Blockchain: A Safe, Efficient Solution for Driver Privacy and Connected Vehicle Transportation Data Sharing

Yingxi Cao, Abdullah Kurkcu, Kaan Ozbay

BLOCKCHAIN: A SAFE, EFFICIENT SOLUTION FOR DRIVER PRIVACY AND CONNECTED VEHICLE TRANSPORTATION DATA SHARING

Yingxi Cao, M.Sc. **Corresponding Author**
Graduate Research Assistant, C2SMART Center,
Department of Computer Science and Engineering
Tandon School of Engineering, New York University (NYU)
Fifteen MetroTech Center, 6th floor, Brooklyn, NY 11201, USA
Tel: 1-(646)-639-9249; E-mail: yingxi.cao@nyu.edu

Abdullah Kurkcu, Ph.D.
Research Associate, C2SMART Center,
Department of Civil & Urban Engineering,
Tandon School of Engineering,
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201, USA
Tel: 1-(646)-997-0538; Email: ak4728@nyu.edu

Kaan Ozbay, Ph.D.
Professor & Director, C2SMART Center (A Tier 1 USDOT UTC),
Department of Civil and Urban Engineering &
Center for Urban Science and Progress (CUSP),
Tandon School of Engineering, New York University (NYU)
Six MetroTech Center, 4th Floor (RM 404), Brooklyn, NY 11201, USA
Tel: 1-(646) 997-3691; E-mail: kaan.ozbay@nyu.edu

Word count: 6,431 words text + 4 tables × 250 words (each) = 7,431 words

Submission Date: **August 16, 2019**

ABSTRACT

Connected and automated vehicles (CAVs) are becoming increasingly prevalent, bringing with them potential for better safety and mobility. However, these vehicles can create many thousands of transactions in a flash, creating a challenge for current technologies that are not capable of transmitting such big data “privately” and “securely”. Distributed ledger technologies such as Blockchain have the potential to address this challenge by using decentralized system. Blockchain-based system allows users to enter into direct relationships with each other following commonly agreed terms with a high degree of trust, eliminating the need for a central authority while retaining security and privacy.

This study investigates the potential for Blockchain to support safer delivery of CAV data. By using an actual simulation based implementation of the proposed architecture, it demonstrates how Blockchain can improve the level of security and privacy of data sharing and attempts to answer two fundamental questions: 1) how to securely and privately get and store data from CAVs and 2) how to find the best method to connect them using Blockchain technology. The proposed eight-layer framework uses Hyperledger Fabric as an underlying Blockchain technology and uses machine learning models for analyzing data collected in chain. The traffic data in the physical layer are simulated using microscopic traffic simulation tool SUMO and then incorporated into the Blockchain platform. The experiments highlight that the CAV system can be effectively combined with Blockchain technologies while enhancing security in a significant manner.

Keywords: Blockchain, autonomous vehicle, security, and privacy, machine learning, simulation

INTRODUCTION AND MOTIVATION

Data analysis for transportation research is experiencing a major disruption in the era of big data. New technologies in transportation systems are generating massive amounts of data, including complex and diverse traffic and vehicle specific information. For instance, connected vehicles generate speed, location, and acceleration records every 0.1 seconds and share them with other vehicles and the infrastructure. IoT supported autonomous vehicles may also have the capability to share the generated data. These changes in transportation systems raise significant concerns regarding privacy and security. Security is critical, as the data must be kept confidential to protect individual privacy. Accounting for these security needs while meeting operational requirements is quite challenging for traditional CAV systems. However, Blockchain can provide a potential solution for efficiently securing data.

In this paper, the advantages of using a decentralized Blockchain's benefits such as high security and scalability are investigated and utilized to secure transportation data exchange systems. The critical feature of Blockchain is that the technology ensures trust. In Blockchain, the ledger is decentralized. It means no single computer or single system has control over the ledger at any one time. To be able to gain access, one needs to coordinate an attack simultaneously using thousands of smart devices. The chain itself is also a complicated security measure. Anyone who tries to modify or forge a transaction would first have to accurately replicate all transactions leading up to that transaction. There are more security features of Blockchain such as the verification of transactions and the usage of cryptographic keys. This study uses and evaluates most of Blockchain's security features, including the chain of sequential blocks, decentralization and cryptographic keys, to design a customized Blockchain for a smart vehicle platform. In addition, value-adding services such as traffic state prediction using machine learning techniques are evaluated as a layer in the proposed system. A Blockchain-based solution can also solve several major problems in current connected vehicle systems, such as data ownership, data collection and accuracy, data exchange protocols, and application infrastructure. With Blockchain, data ownership and access to the data belong to the data provider.

The rest of the paper is structured as follows: the background section illustrates the requirements and underlying technologies that were used to create the Blockchain. In the following section, previous studies using Blockchain technologies in the transportation literature are reviewed. Then, a novel 8-layer Blockchain-based system is proposed and demonstrated with scenarios using microscopic traffic simulation software coupled with the actual implementation of Blockchain technology in the app developed by the research team. How the proposed system can affect privacy and security measures is explained in the security, privacy and scalability section. The top layer, which incorporates machine learning-based prediction algorithms, is evaluated with user-generated data to demonstrate its potential to provide more personalized services. The final section of this paper reports results of scenarios developed for testing the latency of various numbers of Blockchain nodes and for predicting the vehicles/ driving speeds securely obtained from the Blockchain implementation. The paper is concluded with the reporting of the performance of the proposed system in terms of message delay and latency and the discussion of the potential future work.

BACKGROUND

CAV applications and mobility services are one of the most significant innovations that the automotive industry has experienced within the last couple of decades. Most of these applications rely on the automotive industry to equip their vehicles with smart devices such as smart phone. Most car makers have started not only installing CAV capabilities in their vehicles but also exploring ways to use Blockchain to improve transportation. Porsche was the first company to test Blockchain technology in their cars. However, they are not the only company investigating solutions to integrate Blockchain technology into their vehicles. Several other companies like Dovu (*1*) and Streamr (*2*) have started experimenting Blockchain technologies mainly due to its security benefits. Dovu (*1*) adopts Blockchain technology not only to vehicle industry but also to aircraft and railways. Meanwhile, Streamr (*2*) focuses on data sharing field.

While the terms Blockchain and the term distributed ledger are used interchangeably, they do not necessarily mean the same thing. Blockchain is a technology that decentralizes a digital ledger relying on the consensus of a global peer-to-peer network to operate. A distributed ledger is a database that is spread across many smart devices. Information stored on a Blockchain also exists as a shared database. Thus, every Blockchain is a distributed ledger. However, not all distributed ledgers use a chain of blocks to provide security. The uniqueness of the Blockchain comes from the fact that the data is organized in blocks. These blocks are then grouped together and secured using cryptography.

