National University of Computer and Emerging Sciences
(FAST-NUCES)

TRUST MANAGEMENT FRAMEWORK FOR REDUCING ABNORMAL DATA IN IoT USING BLOCKCHAIN

Project Supervisor
Mr. Shoaib Raza

Project Team

M. Ahsan Raza	20K-0137
M. Anas Khan	20K-0324
Hamdan Qureshi	20K-0327

Submitted in the partial fulfilment of the requirements for the degree of Bachelor of Science in Cyber Security.

Department of Cyber Security
National University of Computer and Emerging Sciences
(FAST-NUCES) Main Campus, Karachi

Spring 2024

Project Supervisor	Mr. Shoaib Raza
Project Team	M. Ahsan Raza 20K-0137 M. Anas Khan 20K-0324 Hamdan Qureshi 20K-0327
Submission Date	

Mr. Shoaib Raza ____________________________

Supervisor

Dr. Fahad Samad ____________________________

Head of Department

Department of Cyber Security
National University of Computer and Emerging Sciences
(FAST-NUCES) Main Campus, Karachi

Spring 2024

Acknowledgement

We are incredibly grateful to everyone who played a role in the success of this project. This journey has been a testament to the power of collaboration. Working together, we created a supportive environment where we could learn from each other and grow both individually and as a group. We are especially thankful to our group members for their invaluable contributions.

Our sincere appreciation goes to Mr. Shoaib Raza, our project supervisor, whose guidance, expertise, and patience were instrumental. His insights, feedback, and encouragement constantly motivated us to excel. We are also grateful to the university staff and faculty who provided essential resources, guidance, and facilities. Their willingness to answer our questions, big or small, and their insightful feedback helped us refine our research and approach challenges with confidence. Finally, we extend our heartfelt gratitude to our parents, families, and friends for their unwavering support and encouragement. Their belief in us and willingness to help kept us motivated and focused throughout this journey.

Document Information

Category	**Information**
Project Title	Trust Management Framework for Reducing Abnormal Data in IoT Using Blockchain
Supervisor	Mr. Shoaib Raza
Document Type	Final Year Project Report (FYP-Report)
Authors	M. Ahsan Raza, M. Anas Khan, Hamdan Qureshi
Approver(s)	Mr. Shoaib Raza
Project Category	Research and Development-Based (R&D)
Department	Department of Cyber Security
Issue Date	

Definition of Terms, Acronyms, and Abbreviations

Terms	Description
ML	Short for Machine Learning.
IoV	Short for Internet of Vehicle.
AI	Short for Artificial Intelligence.
RSUs	Short for Road Side Units.
CNN	Short for Convolutional Neural Network.
VANETs	Short for Vehicular Ad Hoc Networks.
Bad data/Abnormal data	A collective term given to malicious, incomplete or erroneous data.
IPFS	Short for Interplanetary File System.

Abstract

Blockchain, known for offering secure and immutable data storage, suffers from the inclusion of inaccurate or incomplete data, threatening data integrity. Prior solutions focused primarily on source authentication, leaving room for inaccurate or tampered data to enter the Blockchain. This work addresses this gap by proposing a Blockchain-based trust management framework. The framework leverages a Convolutional Neural Network model to analyze data characteristics and assess its trustworthiness. To comprehensively assess data trustworthiness before it enters the Blockchain, the framework incorporates a reputation management scheme that tracks the trust levels of vehicles and Road Side Units. While trained on a passive vehicular dataset for Internet of Vehicles systems, the framework's design allows for adaptation to different kinds of data, such as electronic health or transaction records encountered in diverse Blockchain-based applications.

Keywords

Internet of Vehicles, Trust Management, Data Integrity, Blockchain, Machine-Learning, Smart Contracts

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

Amid the ever-evolving landscape of computer science, innovative technologies are emerging as frontrunners. Among them, Blockchain technology has risen to prominence, presenting transformative solutions across diverse domains. Structuring data into discrete blocks linked through cryptographic hash values ensures transparency and immutability, secured by consensus mechanisms [1]. With industries undergoing digital transformation, Blockchain has found its way into various domains like supply chain [2], healthcare [3], [4], and vehicular communication [5].

However, despite its promises, a critical challenge persists—the potential inclusion of erroneous or malicious data into the Blockchain. While Blockchain technology guarantees the originality of data, it falls short when it comes to ensuring data accuracy, leaving room for errors or manipulation that can compromise data trustworthiness. This issue poses a significant risk, for example, leading to inaccuracy in healthcare data [6] that could impact patient care and research integrity or lead to inaccurate decisions, resulting in accidents in the case of the IoV.

Addressing this challenge requires continuous efforts and innovative solutions. While existing approaches focus on validating the source of information, the issue of data accuracy itself remains largely unaddressed. To bridge this gap, we propose a novel approach centred around leveraging ML techniques to validate erroneous data before its inclusion in the Blockchain. By incorporating ML into Blockchain systems, we aim to enhance data integrity and reliability, ultimately advancing the effectiveness of Blockchain technology across diverse domains.

2 RELATED WORKS

The dream of tamper-proof data storage with Blockchain technology faces a harsh reality: bad data inclusion. Inaccurate sensor readings, incomplete data, or maliciously injected information can disrupt Blockchain functionality and erode trust. Initially, researchers explored removing inaccurate data, but these methods conflicted with Blockchain technology's core principle of immutability. Consequently, the focus shifted towards proactive solutions. This literature review examines strategies for preventing bad data in Blockchain systems, specifically in Blockchain-enabled IoV setups. We explore AI and ML-based solutions, covering initial solutions centred around trust management schemes, data verification methods, and the role of trusted data sources. By analyzing these approaches, we aim to identify the most promising strategies for ensuring data integrity within Blockchain-IoV environments.

Continuing with the initial focus on removing bad data, Carvalho et al. [7] explored potential options for dealing with malicious or erroneous data stored on the Blockchain. Their research identified three possible reactive solutions: rollback, overturn, or do nothing. A rollback would essentially rewind the Blockchain to a state before the bad data was inserted. Overturning would involve creating a new transaction to invalidate such data. However, both these approaches encounter significant limitations. Rollbacks raised concerns about valid data transaction loss and the immutability of the Blockchain. Overturning might not be feasible in all scenarios, and it could bloat the ledger with additional data. Finally, the do-nothing approach accepted the presence of bad data, eroding trust in the system's integrity. Additionally, the study failed to explain how to identify such data in the Blockchain. Hence, the reactive approach was considered a failure concerning this issue.

