Secured Data Sharing and Storage System for Intelligent Transportation System via Lightweight Blockchain

Jiarui Zhang*, Wei Lin†, Xiaojun Shang†, Yiming Zeng‡, Yuanyuan Yang*
* Department of Electrical and Computer Engineering, Stony Brook University, USA
{jiarui.zhang.2, yuanyuan.yang}@stonybrook.edu
† Department of Computer Science and Engineering, University of Texas at Arlington, USA
wxl0991@mavs.uta.edu, xiaojun.shang.@uta.edu
‡ School of Computing, Binghamton University, USA
yimingz@binghamton.edu

***Abstract*—Secure data sharing and storage are critical to realizing the ultra-high safety and efficiency promised by intelligent transportation systems (ITS). However, ensuring data integrity under stringent resource constraints of ITS remains an open challenge. Existing approaches either incur substantial computational and transmission overhead or demand resources beyond what ITS environments can provide. To address this dilemma, we propose a lightweight blockchain-based data sharing and storage framework tailored for ITS. The system features a distributed, asynchronous reputation mechanism that enables trustworthy data evaluation and dissemination without burdening critical computing and communication paths. Additionally, a resource-efficient blockchain storage layer ensures low-latency, tamper-resistant access to locally shared data. Extensive simulations demonstrate that our approach outperforms existing solutions in both integrity assurance and resource efficiency.**

***Index Terms*—ITS, Reputation Mechanism, Blockchain**

I. Introduction

Intelligent Transportation Systems (ITS) are revolutionizing modern transportation infrastructure by enhancing safety, mobility, and operational efficiency through the integration of advanced information, computing, and communication technologies [1], [2]. The global ITS market is projected to grow from 50.7 billion in 2024 to 70.7 billion by 2029 [3]. A fundamental enabler of ITS is the pervasive data sharing among connected transportation entities, including vehicles, pedestrians, roadside units, and edge/cloud servers. This data exchange enables extended sensing capabilities, mitigates occlusion, enhances environmental awareness, and facilitates multi-vehicle coordination [4], [5]. These capabilities form the foundation of a wide range of ITS applications, such as collaborative perception, infrastructure-assisted localization, cooperative navigation and routing, adaptive signal control, and road infrastructure planning [6], [7].

The benefits of data sharing in ITS critically depend on the integrity of the shared information, which is inherently vulnerable due to the distributed and collaborative nature of ITS environments [8]. First, data can be compromised at the source, for example, by a malicious participant within the network seeking personal gains. Many Connected and Autonomous Vehicles (CAVs) rely on shared road condition data for real-time routing decisions [9]. As such, falsified or misleading information, such as fabricated accident reports, can be disseminated to divert traffic and optimize routes for the attacker, thereby gaining unfair advantages in travel time or energy consumption. Moreover, data integrity is also at risk during storage [10]. Owing to bandwidth limitations and privacy concerns, ITS data are often stored temporarily or permanently in distributed local repositories with weaker security guarantees than centralized public cloud infrastructures, increasing their susceptibility to tampering and unauthorized modifications [11]. Corrupted local data can have particularly damaging effects when used in critical downstream applications such as training autonomous driving models, potentially leading to unsafe behaviors and severe security vulnerabilities.

Various countermeasures have been proposed to enhance the integrity of shared data in ITS. Lightweight network security mechanisms are typically deployed at intermediate points along the communication path [12] or at receiving ITS entities [13] to filter data from suspicious sources with minimal overhead. However, these methods are ineffective against data manipulation conducted by legitimate yet malicious participants within the collaborative network [9]. To address this challenge, advanced detection and verification techniques have been developed to eliminate falsified information in collaborative ITS environments [14]. While effective in controlled settings, these techniques often introduce latency and computational overheads that are incompatible with the real-time requirements of many ITS applications and may exceed the capabilities of resource-constrained devices like vehicles, roadside units, and pedestrian devices.

At the same time, robust protection mechanisms have been established for securing data storage in centralized cloud infrastructures [15]. However, these solutions are generally not well suited for edge-based storage, which is characterized by decentralized architectures and limited computational and security resources [16]. Consequently, the low security of edge storage undermines the reliability of prior verification efforts, as data can still be altered after validation. Transmitting locally stored data to public clouds for centralized verification and processing enhances data integrity but at the cost of significant communication overhead and computational burden, due to the sheer volume of ITS data.

As a complementary approach to existing data integrity solutions in ITS, we propose a novel blockchain-based system for data sharing and storage. This framework is built

upon Roadside Infrastructures (RSIs) equipped with sufficient communication, computing, and storage capabilities. Representative RSIs include intelligent traffic signal poles, 5G base stations, and micro edge data centers, which have been widely adopted as the communication and computing backbone of modern ITS [17]. Within our framework, RSIs function as local data-sharing hubs, data contributors within their communication range upload shared data to them, while data users retrieve data from them. In addition, RSIs are responsible for storing the shared data and preserving its integrity until it is uploaded to the cloud.

The primary contribution of our data sharing and storage system lies in addressing the fundamental tension between ensuring data integrity, defined as both the trustworthiness of shared data and the reliability of its evaluation, and satisfying stringent resource constraints in heterogeneous ITS environments, a challenge that remains largely unresolved in existing solutions [18], [19]. To overcome this, our framework departs from conventional methods that enforce integrity checks along the critical path of data transmission and processing, often incurring significant overhead, by delegating integrity validation to data consumers through a distributed, asynchronous reputation mechanism. Each user independently evaluates the quality of received data based on task-specific metrics, and these decentralized assessments collectively determine the reputation of data contributors. Data from low-reputation sources is filtered or blocked based on user-defined strategies, thereby preserving data integrity without interrupting time-sensitive ITS operations. Notably, the reputation system can operate in conjunction with traditional on-path integrity mechanisms, reducing their burden by ensuring both trustworthy data and trustworthy data sources.