Blockchain is a decentralized system that exists between all permitted participants. This removes the need to pay intermediaries and the potential for conflicts. Blocks are linked and secured using cryptography in a distributed environment, so they are inherently resistant to deletion and modification of the data. The summary of the comparison between traditional distributed systems, server-centric platforms, and Blockchain-based technologies can be seen in TABLE 1.

TABLE 1 Advantages of Blockchain platform over other platforms (*3*)

	Blockchain-based platforms	**Server-centric platforms**	**Traditional distributed systems**
Control schema	Decentralized control	Centralized control	Partially centralized control
Implementation difficulty	Easy to implement	Easy to implement	Hard to implement
Efficiency	High efficiency	Efficiency depends on the server configuration	Efficiency depends on the master node
Cost	Automated inspection	Large amount of manual inspection	Small amount of manual inspection
Safety (*4*)	Secure	Vulnerable	Vulnerable
Privacy	Private	Information leakage happens	Information leakage happens
Transparency	Transparent	Not transparent	Partially transparent
Application logic	Smart contracts	Central algorithms are needed	Complex algorithms are needed

Another essential concept related to Blockchain is smart contracts. Smart contracts are a set of self-executing instructions written in computer code. They represent the terms of the agreement between drivers, vehicles, and government. Smart contracts are a set of rules that are executed automatically. For instance, passengers are required to pay to use the subway system. It is an example of a smart contract between the passenger and the transit authority. The moment they pay for it at the turnstile, they start receiving the service on the other side of the turnstile. Instead of being written into plain English, smart contracts are written into lines of code. Smart contracts ensure trusted and permissioned data transmission between anonymous parties without the central authority.

There are various tools that can be used to establish the Blockchain framework. Most of them include aforementioned technologies in this section. In our case, Hyperledger Fabric (*5*) is chosen over others, because it is a permission and modular platform. It allows a faster data transmission speed and a higher privacy level, which is suitable for CAV system. Hyperledger Fabric is an open source Blockchain platform, and it supports high security, membership identity services, and pluggable consensus protocols. Hyperledger Fabric is still evolving under the Hyperledger Linux Foundation (*6*) project. Version 1.1 is used for the case study.

Hyperledger Composer (*7*) is an open source toolset that helps developers to quickly and easily develop and deploy their models and application logic. It has been built with JavaScript with multiple interfaces that allow us to efficiently incorporate simulated vehicle trajectory data with Blockchain. Composer version 0.19.10 is being used for the testing environment. Composer supports the existing Fabric Blockchain infrastructure, and it consists of a set of tools that make building Blockchain applications easier. For these reasons, it is selected as the framework for this study. FIGURE 1 shows the relation between Hyperledger Fabric and Hyperledger Composer, and how they are utilized in the CAV network.

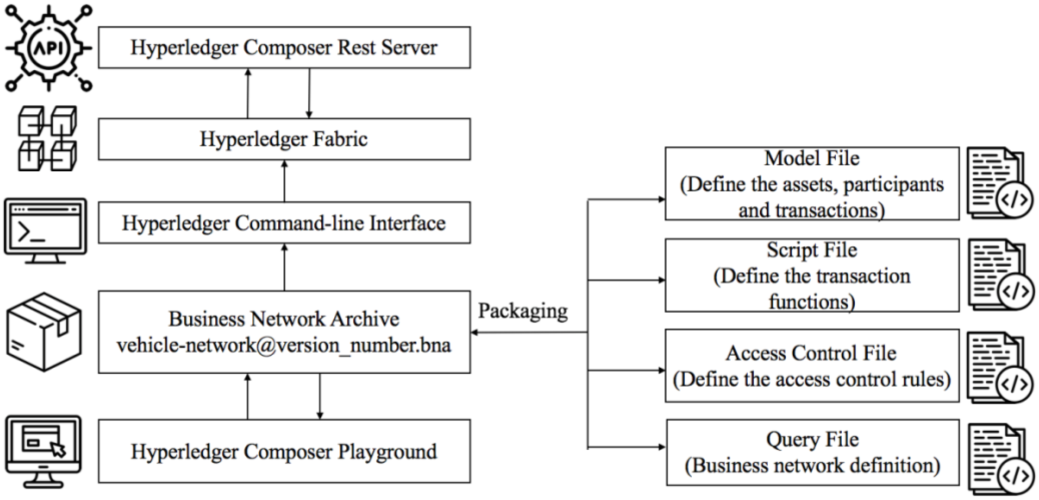

FIGURE 1 The connection between Hyperledger Fabric and Hyperledger Composer

LITERATURE REVIEW

Most conventional cybersecurity algorithms and methods are incapable of providing security to CAVs' data-sharing technologies. Diverse and complex driver behavior involves a high degree of social complexity. Thus, traditional Intelligent Transportation Systems (ITS) are confronted with critical security risks. In order to reduce these risks, a basic Blockchain-based novel ITS (B^{2}ITS) framework (*8*) is introduced. This framework consists of 7 layers from the physical layer to the application layer. It connects the Internet of Things (IoT) devices in the physical layer and includes car sharing schema in the application layer. Besides demonstrating the B^{2}ITS framework, they also analyze the research by artificial societies, computational experiments and parallel (*9*) approach about how to transit B^2ITS to Parallel transportation Management System (PTMS). Interested readers about artificial societies are referred to the related reference (*9*). Finally, a Dapp (Decentralized application) named La'zooz (*10*) is shown to prove that a 7-layer framework is utilizable in a real-time car sharing scenario.

A similar 7-layer Blockchain-based Intelligent Transportation System is proposed by Madhusudan Singh and Shiho Kim (*11*). They show an Intelligent Vehicle Trust Point(IV-TP) element (*12*) that is used to build trust and reliable data communication channel among IVs. Different from the previous framework, this is a reward-based IV communication framework using Blockchain technology. In addition, they incorporate Vehicular Cloud Computing (VCC) (*12*). VCC is a hybrid technology that makes use of vehicle resources to execute computations on the cloud. The innovation of their framework is the rewarding system: if a vehicle wins the consensus competition, then it will get a trust point from the benefiter IV, so its trust point goes up. While Singh and Kim *(11)* did not illustrate the weaknesses of the proposed framework, they theoretically show that the improved Blockchain framework with crypto IV-TP (*13*) can help to improve the privacy of IVs.

Scalability was another important consideration while researching for the most suitable framework for the purpose of this study. Dorri et al. (*14)* propose an optimized Blockchain instantiation for the IoT called Lightweight Scalable Blockchain (LSB) established upon Blockchain technology. A public Blockchain is managed by the overlay nodes, such as smart vehicles. Overlay transactions are broadcast and verified by Overlay Block Managers (OBMs). As a result, scalability is improved. The mission of these OBMs is to verify each transaction's public key and protect the whole network from malicious attacks. In order to reduce latency, they also incorporate a soft handover method that selects new OBM with the lowest delay for IoT devices. LSB can be used for various applications, such as remote software updates, insurance, smart charging service and car sharing schema. The weakness of this framework may include the high overhead caused by the frequent mobility of the vehicles.

