BITS: Blockchain based Intelligent Transportation System with Outlier Detection for Smart City

Shirshak Raja Maskey*, Shahriar Badsha† Shamik Sengupta‡ Ibrahim Khalil§
*†‡Department of Computer Science and Engineering, University of Nevada, Reno
Reno, NV, USA
§Department of Computer Science and Software Engineering, RMIT University,
Melbourne, Australia
Email: *shirshak.maskey@nevada.unr.edu, †sbadsha@unr.edu, ‡ssengupta@unr.edu,
§ibrahim.khalil@rmit.edu.au

***Abstract*—With the rise of smart cities, transportation systems are getting smarter every day. An Intelligent Transportation System (ITS) should be secure, autonomous, capable of discerning safeness levels at the roads, and provide services to improve human experience. To reach the gold standard, the ITS faces several issues such as centralization, trust, and data integrity. The Transportation System and the data generated from the vehicles can be intercepted, manipulated and corrupted with coordinated attacks. Moreover, every system might have bad actors who want to manipulate the system or data to his or her favor by exploiting the system. In order to guarantee data integrity, immutability, and availability for the ITS, we propose Blockchain based architecture with outlier detection to prevent malicious activity by the vehicles while preserving integrity in sharing information. The Outlier Detection is designed to reside before the consensus process, to identify and prevent participation of malicious vehicles in consensus process or block mining. In our proposed Blockchain based Intelligent Transportation system with Outlier Detection for Smart City (BITS), we used machine learning to detect the anomaly in the data. The proposed model can be used in various applications of ITS such as traffic monitoring, criminal activity profiling, accident detection and reporting, etc.**

I. INTRODUCTION

Transportation is the backbone of modern civilization, and we are moving towards smart cities. Therefore, our transportation must also be smart. As people recently have become more dependent on transportation systems, the transportation systems face several opportunities and challenges. An average of 40% of the population spending at least 1 hour on the road each day was estimated [1]. The directive of the European Union 2010/40/EU, made on July 7, 2010, defined ITS as systems in which information and communication technologies are applied in the field of road transport. This includes infrastructure, vehicles and users, and in traffic management and mobility management, as well as for interfaces with other modes of transport [2].The vast application and scope of ITS requires a good security measure in its implementations. However, there are many vulnerabilities and issues in the ITS, which require our attention and effort in order to explore its fullest potential. According to the authors in [3] there are two critical issues with ITS: security risk caused by centralization in ITS and lack of necessary mutual trust among ITS entities. The centralization exploited attack can lead to targeted attacks which can lead to unavailability of the system, poor performance, or malicious action. The lack of mutual trust in ITS makes it harder to add functionalities such as transactions without establishing hierarchical structures and increasing system complexity, which may slow down the system and enable single point of failure.

In order to eliminate such issues, we need a robust, secure, and trustless architecture, which can provide a high level of confidence. To achieve such confidence on the system, we need Blockchain technology for data storage, for Blockchain is an open, distributed ledger that can record transactions between two parties efficiently, verifiably, and permanently. [4]. Although Blockchain has many advantages over classical data storage and sharing, it has its fair share of vulnerabilities and threats [5]. One of the threats in Blockchain based ITS is the 51% attack where the attacker nodes collectively control more computational power than the good nodes [6]. The Blockchain based applications have been applied by authors in [7] [8] [9] [10] [11] [12] [13], which range from ride sharing, firmware upgrade of IoV to crowd funding of cyber-product insurance and decentralized storage network. In order to use the advantages of Blockchain technology in Blockchain based ITS while defending it through vulnerabilities and attacks, we need a system in place which can differentiate between normal and malicious events. Therefore, we propose an Outlier Detection system on top of Blockchain based ITS to reach our required design goals. Outlier Detection is the identification of rare items regarding events or observations which raise suspicions by differing significantly from the majority of the data. [14]. By integrating all three components (Outlier Detector, Blockchain for ITS and Reputation Model) under a single architecture, we get a versatile, highly secure, and efficient architecture for BITS.