In parallel, our system introduces a lightweight blockchain-based storage layer to ensure the immutability of locally shared data across distributed nodes, while maintaining high resource efficiency and enabling low-latency access for real-time ITS applications. While prior studies have explored blockchain-based data management in ITS settings [11], [20], they often neglect the stringent computational and bandwidth constraints inherent to CAVs, limiting their practical deployment. Our design addresses this gap by incorporating a novel off-chain storage mechanism that supports high-integrity data sharing across the network while significantly reducing on-chain computational and storage overhead.

The rest of this paper is organized as follows. In Section II, we formulate the problem and model. In Section III, we introduce the reputation mechanism for ITS. In Section IV, we proposed the lightweight blockchain design with RSIs and reputation mechanism. Simulation results are presented in Section V. Section VI briefly reviews the related work, and the last section concludes this paper.

II. Problem Statement and Model

We focus on the integrity challenges of data sharing in ITS. Existing solutions still face several unresolved issues. First, they often fail to filter out malicious participants. Users or vehicles that act dishonestly should be excluded from future interactions with honest participants. Second, computing and network resources are not efficiently utilized. Due to the heterogeneity of devices in ITS, resource capabilities vary significantly, making it essential to support participation from users with limited resources. Additionally, users may need to prioritize critical tasks such as driving, which further limits the resources available for data-related operations. Therefore, we aim to ensure integrity of data sharing in ITS while maintaining low resource consumption.

Mobile users in ITS, such as vehicles and pedestrians, continuously share data within the ITS. Each Internet-connected device used by these users can be uniquely identified or authenticated, for example, through a vehicle's VIN or a mobile phone's IMEI number. For simplicity, we use the term user equipment (UE) to refer to all devices participating in data sharing, including CAVs and mobile phones. The set of all UEs is denoted as $\mathcal{U} = U_1, U_2, ..., U_i, ...$, and RSIs are denoted by the set $\mathcal{S} = S_1, S_2, ..., S_i, ...$ UEs share data through the network established by these RSIs. The shared data set is denoted by $\mathcal{T} = T_1, T_2, ..., T_i, ...$, where each data item is represented as a transaction in the form $T_i = (p_i, t_i, d_i)$. Here, p_i denotes the proposer of transaction T_i, t_i is the timestamp when T_i is proposed, and d_i contains the actual data content. In the remainder of this paper, we use the term transaction to refer to both the shared data and its associated information.

The goal is to improve data quality while reducing resource consumption. To address this, we propose a blockchain-based reputation system. Each UE can evaluate the reputation of other UEs, and these reputations are recorded on a blockchain maintained by UEs, RSIs and cloud servers. UEs consult the reputations stored on the blockchain to make informed decisions when selecting data from trusted peers.

III. ITS Reputation Mechanism

In this section, we introduce the reputation mechanism for ITS. Reputation serves as an effective approach to evaluate the behavior of UEs. We first describe the method for computing the reputation of transactions, and then explain how these values are used to infer the reputations of UEs.

A. Trust Management Solution

Currently, various machine learning methods have been applied to address trust management and security issues. However, due to the limited computing resources in the ITS environment, these methods are not proper, as training and deploying learning models require significant computational overhead.

Reputation is an effective solution for verifying the behavior of UEs in ITS. It offers two main advantages. First, reputation-based methods are more cost-efficient than machine learning approaches. They avoid the high training and storage overhead associated with learning models, requiring only a limited amount of historical evaluation data. Second, reputation methods offer better scalability. They can be deployed across a wide range of devices with varying capabilities, making them especially suitable for UEs with limited computing and network resources.

B. Design Goal

The primary objectives of ITS focus on two key aspects: traffic safety and traffic efficiency [21]–[23]. Simply speaking, the development of ITS aims to improve traffic efficiency while maintaining traffic safety. However, these two objectives often involve trade-offs. For example, driving at lower speeds generally improves traffic safety but may reduce traffic efficiency, while higher speeds can increase traffic efficiency at the cost of greater collision risk.

The trade-off between traffic safety and traffic efficiency creates opportunities for malicious users to prioritize their own efficiency at the expense of others' safety. For example, a user might deliberately conceal accident information on road A to redirect traffic from road B to road A. This manipulation reduces congestion on road B, improving their own traffic efficiency, while increasing the risk of accidents for vehicles diverted to road A. In this case, the safety of others is compromised to benefit the efficiency of an individual.

To prevent malicious UEs from reporting false information, the reputation of the data they provide serves as a key factor in evaluating the reputation of the UEs. In general, UEs that consistently report accurate traffic information earn higher reputations than those that provide inaccurate data.

C. Transaction Reputation

1) Personal Reputation: UEs evaluate transactions by expressing their own opinions, referred to as personal reputations. Each UE that receives a transaction can independently assign a reputation score based on its evaluation strategy. The system does not enforce a fixed reputation criterion, allowing UEs to adopt their own methods for assessment. The evaluation is represented by a numerical value, where p_{ij} denotes the personal reputation assigned by UE U_i to transaction T_j. A higher value of p_{ij} indicates a higher level of trust in the transaction.

2) Standardization: To ensure fairness in aggregation, personal reputation scores must be standardized before they are combined. Since UEs may adopt different reputation algorithms [24], [25], the resulting values can vary widely in scale. Without standardization, UEs could influence the system by assigning excessively high reputation values, skewing the aggregated results.