Recognizing this limitation, Powell et al. [8] categorized this problem as a 'Garbage In Garbage Out' (GIGO) problem. To address this, they proposed a consensus-based approach for identifying faulty data within the Blockchain. Their solution leveraged data sharing among participating nodes. By comparing block hashes and Merkle tree root hashes, inaccuracies in the historical record can be detected. Building upon this concept, the research in [9] suggested employing multi-signature (multi-sig) protocols on the smart contract level. This approach shifted focus from purely technical consensus to a more social form of consensus, potentially involving multiple stakeholders in data validation.

However, even the most robust data consensus methods have limitations, as pointed out by Powell et al. [10]. The entire system relied on the data collected by peripheral sensors. However, these sensors are vulnerable to damage and tampering, which means that the captured data may not always be accurate. As a result, there is a risk of wrong data being considered correct due to consensus.

Recognizing the vulnerabilities of data sources, researchers proposed using oracles to feed data into the Blockchain [11], [12], [13]. These oracles collect data from trusted off-chain sources for on-chain processing using smart contracts. Voting-based and reputation-based methods are employed to ensure data reliability and accuracy. While oracles offer a way to reduce tampered or erroneous data, they introduce new challenges. A recent Blockchain oracle survey by Hassan et al. [14] highlighted these concerns, including single point of failure, data confidentiality issues, and additional dependencies.

While oracles might be practical for domains using static data, they fall short in fields with dynamic data, like the IoV. In such cases, the inherent risks associated with sensor data become unavoidable. Therefore, despite limitations, sensors remain the primary source of data in dynamic Blockchain applications.

Building upon the limitations of oracles for dynamic data, researchers explored trust schemes as an alternative approach. These schemes assigned trust scores to each node within the network. These scores were updated dynamically based on a node's behaviour and participation in network activities. The research [15] explored various trust factors, such as the environment and the goals achieved, that could be utilized for such updates. However, as highlighted in [16], relying solely on a single source for trust assignment can be unreliable. Therefore, trust updates should only occur when multiple sources provide evidence, acting as a form of collective verification. To achieve this, the study proposed a distributed trust metric-sharing mechanism among network participants. This allowed trust scores to be updated based on event validation and interaction with other nodes.

Yang et al. [17] and Shrestha et al. [18] presented two notable approaches for event validation to update trust scores in VANETs using Blockchain. Yang et al. [17] propose a two-pass validation approach where both RSUs and vehicles participate in validating an event, each using a different validation algorithm. This added an extra layer of security compared to relying on a single source. In contrast, Shrestha et al. [18] introduced a consensus-based approach where all vehicles within a designated proximity of the event, determined by the

RSU, participate in the validation process. This approach aimed to achieve consensus through collective verification by multiple trusted sources.

Similar to the two-pass validation approach by Yang et al. [17], Pu [19] proposes a multi-criteria decision-making framework that incorporates both vehicles and RSUs for trust evaluation. This framework involved vehicles checking the received message originator's trust value against a pre-set threshold. Based on the outcome, the vehicle decided whether to consider the source trustworthy and broadcast the message or not. Receiver vehicles then transmitted the source's trust value to the RSU, which recalculated the originator's reputation score based on the received trust values. This approach aimed to reduce the spread of fictitious messages.

Despite advancements in securing event data, both Yang et al. [17] and Shrestha et al. [18] solutions face distinct challenges. The two-pass validation approach introduces a potential single point of failure if a malicious actor compromises the RSU responsible for the event area. Additionally, running separate validation algorithms can strain computational resources on resource-constrained vehicles. Conversely, the consensus-based approach faces scalability issues and potential delays as the number of vehicles increases. Furthermore, it can be susceptible to data manipulation by a large number of malicious actors within the designated proximity [20]. While Pu's framework [19] addresses the single point of failure issue by utilizing distributed RSUs, the possibility of data tampering remains a concern in this approach as well.

Continuing the discussion on the limitations of existing approaches, Mao et al. [21] proposed a Software-Defined Vehicular Network (SDVN) to address communication delays and overcome the limitations of current trust management schemes in the dynamic environment of the IoV. This approach leveraged a hierarchical hybrid trust management architecture, essentially providing two methods for trust calculation. Vehicles within the coverage area of an RSU utilized a hybrid trust calculation that combined information from both the vehicle and the RSU's infrastructure-based trust management system. Conversely, vehicles outside the RSU range rely on a distributed trust management approach. This method evaluated the dynamic changes in trust based on a combination of factors, including trust between vehicles (direct interaction), auxiliary trust provided by the infrastructure (e.g., RSU) to the vehicle, static and dynamic information about the vehicle, and other relevant indicators.

While Mao et al.'s approach effectively overcomes communication delays associated with centralized models, it still suffers from limitations like a single point of failure and scalability issues, as highlighted in reviews of security techniques for Blockchain-IoV setups [22].

With the advent of ML and AI, new possibilities emerged for tackling data integrity and accuracy issues within Blockchain-based IoV solutions. Recent works showcase the potential of ML for trust evaluation and data integrity. For instance, Guleng et al. [23] present a fuzzy-logic-based approach coupled with Q-learning to evaluate direct trust based on factors like cooperativeness, honesty, and responsibility. This approach moves beyond static trust models by incorporating dynamic trust assessment.

Another promising direction involves leveraging random forests and clustering for workload management and improved accuracy, as presented in [24]. This approach grouped vehicles into clusters with designated representatives responsible for communicating with RSUs.

While it promotes joint decision-making to identify malicious actors, the reliance on the trustworthiness of cluster heads remains a significant drawback.

Building upon the potential of ML, Wang et al. [25] propose a system that leverages deep learning techniques for data authenticity verification in the context of the IoV. This approach takes advantage of the tamper-evident and transparent ledger provided by Blockchain technology.

Wang et al.'s system utilizes external factors like speed, location, and potentially other relevant data points to calculate trust scores and perform anomaly detection. While this approach effectively helps to identify the authenticity of the data source, it has limitations. Specifically, it focuses on source verification and may not necessarily guarantee the accuracy of the information itself. Therefore, the risk of inaccurate data entering the network persists.

Pre-filtering approaches focused solely on source authentication, like the work by Wang et al. [25], offer valuable benefits in terms of identifying malicious actors. However, they leave room for inaccurate data originating from compromised or malfunctioning sources to enter the Blockchain. This emphasizes the need for pre-filtering techniques based on data content to classify it into malicious and non-malicious categories. With this regard, the work of Ratnayake et al. [26] presents a significant advancement.

Ratnayake et al.'s [26] system employs a multi-step process for content verification, shifting focus from source authentication to data content analysis. First, incoming data is passed through the pre-processing stage to prepare it for efficient analysis by the CNN model. The pre-processed data is then fed into the trained CNN model with double validation being performed. CNN categorizes data as trustworthy or untrustworthy for potential exclusion from the Blockchain ledger. This solution addresses the problem of bad data by focusing on the content, ensuring that only reliable information is stored on the Blockchain.

