IETE Technical Review

ISSN: 0256-4602 (Print) 0974-5971 (Online) Journal homepage: https://www.tandfonline.com/loi/titr20

Blockchain For Intelligent Transport System

Anandkumar Balasubramaniam, Malik Junaid Jami Gul, Varun G. Menon & Anand Paul

IETE TECHNICAL REVIEW
https://doi.org/10.1080/02564602.2020.1766385

Blockchain For Intelligent Transport System

Anandkumar Balasubramaniam[1], Malik Junaid Jami Gul [1], Varun G. Menon[2] and Anand Paul[1]

[1]Kyungpook National University, Daegu 41566, South Korea; [2]Department of Computer Science and Engineering, SCMS school of Engineering and Technology, Kochi 683576, India

ABSTRACT

Intelligent Transportation System (ITS) is gaining attention but at the same time, road accidents, congestion, delays, etc. have also increased. Relative information about such events is vital. Such information can be presented in legal processes as digital proof. Availability of the information is not a problem as multidimensional data have been recorded all the time by ITS. Recording all the information in ITS arises the problem of fetching relevant information and removing other facts and figure that are not required to describe certain situations such as an accident. To address this issue, we analyze road accident data and reduce various dimensions with Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Non-negative Matrix Factorization (NMF). We conduct comparative analysis with three datasets where error rate for PCA is 32% with Dataset1. Likewise, error rate for LDA and NMF are 36% and 35%, receptively. While keeping in mind that such reduced data is helpful in many legal processes, we introduce Blockchain in the framework. Blockchain can make data immutable thus can be considered as digital proof. Blockchain also requires a smart contract in this situation between insurance companies to collect data in case of any uncertain situation. Such analysis can offer a different point of views and trends in data. Information can be more explainable to define the situation and helps to develop a friendly environment for day-to-day customers. The proposed framework provides dimensionality reduction of data that eventually reduce the data dimension to store in Blockchain.

KEYWORDS

Accident analysis; Blockchain; ITS; Principal component analysis; Smart transportation system; Traffic surveillance

1. INTRODUCTION

Blockchain is a promising technology and intelligent transportation system (ITS) can be greatly benefited from this rising technology to preserve data as digital proof. Blockchain is distributed database that is shared among the Blockchain nodes or participating organizations. At the beginning, Blockchain was only adapted by organizations that are involved with finance but new possibilities are being explored to utilize Blockchain in other areas. Data in Blockchain is stored in a distributed database after verification. The process of data verification in Blockchain is carried out by consensus algorithms. Proof work and proof of stack are examples of consensus algorithms that are adapted by the Blockchain. Blockchain is also getting popular among other domains and the aim of our study is to explore possibilities of Blockchain in intelligent transport system (ITS). Blockchain is used to record transaction information among trusted parties likewise data from ITS is also a digital asset and can be used in many legal processes. Some possible application domain of our study can be insurance-based services provided by insurance companies or court hearings where digital assets that is data stored in blockchain can be presented as digital proof. Blockchain is distributed, secure, transparent, and immutable. The information about the accident or congestion of the road in certain conditions can be helpful for involved parties from driver to road and also other entities that have contracts with the vehicle or the driver. The problem is: Blockchain is not feasible for data with a higher number of dimensions. Our analysis is also aimed at reducing this problem so that only relevant data goes into Blockchain. The reduced data can be directly be presented to the authorities because Blockchain has no central authority for verification as the data in the Blockchain is verified via consensus which saves further cost and provides more security. ITS can hold different types of data, *e.g.*, weather data, road condition data and other information that can be used for road safety and management. Real-time weather information and traffic condition can prevent accidents and can save fatalities. Weather conditions like fog, rain, snow, etc. are very helpful for the driver meanwhile in certain conditions such as an accident. This information should be saved so it can save time finding and prove all the information from different sources. Information about heavy traffic and road condition can save unfortunate event meanwhile can provide insight for relevant organizations to save the expenditures on road infrastructure. Traffic control is a big issue that is being eased out by ITS but there are other issues that can be maintained by ITS such as parking management, gathering traffic data, etc. Parking management involves a smart system that can check the license plate of the vehicle parked is legal or not. These data became important for the legal process and require verification from the central authorities. Data that is gathered for parking scenario is different but there are certain other irrelevant data that is not required and take time to get information from the data and verify to put in front of the court or insurance company. Principal component analysis (PCA) is a technique that can solve issues in reducing dimensions of the data, so data become more relevant as only correlated variables are selected. Such features can save the time of the involved organization to whom time and cost matters. It also enables road safety management teams to make decisions according to the fact from data. This can decrease the congestion and accident rate while the customer can prove the scenario if they get involved in the legal process.