Standardization is applied by restricting the total reputation values assigned by each UE as follows [26]:

$$s_{ij} = k \times \frac{p_{ij}}{\sum_j p_{ij}}. \tag{1}$$

Simply speaking, the total weight of reputation values assigned by each UE is normalized to 1, excluding the factor k. Each UE assigns non-negative reputation values to the evaluated transactions such that their sum equals 1. However, since the number of transactions evaluated by each UE may vary, giving every UE the same total weight would diminish the influence of those that contribute more evaluations. To address this imbalance, we adjust the total weight from 1 to k, where k represents the number of evaluations submitted during a specific time period. Consequently, each standardized value s_{ij} is scaled by k.

Note that if a UE gives evaluations that total reputation values is less than k, standardization is not required. This is because UEs do not intentionally assign lower total reputation values to reduce the impact of their evaluations. Instead, lower values generally reflect the UE's real judgmen that the evaluated transactions are of low quality. In such cases, the standardized value remains unchanged, i.e., $s_{ij} = p_{ij}$.

3) Aggregated Reputation: With standardized personal reputation values, RSIs and UEs can compute the aggregated reputation of each transaction by a weighted sum approach [24]. In our reputation mechanism, UEs with higher reputations have more weight during aggregation. The aggregated reputation a_j of transaction T_j is computed as follows:

$$a_j = \frac{\sum_j r_i(t) \cdot s_{ij}}{\sum_j r_i(t)}. \tag{2}$$

Equation 2 is the weighted average reputation value of transaction j. In this equation, $r_i(t)$ is the reputation of UE U_i in the t-th block interval, which would be presented in the following subsection.

D. UE Reputation

Existing works [24], [27] consider multiple factors in UE reputation computation, such as data accuracy, reliability, and driving behavior. Our reputation mechanism focuses on data quality and evaluation accuracy, defining the UE reputation as a combination of transaction reputation and evaluation accuracy.

As previously stated, the reputation of UE U_i during the t-th block interval is denoted as $r_i(t)$. The inclusion of the time variable t reflects the fact that a UE's behavior and performance may vary over time, making reputation a dynamic attribute. Additionally, using block intervals provides a temporal boundary for evaluations, allowing the system to compute reputations based on recent interactions rather than aggregating over the entire history of the network. The t-th block interval refers to the time period after the $(t-1)$-th block has been generated but before the t-th block is created.

Before presenting the equation of UE reputation, we define two metrics related to U_i. The first is the transaction score, denoted as $TS_i(t)$, which is calculated as follows:

$$TS_i(t) = \frac{\sum_j a_j}{|\mathcal{T}(t,i)|}, \forall j \in \mathcal{T}(t,i). \tag{3}$$

$\mathcal{T}(t,i)$ denotes the set of all transactions proposed by U_i during the t-th block interval, and $|\mathcal{T}(t,i)|$ represents the number of transactions in that set. The transaction score $TS_i(t)$ is defined as the average aggregated reputation value of all transactions in $\mathcal{T}(t,i)$. This score reflects the overall quality of data shared by U_i during the given interval. A high $TS_i(t)$ indicates that U_i is behaving honestly and contributing high-quality data. The second metric is the evaluations score, denoted by $ES_i(t)$, and is calculated as follows:

$$ES_i(t) = \frac{\sum_j (a_j - s_{ij})^2}{|\mathcal{E}(t,i)|}, \forall j \in \mathcal{E}(t,i). \tag{4}$$

$\mathcal{E}(t,i)$ contains all evaluations made by U_i during the t-th block interval, and $|\mathcal{E}(t,i)|$ denotes its size. The evaluation score $ES_i(t)$ measures the quality of these evaluations. Specifically, it is defined as the expected value of the squared deviation between the standardized personal reputation s_{ij} and the aggregated reputation a_j. A lower $ES_i(t)$ indicates that U_i's evaluations are closer to the consensus reputation values, reflecting more reliable evaluation behavior."

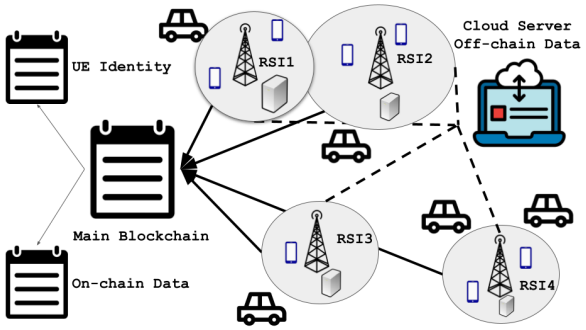
By combining the performance of proposing transactions and evaluating reputations, the reputation of U_i in the t-th block interval, denoted as $r_i(t)$, is defined as follows:

$$r_i(t) = \alpha \times TS_i(t) + \beta \times ES_i(t) + \gamma \times r_i(t-1). \quad (5)$$

Equation 5 consists of three terms, each representing a weighted component of the reputation value. The parameters α, β, and γ denote the weights assigned to $TS_i(t)$, $ES_i(t)$, and $r_i(t-1)$, respectively. In general, an honest UE is expected to receive a higher $r_i(t)$, indicating a higher level of trust within the system. While $TS_i(t)$ and $ES_i(t)$ capture the behavior of U_i during the current block interval t, the reputation value from the previous interval, $r_i(t-1)$, is also considered to maintain continuity. The term $\gamma \times r_i(t-1)$ preserves the impact of past behavior, with γ controlling the degree of this impact.

IV. Blockchain System Architecture

This section presents the blockchain system architecture that supports both data sharing and the reputation mechanism for ITS. Fig. 1 illustrates an overview of the system structure.