The ongoing challenge of securing data within Blockchain-based IoV applications has fueled significant research efforts. While existing pre-filtering approaches offer valuable contributions, a crucial limitation lies in their primary focus on source authentication. In this context, the work of Ratnayake et al. [26] presents a groundbreaking advancement. Their research emphasizes the importance of data content analysis. This underscores the need for pre-filtering techniques that delve deeper than source verification and analyze the actual content of the data itself.

Considering the previous trust management frameworks for Blockchains, this research project focuses on data integrity within Blockchain-IoV systems via trust schemes. Our primary objective is to create a functional prototype that leverages a CNN model for data analysis within a trust management framework. This work aims to assess the effectiveness of the CNN model in identifying inaccurate or incomplete data before it enters the Blockchain. By evaluating the effectiveness of the CNN-based validation framework within a simulated Blockchain-IoV environment, this research contributes to advancements in securing data within these networks.

2.1 PROBLEM STATEMENT

Blockchain, while promising for Internet of Vehicles applications, suffers from data integrity issues. Existing solutions primarily focus on verifying the data origin, leaving the data content vulnerable to tampering. This lack of validation checks can lead to inaccurate or incomplete data entering the Blockchain, jeopardizing its trustworthiness and potentially causing real-world safety hazards. This research addresses this gap by leveraging Machine Learning for data content analysis within a trust management framework for Blockchain-enabled Internet of Vehicle applications.

3 METHODOLOGY

The cornerstone of Blockchain technology lies in trust and transparency, which are fundamentally dependent on the integrity of the data it stores. Inaccurate data on a Blockchain can disrupt trust and hinder its effectiveness. This section details the approach taken in this framework to address this critical challenge. We propose a solution that utilizes a permissioned Blockchain network, specifically a consortium Blockchain model, to ensure data integrity within an IoV environment. This framework leverages a two-stage validation process, where an ML model is employed to identify and rectify inconsistencies in sensor data transmitted by vehicles.

3.1 DATASET AND MODEL

This framework utilizes the "PVS - Passive Vehicular Sensors Datasets" from Kaggle [27], which offers sensor data collected from three different vehicles driven by three distinct drivers under three different conditions. This dataset is a collection of accelerometer data, gyroscope data, and speed.

We leverage the best-performing model identified in the study [28] for road-surface classification using the CNN model as our pre-trained model. The best model uses the six datasets under PVS for training purposes, while the remaining three serve as test data. The model architecture comprises three convolutional layers, each with 128 filters and a kernel size of 5 with Rectified Linear Unit (ReLU) activation function with SoftMax activation function in the output layer.

3.2 SOLUTION OVERVIEW

This section details the core functionalities of the proposed framework designed to ensure data veracity in an IoV scenario. An overview of the solution is presented in Figure 1, which shows how different components interact to mitigate the inclusion of inaccurate data. The framework leverages a two-stage validation process facilitated by a pre-trained ML model deployed on edge servers and a validator. Initially, all vehicles and edge servers have a moderate reputation score. The data validation process within the framework follows a series of steps which collectively form the two-stage validation framework:

1. **Data Transmission:**
   Vehicles transmit sensor data to nearby edge servers acting as RSUs, simulating real-world data collection. Each edge server prioritizes data from vehicles within its limited communication range to optimize network resources.

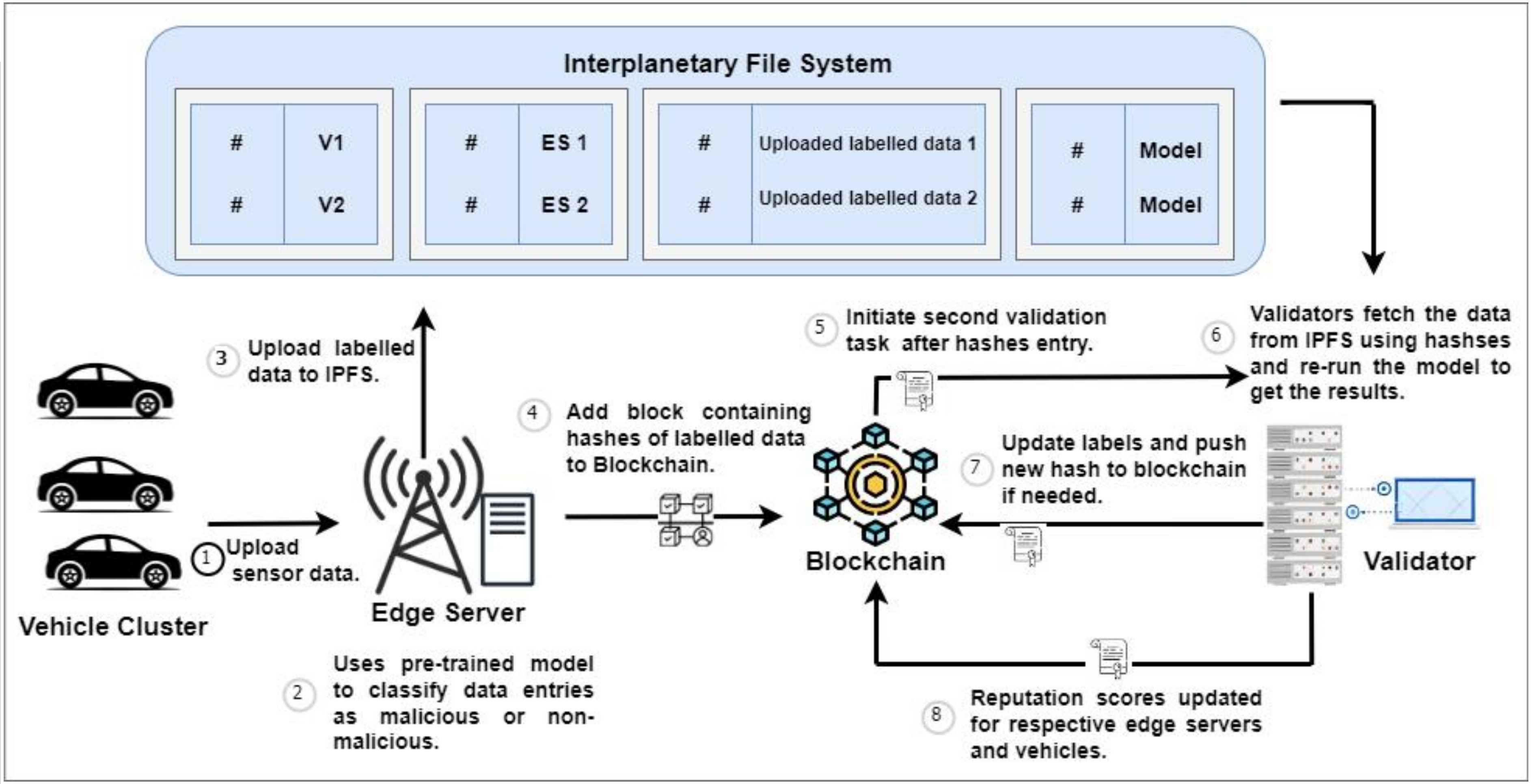
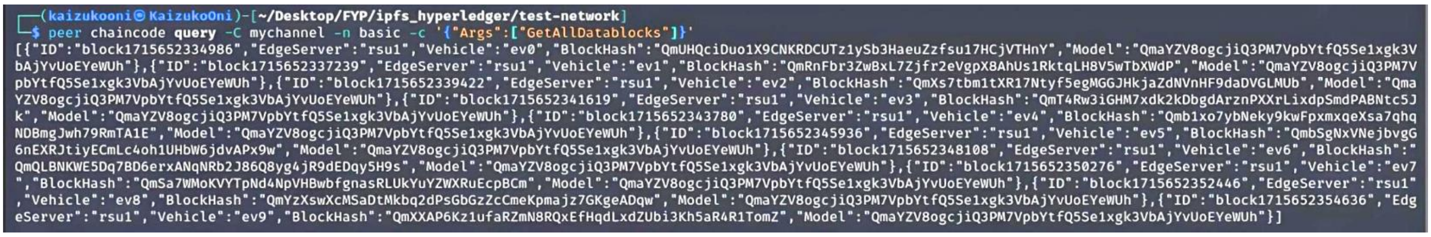

