSNs. In [26], the authors try to address the location privacy issue in VSNs by obscuring the location of original sender of information. Jiang *et al.* [27] proposes an authentication scheme to protect privacy for thin-client in blockchain-based *Public Key Infrastructure* (PKI). However, there are little literatures focusing on the storage cost of generated data in VSNs. As a complement, we design a lightweight blockchain system to store data for VSNs with a low storage overhead.

2) Data Reduction in Blockchain: Many existing solutions to address the security and privacy of VSNs requiring data encryption [22] which brings extra overhead. Blockchain takes advantage of cryptographic hash to ensure security of data. On the other hand, with the assistance of anonymity, blockchain can provide a great privacy protection of VSNs. However, due to the high storage overhead of Blockcahin, it is urgent to address the storage challenge of blockchain. Recently, some works are trying to alleviate the storage requirement in blockchain system.

Based on the different security level of blocks in different terms, Jia *et al.* [28] propose a duplicate ratio mechanism to store different blocks in different ratio in order to achieve low storage cost. The authors believe that the older block can store a small number of blocks compared to the newer blocks because the requirement of computation is larger when modifying a old block. To avoid data loss of blocks with less duplicates, they present a node reliability verification method to ensure the old blocks are stored in reliable nodes. However, the approach introduces a extra chain to store reliable information which increases the storage overhead. Furthermore, the approach needs a master node to calculate the reliability which is impracticable in P2P network.

Xu *et al.* [29] try to address the storage problem by organizing several nodes into a *Consensus Unit*. However, their approach is based on strong trust assumptions between nodes in *Consensus Units*. But it is so difficult to achieve above conditions in hostile VSNs environment.

In [21], the authors present a jigsaw-like data reduction approach, which each node only store relevant data of themselves and uses the merkle path to verify the authenticity of transactions. Although this approach can achieve low storage overhead by only store a few relevant data, it brings a certain number of communication cost when requesting additional data. Moreover, the approach can only apply to the blockchain systems with merkle tree such as Bitcoin. However, the Bitcoin system has a poor scalability in terms of throughput which is unsuitable for VSNs with rapid data generation.

In summary, to the best of our knowledge, this is the first work to study the problem of high storage cost in DAG-based blockchain systems.

III. LDV Design

In this section, we demonstrate the design of LDV. Firstly, we analyze the situation of VSNs in depth and come to several insights. Based on these insights, we give the design of social-based data reduction approach. Then, to further reduce the storage overhead, we enhance the basic design by pruning the historical data within a topic group. Furthermore, there exists several challenges during the design, which are tackled subtly.

A. In-Depth Analysis of VSNs

In VSNs, due to the introduction of social networks, people with common interests can share data with each other and form virtual social relationship. For example, on the road, commuters can share information about traffic or entertainment through VSNs with others during their trips. By utilizing the information obtained, drivers will know the current traffic condition on specific road and make decisions about their optimal routes. In order to obtain information from commuters with common interests, drivers can participate in the specific topics they like such as the traffic condition of a specific road, while paying little attention to traffic condition of other roads they do not pass. After joining the topic group, the members in the same group are able to publish some useful information to the group for the convenience of others, and meanwhile receive the information from others in the group.

In fact, drivers are more likely to communicate with people with similar interests frequently [5], which means they usually pay more attention to topics of interest and care less about irrelevant topics. More generally, commuters are usually interested in the topics of roads between their location and destination, and are unlikely to receive or publish information about traffic on other roads. Besides, the trip routes of drivers are usually fixed and thus the topics they focus on are often regular, which means the social relationships are usually stable compared to dynamic network topologies.