Smart transportation system is one of the emerging research topic in a current era [1–4] for smart city environments [5–9]. The global smart transportation market is growing alongside these developments, while the compound annual growth rate is projected to be around nineteen percent from the latest statistical perspective [10]. This statistical analysis, in combination with the heavy demand for uncompromising technological improvement for efficient smart transport in smart city environments, indicates that there is a high demand for smart transport in the future [4, 11]. The advantages of ITS can be countless: starting with traffic optimisation, recording, tracking and transport monitoring services linked in complete automation. These technological improvements, on the one hand, pave the way for smoother and more hustle and bustle free transport, but on the other, they also need further refinements in order to reduce accident rates, congestion rates and so on in road transport environments. In order to improve road safety in ITS, the two main problems in the transport sector, such as road accidents and traffic congestion must be taken seriously. Bogota enters the top list of congestion cities worldwide in 2018 based on the INRIX report [12] evaluating congestion and mobility patterns in more than 200 cities across 38 countries. Figure 1 shows that Bogota drivers lost 272 h as compared to any other city in the world in 2018, according to the study.

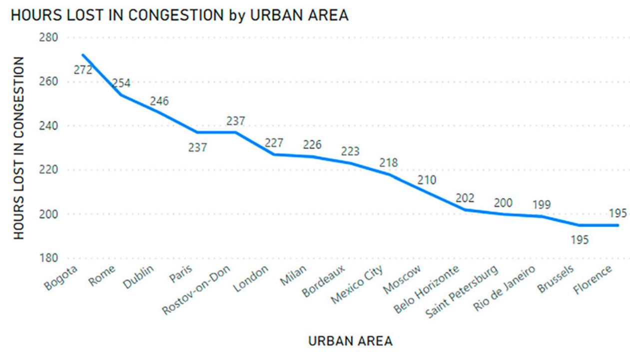

Figure 1: Hours lost due to congestion in worldwide cities in 2018

There may be various reasons for traffic congestion, for instance, lack of an integrated traffic control/monitoring system, inefficiency in traffic surveillance, improper driving, accidents / collisions, human driving errors, etc. Nonetheless, there is a tremendous need for reliable, dense urban traffic surveillance systems to track or monitor all city roads day and night. Parallel to this, the recent articles [13] report that the number of road traffic accidents is increasing. Road traffic incidents and vehicle collisions could have different reasons. To evaluate the information on traffic accidents, a country-wide dataset [14] of traffic accidents across 49 United States is analysed. The dataset is the accident data record compiled from February 2015 to March 2019.

In the dataset, the field severity shows a number between 1 and 4, where 1 indicates the least impact on traffic such as short delay as a result of the accident and 4 indicates a significant impact on traffic such as long delay. From Figure 2, the sum of severity is in the increasing phase from 2016 to 2018. It clearly indicates that there is a greater number of accidents, even though the transport system is adapted to the ̈intelligent ̈transport

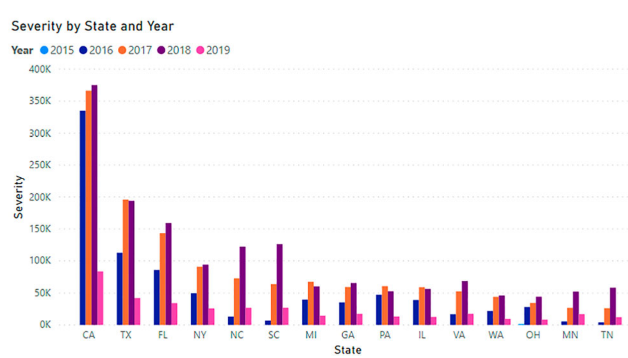

Figure 2: Severity count due to road-traffic accidents in major US cities from 2015 to 2019