A. *Motivation and Contribution*

Assessment of the driving risks for drivers, legal authorities, and insurance companies are critical. We can assess the risk associated with drivers by analyzing different policies such as different driving lanes, areas, speed, premium costs, etc. enforced on the road. Additionally, this could fall victim to false information injection attacks to evade system policies or random anomalies transmitting erroneous data. We needed an architecture, which could guarantee data integrity, robust security, and the ability to detect outliers to prevent malicious activity with the reputation system, making transportation safer without the need of more law enforcement officials on the road. This paper is motivated to solve this problem by developing an integrated architecture, which will compute reputation for each member through the Reputation Model, store the data by ensuring integrity and immutability using Blockchain, and make the system malicious activity tolerant using an Outlier Detector.

Motivated from our scenario in the previous paragraph, we propose the BITS architecture, which can be used in multiple prospective applications to improve and ensure road safety. Blockchain and Outlier Detection is implemented using edge and core nodes with the PBFT consensus mechanism and machine learning respectively. We chose the PBFT consensus mechanism because vehicles are resource constrained devices, and it is not a compute intensive consensus mechanism like Proof of Work. Also, it has very low latency and high throughput in the network. The main contribution of this paper is the BITS architecture, which is capable of assigning reputation to drivers by considering different input factors such as average speed, top speed, top deceleration, GPS coordinates, and criminal record index, and it detects anomalies in the data using machine learning. A criminal record index is a conceptual index that is comprised of different factors such as drivers' previous tickets, DUI offences, accidents, etc. To our best knowledge, we are the first to pursue this integrated architecture model that includes three different technologies(Reputation Model, Blockchain, and Outlier Detection) on a distributed edge and core networking model that compliments one's strengths and weaknesses.

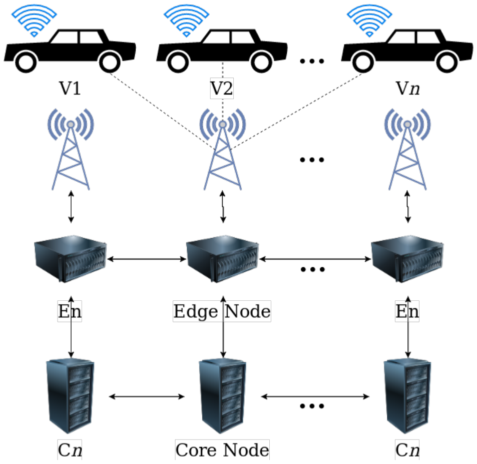

Fig. 1: System Architecture

B. *Related Study*

Our proposed architecture can be used in multiple applications ranging from traffic monitoring, criminal activity profiling, and accident detection and reporting. These applications are more efficient and secure than other available methods [3] [15]. The author in [16] has explored a similar concept where the witnesses were in charge of reporting an accident. This paper has used a simple architecture where Blockchain is applied to only store the end result. It did not have any security measures applied nor any mechanisms to further analyze the data stored. Also, this proposed model will only work if there is at least one witness in an accident's vicinity. Our proposed model does not require human input in the architecture as the nodes communicate with vehicles directly through distributed edge and core nodes. The authors in [3], [17] and [15] have put forth their ideas in order to detect and quickly respond to accidents, but their models do not take anomalous data and false data injections into account. In contrast, we have incorporated an Outlier Detector in the system model. After the anomaly in the data is removed, the data is then recorded into the respective Blockchain. With all the accumulated data, we developed a model that uses the data to create a reputation index that can help create safer roads. The reputation can be linked to incentive mechanisms such as insurance premiums, vehicle loan EMI, or tax credits according to yearly driving performances.

II. System Design

Figure 1 illustrates the proposed architecture of our BITS with a goal to monitor traffic data for anomalies and assign reputation index to drivers on top of an