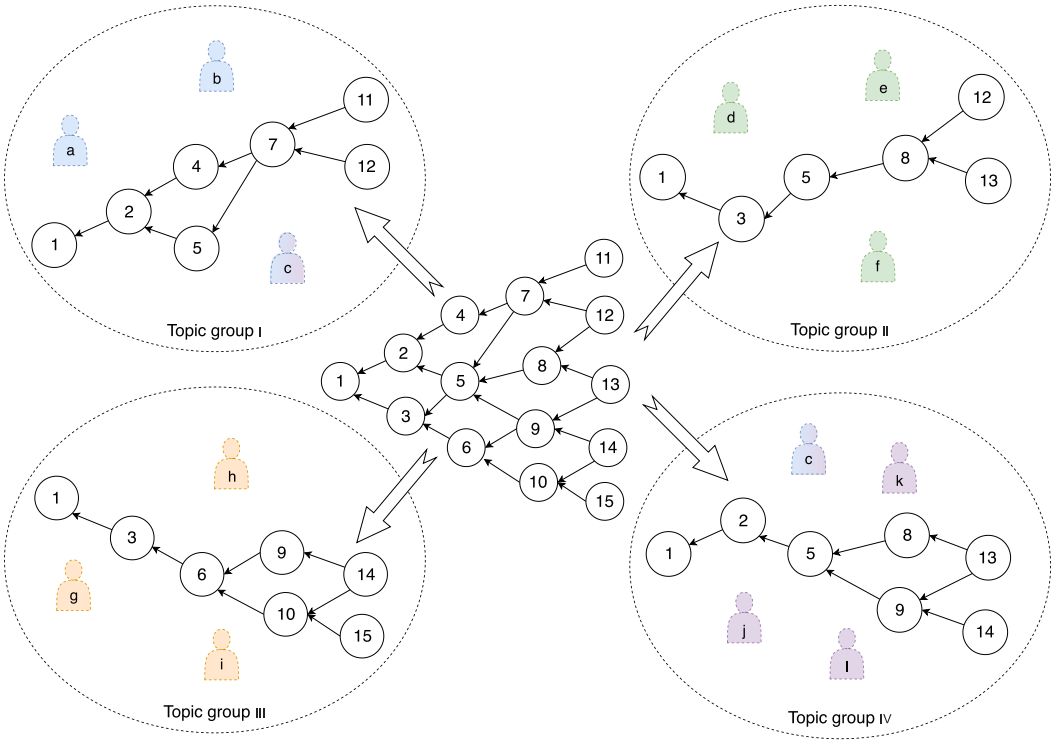

Fig. 3. Overview of the LDV design.

Insight 1: In terms of social relationship in VSNs, people usually focus on the information about topics of interest and have little requirement for other data that are of no interest to them.

Social features in VSNs allow people to get real-time news on relevant topics in order to make the accurate arrangement for the next travel. However, with the rapid generation of data in VSNs, it it hard to identify useful data from vast quantities of data. Due to the timeliness of traffic data, people are more inclined to choose the latest data because the old historic data often has failed to provide useful information. Specifically, in Waze,[3] the validity of the information reported about an incident is only for a while [30]. Besides, the limited resource of vehicles makes it difficult to assist drivers in decision-making by analyzing historical data. Therefore, the older the data, the less value the data can provide. For example, two hours ago, congestion occured on a certain road due to a traffic accident. In this case, the value of this information may be limited as the traffic jam may have recovered. As a result, people tend to pay more attention to real-time traffic data.

Insight 2: The ancient historical data usually has a limited contribution to real-time decision-making compared to real-time data. The real-time information is usually more significant for drivers on the road.

B. System Overview

1) Social-Based Data Reduction: Based on Insight 1, we propose the data reduction approach to reduce storage cost for DAG-based blockchain used in VSNs as shown in Fig. 3. We adopt DAG-based blockchain for VSNs because of its high throughput, which is suitable for VSNs with a rapid generation of data. Different from the block structure of blockchain, DAG-based blockchain adopts transaction as vertex of graph without packing transactions to blocks. The fine-grained transaction structure improves the efficiency of blockchain and facilitates the data management of VSNs. Each piece of data is included in a transaction. For simplicity, the terms transaction and data are used interchangeably in this paper.

Since nodes of blockchain in VSNs mainly care about the data of interest and have no interest in irrelevant data, the nodes only need to store relevant data (i.e., data on topics of concern) to save storage space. Taking the topic group I in Fig. 3 as an example, assuming that transactions numbered 1, 2, 4, 5, 7, 11 and 12 contain data for topic I, thus the node in topic group I only needs to store these relevant transactions while other irrelevant transactions are ignored for reducing storage capacity requirement. Meanwhile, each node can join multiple topic groups to receive information from different interested groups, such as node *c*.

2) Generation and Broadcast of New Transactions: As a vehicle investigates some valuable information about a certain topic, the driver can issue a transaction containing the information to blockchain for the convenience of others. In LDV, the generation of new transaction need to satisfy the *Proof of Work* (POW), which is effective to avoid the spam information and resist to sybil attack. Specifically speaking, the following cryptographic puzzles (i.e., Formula 1) need to be fulfilled in the calculation of POW.

$$\text{Hash}\ \langle \text{transaction, nonce} \rangle < \text{target}. \tag{1}$$

[3]https://www.waze.com/

The *nonce* field represents a random number that can satisfy the puzzle and the *transaction* field represents the rest of components in transaction including hash of previous transactions, data, signature, etc. The difficulty of POW is set to small enough that can be accepted by resource-constrained vehicles in our system. And it can be adjusted dynamically by setting different *target* value.

After the POW of new transaction is completed, the new transaction consisted of valuable information can be issued and further broadcast. To be noticed, the new issued transaction is valid until it achieves the consensus among the vehicles. The consensus process is discussed in Section III-B4. The data inside the valid transaction can provide drivers with useful information to plan their journeys. As the vehicular nodes receive new transactions, vehicles can selectively store related transactions locally according to its interests. Compared to storing the entire data of blockchain, this storage mechanism that only store relevant data can reduce the storage overhead largely.

3) The Roles in LDV: Before discussing the consensus of LDV, we first give an introduction to the roles in LDV design. According to the different functions of nodes, LDV includes two categories of nodes.