(a)

Fig. 1. An overview of our proposed blockchain structure.

A. Vehicle communications

During a block interval, vehicles that communicate with the same RSI are regarded as committees in a sharding blockchain.

The United States Department of Transportation defines three types of connected vehicle communications for CAVs, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-device (V2X) [23]. V2V communication facilitates private interactions between vehicles and typically depends on the capabilities of the involved vehicles. These interactions are generally not relevant to other vehicles. V2X offers broad scalability, as 'X' can represent various entities such as passengers, onboard devices, or cloud servers. Because V2X is highly specific to individual vehicle capabilities and use cases, we do not consider it a core component of our system. Our system focuses on V2I communications, which include the following functionalities:

- **Uploading transactions:** UEs can upload transactions, including traffic data such as real-time photos.
- **Downloading transactions:** UEs can access the most recent transactions from nearby RSIs.
- **Uploading evaluations:** UEs can submit their evaluations of the transactions they have downloaded.
- **Downloading evaluations:** UEs can retrieve evaluations from trusted UEs, or compute the reputations of other UEs based on available evaluations.

B. Evaluator of Data

Traffic information is sensitive to both time and location, and it varies significantly across different cities worldwide [28]–[30]. Therefore, most evaluations of a given piece of data are provided by UEs located within the same RSI coverage area during the same block interval. By aggregating all evaluations of the data, its quality can be verified. Based on the data shared by a UE, other UEs can rate its reputation.

To manage and distinguish temporal information, we use blockchain to maintain both reputation and traffic data in chronological order. The blockchain structure allows UEs to access real-time data contained in transactions. In addition, it enables UEs to trace historical data accurately, as the blockchain ensures all records are traceable and untamperd.

The entire area is covered by RSIs which divide the city spatially. This setup functions similarly to a cellular network, where mobile devices connect to the nearest base station. In each block interval, a UE communicates with the closest RSI. During this interval, the RSI manages the traffic data generated by UEs within its coverage area.

A UE may move rapidly across areas covered by different RSIs within the same block time interval. Therefore, a UE is allowed to communicate with multiple RSIs that cover different regions. During a specific period, a UE connects to the nearest RSI to carry out communications, including submitting transactions and evaluations. Since each RSI does not verify transactions recorded by other RSIs, identical traffic data may lead to the creation of separate transactions in different RSIs. These transactions can be evaluated by different UEs within the coverage of those respective RSIs.

C. Permissioned Blockchain

The advantages of blockchain make it a secure, immutable, and decentralized ledger. While keeping the security, the data stored by the blockchain can be simply verified.

A permissioned blockchain is suitable for ITS. Unlike permissionless blockchains, permissioned blockchains are specifically designed to foster mutual trust and enhance collaboration efficiency among multiple parties in industrial applications. These systems implement access controls to ensure that only authorized participants can engage, thereby enhancing security, scalability, and regulatory compliance [31]. In contrast, permissionless blockchains allow users to join and contribute freely without verification. Although this supports

anonymity, it also introduces potential security risks. Permissioned blockchains, by identifying participants, reduce these risks and provide better privacy protection mechanisms [32]. A UE needs to register in the system to propose transactions, while UEs without a legal indentity cannot join the system, ensuring that all participants are legitimate and accountable for the validity of their transactions.

All RSIs maintain permission control. While a new vehicle enters the network, it connects the closest RSI to register by verifying the identity. After verifying, the RSI generates a grant transaction and broadcasts it to the network. The UE would be approved after the grant transaction is recorded in the main blockchain.

D. RSI-based Sharding Blockchain

We develop an RSI-based sharding blockchain to manage shared data and support the reputation mechanism. Traditionally, sharding is introduced to enhance the scalability of blockchain systems. Following this approach, each RSI functions as a shard within the blockchain network. Each shard includes all UEs located within the coverage area of the corresponding RSI during a given block interval.

During each block interval, all vehicles around a single RSI construct a shard. Since vehicles keep moving, an RSI may communicate with different sets of vehicles at the beginning and end of a block interval. A vehicle is included in the shard that includes the closest RSI at the end of the block interval.

As previously stated, a UE could appear in multiple areas covered by different RSIs within the same block interval. This differs from traditional sharding blockchains, where each user belongs to a single shard. In our design, a UE is allowed to propose transactions in multiple shards, with each shard independently maintaining its own set of transactions.

The main chain is formed through the contributions of all shards. At the end of a block interval, every shard selects one UE to generate a shard block. This shard block is supposed to contain all traffic data transactions and evaluations submitted to the RSI during that block interval. Each block on the main chain records the collection of shard blocks from all RSIs. To decrease the size of the main chain, the shard blocks do not fully record the traffic data transactions and evaluations. Instead, the shard block stores metadata referencing the transactions and evaluations, which are kept in off-chain storage [33], [34]. This off-chain storage may reside on the RSI's local storage or cloud servers. UEs can retrieve specific transaction evaluations in different shards by following the metadata and requesting the corresponding data from the off-chain storage.

E. Two-Layer Communication and Storage Strategy

Based on the time-sensitive property of the transactions and evaluations, we propose a two-layer communication and storage strategy.

1) First layer: RSI pool: We introduce the concept of an RSI pool, where each RSI provides a high-speed storage space that functions as a simple data center. This pool allows nearby UEs to upload and download transactions and evaluations efficiently. It temporarily stores all shared data generated during recent block intervals, including transactions and evaluations.