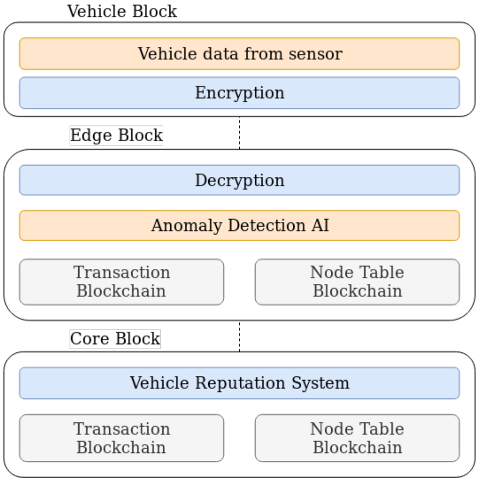

Fig. 2: System Model

ironclad architecture for a secured, transparent, and flexible IoV system. In this layer, we have three major components: IoV $[V_1, V_2, ...V_n]$, distributed edge nodes $[E_1, E_2, ...E_n]$ and the core nodes $[C_1, C_2, ...C_n]$ as shown in figure 1. The different distributed networks are designed as a multi-ring system inside a concentric circle. The IoVs are distributed along the edge network, so IoVs are connected to one or multiple edge nodes. The edge nodes are then connected to core nodes in the distributed core network. We plan to use weightless technology for communication from IoV to edge nodes and use wired mediums for communication between all the server nodes. The security aspect of the design is handled by AES encryption [18] at all nodes, IPsec protocol for data transfer, and Blockchain for data immutability. The ANN resides on top of a distributed edge and core network, looks for any anomalous data in the edge network, and creates a reputation model in the core network. The system block model is illustrated in figure 2. The ANN fetches input parameters from the node table including average speed, top speed, top deceleration, GPS coordinates, and criminal record index, which processes them to get the output as shown in figure 3. The output from the ANN assigns the reputation of each driver in the system. The IoV generates data such as device information, user information, speed, direction, etc. The generated data is processed and checked for anomalies before disseminating to multiple edge nodes. When edge nodes receive the data from an IoV device, it decrypts the data and runs an integrity test. The edge node creates a block of all recently received data and appends to the BAIT. The distributed core network provides the reputation index for users and acts as a miner and validator in the system along with the edge nodes.

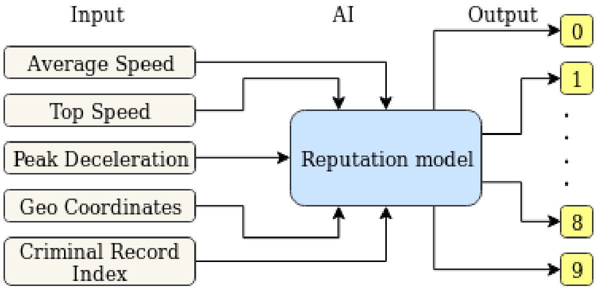

Fig. 3: Reputation Model

III. Methodology

This paper proposes a robust and secure architecture to monitor traffic in smart cities while looking for anomalies in data. It implements an Outlier Detector to detect anomalies and a reputation model to create a reputation index of the driver. A reputation index is an output of the reputation model, ranging from 0 to 9. We have used Isolation Forest as the model for the Outlier Detector, which explicitly identifies anomalies instead of profiling normal data points. Isolation Forest, like any tree ensemble method, is built on the basis of decision trees. In these trees, partitions are first created by randomly selecting a feature and then selecting a random split value between the minimum and maximum value of the selected feature [19]. As with other outlier detection methods, an anomaly score is required for decision making. In the case of Isolation Forest, it is defined as:

$$S(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

where h(x) is the path length of observation x; c(n) is the average path length of unsuccessful searches in a Binary Search Tree, and n is the number of external nodes. We have used simulated data from 10 IoV devices with 5 features for 50,000 unit cycles to train the Outlier Detector. In our system model, the data is sent to the nearest edge node in each cycle through weightless technology with AES encryption for preliminary computation such as an integrity test and outlier detection. The results obtained from the Outlier Detector are stored in the node table. After passing through the detector, the data is then transmitted to other edge nodes in the network through IPsec protocol. Then, the edge network creates and appends individual blocks to Blockchain, which is then transmitted to the distributed core network for further processing.

The core network consists of a reputation model which assigns each driver a reputation index ranging from 0 to 9. The input parameters will be in accordance to node table attributes. The output of our reputation model is explained in the forthcoming equations.