- *Normal node:* In addition to the broadcast and verification of new transactions, the normal node is also responsible for the data management such as providing storage service for data in VSNs. The normal node can be further divided into two subclasses depending on the type of device.
  - Vehicular node: Vehicular nodes refer to general vehicles (e.g., cars, buses). These nodes are usually highly mobile, whose locations are always changing dynamically.
  - Road side unit node (RSU node): Different from the mobile characteristic of vehicular node, the location of RSU node is always fixed and stable relatively (e.g., traffic lights).
- *Monitoring node:* Apart from the duties of normal nodes, the monitoring node is the regulator of blockchain in each topic group. It plays important roles in the process of consensus. These nodes can be served by transportation departments owing to its authority. Besides, the transportation departments are able to access accurate traffic information in time through surveillance cameras.

4) Verification and Consensus of New Transactions: After the node receives the new transaction from network, the node will verify the validation of this transaction. The process of verification includes the check of signature and the validation of data inside the transaction. After the completion of verification, the new transaction can be stored locally and broadcast further to other nodes for verification by other nodes.

For the stake of simplicity, we first discuss the consensus in each topic group. When the validity of a new transaction is verified by a node, the node can issue other transactions to the network by refering to these valid transactions. The reference relationship indicates that other nodes agree with the information in this new transaction. For example, in topic group C of Fig. 4, the reference relationship between transaction 5 and 8 indicates transaction 8 agrees with the validation of transaction 5. The final

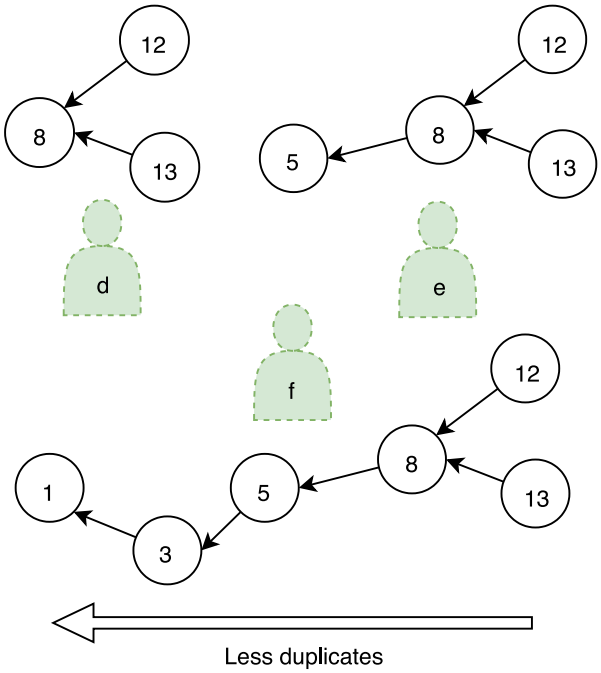

Fig. 4. The older history with fewer duplicates.

TABLE I
NOTATIONS

Symbol	Meaning
CW_i	The cumulative weight of transaction i
Γ_i	The transaction set in which cites the transaction i directly or indirectly
τ	A transaction belong to the transaction set Γ
ω_τ	The weight of transaction τ
tx	The transaction
$Citation_i$	The citation of transaction i

validation of a new transaction is determined by the *cumulative weight* [18] of the transaction, which is proportional to amounts of nodes that agree with this transaction. The cumulative weight of transaction *i* is defined as follow:

$$CW_i = \sum_{\tau \in \Gamma_i} \omega_\tau \left(\Gamma_i = \{tx | tx \in \text{Citation}_i\}\right) \tag{2}$$

The meaning of symbol in Formula 2 is described in Table I. As shown in Formula 2, the cumulative weight of transaction *i* is defined as the weight sum of transactions citing the transaction *i*. The more computational power it consumes in POW, the higher the weight of the transaction. The greater the cumulative weight of the transaction, the more likely the transaction is valid and final because the computational power consumed is larger. When the cumulative weight of transaction comes to a certain level, the transaction is believed to valid. Additionally, the cumulative weight is one of the determining factor in distinguishing between honest transactions and illegal transaction issued by malicous nodes. In the normal case, the transaction issued by honest nodes will be verified and cited by other honest nodes, therefore, the

cumulative weight will keep increasing and larger than illegal transactions. Therefore, the transaction issued by honest nodes will be valid finally and provide useful information for other nodes.

To prevent malicious nodes from destroying the entire system when the total computational power of malicous nodes is larger than honest nodes, we introduce the monitoring nodes mentioned above to solve this security problem. Meanwhile, we assume the total computational power of malicious nodes is less than 50%. The transactions cited by a transacton issued by monitoring nodes is valid and the calculation of cumulative weight is unnecessary for it, because the monitoring node is honest and it can distinguish the authenticity of the information inside transactions through surveillance cameras. Once the validation of two conflict transaction is difficult to confirm by both cumulative weight and monitoring nodes in a topic group, in this case, the consensus of these conflict transactions need to be carried out by the entire network by combining the data and the nodes in other topic group. But the possibility of this situation is so low that can be ignored. In the normal case, the consensus of transactions is carried out in each topic group in order to reduce the communication overhead and the broadcast time in VSNs with a poor communication capabilities.