Interplanetary File System
#
V1
#
V2
#
ES 1
#
ES 2
#
Uploaded labelled data 1
#
Uploaded labelled data 2
#
Model
#
Model
Vehicle Cluster
1 Upload sensor data.
Edge Server
2 Uses pre-trained model to classify data entries as malicious or non-malicious.
3 Upload labelled data to IPFS.
4 Add block containing hashes of labelled data to Blockchain.
Blockchain
5 Initiate second validation task after hashes entry.
6 Validators fetch the data from IPFS using hashses and re-run the model to get the results.
7 Update labels and push new hash to blockchain if needed.
Validator
8 Reputation scores updated for respective edge servers and vehicles.

Figure 1: Overview of the proposed solution for the IoV context.

2. **Initial Validation by Edge Servers:**
   Edge servers are responsible for checking the validity of incoming data. They perform this check using a pre-trained ML model that analyzes sensor data features. The model classifies road surfaces into three types: Dirt, Asphalt, or Cobblestone. Based on the majority classification of data entries, the model decides which data entries are malicious or non-malicious. For instance, if most data entries classify a road as dirt, those entries are considered non-malicious, whereas others are marked as malicious.

3. **Data Storage:**
   Based on the edge server's classification, the data is labelled accordingly. Labelled data is then pushed to a decentralized storage such as IPFS to ensure data immutability and make it tamper-proof. For efficiency, only the cryptographic hash of the uploaded data is added to the Blockchain. This hash acts like a fingerprint that verifies the data's existence, as illustrated in Figure 2. The figure depicts a simplified representation of a Blockchain ledger with several transactions. Each transaction includes a data hash reference, not the entire data itself.



```
┌──(kaizukooni㉿KaizukoOni)-[~/Desktop/FYP/ipfs_hyperledger/test-network]
└─$ peer chaincode query -C mychannel -n basic -c '{"Args":["GetAllDatablocks"]}'
[{"ID":"block1715652334986","EdgeServer":"rsu1","Vehicle":"ev0","BlockHash":"QmUHQciDuo1X9CNKRDCUTz1ySb3HaeuZzfsu17HCjVTHnY","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3V
bAjYvUoEYeWUh"},{"ID":"block1715652337239","EdgeServer":"rsu1","Vehicle":"ev1","BlockHash":"QmRnFbr3ZwBxL7Zjfr2eVgpX8AhUs1RktqLH8V5wTbXWdP","Model":"QmaYZV8ogcjiQ3PM7V
pbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652339422","EdgeServer":"rsu1","Vehicle":"ev2","BlockHash":"QmXs7tbm1tXR17Ntyf5egMGGJHkjaZdNVnHF9daDVGLMUb","Model":"Qma
YZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652341619","EdgeServer":"rsu1","Vehicle":"ev3","BlockHash":"QmT4Rw3iGHM7xdk2kDbgdArznPXXrLixdpSmdPABNtc5J
k","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652343780","EdgeServer":"rsu1","Vehicle":"ev4","BlockHash":"Qmb1xo7ybNeky9kwFpxmxqeXsa7qhq
NDBmgJwh79RmTA1E","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652345936","EdgeServer":"rsu1","Vehicle":"ev5","BlockHash":"QmbSgNxVNejbvgG
6nEXRJtiyECmLc4oh1UHbW6jdvAPx9w","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652348108","EdgeServer":"rsu1","Vehicle":"ev6","BlockHash":"
QmQLBNKWE5Dq7BD6erxANqNRb2J86Q8yg4jR9dEDqy5H9s","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652350276","EdgeServer":"rsu1","Vehicle":"ev7
","BlockHash":"QmSa7WMoKVYTpNd4NpVHBwbfgnasRLUkYuYZWXRuEcpBCm","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652352446","EdgeServer":"rsu1"
,"Vehicle":"ev8","BlockHash":"QmYzXswXcMSaDtMkbq2dPsGbGzZcCmeKpmajz7GKgeADqw","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"},{"ID":"block1715652354636","Edg
eServer":"rsu1","Vehicle":"ev9","BlockHash":"QmXXAP6Kz1ufaRZmN8RQxEfHqdLxdZUbi3Kh5aR4R1TomZ","Model":"QmaYZV8ogcjiQ3PM7VpbYtfQ5Se1xgk3VbAjYvUoEYeWUh"}]
```

Figure 2: Blockchain ledger state after uploading data hashes.

4. **Smart Contract Triggered Re-validation:**
   A smart contract on the Blockchain automatically triggers a second validation stage whenever a new data hash is uploaded.

5. **Validator Re-assessment and Resolution:**
   The validator accesses the pending validation requests generated by smart contracts to perform a second validation. It retrieves the actual data from IPFS using the stored hashes and then re-evaluates the classification using the same model.