$$z = \sum_{n=1}^{5}(x_n \times w_n) + b \times 1 \tag{2}$$

Here in equation one, z is the output at any node, which is obtained by adding every multiple of weight and inputs. x_1, x_2, x_3, x_4, x_5 are the inputs in accordance to the node table attributes. w_1, w_2, w_3, w_4, w_5 are the individual weights for each input as mentioned earlier. We have b as bias weight to align the system towards our favourable outcome and 1 as bias value in the system. Even if we have our output z, we still need to fit it within a range, which is done by activation function or transfer function, $sigmoid(z)$

$$\widehat{y} = a_{out} = sigmoid(z) = \frac{1}{1+e^{-z}} \tag{3}$$

$\widehat{y}$ is the output at any individual neuron. The output of sigmoid ranges from $0to1$, which will activate one of ten output nodes. Therefore, the final output is the reputation index of each driver ranging from 0 to 9 where 0 is the lowest and 9 is the highest reputation. Each new user will have a default reputation value of 4, and based on its performance, it can either reach 0, where the user is banned and prosecuted accordingly, or 9, where the user has the highest reputation. A driver's reputation will decrease if he has a high criminal index, many over-speeding flags, rough driving, etc. and it will increase if a driver drives while following road traffic rules. After the new reputation index is generated, it is updated in the node table, and a block is created in two chains. One chain is for the node table and another chain is for time series data transactions. Afterwards, the PBFT consensus protocol is initiated by the core nodes who validate the blocks. After the consensus is reached, the block is added to the Blockchain, and the chain is updated. In algorithm 1, we have a connection from the connected vehicle to the edge nodes. The inputs here are the input groups as listed in algorithm 1. All the functions are contained within an infinite while loop to always read data at certain intervals. The connected vehicle will then connect to all the towers in range. Before sending the data to the edge nodes, it will encrypt the data for privacy and security.

Algorithm 2 deals with the edge node part of communication. Similar to algorithm 1, it has infinite while loop to always listen to incoming data from the connected vehicle. After the data is received, the node tests the integrity of the data and passes the data through the Outlier Detector. After the data is screened through the Outlier Detector, outlier data is rejected and the node table is updated. It then distributes the data to all edge nodes. After this is done, the nodes organize the data it has. In addition, they create and append a block to the Blockchain.

Algorithm 1: Connected Vehicle Algorithm

```
Data: position = $p$, average speed = $v$, basic
      vehicle information = $bvi$, basic driver
      information = $bdi$, Number of Towers =
      $i$;
while true do
    if any tower in range then
        while $i \geq 1$ do
            Connect to $i^{th}$ tower;
            $i - 1$;
        end
        Gather all data;
        Encrypt the data with AES;
        Send the data to all towers;
    end
end
```

Finally, we reach to the third algorithm, which deals with core nodes. Similar to before, the infinite while loop looks for data while if statement checks for integrity. All the Blockchain are now fetched from each node along with the node table, which has node attributes. The attributes are processed in the reputation model, which produces output as a new reputation index, which is then updated in the node table. Then the core nodes validate the blocks and participate in the consensus mechanism where the new block is added to the blockchain upon reaching consensus.

IV. Experiments and Results

In this experiment, we analyzed different outlier detection models using both supervised and unsuper-

Algorithm 2: Edge Network Algorithm

```
Data: Encrypted data from connected vehicle
while True do
    Decrypt the data;
    if integrity test then
        Pass through Outlier Detector;
        while data != anomalous do
            Distribute the data to all edge nodes;
            Create Block for block chain;
            Add Block to Blockchain;
        end
        Update node table;
    end
end
```

Algorithm 3: Core Network Algorithm

```
Data: Blockchain from Edge Node
while True do
    if integrity test then
        Fetch data from node table;
        Run reputation model;
        Update reputation index;
        Fetch Blockchain from nodes;
        Run validation and reach consensus;
        Update new block;
        Reach consensus and update
          Blockchain;
    end
end
```

vised machine learning techniques in order to find the optimum method. Therefore, we created a data-set with five features assuming the five vehicular inputs: avg. speed, top speed, top deceleration, GPS coordinates, and criminal record index. We generated a normalized non co-related synthetic dataset with $50,000$ unit time entries of these five features. After creating the dataset, we randomly injected approximately 10% or $4,878$ outlier values into the dataset, so the outlier value points are distributed far away from clusters in order to simulate the anomaly values in the real world systems. We started with a dendogram creation to visualize the number of clusters in the dataset. The number of clusters is taken at the point of intersection, which is 50% value of the longest branch in the dendogram. Since we can now visualize the number of clusters, we need to label the outlier cluster in order to run supervised machine learning models. To label the clusters, we used k-means clustering with the cluster value obtained from the dendogram. We added the label as features in our dataset. We used the labeled dataset in three supervised machine learning models (Random Forest Classifier, Support Vector Machine and Multi layer perceptron) by classifying in such a way that one cluster is assumed as an outlier. In addition to this, we contemplated that the outlier can also not form a cluster and should be randomly distributed in the dataset. We shifted towards unsupervised clustering as this type of dataset cannot be labeled for supervised learning. We opted for the Isolation Forest model, which is a tree ensemble method that explicitly identifies anomalies instead of profiling normal data points. While setting up parameters for this experiment, we set up the contamination value to 0.1, with 4 parallel jobs running. We have used a new behavioral model for our experiments, and the value of estimators and sample size differs from 100 to $1,400$ and $1,000$ to $25,000$ respectively.