5) Storage Cost of Large-Scale Topic Group: Until now, the design mentioned above can well deal with the storage overhead challenges of blockchain in VSNs when the scale of the topic group is uniform. However, as the number of transactions in each topic group increases, especially the hot topic, the storage overhead is also unacceptable once the size of these transactions in the hot topic group come to a high level.

Challenge 1: How to deal with the storage cost problem in a hot topic group with a large number of transactions?

C. Data Reduction Within a Group

To address the Challenge 1, we further present the data reduction approach in a topic group based on the Insight 2. Through the in-depth analysis of VSNs, we learn that the drivers are more likely to choose the latest information when making decision on the travel route. The fresh data often provides more valuable information compared to historical data in the fast changing traffic scenario. In addition, the older transactions have high cumulative weights, which are difficult to be tampered with. Besides, the storage cost of historical data with a large amount of data is so expensive for vehicles with a limited resource.

1) Overview of Data Reduction Approach: As a result, inspired by [28] focusing on data reduction on chain-based blockchain, we reduce the number of duplicates of historical data on DAG-based blockchain instead of deleting historical data of all nodes directly. Meanwhile, the remaining copies are significant to data integrity and traceability of blockchain. And the historical data can also be used for data analysis to discover potential value in VSNs. The overview of enhanced design is depicted in Fig. 4, the older historical data contains fewer duplicates in blockchain network. Taking the topic group II in Fig. 3 as an example, as shown in Fig. 4, the oldest historical transactions numbered 1 and 3 have only one copy around these nodes in this group while the transaction numbered 5 has two copies. Correspondingly, the latest transactions numbered 8, 12, 13 are stored on each node. It further reduces the storage overhead by reducing the number of duplicates of unnecessary historical data.

Although reducing duplicates of historical data can save storage space, the few data replicas bring a serious effect to the data integrity and security of blockchain. A good data allocation strategy is not only conducive to data reduction, but also helpful to data integrity.

Challenge 2: How to allocate the amount and the storage location of historical replicas reasonably?

2) Allocation of Duplicates: To guarantee the data integrity and security of blockchain, we give the allocation strategy of replicas in this section. In the normal blockchain system, each full node need to store the full copies of data to ensure integrity and security of data. But in LDV, to save storage space, only the latest data needs to be stored on each node due to its low cumulative weight and high value. The historical data can be pruned for saving storage space. As a result, only a part of nodes need to store the full data. Each node can adjust the range of historical data to be stored according to their demands and owned resources. To prevent data loss caused by machine failure, the full duplicates including historical data need to be stored on monitoring nodes in each topic group. The monitoring nodes are usually the server machines with large storage capacity, which belongs to traffic control department. Besides, the data stored in monitoring nodes is beneficial to data security, which is also effective to prevent these small amounts of duplicates of historical data from being controlled by malicious nodes.

D. Complements of Design

At present, the above-mentioned design is able to provide a lightweight blockchain system for VSNs. However, there remain some challenges, such as data query of cross-group. Therefore, we advance the design of LDV in terms of data integrity and query in this subsection.

1) Data Integrity of a Single Group: As discussed in Section III-C, increased nodes and transactions will further deteriorate the storage problem, especially in the large-scale groups. On the contrary, the decrease in number of nodes will affect the integrity of data because each node only stores relevant data in our design. More seriously, when no one pays attention to a topic, the data on that topic will be at risk of loss.

Challenge 3: How to ensure the data integrity of the topic groups with a small number of nodes?

Fortunately, the RSU nodes of VSNs are very useful for ensuring the data integrity of groups with a small amount of nodes. In general, the RSUs are highly common on the roads and are a part of road such as the traffic light. Naturally, the RSU node is a member of topic group about that road. As a result, we can use the stability and university of RSU nodes to store data for frosty topic groups. For example, a topic about

Road A is less concerned. In this way, the RSUs around this road will join this topic group automatically, which can be used to store information about that topic in order to avoid data loss. Furthermore, to avoid the data loss incurred by the failure of RSU nodes in one road, the RSU nodes near that road will join the topic group of that road automatically when the number of RSU nodes drops to a certain threshold. The threshold can be set flexibly according to different situations.

2) Data Query of Cross-Group: In addition to information acquisition of topics of interest, sometimes it is also significant to get information on other topics. Apart from joining this topic group to retrieve data, querying the data of this topic directly is also a way for those who do not want to join this topic and just require the data temporarily. As stated in III-A, because the trips of commuters are usually stable, thus they just need to join the topics about their trips. When the commuters have requirements for data in other topic groups, they can query this data directly from the nodes in other topic groups.

Challenge 4: How to query data from other topic groups?