Conventionally, private keys are used to digitally sign safety messages. Another feasible option is to make use of public key infrastructure (PKI) with centralized management for creation and revocation of digital certificates in order to ensure security. Those methods usually impose significant overhead on vehicles. A new Blockchain based scheme that could alleviate the computation overhead and enhance the response time while improving the overall system security is proposed by Lasla, N., M. Younis, W. Znaidi, and D. B. Arbia (*15*). The core idea of this mechanism is to make use of PKI and guarantee the security of every transaction. Each safety message that transmits in the Blockchain has to be signed by a private key, unsigned messages will be refused as it lacks of authentication. Meanwhile, the public key is known by other vehicles and is used to verify the message integrity in a decentralized manner. So, in contrast to the traditional PKI, the certificate is no longer included in safety messages, and the verification is replaced by a simple lookup function, which is much faster than traditional signature verification.

Having learned several Blockchain based transportation systems and frameworks, it is clear that how to make full use of those data and create valuable analysis is another interesting topic. For example, Uber developed a machine learning platform to predict the end to end delivery duration during complex multi-stage process. (*16*). It is an internal machine learning (ML)-as-a-service platform that adopts machine learning methods and algorithms to facilitate people's lives while meeting business requirements. For example, an application called UberEATS (*17*) is designed to estimate food delivery time. The idea to combine Blockchain technology with machine learning and create our unique platform emerged because the data is collected in the chain in a very secure and decentralized way, so the analysis generated by the ML platform will be even more valuable.

Blockchain attracts the interest of many researchers conducting studies and experiments in this field. Nevertheless, most of the studies are at the very beginning of the Blockchain cycle. TABLE 2 lists some of the most recent studies and implementations using Blockchain in transportation. There are many Blockchain projects that are currently being developed but they are not widely used for applications in transportation. Blockchain is commonly used for cryptocurrency, banking (*4*), biotechnology, pharmacy, life sciences (*18*), crowdsourcing (*19*) and IoT (*20*) applications.

TABLE 2 Existing Blockchain studies and their implementation

Existing studies	**Domain**	**Implementation**
Dovu (*1*)	Mobility/Transportation	Implemented
Streamr (*2*)	Data sharing/Market place	Implemented
Vehicle data sharing framework (*11*)	Transportation	Conceptual

DESCRIPTION OF THE PROPOSED BLOCKCHAIN FRAMEWORK ARCHITECTURE

Based the existing Blockchain studies in the literature, we design an eight-layer framework based on Blockchain-based Intelligent Transportation Systems (*8*) and Hyperledger Fabric (*5*). This 8-layer network ranges from a low-level hardware data communication layer to a high-level services layer. The data generated from smart devices, including cars, sensors, and IoT devices, will be sent to our local Hyperledger Fabric platform by customized programs written using Golang (*21*). Golang is a programming language invented by Google in 2009 that highly supports concurrency. All of the source codes, tools and compilers are open source.

The CAV network may generate both sensitive and insensitive data involving different stakeholders ranging from government to public clients. Hence, the Hyperledger access control system is used to assign various access levels to different participants. Moreover, channels in Hyperledger protect data privacy. Channels can be defined as sub-networks. If two organizations use different channels, then they will not be able to communicate with each other, and each node holds a separate ledger. In other words, if one node belongs to two different channels, then it holds two different ledgers. However, if two nodes belong to the same channel, they can quickly view the transactions that occurred within that channel. The information exchange speed is faster when the same channel is used. Each connected vehicle will generate ten messages per second. Because of this high throughput requirement, a Byzantine fault-tolerant system (BFT) (*22*) is not suitable for our system. Hyperledger Fabric provides three different types of consensus mode (orderer service): SOLO, Kafka, and BFT. The SOLO mode can be considered centralized because the entire fabric network relies on a single orderer node. The orderer node is a node that running the communication service. It can guarantee a delivery, broadcast proposals and results to other nodes in network, collecting responses in order to realize consensus. Kafka mode is a semi-centralized mode which relies on a Kafka cluster (*23*). The BFT mode represents a decentralized orderer cluster. For simplicity, the SOLO order node is used to fulfill the consensus task of the testing system. In the beginning, the application generates a transaction proposal and sends the proposal to corresponding peers for endorsement. Each peer executes the Chaincode separately. If the peers agree on the proposal, they respond by adding their digital signatures, and signing the entire payload using their private keys without updating the ledger. Once the orderer node receives enough responses and achieves consistent endorsement by all relevant organizations, it will package the transactions into blocks and send them back to each peer. Finally, new transactions append to the ledger.

All the application logic is defined in a smart contract written by the authors in JavaScript. FIGURE 2 shows the high-level 8-layer network architecture. The bottom layer is the physical layer. It consists of diverse IoT devices including connected vehicles, smart phones, sensors, and so on. Those devices could post real time information to the data encryption layer. They wait for the real-time data to be hashed then to be transmitted in different channels. Each organization has a membership service. Members who want to enter the network have to be preregistered, so this process happens on the authentication layer. In order to realize consensus, the orderer service is implemented on the consensus layer. It collects corresponding responses and return consensus results. Finally, smart contracts are designed on the contract layer and data exchange schema is created on the application layer to fulfill the application logic. The top layer is the service layer, where two machine learning models are created to forecast future vehicle speed.