   - If the edge server's classification matches the validator's assessment (both malicious and non-malicious), no changes related to classification take place.
   - If they differ, the validator's assessment takes precedence to ensure data accuracy. In such cases, a new data entry with the corrected label and the original data is uploaded to IPFS, and its corresponding hash is added to the Blockchain, creating an audit trail.

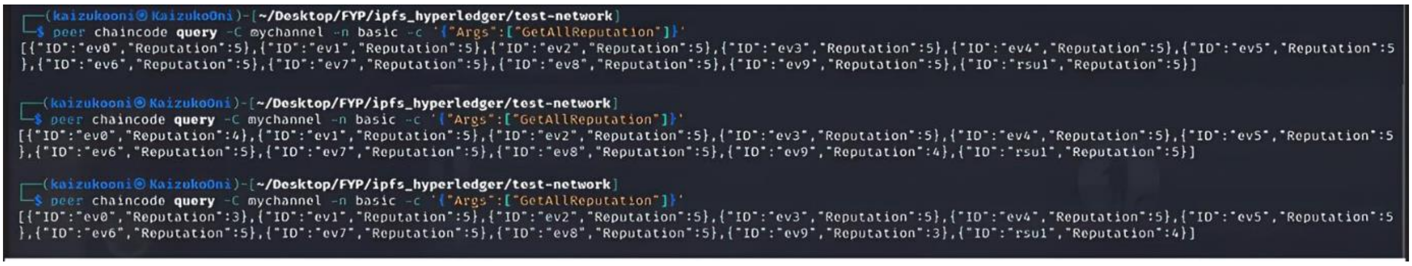
6. **Trust Score Updates:**
   The trust scores of both vehicles and edge servers are dynamically adjusted using a smart contract, considering the outcome of step 5. The reputation score is adjusted for vehicles based on the quality of the data they provide and for edge servers

according to the accuracy of the labels and classifications they assign. This dynamic adjustment results in three possible trust update scenarios:

- **Correct Data and Consistent Validation Results:** No change in trust score for either vehicles or edge servers.
- **Inaccurate Data but Consistent Validation Results:** The trust score decreases for vehicles with incorrect data, while the RSU score remains unchanged.
- **Inaccurate Data and Contradicting Validation Results:** Trust score decreases for both vehicles and the RSU.

Building upon the trust update scenarios mentioned above, Figure 3 visually depicts the dynamic adjustments in reputation scores for ten sample vehicles with an initial reputation equal to 5. This figure showcases the impact of different validation outcomes on their trust scores, with each chain code query providing the trust score frame for three different situations.

- **Scenario 1 (All Correct):** The first chain code query retrieves a trust score frame, indicating a successful validation for the vehicle data and RSU classification. Consequently, vehicles and RSU maintain their initial score of 5.
- **Scenario 2 (Incorrect Data Provision):** The second chain code query retrieves a trust score frame that reflects a discrepancy between the data provided by specific vehicles (EV0 and EV9 in this case) and the majority classification. As a result, the trust score for these vehicles decreases (from 5 to 4), while the scores of the vehicles with accurate data and RSU remain unchanged.
- **Scenario 3 (Inaccurate Data and Mismatched Validation):** The third chain code query retrieves a trust score frame that reveals both inaccurate data from vehicles (EV0 and EV9 in this case) and a failing initial validation by the RSU (RSU1). Due to this error, the trust score for these vehicles experiences a decrease of one. Additionally, the RSU's trust score is also decremented by one due to its failure to identify the discrepancy during initial validation.



```
┌──(kaizukooni㉿KaizukoOni)-[~/Desktop/FYP/ipfs_hyperledger/test-network]
└─$ peer chaincode query -C mychannel -n basic -c '{"Args":["GetAllReputation"]}'
[{"ID":"ev0","Reputation":5},{"ID":"ev1","Reputation":5},{"ID":"ev2","Reputation":5},{"ID":"ev3","Reputation":5},{"ID":"ev4","Reputation":5},{"ID":"ev5","Reputation":5
},{"ID":"ev6","Reputation":5},{"ID":"ev7","Reputation":5},{"ID":"ev8","Reputation":5},{"ID":"ev9","Reputation":5},{"ID":"rsu1","Reputation":5}]

┌──(kaizukooni㉿KaizukoOni)-[~/Desktop/FYP/ipfs_hyperledger/test-network]
└─$ peer chaincode query -C mychannel -n basic -c '{"Args":["GetAllReputation"]}'
[{"ID":"ev0","Reputation":4},{"ID":"ev1","Reputation":5},{"ID":"ev2","Reputation":5},{"ID":"ev3","Reputation":5},{"ID":"ev4","Reputation":5},{"ID":"ev5","Reputation":5
},{"ID":"ev6","Reputation":5},{"ID":"ev7","Reputation":5},{"ID":"ev8","Reputation":5},{"ID":"ev9","Reputation":4},{"ID":"rsu1","Reputation":5}]

┌──(kaizukooni㉿KaizukoOni)-[~/Desktop/FYP/ipfs_hyperledger/test-network]
└─$ peer chaincode query -C mychannel -n basic -c '{"Args":["GetAllReputation"]}'
[{"ID":"ev0","Reputation":3},{"ID":"ev1","Reputation":5},{"ID":"ev2","Reputation":5},{"ID":"ev3","Reputation":5},{"ID":"ev4","Reputation":5},{"ID":"ev5","Reputation":5
},{"ID":"ev6","Reputation":5},{"ID":"ev7","Reputation":5},{"ID":"ev8","Reputation":5},{"ID":"ev9","Reputation":3},{"ID":"rsu1","Reputation":4}]
```

Figure 3: Dynamic trust score updates.

To sum up, this two-stage validation framework with a reputation system offers a promising approach to improving data quality in the Blockchain over time. The framework divides the validation process into two stages. During the initial stage, edge servers equipped with a pre-trained ML model classify incoming data and label them based on the majority. Subsequently, a second validation stage triggered by a smart contract leverages an external validator for re-evaluation. As the vehicles and edge servers consistently contribute accurate data, they maintain their reputation. Conversely, contributions from vehicles and edge servers with low reputation scores are less likely to be trusted.

As established, the two-stage validation process effectively updates reputation scores based on the results obtained during data validation. The framework implements a data access threshold based on reputation scores. This threshold acts as a safeguard, preventing entities with consistently unreliable data contributions from jeopardizing network integrity. When a participating entity, such as a vehicle or an RSU, reputation score falls below a predefined threshold, the framework restricts its ability to submit new data entries to the Blockchain.

4 RESULTS

To assess the effectiveness of the proposed validation framework in achieving data integrity a prototype was developed using Hyperledger Fabric 2.5 and Golang. Additionally, Python was used to implement functionalities related to validator and edge server processing during the validation stages. This approach facilitated the evaluation process by simulating real-world interactions between vehicles transmitting sensor data, edge servers performing initial validation, and validators participating in the second validation stage (refer to Figure 4 for the prototype model structure).