sector. In addition to road accidents, problems with traffic congestion still play a major role in the transport sector. Traffic congestion problems can be avoided by introducing effective traffic forecasts or efficient traffic surveillance systems. This paper analyses road traffic incident data and uses dimensionality reduction techniques, such as PCA, linear discriminant analysis (LDA), and non-negative Matrix Factorization (NMF) to evaluate data patterns by reducing high-dimensional data to reduce components and predicting traffic data. To deal with high-dimensional datasets, PCA [15] is used since it is a fantastic tool for data exploration and data reduction/compression. It is used for various approaches in terms of statistical analysis and machine learning perspectives. Though PCA helps to reduce the dimensionality, variation present in all the features are retained in the initial few principal components. PCA reduces the number of variables of the dataset and makes it easier and faster to perform various machine learning algorithms while preserving as much information as possible from the original high-dimensional dataset. There are various other research articles on PCA. In [16], the author utilizes the concept of PCA on high-frequency data and develops the necessary tools for implementing PCA at high-frequency data. They found that the first principal component explains about 60% variance of the data during the financial crisis, whereas the second principal component drives the common variation of the financial stocks. In [17], the author used PCA to observe the relationship between the proposed soundscape categories and to the quadrant classifications in the two-dimensional core effect. It is clearly noted that the foregrounded soundscape categories in the plane spanned by the two principal components such as the first principal component explains 71% of the variance and the second principal component explains 22% of the variance, approximately. In [18], the author used PCA for analyzing and mining the data from *in vitro* fertilization (IVF) in order to predict the pregnancy. In this different types of IVF in particularly, 805 types of IVF were analyzed to classify the pregnancy. PCA is applied in various industries and different fields of sciences. This application of PCA in the medical sciences allowed the models to predict pregnancy in the most effective way using PCA. The various analyses are done on the road-traffic accident data, analysis using various dimensionality reduction techniques such as PCA, LDA, and NMF on the road-traffic data, are elaborated in the following sections of the paper. Also, the dataset used in PCA analysis is parallelly analyzed using LDA [19, 20] and NMF [21–23], which are another dimensionality reduction technique used in the machine learning field. LDA and NMF work like PCA such as reducing the variables while retaining the information

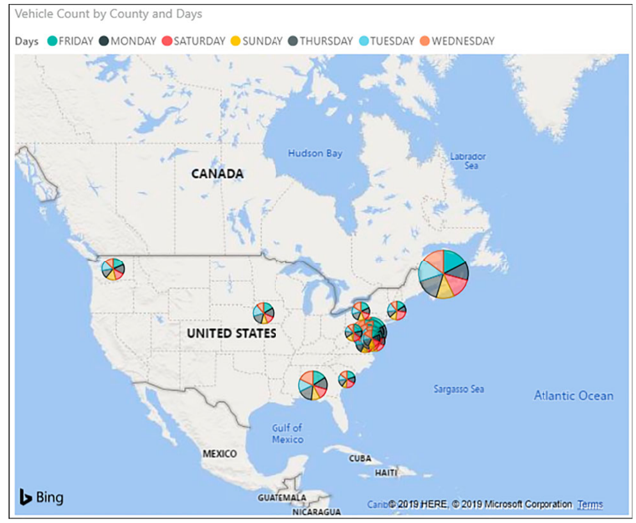

Figure 3: Location-based accident data

as much as possible. The paper discusses the different analyses on road accident data, and analysis using various dimensionality reduction techniques for road traffic dataset. In the current transport sector, road accidents and traffic congestion are increasing. There are many reasons behind road accidents and congestion such as distracted driving, unsafe lane changes, congestion collision, unsuitable turns, crossing of animals/pedestrians, weather, etc. A lot of collision avoidance methods are implemented in recent years to avoid or reduce those unwanted collisions and congestion. Specific analyses are performed on the dataset [24] to test road traffic accident scenarios. The diagram below shows the results of those tests. Figure 3 displays location-based visualizations of road traffic incident data collected during 2012 by the Maryland Police Department. The diagram below is based on the analysis of the data collection for the accident which is downloaded from the data archive of the US government [24]. The display shows that, in several cities in the US State of Maryland, there are a greater number of road accidents such as car crashes, during the weekend from Friday to Sunday, whereas the number of traffic accidents ranges from smaller to larger shapes, respectively. Similar results will apply to most locations in different countries as the amount of road traffic is increased during the weekends.

Blockchain is essentially a storage medium that holds transactions from legitimate users with respect to rules defined in a smart contract by using hashing algorithms. However, smart contracts can come under attack, and to resolve this issue, researchers have determined a method for securing it [25]. The concept of Blockchain was coined by a Japanese hacker. Blockchain was initially introduced by a Japanese hacker to authorize information in a decentralized way so data became immutable. Blockchain came with many issues, especially for governments. There are many regulatory issues that can be problematic for governments such as sending money abroad without government consent. The EU and the USA are already working on "smart regularity handoff" based approaches to tackle this problem [26]. However, Blockchain is evolving, and the healthcare industry is also looking to adopt Blockchain technology. Zand *et al.* [27] stated that Blockchain-based software in the healthcare industry could save as much as 15–20 billion annually. For this purpose, a Blockchain-based ledger has been developed for biomedical and healthcare applications [28]. Now, Blockchain must be generalized for all organizations, and such opportunities are discussed by [29].