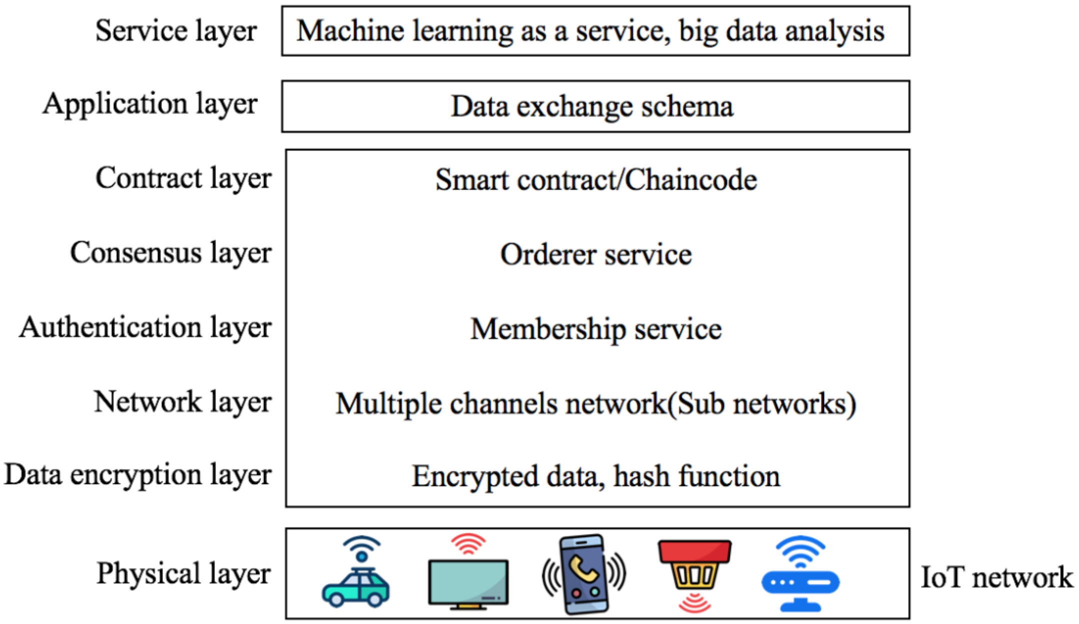

FIGURE 2 Blockchain based 8-layer network architecture

For testing purposes, real-time vehicle trajectory data from a calibrated microscopic traffic simulation model is generated and transferred to the Blockchain system using the developed code. Traffic simulation model is created using SUMO, an open source traffic simulation tool (*24*), and the trajectory data is collected at a certain section of the model for an hour. After the data collection, trajectories containing data for every 0.1 second interval are post-processed and posted to the Blockchain. The primary function of the developed codes is to format the raw data, record vital timestamps such as posting time and receiving time, and POST converted JSON data corresponding to the Blockchain platform. Each time a valid message is disseminated in the network, a transaction will be created in Blockchain along with a cryptographic hashed transaction ID. Overall, it is valid to state that the activities taking place in the Physical layer of FIGURE 2 are simulated in SUMO and then they are incorporated into the Blockchain framework. The data flow is shown in FIGURE 3.

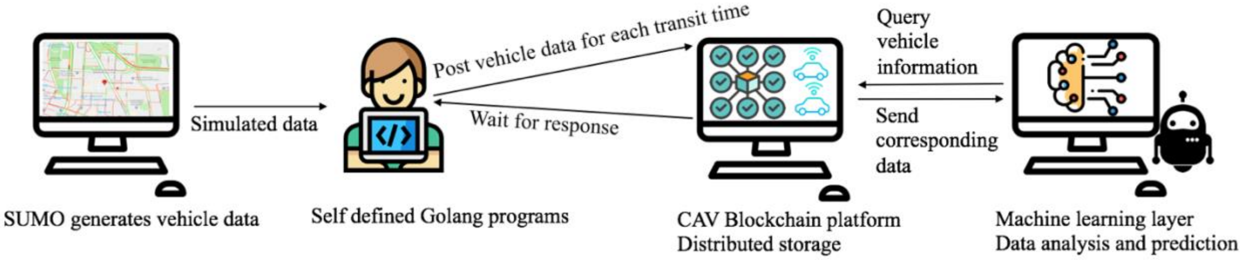

FIGURE 3 The overall implementation framework for the proposed Blockchain Architecture for CAVs

The other two crucial components of Blockchain are participants and assets. Participants are the manipulators who are involved in the network. Assets can be real estate or conceptual property. In our system, drivers and decision makers are participants. Trips and vehicles are assets. All clients can report themselves as a driver with their driver ID and create a trip with their driver's license. The vehicle information such as the ID, vehicle state, manufacturer, model type, and color will also be transmitted to the proposed system and stored in Hyperledger Fabric. The vehicle state could be "ACTIVE," "OFF_THE_ROAD" or "IN_INCIDENT." The trip state could be "CREATED," "DRIVER_ASSIGNED," "VEHICLE_ASSIGNED," "DEPARTED" or "ARRIVED." The status of both vehicle and trip will be updated by participants. However, if a vehicle is in an accident, only the regulators have the control to change its status to normal. For the access control of the system, regulators will have the highest authority, while drivers will only have limited access.

FIGURE 4 shows the topology of the Blockchain network. There are ten peers and 1 SOLO orderer node. Peers can be considered as nodes in Blockchain, which can hold both ledger and smart contracts. They can be separated by physical machines or different Docker containers (*25*) that reside on the same machine. There are three organizations (driver organization, academic institution, and government agency) in the CAVs Blockchain platform. Each organization has an administrator. Vehicles can post real-time information through an authorized registered port in the network. All participants must ask the organization for a valid key to enter the network. In Hyperledger Fabric, this mechanism is called membership service. Membership service requires that each participant entering the network should have a valid identity and get approved by the administrator of the corresponding organization. The driver organization, academic institution and government agency will post data through different ports. However, they all store the data inside the same Blockchain network and hold the same ledger.

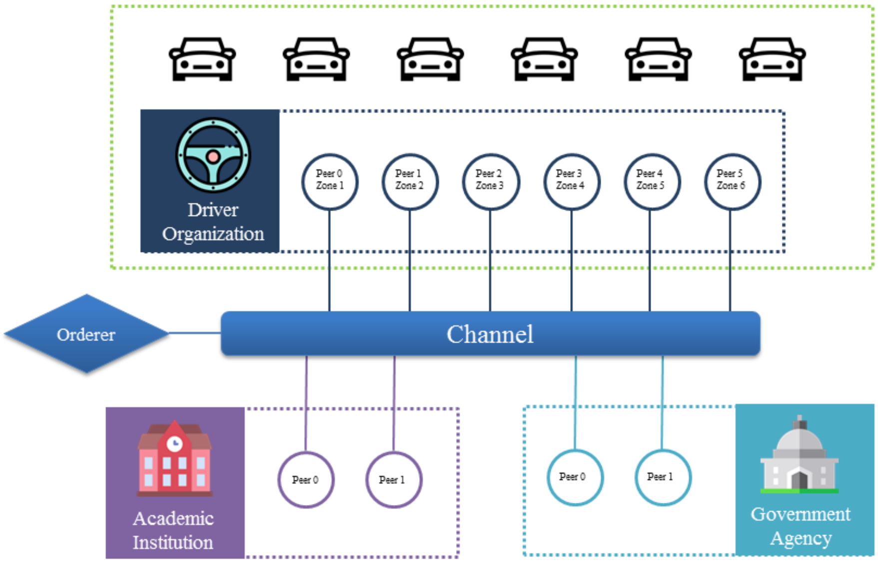

FIGURE 4 Topology of the proposed Blockchain network

FIGURE 5 presents a more detailed scenario where there is a person who drives to a city every day. She is a new user of this CAVs Blockchain platform. First of all, she is required to register herself in Blockchain as a legal driver with a driver's license. Then, she will be allowed to create a trip, register her car or wait for an automatic car attribute assigned by the system (There are 535 pre-registered vehicles in system). After receiving the car attributes and beginning the journey, the vehicle's status will turn to ACTIVE, and its trip status turns to DEPARTED. If the vehicle has a problem in the middle of the trip, the vehicle's status changes to OFF_THE_ROAD. All these updated statuses will be stored in Blockchain as transactions. Decision-makers will receive these updates. If they think the problem is valid, then they will update the corresponding status to IN_INCIDENT. When the problem is resolved, the vehicle's status will be switched back to ACTIVE. The driver can continue her journey and finally report her arrival to the Blockchain platform. All the transactions such as status changes and information registrations are stored in the CAVs Blockchain platform. Transactions move through the previously mentioned 8-layer framework.