As aforementioned, only the relevant data is stored locally. If the commuters need the data of other groups, the commuters can issue a transaction including the data request of relevant groups and broadcast it out to wait for the response of data. When a node in that group receives the request, it returns the request data to the commuter. Nevertheless, the correctness of data obtained from other groups is questionable. Therefore, ensuring the validity of data is a challenge for data query of cross-group. As described in III-B4, the validity of transaction is determined by the cumulative weight of it. Therefore, we utilize the cumulative weight to ensure the validity of data received from other groups. As the cumulative weight is attached to the requested data, the commuters can verify the validation of the returned data easily through the cumulative weight. To prevent the request data and its cumulative weight from being tampered with by malicious nodes, the data query request of the topic group will be responsed by the monitoring nodes to guarantee the correctness and security of the requested data.

IV. Evaluation

We have implemented a prototype DAG-based blockchain system called DAGChain for VSNs, and LevelDB[4] is taken as the underlying database of DAGChain. DAGChain adopts DAG structure instead of chain structure for efficient parallelism. The transaction is taken as the vertex of graph to achieve more efficient data processing inside transactions. Based on DAGChain, we implement LDV to evaluate the effect of the proposed data reduction approach, and conduct several experiments on the servers to simulate the situation of VSNs. The servers are used to simulate the vehicular nodes to send, broadcast and store data. Each machine has two 24-core Intel Xeon 8260 2.4 GHz CPUs, 128 GB DRAM, and 7.2 TB HDD, with CentOS 7.6 operating system. To ensure the uniformity of storage in different nodes with different interested topics, the size of data inside all transactions is set to same.

[4]https://github.com/syndtr/goleveldb

TABLE II
The Number of Transaction in Different Topics and Its Followers

Topics	The number of transactions	Followers
topic1	100	Node A, F
topic2	500	Node B, F
topic3	1000	Node C, F
topic4	2000	Node D, F
topic5	6000	Node E, F

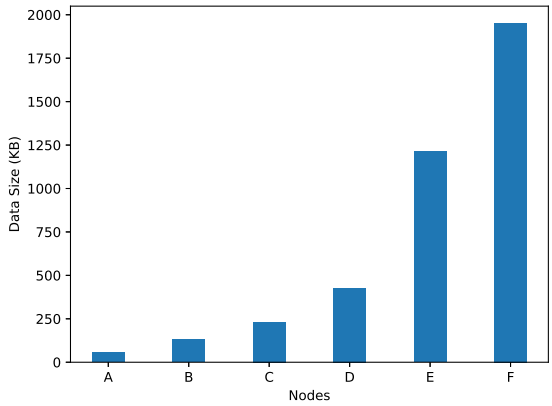

Fig. 5. The data size of different nodes using social-based data reduction approach.

A. Effects of Social-Based Data Reduction Approach

We first evaluate the social-based data reduction approach described in III-B1. Each node only needs to store the data of interested topics. A node can join multiple topic groups freely to get the interested information, and leave freely if it is no longer interested. We first study the storage cost of nodes with one interested topic, and the cost of nodes with multiple interested topics will be analyzed in IV-C. The storage space used by different nodes is measured by generating different numbers of transactions in different topic groups. For the sake of simplicity, the transaction only contains information that belongs to one topic. The number of transaction in different topic groups and the member of groups are listed in Table II. Taking the data of last line as an example, *topic5* has 6,000 transactions containing the data about this topic, and node E, F are interested in this topic.

Fig. 5 shows that the storage cost of different nodes with different topics interested. In particular, node F does not adopt data reduction approach, i.e., it follows all the topics, which can be considered as a full node in normal blockchain system. As we can see from the experimental results, compared to full node F, other nodes have less storage cost when only joining the topic group of interest. Node A consumes the minimal storage space due to the topic it follows has the minimum transations. Specifically, it saves 97.13% storage space compared to node F, which is beneficial to resource-constrained VSNs.

B. Effects of Data Reduction Within a Single Group

To evaluate the effect of storage reduction described in Section III-C, we reset the experimental setting. Specifically, the blockchain of this experiment only contains one topic to