2) Second layer: Off-chain Storage: The system incorporates an off-chain storage layer to manage traffic information efficiently [33], [34]. Since traffic data is time-sensitive, most access requests occur shortly after the information is generated. As the data becomes less relevant over time, historical records can be stored off-chain, where they are retained for infrequent access while reducing the load on the main chain.

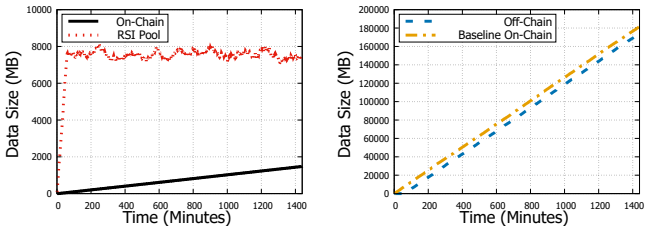
The relationship between the first layer and the second layer resembles that of memory and hard disk storage in computer architecture. The RSI pool functions as a short-term memory, storing recent transactions that are frequently accessed within a limited time. In contrast, the off-chain storage acts as long-term storage, preserving historical records for occasional access when needed. The RSI pool retains transactions and evaluations from several recent block intervals, based on system requirements. As the system transitions to a new block interval, transactions and evaluations from the oldest interval in the pool are removed and archived in the off-chain storage.

F. RPoR Consensus Mechanism

The consensus mechanism is a critical component of the blockchain, as it determines both the security and efficiency of the system. Traditional Proof-of-Work (PoW) consumes significant computational resources, making it unsuitable for the resource-constrained ItS environment. Instead, we adopt Proof-of-Reputation (PoR), leveraging the reputation system, where UEs that consistently provide reliable traffic data are assigned higher trust levels. However, relying on a single UE to calculate reputation updates for all participants in the network in real time is impractical due to computational and communication limitations. In addition to managing reputation updates, the selected UE is also responsible for generating and appending a new shard block to the blockchain.

We propose an RSI-based Proof of Reputation (RPoR) mechanism to select a capable UE for shard block generation. Using the locations of RSIs, the system organizes UEs into multiple shards. Each shard is associated with a specific RSI, and every UE is assigned to the shard that includes its nearest RSI. In each block interval, the RSI selects a leader UE with the highest reputation within its coverage area to generate the shard block. This leader is responsible for computing updated reputation values based on the evaluations submitted by UEs within the same shard. The traffic transactions and evaluations are included in the form of metadata, while the raw data is stored in the off-chain storage. If the highest reputation UE is unavailable, either due to limited capacity or having moved out of the RSI's range, the RSI will continue selecting from the remaining UEs in order of reputation until one agrees to serve as the leader.

The selected leader may act maliciously by creating a shard block that contains incorrect or manipulated content. To mitigate this risk, other UEs connected to the RSI serve as verifiers to supervise the validity of the shard block. Once the



(a) (b)

Fig. 2. During the same simulation, the change of data size over time. (a) shows the size of on-chain data and total size of all RSI pools, while (b) shows the size of off-chain data and total size of baseline on-chain data.

shard block is generated, the RSI broadcasts it and initiates a timer for the verification process. During this period, any UE belonging to the shard can verify the shard block and cast a vote regarding its validity. If more than half of the votes approve the shard block, the RSI confirms it and broadcasts the validated shard block to the RSI network for inclusion in the main chain. If the shard block fails to receive majority approval, the RSI will select a new leader and repeat the process until a shard block is approved by a majority of votes.

V. Simulation

In this section, we present the simulation results of the proposed system. To evaluate the design goals of the system, we focus primarily on the size of the blockchain data and the change of UE reputations.

A. Simulation Settings

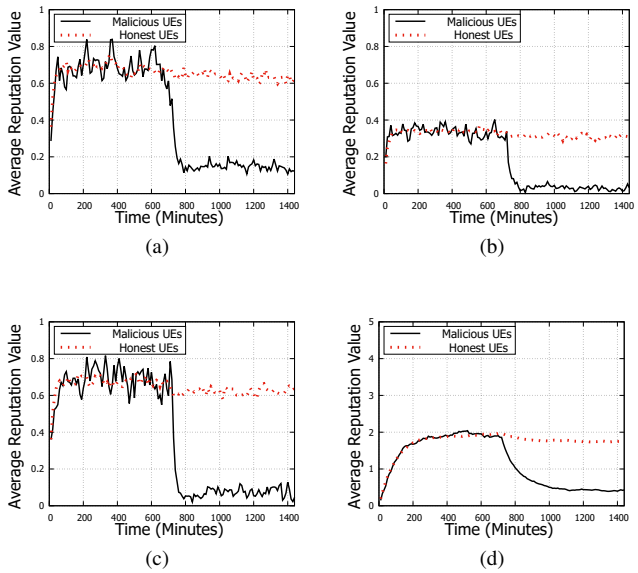
The simulations are based on the T-Drive trajectory dataset [35], [36], which contains one week of GPS trajectories from 10,357 taxis. We use data collected on February 3, 2008, within the geographic region bounded by [39.6, 116.3] and [40.1, 116.8] [37]. Each RSI is assumed to cover an area spanning 0.005 degrees in both latitude and longitude.

Traffic data transactions are generated for each UE using a Poisson distribution, assuming an average of 1 traffic data transaction per UE per hour. Similarly, each UE evaluates an average of 10 traffic transactions from the nearest RSI per hour, also modeled using a Poisson distribution. RSI pools retain recent data for one hour, and data older than one hour is stored in off-chain storage. Each traffic data transaction ranges in size from 0.5 MB to 1 MB, while each evaluation transaction is 1 KB.