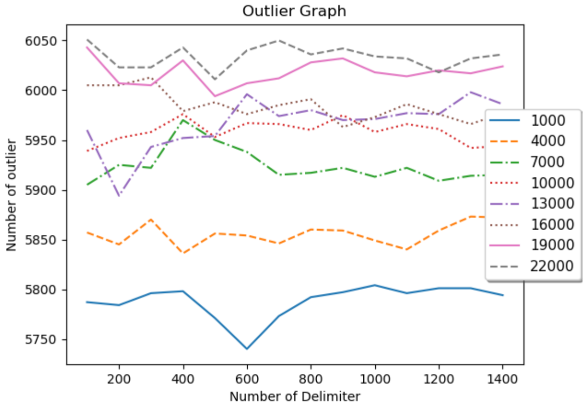

Fig. 4: Outlier Detection

For the first set of experiments, we got the cluster value of 4 when we created the dendogram. We inserted that value as a number of clusters in the k-means clustering and obtained the result. And the labeled data obtained from the k-means algorithm is passed to multiple supervised learning models. We obtained a precision and recall score of 1.00 in all of our supervised models. For the second set of experiments, we randomized the outliers while keeping the number of outliers the same. With the k-means cluster visualized, we were not able to run supervised learning on the labelled dataset. Instead, we opted for an unsupervised learning model,the Isolation Forest model. We obtained the following result graph from our experiment, which is shown below in figure 4. From our results, we observed that the best average value is at 1000 for the sample size and 600 for the number of estimators. This configuration yielded us a false positive rate of 21.03% and an accuracy of 79.87%.

V. Example Scenario

Consider an example scenario with vehicles V_1, V_2, V_3 connected to the edge nodes, E_2 and E_2, which are a couple of the interconnected edge nodes $E_n = (E_1, E_2, E_3)$. E_n is connected to the core nodes, $C_n = (C_1, C_2)$. The Outlier Detector is implemented at E_n, and the Reputation Model is implemented at C_n as Bob and Alice respectively. V_2 and V_3 are moving from left to right and are transmitting input data parameters, but they are not only limited to the example as shown in figure 5. Let us assume that all the data that is coming from V_2 and V_3 are valid. When the data reaches node E_2, it is passed through the Outlier Detector where data from V_2 is rejected due to an unrealistic value of speed. The data for V_3 is passed through the Outlier Detector but with an overspeeding flag. Both entries are recorded in the node

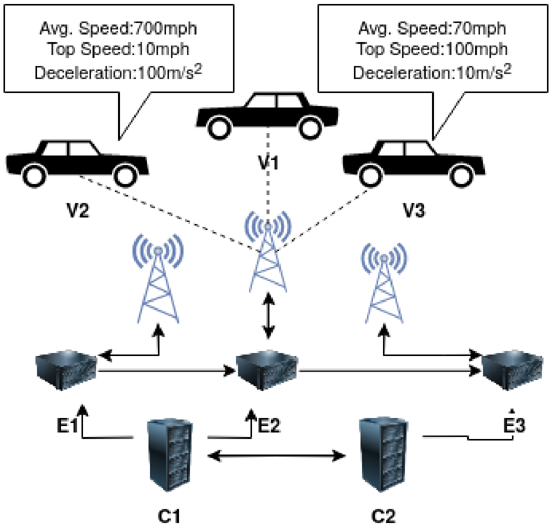

Fig. 5: Example Scenario

table, and the data is recorded as transactions in the blocks. The copy of this data is disseminated to E_n. The blocks are fetched at C_n for validation along with the node table for updating the reputation index with inputs from the node table attributes. At the core, the reputation index has been updated by Alice, and a new block is created after validation and the consensus process. The block is appended in the Blockchain, and the data is available for everybody to use. After the block is created and added, respective authorities can be notified about vehicle V_3, and necessary steps will be taken by the authorities after analyzing the data.