Proof of work is very popular in Blockchain consensus technology. Proof of work is designed to make systems secure from certain types of attacks such as Denial-of-Service (DoS). Previously, DoS attacks are mainly of concern in networks. Many researchers have contributed to the deterring of DoS attacks with respect to network technology [30]. Proof of work has many functions that can be utilized. [31] studied proof of work and modified it with their model in which they calculate the computational cost, bandwidth cost, and aim to maintain speed and security at the same time.

Proof of stack was introduced by Sunny King and Scott Nadal in 2012 to reduce the energy consumption of mining [32]. However, proof of stake now has a newer version known as delegate proof of stake. In this version, a stockholder chooses a delegate to verify and validate their block [33].

The number of IoT devices is increasing exponentially. Because of this real-time processing, it is required to place computing servers near to these devices at a very initial level to get some load off the cloud. As far as Cisco is concerned, they define fog computing as an "Extension of the cloud to the edge of the network" [34, 35]. It provides services like computing, storage, and networking [36]. Until now, many frameworks have been developed to deploy fog computing in the real world. Symantec Internet Security Threat Report – ISTR [37] proposed a framework enabled with an IP scheme to facilitate mobile users along with the fog. In their framework, they place the "IP core network" above the fog layer to provide IP services to the fog layer. The fog layer also provides services like security and privacy [38–44]. Gervais *et al.* [45] placed a VPN, firewalls, analysis unit, and response unit in the fog to secure IoT data. The rest of the paper is organized as follows: Section 2 provides brief information about the proposed framework, while Section 4 provides experiment and analysis results, proposed algorithm and discusses the results. In the end, the authors provided future directions and conclude the work while providing insights and facts that PCA can resolve the feasibility issue for organizations where blockchain can not hold large amount have to be reduced before storing in blockchain ledger.

2. PROPOSED FRAMEWORK

Figure 4 shows an overview of the proposed framework but there are certain other factors that are working and are shown in Figure 5. The proposed framework starts by fetching data from real-world and going through network or data connections to ITS. ITS holds a broad spectrum of data that can be sent to the consensus layer if there is a smart contract among the involved parties. ITS can hold parking data, weather data, accident data, vehicular details, and even driver detail according to the vehicle number plate. Once ITS stores the data it could be sent for PCA analysis after which data can be sent for consensus and finally storing it in Blockchain.

Data that has been stored in Blockchain is now immutable thus cannot be modified. If the data is modified, our framework will alert that Blockchain is invalid. The valid block entries can be used as legal proof. It saves the cost of verification and time.

Data collected from ITS layer will go to the PCA analysis layer where PCA do the reduction and send correlated data for consensus. The consensus layer verifies the information and sends it for Blockchain storage.

2.1 Experimental Setup

We used R-Studio with Python for our study. We utilize the R-studio to analyze out datasets while Blockchain is built in Python incorporated with R-studio. We build a private Blockchain to review our results for Blockchain validity.

3. RESULTS AND DISCUSSION

3.1 PCA Analysis on Road-Traffic Data

PCA is the traditional data reduction/compression technique for finding patterns in the high-dimensional dataset, since PCA is a convenient way of reducing high-dimensional data to a smaller component number. It is also a linear combination of the original dataset which