B. Blockchain Size Test

We simulate the proposed system over a duration of 1440 minutes. During the simulation, we record the size of on-chain data, off-chain data, and the total data stored across all RSI pools. For comparison, we also simulate a traditional solution in which all transactions and evaluations are stored on-chain, following the approach described in [38], and use it as the baseline. The results are presented in Fig. 2. Due to the significant variation in data sizes, the larger values are shown in (a), while the smaller values are shown in (b).

Overall, the on-chain data, off-chain data in our system, and the baseline on-chain data all increase linearly over time. This trend aligns with expectations, as UEs continuously generate



(a) (b) (c) (d)

Fig. 3. The change of average reputation of different types of UEs over time in different parameters. (a) $\alpha = 0.5, \beta = -0.1, \gamma = 0.5$, (b) $\alpha = 0.25, \beta = -0.1, \gamma = 0.5$, (c) $\alpha = 0.5, \beta = -0.2, \gamma = 0.5$, (d) $\alpha = 0.5, \beta = -0.1, \gamma = 0.9$.

data at a consistent rate, and the blockchain stores newly generated transactions accordingly. In our system, the size of off-chain data is approximately 95% of the baseline on-chain data, indicating that the majority of information is stored off-chain. The on-chain data size in our system is significantly smaller than the off-chain portion. Notably, the ratio of on-chain to off-chain data decreases over time, reaching approximately 0.85% at 1440 minutes. This demonstrates that the proposed lightweight blockchain system substantially reduces on-chain storage requirements and network bandwidth usage compared to traditional fully on-chain approaches.

The data stored in RSI pools increases rapidly at the beginning of the simulation and then stabilizes, fluctuating around 7500 MB. RSI pools are designed to store only frequently accessed data, which represents a small fraction of the total dataset. The relatively small size of this real-time data, compared to the complete data volume, suggests that the storage and access burden on RSIs and UEs during real-time interactions remains low.

C. Reputation Test

In the reputation tests, we define two types of UEs: honest and malicious. 0.9 of the total UEs are designated as honest, while the remaining ten percent are malicious. Honest UEs generate correct transactions and evaluations with a probability of 0.9. Malicious UEs behave similarly during the first 720 minutes, submitting correct data with the same probability of 0.9. After this initial period, their behavior changes, and the probability of submitting correct transactions and evaluations drops to 0.1. We vary the parameters in Equation 5 to assess the effectiveness of the proposed reputation mechanism and to examine the impact of different parameter settings. All UEs start with an initial reputation value of zero.

Fig. 3 illustrates the evolution of the average reputation of different types of UEs over time under various parameter settings. For brevity, we omit the term "average" in the following discussion when it does not lead to ambiguity. Four distinct parameter settings are selected to highlight the characteristics of the reputation mechanism. Fig. 3(a) represents the baseline setting with parameters $\alpha = 0.5$, $\beta = -0.1$, and $\gamma = 0.5$. In comparison, Fig. 3(b) shows the effect of reducing α to 0.25, Fig. 3(c) explores the impact of decreasing β to -0.2, and Fig. 3(d) presents results with an increased γ value of 0.9.

Overall, the simulation results display consistent patterns across all configurations. At the beginning, the reputations of all UEs increase rapidly and then stabilize around a certain value. After 720 minutes, when malicious UEs begin submitting low quality data, the reputation of honest UEs shows a slight decline, while the reputation of malicious UEs drops significantly. During the initial 720 minutes, the reputation of malicious UEs fluctuates more noticeably due to their smaller population and the inconsistency in the quality of their submitted data. The minor decrease in the reputation of honest UEs after 720 minutes can be attributed to the overall decline in data quality throughout the network.

Next, we analyze the impact of different parameter settings. When the α is reduced from 0.5 to 0.25, the average reputation value before 720 minutes decreases from approximately 0.7 to 0.35. After 720 minutes, the reputation of malicious UEs drops to 20 percent of the reputation of honest UEs in Fig. 3(a), whereas the ratio decreases further to 15 percent in Figure 3(b). When the parameter β is changed from negative 0.1 to negative 0.2, the reputation of honest UEs remains largely unchanged, while the reputation of malicious UEs drops to 10 percent of that of honest UEs, as shown in Figure 3(c). When γ is increased from 0.5 to 0.9, the most noticeable effect is a reduction in reputation fluctuations and a slower stabilization rate. Additionally, the overall reputation values are higher. However, this increase can be partially attributed to the fact that α and β remain unchanged and are not adjusted in accordance with the increase in γ.

These results demonstrate that the reputation mechanism effectively distinguishes between honest and malicious UEs. The parameters α and β determine the weight of transactions and evaluations in the reputation calculation, primarily influencing the overall reputation values. γ controls the influence of historical data, thereby affecting the stability of reputation over time. By adjusting these parameters appropriately based on the environment, the effectiveness and responsiveness of the reputation mechanism can be optimized.

VI. Related work

ITS and CAV have attracted researchers to develop techniques to resolve potential demands, especially cooperative perception and data sharing [39], [40]. Zhang et al. [9] study the threats on sensor data integrity in CAV networks. They propose a system deployed on CAVs that detects malicious data in real time. In addition to real-time data detection, the storage and protection of traffic data also need to be studied. Blockchain has been deployed in edge computing and IoT networks to support data storage and management [11].