The Passive Vehicular Sensors (PVS) Datasets [27] were used to train and evaluate a pre-trained CNN model [28] for the road surface-type classification task. Extracting subsets of field, normalizing values and reshaping data to suit the model formed the preprocessing stage for data preparation. Smart contracts were implemented to carry out the following tasks:

- To activate validation tasks on data entries added to the Blockchain.
- To add new corrected entries to the Blockchain when mismatches are found.
- To update the reputation scores of vehicles and edge servers.

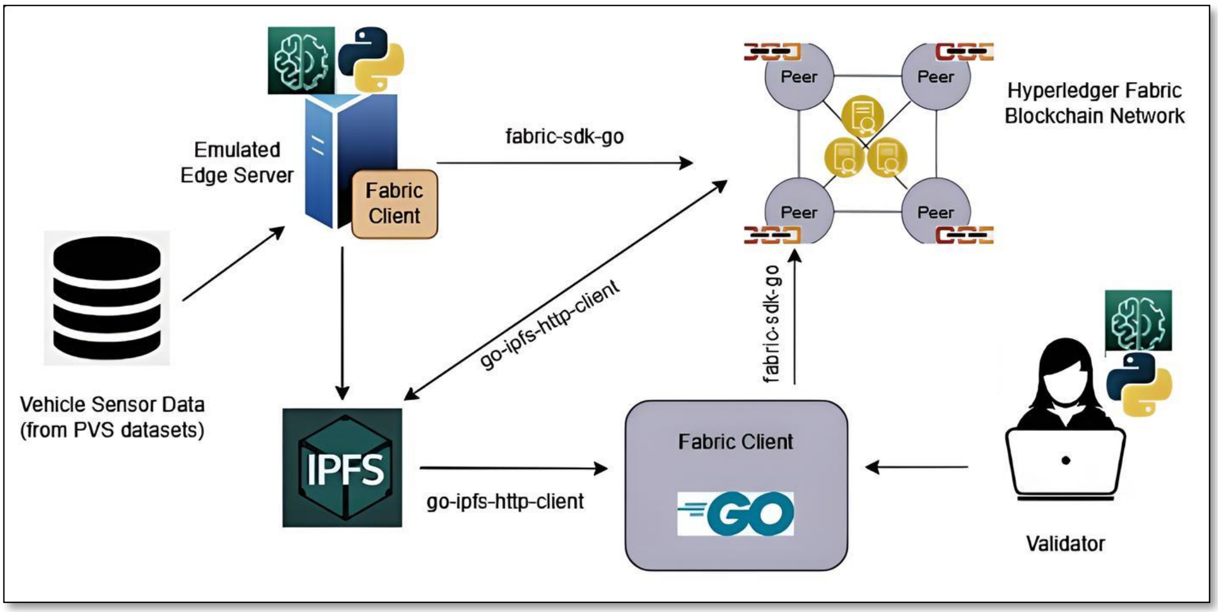


Figure 4: Prototype implementation.

4.1 Key Findings

Building upon the prototype described above (refer to Figure 4), this section delves into the key findings obtained during the evaluation of the proposed trust management framework. Our primary focus lies on assessing the effectiveness of the framework in achieving data integrity within the network, with a particular emphasis on the trust update mechanism.

4.1.1 TWO-STAGE VALIDATION: IMPACT ON REPUTATION SCORES

The evaluation investigated the framework's performance under different concentrations of malicious actors within simulated vehicle clusters. Three scenarios were evaluated: clusters with 10, 25, and 50 vehicles, where a portion of the vehicles were designated as malicious, intentionally transmitting incorrect data. The trust scores for vehicles and edge servers either remain unchanged or are altered by subtracting an arbitrary value to reflect their contribution in the simulated environment. These trust score updates followed the patterns expected based on the two-stage validation process, as outlined in Table 1.

Table 1: Trust Score Update Table.

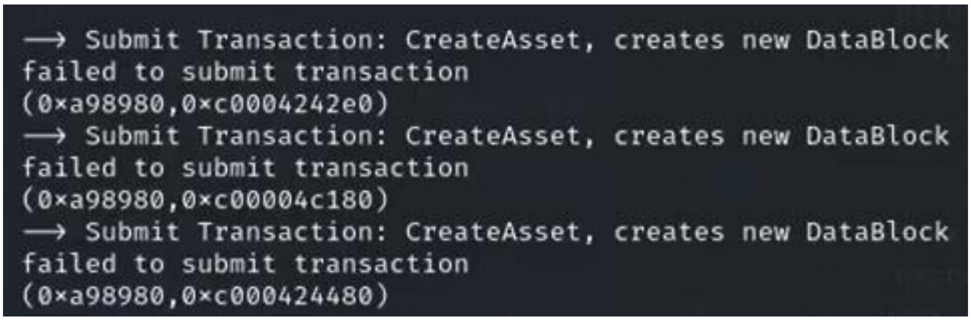
Vehicle's Sensor Data		**Edge Server's Validation** (Majority Favors)	**Validator's Validation Result**	**Vehicle's Reputation**		**Edge Server's Reputation**
A	**B**			**A**	**B**	
Malicious	Malicious	Malicious	Malicious	Unchanged	Unchanged	Unchanged
			Non-Malicious	Decreases	Decreases	Decreases
		Non-Malicious	Malicious	Unchanged	Unchanged	Decreases
			Non-Malicious	Decreases	Decreases	Unchanged
Malicious	Non-Malicious	Malicious	Malicious	Unchanged	Decreases	Unchanged
			Non-Malicious	Decreases	Unchanged	Decreases
		Non-Malicious	Malicious	Unchanged	Decreases	Decreases
			Non-Malicious	Decreases	Increases	Unchanged
Non-Malicious	Malicious	Malicious	Malicious	Decreases	Unchanged	Unchanged
			Non-Malicious	Unchanged	Decreases	Decreases
		Non-Malicious	Malicious	Decreases	Unchanged	Decreases
			Non-Malicious	Unchanged	Decreases	Unchanged
Non-Malicious	Non-Malicious	Malicious	Malicious	Decreases	Decreases	Unchanged
			Non-Malicious	Unchanged	Unchanged	Decreases
		Non-Malicious	Malicious	Decreases	Decreases	Decreases
			Non-Malicious	Unchanged	Unchanged	Unchanged

Table 1 presents how reputation scores are updated for two example vehicles, A and B, based on the assessments conducted during the two-stage validation process. The table shows vehicles A and B sending malicious and non-malicious data to the edge server. Edge server passes a ruling in favour of either malicious or non-malicious data based on the majority classifications obtained during the initial validation stage. Following this is the validator's assessment, which leads to the reputation score updates via smart contract if required. It highlights the impact of both the initial edge server assessment and the final validator's verdict on the reputation scores of the vehicles and the edge server involved.