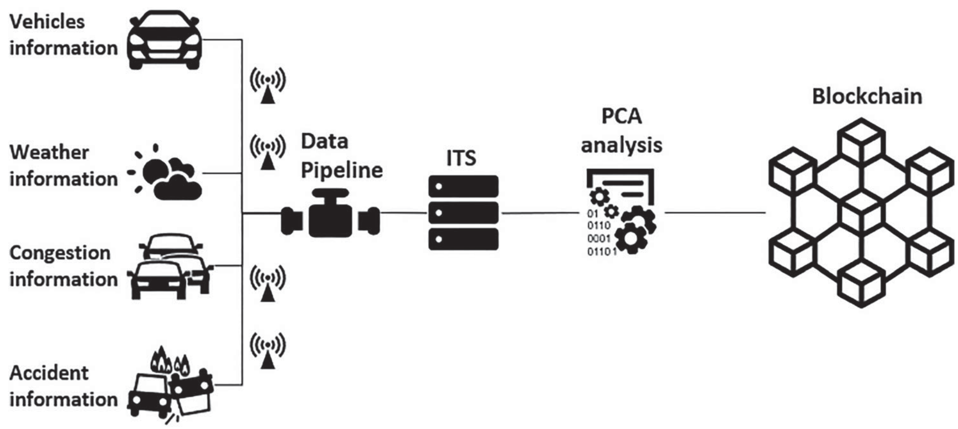

Figure 4: Proposed framework

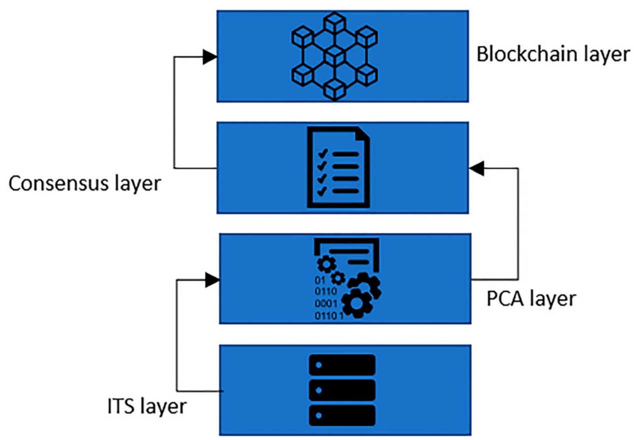

Figure 5: Blockchain workflow

transforms the data into a new system of coordinates. The first phase of the analysis is about analyzing the vehicular traffic count dataset with PCA. The dataset [23] is downloaded from the Kaggle domain and consists of the vehicular count recordings in New York City. The dataset is cleaned initially as a process of data pre-processing and then performed PCA analysis using R Studio. The entire experimental setup is carried out using the R Language. The dataset has 24 features and these features are reduced using the PCA process. These reduced features are termed to be as principal components. Using one of the test prediction models such as multinominal logistic regression, the principal components obtained using PCA analysis are used for simple traffic count prediction. The dataset is to be centered and scaled to perform the PCA operation. The centering and scaling are carried out in the dataset as mathematically as follows:

$$X = [x_1, x_2, x_3, \ldots\ldots, x_N] \tag{1}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

$$S = \sum_{i=1}^{N}(x_i - \mu) \tag{3}$$

whereas X and N are the variables from the input and the number of samples, respectively. The second equation provides the mean of X values and x_i is the ith sample.

3.2 LDA and NMF Analysis on Road-Traffic Data

Alongside to PCA analysis, the same dataset is analyzed using LDA and NMF. LDA is a powerful linear classifier. LDA can be mainly used for classification purposes and used for the dimensionality reduction technique. It is used to move features in a larger dimension space into a smaller dimension space. LDA provides class separation by drawing a decision area between the various classes. It tries to optimize the ratio of the variance between classes and the variance within classes. NMF is a dimensionality reduction that imposes nonnegative constraints in the process leading to the generation of lower-dimensional representation without negative values. Due to the avoidance of negative values, the quality of Matrix Factorization is improved. In NMF, the original bigger matrix is factored into two smaller factor matrices. Say for a matrix $\boldsymbol{X} \epsilon \mathbb{R}^{(m\times n)}$, it can be factorized into two factor matrices $\boldsymbol{W} \epsilon \mathbb{R}^{(m\times r)}$ and $\boldsymbol{H} \epsilon \mathbb{R}^{(n\times r)}$. It can be defined as,

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}^T \text{ s.t, } \boldsymbol{W} \geq 0 \text{ and } \boldsymbol{H} \geq 0 \tag{4}$$

and the optimization problem can be formulated as a minimization problem with Euclidean distance as the cost function as,

$$\left\| \boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}^T \right\|_2 \text{ s.t, } \boldsymbol{W} \geq 0 \text{ and } \boldsymbol{H} \geq 0 \tag{5}$$

The above optimization problem can be solved using any of the optimization algorithms like Alternating

Least Squares (ALS), Multiplicative Update Rule (MU), and Gradient Descent (GD). It is worth noting that the factor $\boldsymbol{W}$ is the lower dimensional representation of mode-1 of $\boldsymbol{X}$ and $\boldsymbol{H}$ is the lower dimensional representation of mode-2 of $\boldsymbol{X}$ and depending on the application, any one of these can be used.