Li et al. [18] propose a fundamental blockchain scheme with basic protocols to support video transaction traceability in IoV. The scheme claims it provides evidence for disputes, but more details on costs and dispute resolution are needed before it can be put into practical use. Considering the security, Liang et al. [41] apply macro-micro blockchain to record the data feature to resist potential illegal data from attackers. While the blockchain grows rapidly, the efficiency of the blockchain may not match the limited capability and bandwidth of vehicles. Tan et al. [19] enable digital twins to support data sharing for IoV. The digital twins aggregate data by federated learning. Their proposal remains concerned with computing resources. Real-time federated learning requires a lot of computing resources. To alleviate this issue, digital twins are imported, but communication overhead becomes another issue. Kang et al. [42] propose a fundamental blockchain solution to data sharing in IoV. They propose a blockchain protocol for edge devices, and a reputation mechanism for vehicles to evaluate each other. However, the high storage and bandwidth of blockchain are not discussed, and the details of the reputation mechanism lack clarity, which may lead to problems such as reputation calculation being easily manipulated. Firdaus et al. [38] deploy a consortium blockchain with a reputation mechanism to incentivize vehicles to share data. Their solution stores all data on-chain, which creates huge overhead and bandwidth in the blockchain development process.

VII. Conclusion

In this paper, we have presented a practical and scalable framework for secure information management in future intelligent transportation systems. We have integrated a distributed reputation mechanism to support trustworthy data evaluation and dissemination within vehicular networks. Additionally, we have designed a hybrid blockchain storage architecture combining on-chain and off-chain layers to ensure low-latency access while maintaining data integrity and tamper resistance. Extensive simulations have been conducted to validate the proposed approach, demonstrating significant improvements in both data integrity and resource efficiency compared to existing solutions.

Acknowledgement

This work is supported in part by US National Science Foundation under grant number 2230620.

References

[1] Christian Creß, Zhenshan Bing, and Alois C Knoll. Intelligent transportation systems using roadside infrastructure: A literature survey. *IEEE Transactions on Intelligent Transportation Systems*, 25(7):6309–6327, 2023.

[2] Tingting Yuan, Wilson da Rocha Neto, Christian Esteve Rothenberg, Katia Obraczka, Chadi Barakat, and Thierry Turletti. Machine learning for next-generation intelligent transportation systems: A survey. *Transactions on emerging telecommunications technologies*, 33(4):e4427, 2022.

[3] MarketsandMarkets. Intelligent transportation system market by offering, system, mode of transport, application and region - global forecast to 2029. https://www.marketsandmarkets.com/Market-Reports/intelligent-transport-systems-its-market-764.html, 2024. Accessed: 2025-05-22.

[4] Jie Cui, Fenqiang Ouyang, Zuobin Ying, Lu Wei, and Hong Zhong. Secure and efficient data sharing among vehicles based on consortium blockchain. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):8857–8867, 2021.

[5] Jinbo Xiong, Renwan Bi, Mingfeng Zhao, Jingda Guo, and Qing Yang. Edge-assisted privacy-preserving raw data sharing framework for connected autonomous vehicles. *IEEE Wireless Communications*, 27(3):24–30, 2020.

[6] Xumiao Zhang, Anlan Zhang, Jiachen Sun, Xiao Zhu, Y Ethan Guo, Feng Qian, and Z Morley Mao. Emp: Edge-assisted multi-vehicle perception. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 545–558, 2021.

[7] Yuze He, Li Ma, Jiahe Cui, Zhenyu Yan, Guoliang Xing, Sen Wang, Qintao Hu, and Chen Pan. Automatch: Leveraging traffic camera to improve perception and localization of autonomous vehicles. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 16–30, 2022.

[8] Xianhao Shen, Yufang Lu, Yi Zhang, Xiaoyong Liu, and Lieping Zhang. An innovative data integrity verification scheme in the internet of things assisted information exchange in transportation systems. *Cluster Computing*, 25(3):1791–1803, 2022.

[9] Qingzhao Zhang, Shuowei Jin, Ruiyang Zhu, Jiachen Sun, Xumiao Zhang, Qi Alfred Chen, and Z Morley Mao. On data fabrication in collaborative vehicular perception: Attacks and countermeasures. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 6309–6326, 2024.

[10] Feilong Wang, Xin Wang, and Xuegang Jeff Ban. Data poisoning attacks in intelligent transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 165:104750, 2024.

[11] Yongjun Ren, Yan Leng, Yaping Cheng, and Jin Wang. Secure data storage based on blockchain and coding in edge computing. *Math. Biosci. Eng*, 16(4):1874–1892, 2019.

[12] Tiffany Hyun-Jin Kim, Cristina Basescu, Limin Jia, Soo Bum Lee, Yih-Chun Hu, and Adrian Perrig. Lightweight source authentication and path validation. In *Proceedings of the 2014 ACM Conference on SIGCOMM*, pages 271–282, 2014.

[13] Sahbi Boubaker, Faisal S Alsubaei, Yahia Said, and Hossam E Ahmed. Lightweight cryptography for connected vehicles communication security on edge devices. *Electronics*, 12(19):4090, 2023.

[14] Srivalli Boddupalli, Akash Someshwar Rao, and Sandip Ray. Resilient cooperative adaptive cruise control for autonomous vehicles using machine learning. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):15655–15672, 2022.

[15] B Thirumala Rao et al. A study on data storage security issues in cloud computing. *Procedia Computer Science*, 92:128–135, 2016.

[16] Xin Jin, Charalampos Katsis, Fan Sang, Jiahao Sun, Ashish Kundu, and Ramana Kompella. Edge security: Challenges and issues. *arXiv preprint arXiv:2206.07164*, 2022.

[17] Shuyao Shi, Neiwen Ling, Zhehao Jiang, Xuan Huang, Yuze He, Xiaoguang Zhao, Bufang Yang, Chen Bian, Jingfei Xia, Zhenyu Yan, et al. Soar: Design and deployment of a smart roadside infrastructure system for autonomous driving. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 139–154, 2024.