The information presented in Table 1 applies to all vehicles within the framework. Each vehicle's reputation score is updated based on the assessments conducted during the two-stage validation process, following the patterns outlined in the table. It is important to note that the two-stage validation framework relies on a majority voting approach, making it independent of the chosen model for the initial validation stage.

4.1.2 DATA ACCESS THRESHOLD

To assess the practical impact of data access thresholds, we simulated the behaviour of an RSU (Roadside Unit) with a trust value of 0. As depicted in Figure 5, the simulation resulted in failed transactions for the RSU. This scenario reflects the consequence of a trust score falling below the predefined data access threshold within the framework.



```
⟶ Submit Transaction: CreateAsset, creates new DataBlock
failed to submit transaction
(0×a98980,0×c0004242e0)
⟶ Submit Transaction: CreateAsset, creates new DataBlock
failed to submit transaction
(0×a98980,0×c00004c180)
⟶ Submit Transaction: CreateAsset, creates new DataBlock
failed to submit transaction
(0×a98980,0×c000424480)
```

Figure 5: RSU transaction failures due to Data Access Threshold.

The specific value of the data access threshold is a critical parameter that can be fine-tuned based on the specific network application's needs. A stricter threshold (higher threshold value) allows only highly reputable entities to submit data, maximizing data quality but potentially limiting participation. Conversely, a less strict threshold (lower values) means more data submissions but introduces a risk of unreliable data entering the system. Finding the optimal balance between data quality and participation is crucial for maximizing the framework's effectiveness.

4.2 DISCUSSION

4.2.1 COMPARISON

ML-powered data validation with a trust scheme presents a potential approach to address the data integrity issue. To evaluate the effectiveness of this approach by comparing our prototype (based on the ML-based validation approach presented in [26]), which implements this technique, to other pre-filtering solutions. This comparison, detailed in Table 2, focuses on characteristics relevant to bad data reduction within Blockchain-enabled IoV systems.

Table 2: Comparison of Solutions.

	Techniques						
Characteristics	**Oracles-based Solutions** ([11], [12], and [13])	**Two-Pass Validation** ([17])	**Consensus-based Validation** ([18])	**Multi-criteria Decision** ([19])	**Clustering with Random Forest** ([24])	**Data Authenticity Verification** ([25])	**Content Validation with Trust Scheme** ([26])
Includes Bad Data Identification		✓				✓	✓
Source Verification	✓	✓	✓	✓		✓	✓
Content Analysis		✓				✓	✓
Bad Data Reduction method	✓	✓	✓	✓	✓	✓	✓
IoT Compatible					✓	✓	✓

Another important aspect is to consider the scalability and resource consumption tradeoff. Different pre-filtering techniques offer varying levels of scalability and resource consumption. Techniques like oracles provide excellent scalability but lack advanced anomaly detection capabilities. Conversely, solutions prioritizing robust anomaly detection, such as content verification with CNNs, demand significant resources. This shortcoming highlights the need to balance scalability and resource consumption to ensure the widespread adoption and real-world viability of secure data management solutions within resource-constrained Blockchain-enabled IoV applications.

4.2.2 LIMITATIONS

Despite demonstrably improving data integrity, this framework has limitations. Pre-filtering may not be foolproof, requiring further mitigation strategies against sophisticated attacks. Additionally, the majority vote-based validation assumes a network dominated by honest actors, highlighting the need for robust network security. Finally, the security of smart contracts used for trust management is critical, as vulnerabilities could undermine trust. Addressing these limitations is crucial for real-world deployments.

5 CONCLUSION

Blockchain technology has ushered in a paradigm shift in data storage, promising immutability and transparency. However, the susceptibility of bad data remains a significant roadblock hindering its full potential. The proliferation of this roadblock compromises the foundations of trust and security that Blockchain aims to establish. Leveraging a pre-trained CNN model and a reputation management scheme, our proposed trust management framework contributes to this ongoing effort by ensuring data integrity within Blockchain ecosystems.

The ability to ensure data integrity within Blockchain systems opens the door to a new era of transformative decentralized applications characterized by inherent data accuracy and integrity. Further exploration, as outlined in the future work section, is essential for fully realizing the framework's potential and fostering wider adoption of secure and transparent Blockchain-based systems.

6 FUTURE WORK

While a functional prototype of the trust management framework has been developed and tested using a dataset, real-world evaluation is crucial for assessing its effectiveness in practical scenarios. Integrating the framework with an IoV testbed simulating real-time vehicle data streams will allow us to assess its accuracy, practicality, and efficiency in a more realistic setting. Additionally, exploring the value of incorporating data from additional sensor readings, such as weather forecasts, GPS coordinates, and other vehicle sensor data, holds promise for improving the framework's ability to distinguish trustworthy from erroneous data. A pilot study can be conducted to assess the effectiveness of incorporating a broader range of sensor data in the framework's analysis process.

As the framework expands to accommodate more vehicles, privacy concerns will become apparent. Federated learning offers a promising solution, enabling the pre-trained CNN model to learn from a broader data pool without compromising individual vehicle data privacy. Finally, for large-scale deployments managing vast vehicle networks, Cloud integration can provide the necessary scalability and flexibility. Exploring Cloud-based platform integration will pave the way for the framework's deployment in real-world, large-scale Blockchain applications.

By pursuing these future work directions, we can further refine and strengthen the proposed framework, ultimately contributing to the development of more robust and secure Blockchain-based systems, and fostering wider adoption of this transformative technology across various domains.

7 REFERENCES

[1] H. K. Cheng, D. Hu, T. Puschmann, and J. L. Zhao, "The landscape of Blockchain research: impacts and opportunities," Information Systems and e-Business Management, vol. 19, no. 3, pp. 749-755, Sep. 2021, doi: 10.1007/s10257-021-00508-6.

[2] R. Ahmad, H. Hasan, R. Jayaraman, K. Salah, and M. Omar, "Blockchain applications and architectures for port operations and logistics management," Research in Transportation Business & Management, vol. 41, p. 100620, 2021, doi: 10.1016/j.rtbm.2021.100620.

[3] J. Lee, C. Chew, J. Liu, Y. Chen, and K. Tsai, "Medical Blockchain: Data sharing and privacy-preserving of EHR based on smart contract," Journal of Information Security and Applications, vol. 65, p. 103117, 2022, doi: 10.1016/j.jisa.2022.103117.