The prediction output results are graphically represented in Figure 6. The graphs show the actual and predicted values of the vehicle count in the specific geographical regions. The actual and predicted values using the dimensionality reduction techniques are portrayed as graphical representations. The graphs conclude that there are a countable number of misclassifications. However, the results from PCA analysis is comparatively good than the other dimensionality reduction techniques. Also, the dimensionality reduction techniques are applied to do analyses on another dataset [24] and the prediction results are graphically represented. The dataset consists of the vehicular classifications such as "van," "bus," "opel," and "saab," which is classified based on the given 18 variables and 846 observations named to be as features. The dataset has been pre-processed to check any missing values. The dataset is analyzed using the above-mentioned dimensionality reduction techniques and the prediction is carried out using the multinomial logistic regression model. The analysis results of the actual and predicted values are graphically represented in Figures 6–8. This dataset is the classification-based and hence the predicted value using LDA is comparatively fair with PCA and NMF analysis.

Though there is misclassification compared with the actual data, the importance of the analysis using the dimensionality reduction techniques are portrayed. PCA helps to mitigate the prediction errors compared with other dimensionality reduction techniques.

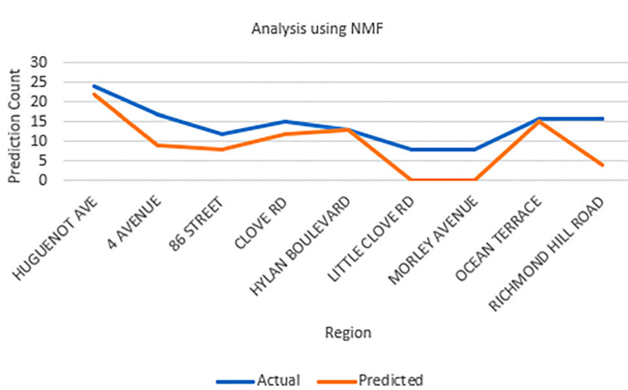

Figure 6: Graphical representation of the actual versus predicted traffic count output PCA (dataset1)

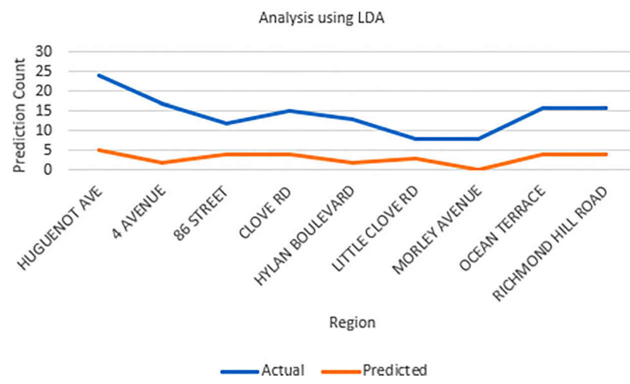

Figure 7: Graphical representation of the actual versus predicted traffic count output LDA (dataset1)

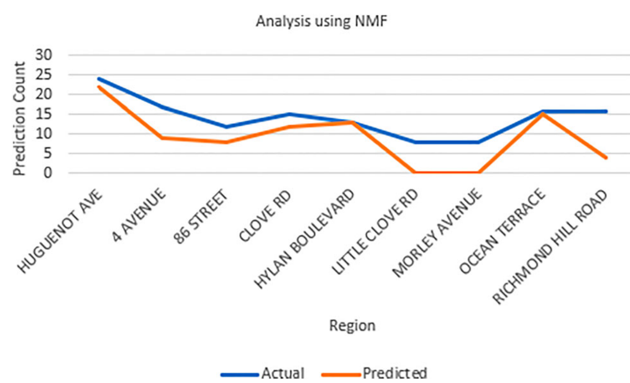

Figure 8: Graphical representation of the actual versus predicted traffic count output NMF (dataset1)

In Figures 9–11, the misclassification error rate comparison is represented graphically which are numerically presented in Table 1. It is clear that the misclassification error rate is comparatively less to the classification dataset than the vehicular traffic count dataset. Also, the error rates during the PCA analysis is comparatively reduced. Thus, with the given vehicular dataset as input, the entire analysis is carried out using dimensionality reduction

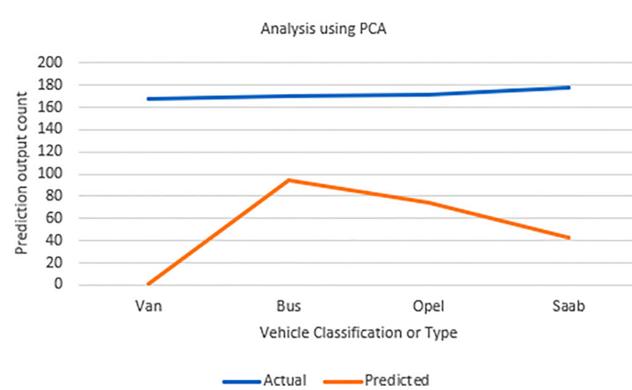

Figure 9: Graphical representation of the actual versus predicted traffic count output PCA (dataset2)