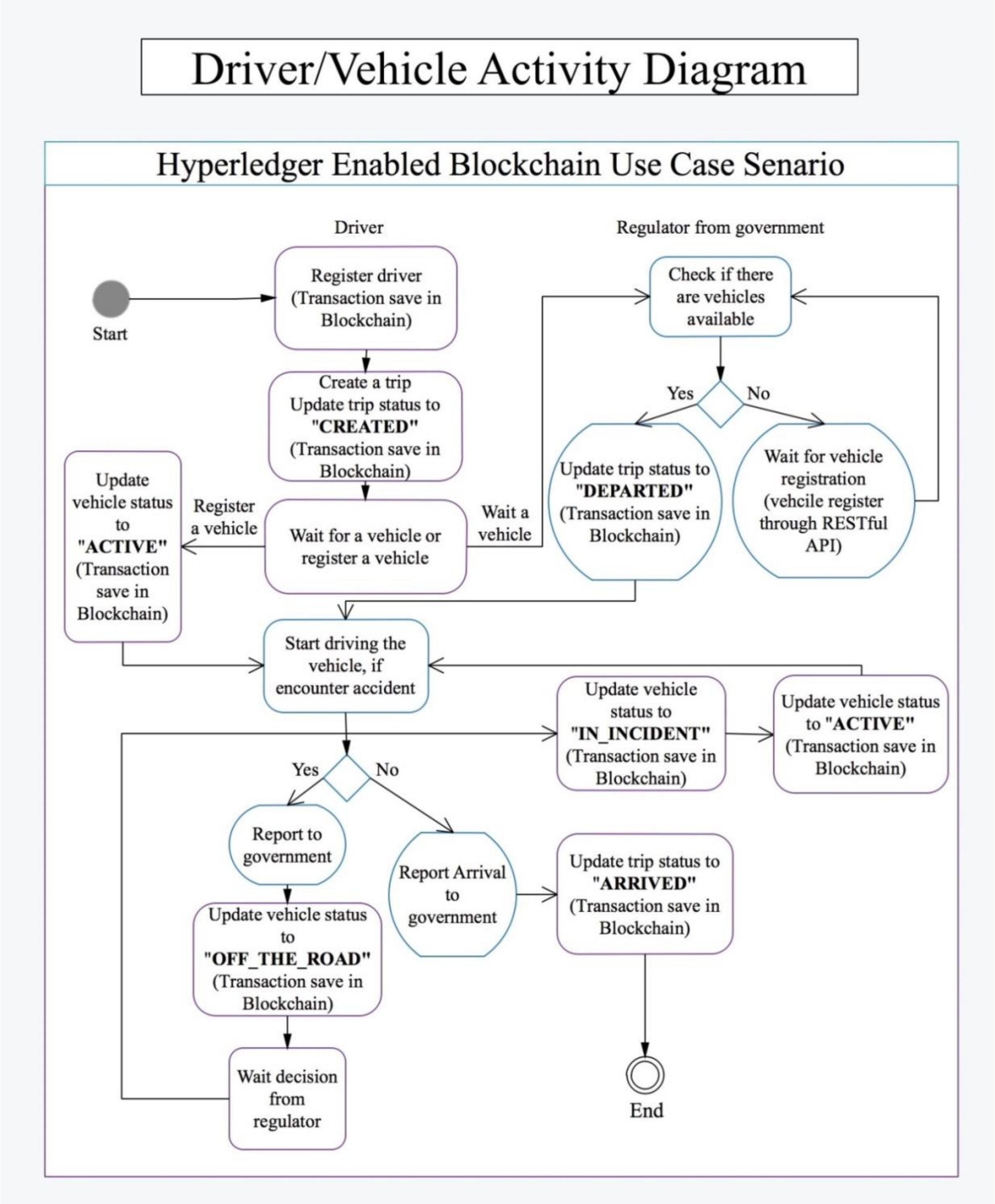

FIGURE 5 Driver-vehicle activity diagram example

Security, Privacy, and Scalability

The proposed 8-layer platform focuses on security, privacy and scalability aspects of CAV data transactions. It is necessary to introduce state-of-the-art technologies to appropriately address the increasing need for data exchange while ensuring safety and security. To increase safety, different numbers of subnetworks can be designed to establish private communication channels in the proposed platform. All the identifiers of transactions are encrypted and kept anonymous. This double layer of security makes the system challenging to hack. Several additional methods are implemented in the proposed CAVs Blockchain platform to best secure the data stored in Blockchain while guaranteeing user privacy, including:

- Incorporating a Data Encryption Layer
    - On a data encrypted layer every transaction and block's information is encapsulated into a hash code.
- Key Pairs
    - Every user needs to register before entering the network. Member services provide a generated key pair to let them join the network. A certificate (key) needs to be provided before getting access to the Blockchain network.
- Identity
    - Everyone in each organization is required to confirm their identity to access data from Blockchain.

It is challenging for attackers to break into the Blockchain. Cracking multiple complex processes and authentication layer of this platform is not only time consuming but it also requires an intensive source of computing power. A potential attack may include introducing a fake node or a computer, from which attacker may try to create a terminal with a wrong pseudonym certificate in order to get access to the CAVs data (*26*). With the proposed platform, such scenario is less likely to happen due to the existing authentication layer. All the members of the platform should be preregistered in the secured database. A pseudonym certificate is unlikely to be allowed to query information from the Blockchain platform.

Besides security and privacy, scalability is also a vital factor for the proposed platform. A channel is a private sub-network that connects different network members willing to share the same information and store confidential transactions. Using the same channel enhances the transmission speed of data exchange. Multiple channels can be added in the same network. Peers that belong to the same channel build up a subnetwork which holds the same ledger, while different channels cannot connect with each other directly without the authorization of the Blockchain network. The authorization policy could be given by a smart contract that is designed based on the application logic. This feature of the platform provides the capability of dividing the CAV network into several parts and makes it scalable. For example, different states may impose different speed limits, thus, each state may possess their own channel. In other words, they can create their own subnetworks to execute corresponding traffic rules, guarantee privacy, and enhance the transmission speed at the same time.

The hierarchical CAVs Blockchain framework is designed not only to enhance the security and privacy but also improve the scalability levels of CAV systems. Meanwhile, it could only store data that is frequently requested in the system while discarding uncommonly used data in a certain period.

Machine Learning as a Service Layer

It is possible and efficient to utilize machine learning algorithms to predict and identify future data. ML is commonly used in transportation studies to discover the hidden patterns of transportation data for traffic state estimation and prediction. Some applications include speed prediction, peak hour travel forecasting, accident analysis and prevention, incident management and response.