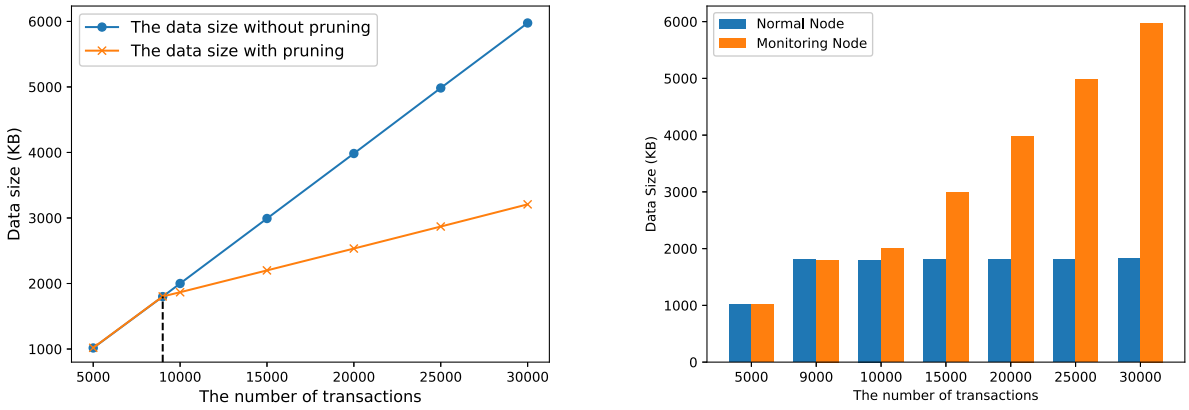

Fig. 6. The data size within a topic group. (a) The average data size. (b) The comparison of data size of normal node and monitoring node.

verify the consumed storage space within a topic group, using the historical data pruning method. In detail, each normal node only needs to store recent data, and the ancient historical data can be pruned to save storage space. Consecutive transactions containing the fixed size data about this topic are generated and stored in six nodes to measure the average storage cost of the six nodes. The six nodes are numbered as A, B, C, D, E and F. Specifically, two of the nodes serve as monitoring nodes (e.g., node E and F) to keep the full duplicates of historical data for the integrity introduced in Section III-C2. The data reduction strategy of historical data is pruning data before a certain days, which can be adjust according to the requirement of each node. In this experiment, for simplicity, we set it to 3 days for all the four normal nodes. We assume that 3,000 transactions are issued every day. More specifically, we prune the transactions more than 9,000 transaction away from the latest transaction and preserve the recently 9,000 transactions to simulate the transactions of 3 days ago in reality.

Fig. 6 demonstrates the storage cost of the pruning method within the topic group. The average data size of all the six nodes is shown in Fig. 6(a). The storage cost without pruning is similar to the counterpart when the number of transactions is low. However, it increases larger and faster when the scale of transactions becomes larger. The storage cost with pruning keeps increasing because the data size of the two monitoring nodes is increasing all the time. Fig. 6(b) compares the data size of normal node and monitoring node respectively. The size of monitoring node keeps increasing. By contrast, the size of a normal node remains unchanged, as it only persists the recently data, which confirms the efficiency of the historical data pruning approach in large-scale groups.

C. Scalability of LDV

To analyze the scalability of LDV, we conduct several experiments to evaluate the storage space influenced by different number of transactions and interested topics. Specifically, to evaluate the storage cost of nodes with multiple interested topics, the total amounts of transactions is uniform in each group of experiments and the specific numbers are listed in Table III.

TABLE III
THE NUMBER OF INTERESTED TOPICS AND CONTAINED TRANSACTIONS

#Total transactions	60,000				
#Interested topics	1	2	3	4	5
#Txs per topic	60,000	30,000	20,000	15,000	12,000

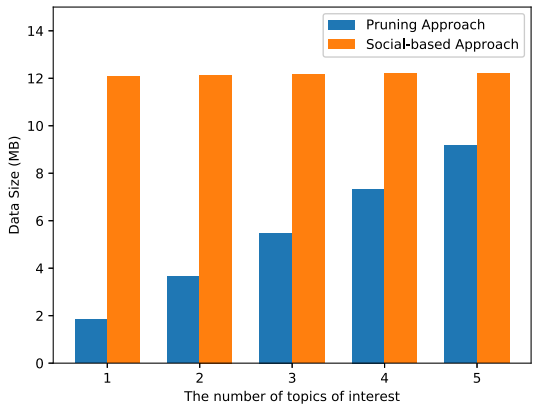

Fig. 7. Consumed storage space affected by the number of topics.

Six nodes are deployed to test the consumed storage space of different number of topics. Additionally, in another experiment about the variation of the number of transaction, for fairness, the number of transactions in each topic group is the same and the number of topics is set to 3. Three nodes without data reduction, with social-based reduction and with social-based combining pruning approach are running respectively to measure the storage cost of different methods. Similarly, the number of historical data stored in the two experiments is set to 9,000.

Fig. 7 depicts the used storage space varying with the number of interested topics. Since the total number of transactions is the same, the storage overhead with different topics of interest is similar. On the contrary, the total data size is increasing when the number of interested topics increases. That is because the pruning method runs in each topic group. Although the storage cost with pruning method is unchangeable in each group, it increases as the number of interested topics increases. Fig. 8 shows the consumption of storage space influenced by the number of transactions. As shown in the result, the LDV with the

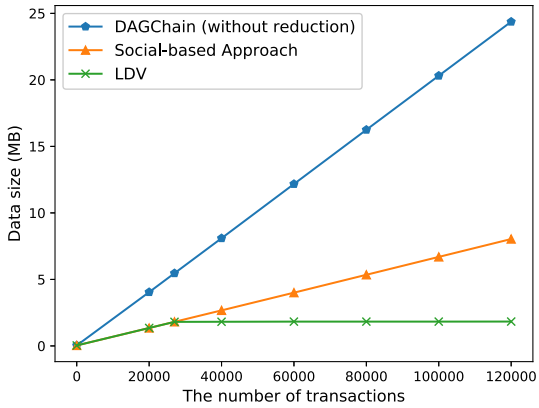

Fig. 8. The storage cost of different methods.

combination of social-based and pruning approaches performs best. The data size increases largely when both the two data reduction approaches are not adopted. Meanwhile, the consumed storage space of LDV keeps the same after 27,000 transactions are issued, which achieves good scalability.