[18] Xinghao Li, Chenchen Tan, Minghao Liu, Tom H Luan, Longxiang Gao, and Youyang Qu. A blockchain-based cooperative perception in internet of vehicles. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pages 1–6. IEEE, 2021.

[19] Chenchen Tan, Xinghao Li, Longxiang Gao, Tom H Luan, Youyang Qu, Yong Xiang, and Rongxing Lu. Digital twin enabled remote data sharing for internet of vehicles: System and incentive design. *IEEE Transactions on Vehicular Technology*, 72(10):13474–13489, 2023.

[20] Rateb Jabbar, Eya Dhib, Ahmed Ben Said, Moez Krichen, Noora Fetais, Esmat Zaidan, and Kamel Barkaoui. Blockchain technology for intelligent transportation systems: A systematic literature review. *IEEE Access*, 10:20995–21031, 2022.

[21] Md Masud Rana and Kamal Hossain. Connected and autonomous vehicles and infrastructures: A literature review. *International Journal of Pavement Research and Technology*, 16(2):264–284, 2023.

[22] Lanhang Ye and Toshiyuki Yamamoto. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Physica A: Statistical Mechanics and its Applications*, 526:121009, 2019.

[23] Hafiz Usman Ahmed, Ying Huang, Pan Lu, and Raj Bridgelall. Technology developments and impacts of connected and autonomous vehicles: An overview. *Smart Cities*, 5(1):382–404, 2022.

[24] Linchao Zhang, Lei Hang, Keke Zu, Yi Wang, and Kun Yang. Dynamic vehicle reputation consensus: Enhancing iov communication with a blockchain algorithm. *IEEE Transactions on Vehicular Technology*, 2024.

[25] Adnan Mahmood, Sarah Ali Siddiqui, Quan Z Sheng, Wei Emma Zhang, Hajime Suzuki, and Wei Ni. Trust on wheels: Towards secure and resource efficient iov networks. *Computing*, 104(6):1337–1358, 2022.

[26] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, pages 640–651, 2003.

[27] Kun Yan, Ping Zeng, Kan Wang, Wenping Ma, Geng Zhao, and Yingjie Ma. Reputation consensus-based scheme for information sharing in internet of vehicles. *IEEE Transactions on Vehicular Technology*, 72(10):13631–13636, 2023.

[28] Jiwon Kim and Hani S Mahmassani. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Procedia*, 9:164–184, 2015.

[29] Yong-Hong Liu, Jin-Ling Ma, Li Li, Xiao-Fang Lin, Wei-Jia Xu, and Hui Ding. A high temporal-spatial vehicle emission inventory based on detailed hourly traffic data in a medium-sized city of china. *Environmental Pollution*, 236:324–333, 2018.

[30] Anthony Stathopoulos and M Karlaftis. Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record*, 1768(1):135–140, 2001.

[31] Huizhong Li, Yujie Chen, Xiang Shi, Xingqiang Bai, Nan Mo, Wenlin Li, Rui Guo, Zhang Wang, and Yi Sun. Fisco-bcos: An enterprise-grade permissioned blockchain system with high-performance. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–17, 2023.

[32] Horst Treiblmaier and Christian Sillaber. The impact of blockchain on e-commerce: A framework for salient research topics. *Electronic Commerce Research and Applications*, 48:101054, 2021.

[33] Jun Wook Heo, Gowri Sankar Ramachandran, Ali Dorri, and Raja Jurdak. Blockchain data storage optimisations: a comprehensive survey. *ACM Computing Surveys*, 56(7):1–27, 2024.

[34] Chaoyang Li, Mianxiong Dong, Jian Li, Gang Xu, Xiu-Bo Chen, Wen Liu, and Kaoru Ota. Efficient medical big data management with keyword-searchable encryption in healthchain. *IEEE Systems Journal*, 16(4):5521–5532, 2022.

[35] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, pages 99–108, 2010.

[36] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 316–324, 2011.

[37] Qinghang Gao, Jianmao Xiao, Zhiyong Feng, Jingyu Li, Yang Yu, Hongqi Chen, and Qiaoyun Yin. Optimization of models and strategies for computation offloading in the internet of vehicles: Efficiency and trust. *IEEE Transactions on Mobile Computing*, 2024.

[38] Muhammad Firdaus and Kyung-Hyune Rhee. On blockchain-enhanced secure data storage and sharing in vehicular edge computing networks. *Applied Sciences*, 11(1):414, 2021.

[39] Tao Huang, Jianan Liu, Xi Zhou, Dinh C Nguyen, Mostafa Rahimi Azghadi, Yuxuan Xia, Qing-Long Han, and Sumei Sun. V2x cooperative perception for autonomous driving: Recent advances and challenges. *arXiv preprint arXiv:2310.03525*, 2023.

[40] Guangzhen Cui, Weili Zhang, Yanqiu Xiao, Lei Yao, and Zhanpeng Fang. Cooperative perception technology of autonomous driving in the internet of vehicles environment: A review. *Sensors*, 22(15):5535, 2022.

[41] Haoran Liang, Jun Wu, Shahid Mumtaz, Jianhua Li, Xi Lin, and Miaowen Wen. Mbid: Micro-blockchain-based geographical dynamic intrusion detection for v2x. *IEEE Communications Magazine*, 57(10):77–83, 2019.

[42] Jiawen Kang, Rong Yu, Xumin Huang, Maoqiang Wu, Sabita Maharjan, Shengli Xie, and Yan Zhang. Blockchain for secure and efficient data sharing in vehicular edge computing and networks. *IEEE internet of things journal*, 6(3):4660–4670, 2018.