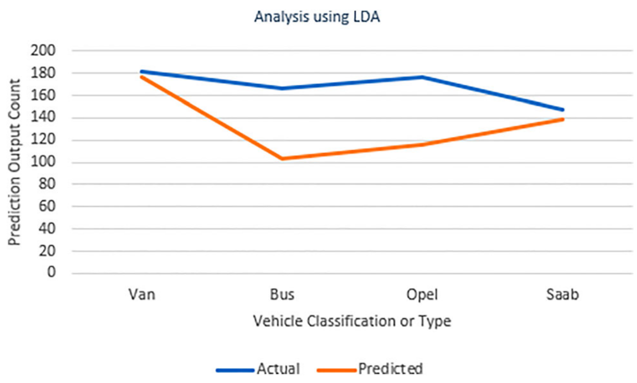

Figure 10: Graphical representation of the actual versus predicted traffic count output LDA (dataset2)

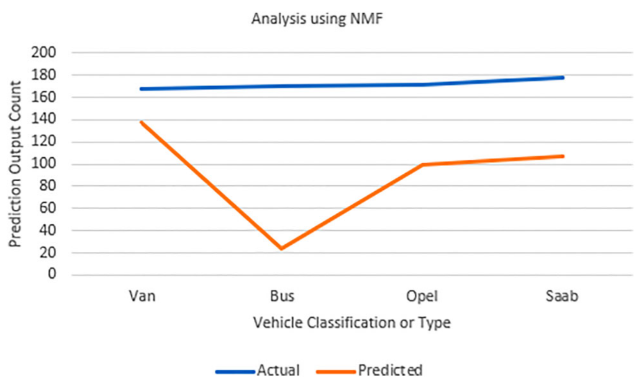

Figure 11: Graphical representation of the actual versus predicted traffic count output NMF (dataset2)

Table 1: Comparative analysis of misclassification error rate using PCA, LDA and NMF in datasets

Methods	Dataset#1	Dataset#2	Average
LDA	36%	38%	37%
PCA	32%	41%	36.5%
NMF	35%	61%	48%

techniques. These analyses and predictions will pave the way for various improvements in road-traffic surveillance. Also, it will lead to the reduction or avoidance of road-traffic accidents as well. Hence, improving the vehicular traffic management systems for a smart and safer environment.

3.3 Blockchain, Consensus, and Smart Contract

Blockchain comes after the PCA analysis. We have developed a generic version of Blockchain that can fit according to our input from PCA.

Figure 12 shows the block structure. We code the block structure that matches according to Figure 12 so we can have required information in the Blockchain. We saved

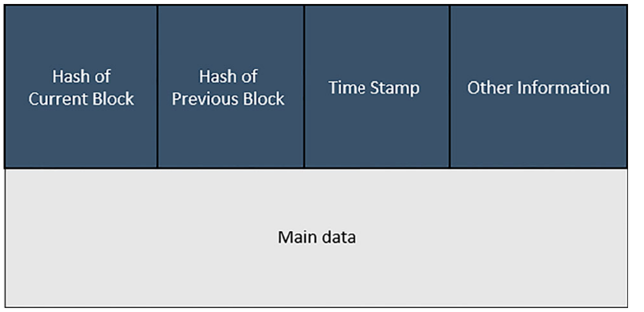

Figure 12: Block structure

the current hash of the block, previous hash, time stamp, and main data that is coming from PCA.

Figure 13 shows the linkage between two blocks in the Blockchain. To create a link between two blocks, the value for the "Hash of previous block" should match. This is already verified in the consensus process so it is reliable. If someone modifies the data, this hash will also change and Blockchain goes to an invalid state (Figure 14).