V. Discussion and Future Work

We will discuss the robustness and communication efficiency of LDV in this section, which are not mentioned in the design. In addition, we give several research points that can be studied in future work.

A. Robustness

The robustness of a system is of significance to ensure the availability of service, especially the online service for VSNs. Although the RSU nodes can dedicate storage for the topic groups with a few nodes to avoid the data loss, the potential failure of RSU nodes (e.g., machine downtime) may still result in data loss. With the prosperity of cloud services, the cloud servers can be used to back up data for VSNs to prevent data loss. To be specific, the wired communication module, which is common for RSUs, is utilized to upload data to cloud servers periodically to avoid losing data. The better fault tolerance mechanism is leaved to our future work to achieve good robustness.

B. Communication

As aforementioned, the data transmission of VSNs relies on the underlying ad hoc network. However, the poor communication capability of ad hoc network may be a bottleneck for the development. Specifically, in our design, the social relationship of topic groups is a virtual link relying on the physical link to communicate. It is possible that the physical distance between two nodes with the direct social relationship is very long. In this case, the data exchange between these nodes is difficult. The emergence of fifth-generation mobile networks may alleviate this problem when used in VSNs due to its high speed. An efficient routing algorithm that takes social relationship in VSNs and the beneficial features of blockchain into consideration may be another possible solution. Since the main focus in the paper is to reduce the storage cost of blockchain used in VSNs, we leave the above challenges to our future works.

VI. Conclusion

In this paper, we design a lightweight blockchain system for VSNs based on the DAG structure. To be specific, a social-based data reduction approach on the whole network, and a pruning method within a single group are proposed to reduce the storage cost of vehicular nodes, respectively. To ensure the data integrity and query ability of cross-group, we further present the relative mechanism in our design. The prototype of LDV has been implemented to evaluate the effect of data reduction. The experimental results demonstrate that LDV can save 97.13% storage space and is scalable.

References

[1] X. Wang, H. Zhang, L. Wang, and Z. Ning, "A demand-supply oriented taxi recommendation system for vehicular social networks," *IEEE Access*, vol. 6, pp. 41 529–41 538, 2018.
[2] S. Smaldone, L. Han, P. Shankar, and L. Iftode, "RoadSpeak: Enabling voice chat on roadways using vehicular social networks," in *Proc. 1st Workshop Social Netw. Syst.*, 2008, pp. 43–48.
[3] D. Camara, C. Bonnet, and F. Filali, "Propagation of public safety warning messages: A delay tolerant network approach," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2010, pp. 1–6.
[4] X. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.
[5] A. Rahim *et al.*, "Vehicular social networks: A survey," *Pervasive Mobile Comput.*, vol. 43, pp. 96–113, 2018.
[6] Q. Yang and H. Wang, "Towards trustworthy vehicular social network," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 42–47, Aug. 2015.
[7] D. Kushwaha, P. K. Shukla, and R. Baraskar, "A survey on sybil attack in vehicular Ad-Hoc Network," *Int. J. Comput. Appl.*, vol. 98, no. 15, pp. 31–36, 2014.
[8] A. M. Vegni, V. Loscrí, and P. Manzoni, "Analysis of small-world features in vehicular social networks," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf.*, 2019, pp. 1–2.
[9] V. R. Neto, D. S. Medeiros, and M. E. M. Campista, "Analysis of mobile user behavior in vehicular social networks," in *Proc. 7th Int. Conf. Netw. Future*, 2016, pp. 1–5.
[10] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf/
[11] X. Xu *et al.*, "A taxonomy of blockchain-based systems for architecture design," in *Proc. IEEE Int. Conf. Softw. Architecture*, 2017, pp. 243–252.
[12] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, no. 2014, pp. 1–32, 2014.
[13] E. Androulaki *et al.*, "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proc. 13th EuroSys Conf.*, 2018, pp. 30:1–30:15.
[14] K. Karlsson *et al.*, "Vegvisir: A partition-tolerant blockchain for the Internet-of-Things," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst.*, 2018, pp. 1150–1158.
[15] K. J. O'Dwyer and D. Malone, "Bitcoin mining and its energy footprint," in *Proc. 25th IET Irish Signals Syst. Conf. China-Ireland Int. Conf. Inf. Commun. Technologies*, 2014, pp. 280–285.
[16] A. Beall, "Bitcoin mining uses more energy than ecuador–but there's a fix," New Scientist, 2017, [Online]. Available: https://www.newscientist.com/article/2151823-bitcoin-mining-uses-more-energy-than-ecuador-but-theres-a-fix/.
[17] E. K. Kogias, P. Jovanovic, N. Gailly, I. Khoffi, L. Gasser, and B. Ford, "Enhancing bitcoin security and performance with strong consistency via collective signing," in *Proc. 25th USENIX Secur. Symp*, USENIX Association, 2016, pp. 279–296.
[18] S. Popov, "The tangle," 2018. [Online]. Available: https://iota.org/IOTA_Whitepaper.pdf
[19] A. Churyumov, "Byteball: A decentralized system for storage and transfer of value," 2016. [Online]. Available: https://byteball.org/Byteball.pdf