[4] H. Al-Aswad, W. M. El-Medany, C. Balakrishna, N. Ababneh, and K. Curran, "BZKP: Blockchain-based zero-knowledge proof model for enhancing healthcare security in Bahrain IoT smart cities and COVID-19 risk mitigation," Arab Journal of Basic and Applied Sciences, vol. 28, no. 1, pp. 154-171, 2021, doi: 10.1080/25765299.2020.1870812.

[5] N. Aung, T. Kechadi, T. Zhu, S. Zerdoumi, T. Guerbouz, and S. Dhelim, "Blockchain Application on the Internet of Vehicles (IOV)," in 2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE), Beijing, China, 2022, pp. 586-591, doi: 10.1109/ICITE56321.2022.10101404.

[6] M. Gaynor, J. Tuttle-Newhall, J. Parker, A. Patel, and C. Tang, "Adoption of Blockchain in health care," Journal of Medical Internet Research, vol. 22, no. 9, p. e17423, 2020, doi: 10.2196/17423.

[7] A. Carvalho, J. W. Merhout, Y. Kadiyala, and J. Bentley II, "When good blocks go bad: Managing unwanted Blockchain data," International Journal of Information Management, vol. 57, p. 102263, 2021, doi: 10.1016/j.iinm.2021.102263.

[8] W. Powell, M. Foth, S. Cao, and V. Natanelov, "Garbage in garbage out: The precarious link between IoT and Blockchain in food supply chains," Journal of Industrial Information Integration, vol. 25, pp. 100261, 2022, doi: 10.1016/j.jii.2021.100261.

[9] W. Powell, S. Cao, T. Miller, M. Foth, X. Boyen, B. Earsman, S. del Valle, and C. Turner-Morris, "From premise to practice of social consensus: How to agree on common knowledge in Blockchain-enabled supply chains," Computer Networks, vol. 200, pp. 108536, 2021, doi: 10.1016/j.comnet.2021.108536.

[10] P. Howson, “Building trust and equity in marine conservation and fisheries supply chain management with Blockchain,” Marine Policy, vol. 115, pp. 108853, 2022, doi: 10.1016/j.marpol.2021.108853.

[11] J. Adler, R. Berryhill, A. Veneris, Z. Poulos, N. Veira, and A. Kastania, “Astraea: A decentralized blockchain oracle,” in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2018, pp. 1145–1152.

[12] R. Kamiya, “Shintaku: An end-to-end-decentralized general purpose blockchain oracle system,” Online: https://gitlab. com/shintakugroup/paper/blob/master/shintaku.pdf, 2018.

[13] J. Guarnizo and P. Szalachowski, “Pdfs: practical data feed service for smart contracts,” in Computer Security–ESORICS 2019: 24th European Symposium on Research in Computer Security, Luxembourg, September 23–27, 2019, Proceedings, Part I 24. Springer, 2019, pp. 767–789

[14] A. Hassan, I. Makhdoom, W. Iqbal, A. Ahmad, and A. Raza, “From trust to truth: Advancements in mitigating the Blockchain Oracle problem,” Journal of Network and Computer Applications, vol. 217, pp. 103672, 2023, doi: 10.1016/j.jnca.2023.103672

[15] R. Iqbal, T. Butt, M. Afzaal, and K. Salah, “Trust management in social internet of vehicles: factors, challenges, Blockchain, and fog solutions,” International Journal of Distributed Sensor Networks, vol. 15, no. 1, p. 155014771982582, 2019, doi: 10.1177/1550147719825820.

[16] T. Gazdar, O. Alboqomi, and A. Munshi, “A decentralized Blockchain-based trust management framework for vehicular ad hoc networks,” Smart Cities, vol. 5, no. 1, pp. 348-363, 2022, doi: 10.3390/smartcities5010020

[17] Y. T. Yang, L. D. Chou, C. W. Tseng, F. H. Tseng, and C. C. Liu, “Blockchain-based traffic event validation and trust verification for vanets,” IEEE Access, vol. 7, pp. 30 868–30 877, 2019.

[18] R. Shrestha, R. Bajracharya, and S. Y. Nam, “Blockchain-based message dissemination in vanet,” in 2018 IEEE 3rd International Conference on computing, communication and Security (ICCCS). IEEE, 2018, pp. 161–166.

[19] C. Pu, “Blockchain-based trust management using multi-criteria decision-making model for vanets,” TechRxiv, 2020, doi: 10.36227/techrxiv. 13106876.v2

[20] I. Ali, A. Hassan, and F. Li, “Authentication and privacy schemes for vehicular ad hoc networks (VANETs): A survey,” Vehicular Communications, vol. 16, pp. 45–61, 2019.

[21] M. Mao, Y. Peng, T. Hu, Z. Zhang, X. Lu, and J. Li, “Hierarchical hybrid trust management scheme in sdn-enabled vanets,” Mobile Information Systems, vol. 2021, pp. 1-16, 2021, doi: 10.1155/2021/7611619.

[22] C. Chen and Q. Shi, “A summary of security techniques-based Blockchain in IoV,” Security and Communication Networks, vol. 2022, pp. 1-14, 2022, doi: 10.1155/2022/8689651.

[23] S. Guleng, C. Wu, X. Chen, X. Wang, T. Yoshinaga, and Y. Ji, “Decentralized Trust Evaluation in Vehicular Internet of Things,” IEEE Access, vol. 7, pp. 15 980–15 988, 2019.

[24] T. Jing, Y. Liu, X. Wang, and Q. Gao, “Joint trust management and sharing provisioning in IoV-based urban road network,” Wireless Communications and Mobile Computing, vol. 2022, pp. 1-18, 2022, doi: 10.1155/2022/6942120

[25] S. Wang, Y. Hu, and G. Qi, “Blockchain and deep learning-based trust management for internet of vehicles,” Simulation Modelling Practice and Theory, vol. 120, p. 102627, 2022, doi: 10.1016/j.simpat.2022.102627

[26] R. Ratnayake, M. Liyanage, and L. Murphy, “Trust Management and Bad Data Reduction in Internet of Vehicles Using Blockchain and AI,” in 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023, pp. 1-5, doi: 10.1109/VTC2023-Spring57618.2023.10200207

[27] J. Menegazzo, “PVS - Passive Vehicular Sensors Datasets,” Kaggle, 2021. [Online]. Available: https://www.kaggle.com/ds/1105310.

[28] J. Menegazzo and A. von Wangenheim, “Road surface type classification based on inertial sensors and machine learning,” Computing, vol. 103, pp. 2143-2170, 2021. [Online]. Available: https://doi.org/10.1007/s00607-021-00914-0.