CAVs’ real-time information is usually complex and highly varible, vehicle speed can change rapidly. It is hard to find a certain pattern of the trend of such data with traditional technologies. Therefore, ML technologies such as simple supervised models (*27*) can be used to predict and analyse future data stored in the Blockchain platform. For instance, traffic speed can be predicted using the historical data stored in the platform real-time. Then, this predicted speed information can also be stored into Blockchain. The authorized user could use those data by querying information from Composer Rest API. A more detailed explanation will be shown in the next section. In this way, at the service layer, an ML platform could be established, and certain APIs will be provided to help others get analysis results from the Blockchain platform with the goal of alleviating traffic pressure and improving people's lives.

PERFORMANCE EVALUATION

In this section, two test case studies and the use case of the ML layer are conducted to test the performance of the CAVs Blockchain platform. The first case scenario is built on a single peer network, and the second case scenario is made of multiple peers. The detailed case scenario parameters are shown in TABLE 3.

TABLE 3 Case Scenario Parameters

Environment		Single peer	Multiple peers
Ubuntu 16.04 LTS operation system 64-bit processor 32GB memory workstation New York University local area network Download/Upload Speed: 100 Mbps	**Vehicle number**	535	535
	Total transactions in the chain	11706	11706
	Report real-time transaction in the chain	11167	11167
	Add asset (Vehicle)	535	535
	Basic setup (add an admin, issue identity, start a network and activate current identity)	4	4

Traffic Micro-simulation Model

The test network is coded using SUMO. The simulation network is extracted from a real network provided by using Open Street Maps (OSM). The details such as the number of lanes, lane width, and speed limit are edited using SUMO’s graphical network editor. FIGURE 6 shows the overall network and selected trajectory data collection section.

FIGURE 6 Selected Section in SUMO

Scenario 1: Single Peer Network

The single peer network consists of one peer, one channel and, one organization. Three programs are developed in Golang (*21*) for formatting and posting SUMO generated vehicle trajectory data to the Blockchain network. The developed program will first read data from a trajectory file, generate JSON array each 0.1 second, and then post them concurrently. There is no posting interval except the response latency from the previous request. FIGURE 7 illustrates the steps in the developed Golang code.

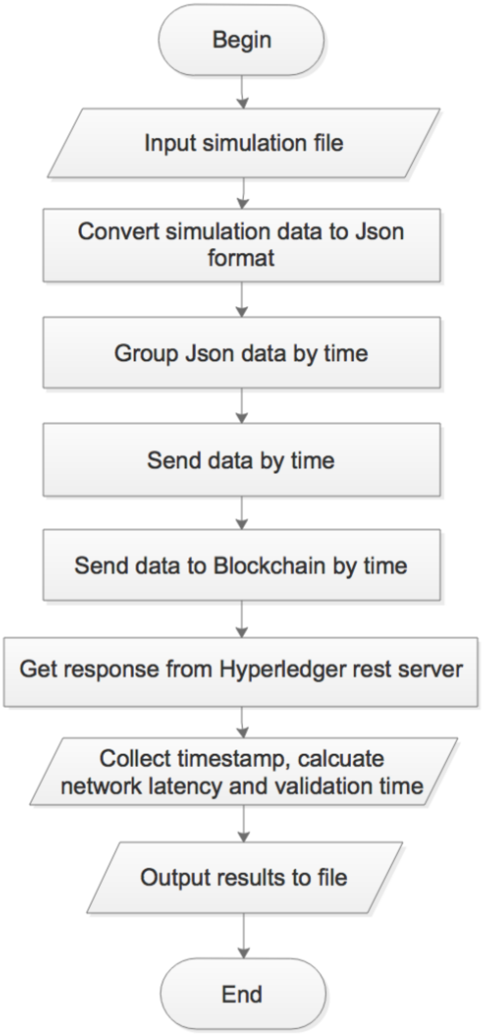

FIGURE 7 Golang Program flowchart

FIGURE 8 below shows the network latency and validation time for each vehicle to post real-time information including speed, acceleration, and location to the Blockchain network with rising block size. The test is conducted for 60 seconds.

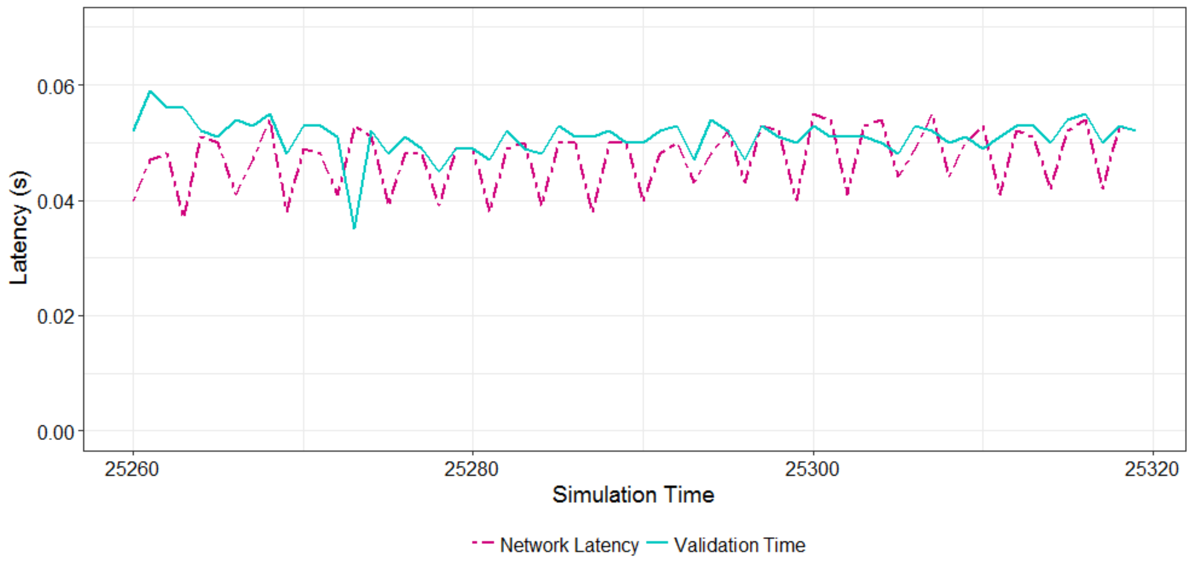

FIGURE 8 CAV single peer Blockchain network latency and validation time

From FIGURE 8, we can see that the overall network latency as well as validation time are all less than 0.1 seconds. Network latency represents the time period from Golang posting data time to the time when Blockchain receives data. While validation time stands for the time period which all transactions generate data in the Blockchain every 0.1 seconds. The latency trend is relatively stable with no apparent ups and downs, confirming the suitability of the proposed platform for real-world traffic applications.