VI. Conclusion and Future Work

A BITS system is proposed to make ITS secure in smart cities by implementing the blockchain based architecture with outlier detection. The proposed architecture can detect the outlier in the BITS and calculate the reputation based on the feature of input. With the help of Blockchain, machine learning, and distributed server architecture, better resilience can be developed against anomalies and attacks. This paradigm can also be used to actively monitor traffic and alert respective authorities in the case of accidents or criminal offences.

Our future work is aimed at completing all the remaining experiments and performing evaluations of the given example scenario in a simulation framework as well as expand the current architecture to a new consensus mechanism, which has low latency and high throughput with a very low network overhead and a high tolerance against malicious activities.

References

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[2] "Directive 2010/40/eu of the european parliament and of the council," Jun 2010. [Online]. Available: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF

[3] Y. Yuan and F.-Y. Wang, "Towards blockchain-based intelligent transportation systems," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 2663–2668.

[4] M. Iansiti and K. R. Lakhani, "The truth about blockchain," *Harvard Business Review*, vol. 95, no. 1, pp. 118–127, 2017.

[5] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, 2017.

[6] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, "Where is current research on blockchain technology?—a systematic review," *PloS one*, vol. 11, no. 10, p. e0163477, 2016.

[7] M. Baza, N. Lasla, M. Mahmoud, G. Srivastava, and M. Abdallah, "B-ride: Ride sharing with privacy-preservation, trust and fair payment atop public blockchain," *IEEE Transactions on Network Science and Engineering*, 2020.

[8] M. Baza, M. Mahmoud, G. Srivastava, W. Alasmary, and M. Younis, "A light blockchain-powered privacy-preserving organization scheme for ride sharing services," *Proc. of the IEEE 91th Vehicular Technology Conference (VTC-Spring), Antwerp, Belgium*, May 2020.

[9] M. Baza, M. Nabil, N. Bewermeier, K. Fidan, M. Mahmoud, and M. Abdallah, "Detecting sybil attacks using proofs of work and location in vanets," *arXiv preprint arXiv:1904.05845*, 2019.

[10] M. Baza, J. Baxter, N. Lasla, M. Mahmoud, M. Abdallah, and M. Younis, "Incentivized and secure blockchain-based firmware update and dissemination for autonomous vehicles," in *Connected and Autonomous Vehicles in Smart Cities*. CRC press, 2020.

[11] S. Kudva, R. Norderhaug, S. Badsha, S. Sengupta, and A. Kayes, "Pebers: Practical ethereum blockchain based efficient ride hailing service," in *IEEE International Conference on Informatics, IoT and Enabling Technologies*, 2020.

[12] I. Vakilinia, S. Badsha, and S. Sengupta, "Crowdfunding the insurance of a cyber-product using blockchain," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2018, pp. 964–970.

[13] I. Vakilinia, S. Badsha, E. Arslan, and S. Sengupta, "Pooling approach for task allocation in the blockchain based decentralized storage network," in *15th International Conference on Network and Service Management. IEEE*, 2019.

[14] A. Zimek and E. Schubert, "Outlier detection," *Encyclopedia of Database Systems*, pp. 1–5, 2017.

[15] H. Guo, E. Meamari, and C.-C. Shen, "Blockchain-inspired event recording system for autonomous vehicles," in *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*. IEEE, 2018, pp. 218–222.

[16] G. S. Praba Devi and J. C. Miraclin Joyce Pamila, "Accident alert system application using a privacy-preserving blockchain-based incentive mechanism," in *2019 5th International Conference on Advanced Computing Communication Systems (ICACCS)*, March 2019, pp. 390–394.

[17] F. Bhatti, M. A. Shah, C. Maple, and S. U. Islam, "A novel internet of things-enabled accident detection and reporting system for smart city environments," *Sensors*, vol. 19, no. 9, p. 2071, 2019.

[18] S. Gueron, "Intel® advanced encryption standard (aes) new instructions set," *Intel Corporation*, 2010.

[19] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.