[20] C. LeMahieu, "Nano: A feeless distributed cryptocurrency network," 2018. [Online]. Available: https://nano.org/en/whitepaper
[21] X. Dai, J. Xiao, W. Yang, C. Wang, and H. Jin, "Jidar: A jigsaw-like data reduction approach without trust assumptions for bitcoin system," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst.*, 2019, pp. 1317–1326.
[22] Z. Yang, H. Yu, J. Tang, and H. Liu, "Toward keyword extraction in constrained information retrieval in vehicle social network," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4285–4294, May 2019.
[23] G. Xu, H. Li, Y. Dai, K. Yang, and X. Lin, "Enabling efficient and geometric range query with access control over encrypted spatial data," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 4, pp. 870–885, Apr. 2019.
[24] H. Ren, H. Li, Y. Dai, K. Yang, and X. Lin, "Querying in Internet of Things with privacy preserving: Challenges, solutions and opportunities," *IEEE Netw.*, vol. 32, no. 6, pp. 144–151, Nov./Dec. 2018.
[25] D. Liao, H. Li, G. Sun, M. Zhang, and V. Chang, "Location and trajectory privacy preservation in 5G-enabled vehicle social network services," *J. Netw. Comput. Appl.*, vol. 110, pp. 108–118, 2018.
[26] B. Ying and A. Nayak, "A distributed social-aware location protection method in un-trusted vehicular social networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6114–6124, Jun. 2019.
[27] W. Jiang, H. Li, G. Xu, M. Wen, G. Dong, and X. Lin, "PTAS: Privacy-preserving thin-client authentication scheme in blockchain-based PKI," *Future Gener. Comput. Syst.*, vol. 96, pp. 185–195, 2019.
[28] D. Jia, J. Xin, Z. Wang, W. Guo, and G. Wang, "Elasticchain: Support very large blockchain by reducing data redundancy," in *Proc. Asia-Pacific Web Web-Age Inf. Manage. Joint Int. Conf. Web Big Data*, 2018, pp. 440–454.
[29] Z. Xu, S. Han, and L. Chen, "Cub, a consensus unit-based storage scheme for blockchain system," in *Proc. IEEE 34th Int. Conf. Data Eng.*, 2018, pp. 173–184.
[30] I. Lequerica, M. G. Longaron, and P. M. Ruiz, "Drive and share: Efficient provisioning of social networks in vehicular scenarios," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 90–97, Nov. 2010.

Wenhui Yang (Student Member, IEEE) received the B.S. degree from the School of Computer Science and Engineering, Northeastern University, Shenyang, China, in 2018. He is currently working toward the M.S. degree with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His current research interests include blockchain system and distributed system.

Xiaohai Dai (Student Member, IEEE) received the M.S. degree from the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2017. He is currently working toward the Ph.D. degree with the School of Computer Science and Technology, HUST. His current research interests include blockchain and distributed system.

Jiang Xiao (Member, IEEE) received the B.Sc. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2009 and the Ph.D. degree from Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, in 2014. She is currently an Associate Professor with the School of Computer Science and Technology, HUST, Wuhan, China. She has been engaged in research on blockchain, distributed computing, big data analysis and management, and wireless indoor localization. Her awards include Hubei Dawnlight Program 2018, CCF-Intel Young Faculty Research Program 2017, and best paper awards from IEEE ICPADS/GLOBECOM 2012.

Hai Jin (Fellow, IEEE) received the Ph.D. degree in computer engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1994. He received German Academic Exchange Service fellowship to visit the Technical University of Chemnitz in Germany in 1996. He was with the University of Hong Kong between 1998 and 2000, and as a Visiting Scholar with the University of Southern California between 1999 and 2000. He is a Cheung Kung Scholars Chair Professor of computer science and engineering with the Huazhong University of Science and Technology, the Chief Scientist of ChinaGrid, the largest grid computing project in China, and the Chief Scientist of National 973 Basic Research Program Project of Virtualization Technology of Computing System, and Cloud Security. He has coauthored 22 books and published more than 800 research papers. His research interests include computer architecture, virtualization technology, cluster computing and cloud computing, peer-to-peer computing, network storage, and network security. He received the Excellent Youth Award from the National Science Foundation of China in 2001. He is a fellow of the CCF and a member of the ACM.