Scenario 2: Multiple Peers Network

Different from the single peer network, the multiple peers network consists of 10 peers, one orderer peer, one channel, and three organizations. All the peers connect to the same channel and hold the same ledger. The orderer peer will package and distribute the valid transactions to each peer.

FIGURE 9 shows the network latency and validation time for the multiple peers network. It can be seen that the network latency is still below 0.1 seconds. The latency did not increase significantly with the increasing number of peers.

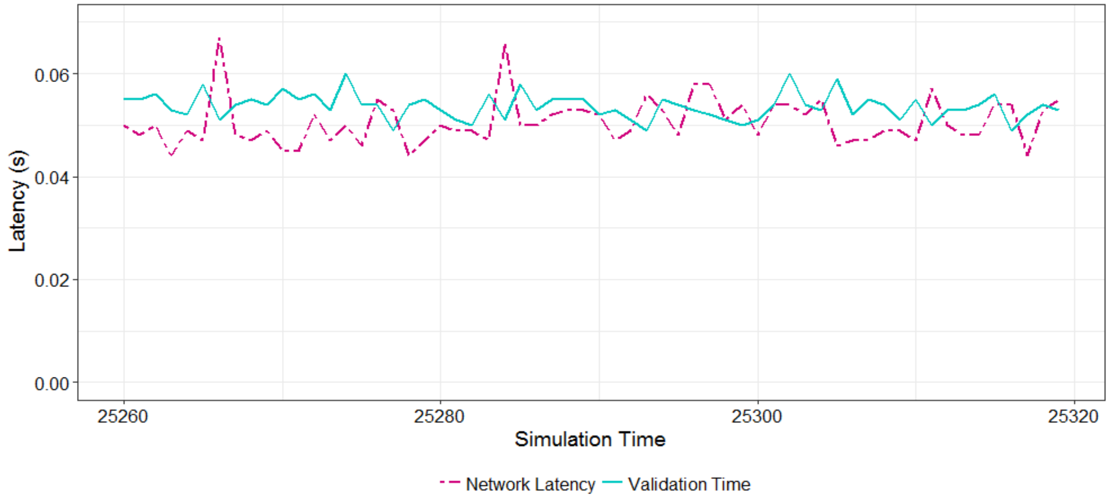

FIGURE 9 Performance for multiple peers Blockchain platform

Example Machine Learning for Data Analytics

This section introduces an example application located at the top layer of the platform to highlight the platform's capabilities. The goal of this experiment is to use the existing data in Blockchain to predict future vehicle speed. Speed is selected because the travel time in this section does not vary too much to properly evaluate the accuracy of the ML model. In the database, there are 999999 total records with features: location, lane, speed and acceleration, link and link id. The dataset is split into the training and testing sets. The first 92,520 transactions, or the first 9 minutes of data, sorted by transit time are used as the training dataset. The goal of this test is to predict the speed for each vehicle in the following 1 minute and compare the accuracy.

The trajectory data is pre-processed to remove the columns that are not used for the prediction to avoid overfitting. After pre-processing, labeled data are collected and formatted. Decision Tree Regression and Linear Regression models are used to fit training features and target training value. Since the training data is dispersed, it is challenging to find a particular trend for speed. However, Decision Tree Regression's max-depth parameter can be accurately set to alleviate this problem.

The accuracy metric used for the prediction is the coefficient of determination R^2 (*28*) in the test. The coefficient is equal to (1 - u/v), where u is the residual sum of squares and v is the total sum of squares. The test results show that using Decision Tree Regression is better than using Linear Regression in our case. Cross-validation score is also used in the experiment to better evaluate the performance. It helps to partition the sample data into complementary subsets, performing the analysis on the training set, and validating the analysis on the testing set. In the test scenario, five rounds of cross-validation are performed. An accuracy score is generated after each round. Then, the average score is taken to give a reasonable estimate of the predictive performance.

To investigate different market penetration levels, we randomly select 20% and 50% of the whole dataset and conduct the same test 10 times. The accuracy levels of the prediction for each run for different market penetration levels can be seen in TABLE 4.

TABLE 4 Accuracy comparison based on sample data

	1 vehicle (ID: 3683)	**535 vehicles 100% market penetration**	**535 vehicles 50% market penetration**	**535 vehicles 20% market penetration**
Training transit time period	~2 minutes (1332 transactions)	~9 minutes (921521 transactions)	~9 minutes (460723 transactions)	~9 minutes (184541 transactions)
Testing transit time period (Last 1 minute)	1 minute (600 transactions)	1 minute (78478 transactions)	1 minute (39277 transactions)	1 minute (15459 transactions)
Cross validation accuracy score	0.69	0.56	0.67	0.55

FIGURE 10 shows the comparison between the predicted value and observed value for 535 vehicles without excluding any data.

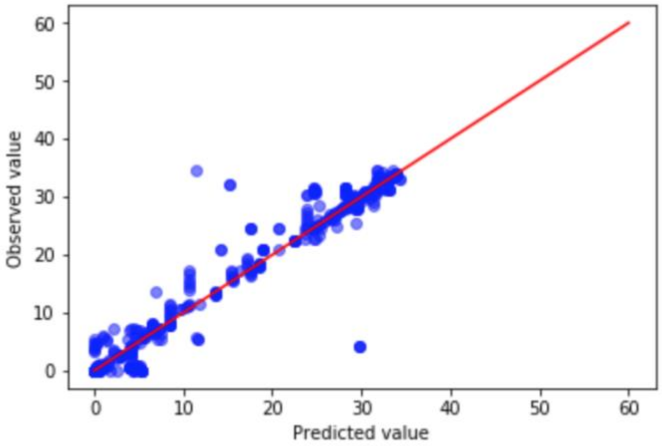

FIGURE 10 Predicted value compared with the observed value for 535 vehicles of the last 1 minute

Although test results show that the accuracy is above 50% with the full sample, the existing dataset is still dispersed. More features are needed to improve the accuracy of the prediction. Other machine learning models and technologies such as a Tensorflow neural network (*29*) and deep learning (*30*) can be used to improve the accuracy. These models are more flexible and more accurate regarding predicting discrete data.

CONCLUSIONS AND FUTURE WORK

The Blockchain is a newly emerging technology that has great potential for applications in the context of transportation systems. This study presents the development and computer based implementation of a simple yet realistic distributed ledger technology for transportation networks. The proposed architecture is based on a permissioned Blockchain (*5*). It is divided into different channels to restrict communication between authorized organizations. Several use case scenarios are evaluated to illustrate the applicability of the proposed framework. The framework is implemented under 8-layer architecture and tested using the trajectory data generated by a calibrated traffic micro-simulation model developed in SUMO. The computational test results show that the CAV system can be effectively combined with Blockchain technologies while enhancing security due to its distributed nature. The privacy of the users is preserved by using a data encrypted layer, key pairs and identity confirmation.

Distributed ledger technologies are more aligned with addressing the security need than traditional systems for more seamless and interconnected transport services. Although there are many early stage and exploratory efforts to use Blockchain in transportation by both industry and academia, more attention and work are needed to move beyond the presentation and discussion of conceptual frameworks towards actual implementation and testing. The first step of this type of implementation oriented approach is to use simulation models similar to the one presented in this paper. These simulations studies will lay the groundwork for more sophisticated and costly field studies with actual cars and users.

More detailed research and use cases are required to understand Blockchain’s scalability, speed, and security by conducting tests at different locations and using larger data sources. In CAV networks, safety data require real-time logging of big datasets and high volume data processing. This may require newer and faster distributed ledger technologies or alternative ways to ensure security. As future work, we will focus on ensuring security by incorporating more Blockchain capabilities such as fault tolerant consensus algorithms, generating automated ways to find network intrusions and forged transactions, and providing benefits such as cryptocurrencies to users to encourage the usage of the platform.

ACKNOWLEDGEMENTS

This research is supported by New York University, Tandon School of Engineering, C2SMART lab. The authors would like to thank Fan Zuo for sharing the trajectory data.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Abdullah Kurkcu, Yingxi Cao, Kaan Ozbay; Trajectory data collection and interpretation: Abdullah Kurkcu; Blockchain data collection and analysis: Yingxi Cao; Analysis and interpretation of results: Yingxi Cao, Abdullah Kurkcu; Draft manuscript preparation: Yingxi Cao, Abdullah Kurkcu, Kaan Ozbay. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Newell, G. F. Dispatching policies for a transportation route. *Transportation Science*, Vol. 5, No. 1, 1971, pp. 91-105.
2. Mohring, H. Optimization and scale economies in urban bus transportation. *The American Economic Review*, Vol. 62, No. 4, 1972, pp. 591-604.
3. De Araújo, M. S. M., and M. A. V. de Freitas. Acceptance of renewable energy innovation in Brazil—case study of wind energy. *Renewable and Sustainable Energy Reviews*, Vol. 12, No. 2, 2008, pp. 584-591.
4. Guo, Y., and C. Liang. Blockchain application and outlook in the banking industry. *Financial Innovation*, Vol. 2, No. 1, 2016, p. 24.
5. Androulaki, E., A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, and Y. Manevich. Hyperledger fabric: a distributed operating system for permissioned blockchains.In *Proceedings of the Thirteenth EuroSys Conference*, ACM, 2018. p. 30.
6. Morgul, E., H. Yang, A. Kurkcu, K. Ozbay, B. Bartin, C. Kamga, and R. Salloum. Virtual Sensors: Web-Based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2442, 2014, pp. 106-116.
7. Carpenter, C., M. Fowler, and T. Adler. Generating Route-Specific Origin-Destination Tables Using Bluetooth Technology. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2308, 2012, pp. 96-102.
8. Yuan, Y., and F.-Y. Wang. Towards blockchain-based intelligent transportation systems.In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, IEEE, 2016. pp. 2663-2668.
9. Wen, D., Y. Yuan, and X.-R. Li. Artificial societies, computational experiments, and parallel systems: an investigation on a computational theory for complex socioeconomic systems. *IEEE Transactions on Services Computing*, Vol. 6, No. 2, 2013, pp. 177-185.
10. Abedi, N., A. Bhaskar, and E. Chung. Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement.In *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*, 2013.
11. Singh, M., and S. Kim. Blockchain Based Intelligent Vehicle Data sharing Framework. *arXiv preprint arXiv:1708.09721*, 2017.
12. ---. Intelligent Vehicle-Trust Point: Reward based Intelligent Vehicle Communication using Blockchain. *arXiv preprint arXiv:1707.07442*, 2017.
13. ---. Crypto Trust Point (cTp) for Secure Data Sharing among Intelligent Vehicles. 2017.
14. Dorri, A., M. Steger, S. S. Kanhere, and R. Jurdak. Blockchain: A distributed solution to automotive security and privacy. *IEEE Communications Magazine*, Vol. 55, No. 12, 2017, pp. 119-125.
15. Lasla, N., M. Younis, W. Znaidi, and D. B. Arbia. Efficient Distributed Admission and Revocation using Blockchain for Cooperative ITS.In *9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2018*, IEEE, 2018. pp. 1-5.
16. Hermann, J., and M. Del Balso. Meet Michelangelo: Uber's machine learning platform. *URL* https://eng.uber.com/michelangelo, 2017.

17. Bonne, B., A. Barzan, P. Quax, and W. Lamotte. WiFiPi: Involuntary tracking of visitors at mass events.In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, IEEE, 2013. pp. 1-6.
18. Chavali, L., N. Prashanti, K. S. G. Rajasheker, and P. K. Kishor. The Emergence of Blockchain Technology and its Impact in Biotechnology, Pharmacy and Life Sciences. *Current Trends in Biotechnology & Pharmacy,* Vol. 12, No. 3, 2018.
19. Li, M., J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, J.-N. Liu, Y. Xiang, and R. H. Deng. CrowdBC: A Blockchain-based Decentralized Framework for Crowdsourcing. *IACR Cryptol. ePrint Arch., Univ. California, Santa Barbara, Santa Barbara, CA, USA, Tech. Rep,* Vol. 444, 2017, p. 2017.
20. Huh, S., S. Cho, and S. Kim. Managing IoT devices using blockchain platform.In *Advanced Communication Technology (ICACT), 2017 19th International Conference on*, IEEE, 2017. pp. 464-467.
21. Pike, R. The go programming language. *Talk given at Google's Tech Talks*, 2009.
22. Castro, M., and B. Liskov. Practical Byzantine fault tolerance.In *OSDI, No. 99*, 1999. pp. 173-186.
23. Garg, N. *Apache Kafka*. Packt Publishing Ltd, 2013.
24. Krajzewicz, D., G. Hertkorn, C. Rössel, and P. Wagner. SUMO (Simulation of Urban MObility)-an open-source traffic simulation.In *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002)*, 2002. pp. 183-187.
25. Merkel, D. Docker: lightweight Linux containers for consistent development and deployment. *Linux J.,* Vol. 2014, No. 239, 2014, p. 2.
26. Petit, J., and S. E. Shladover. Potential cyberattacks on automated vehicles. *IEEE Trans. Intelligent Transportation Systems,* Vol. 16, No. 2, 2015, pp. 546-556.
27. Bhavsar, P., I. Safro, N. Bouaynaya, R. Polikar, and D. Dera. Machine Learning in Transportation Data Analytics.In *Data Analytics for Intelligent Transportation Systems*, Elsevier, 2017. pp. 283-307.
28. Di Bucchianico, A. Coefficient of determination (R 2). *Encyclopedia of Statistics in Quality and Reliability,* Vol. 1, 2008.
29. Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard. Tensorflow: a system for large-scale machine learning.In.
30. LeCun, Y., Y. Bengio, and G. Hinton. Deep learning. *nature,* Vol. 521, No. 7553, 2015, p. 436.