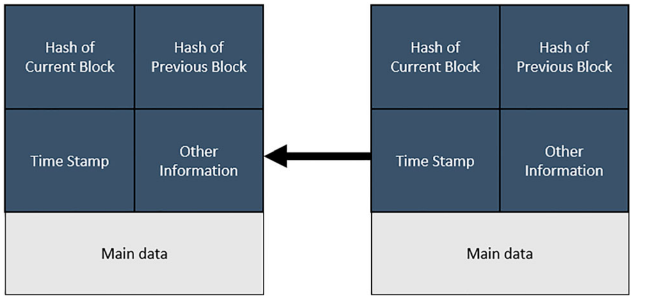

Figure 13: Block linkage

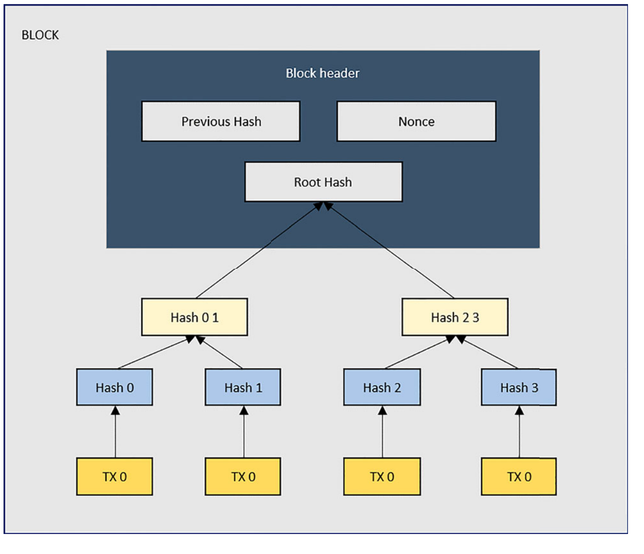

Figure 14: Merkle tree information

Data analysis is done in the PCA analysis layer. All data preprocessing will be done in the ITS layer. This is to take the burden off the PCA analysis layer. This will also check the smart contract for more information about participating parties. Once PCA analysis is done, out is sent to the consensus layer. Verifying the data is the responsibility of the consensus layer. This is a very important stage where data is verified or validated before writing it to the block. We used sha256 for hashing technique and the equation goes like

$$H^{(i)} = H^{(i-1)} + C_{M^{(i)}}(H^{(i-1)}) \tag{6}$$

H is initial value of hash and we compute it sequentially. C represent a SHA256 compression function and M is block of the message that being processed. These blocks can only be computed only one at a time. Blockchain is introduced as a separate layer. The Blockchain layer is responsible for holding the verified transaction and storing data for future use. This layer is also very important because it focuses on our privacy and security goal along with future access to data for the user.

A smart contract is the digital form of a contract. The smart contract saves the business and regulation determined by the organization. Rules and regulations may differ from one to another, which is why these contracts can hold any information stored in them.

As we build private Blockchain, consensus is done via designated member having the smart contract with participating parties. Like in the case of insurance, data will be fetched from ITS for the insurance company that have the smart contact with ITS. Data fetched from ITS will be sent for consensus to the node having the smart contracts. To create the smart contract for ITS and third parties, we assume that manual agreement has been done before this stage and all requirements are fulfilled. Following are the algorithm in pseudo-code to provide an overview of the framework with smart contract and Blockchain.

After applying algorithms, we checked for the validation state of the Blockchain. The validation check is merit on which we are testing our experiments. The status of Blockchain is tested fifty times while modifying Blockchain's information to replicate modification hacking attacks.

Figure 15 shows the results of the Blockchain state. We considered two states, namely the Valid and Invalid states. The valid state is when data is saved in the Blockchain with any problem, meanwhile, if we modify the data after it is saved then our Blockchain goes to the invalid state.

Algorithm 1 Smart contract creation among ITS and Third parties

Step 1: ITS receives third party information as input=$\boldsymbol{O_{info}}$
If $(\boldsymbol{O_{info}} == \boldsymbol{accurate})$
{
Proceed to Step 2 with Smart contract template
}
Else
{
Propmt for accurate information
}
Step 2: Creating smart contract base template on $\boldsymbol{Inf} = \boldsymbol{S_{org}}$
Step 3: Share $\boldsymbol{S_{org}}$ with involved parties
Step 4: Create data pipeline
Step 5: Finish()

Algorithm 2 Smart Contract Algorithm PCA to Blockchain

Step 1: Contract between involved parties= $\boldsymbol{S_{org}}$
Step 2: Contract between layers = $\boldsymbol{S_{LC}}$
Step 3: Send data packet $\boldsymbol{P}$ with contract $\boldsymbol{S_{UC}}$
Step 4: Sink node receive and check P **for** $\boldsymbol{S_{org}}$
If ($\boldsymbol{S_{org}} == \boldsymbol{true}$)
{
Forward it for Concensus
}
Else
{
Goto Step 6
}
Step 5: Consensus check and validate $\boldsymbol{S_{org}}$ again
If ($\boldsymbol{S_{org}}$ ***is validated***)
{
Forward packet to Blockchain
}
Else
{
Prompt and finish execution
}

Step 6: Finish()

The invalid state starts from the block where the previous hash of the next block does not match with the current hash of the current block.

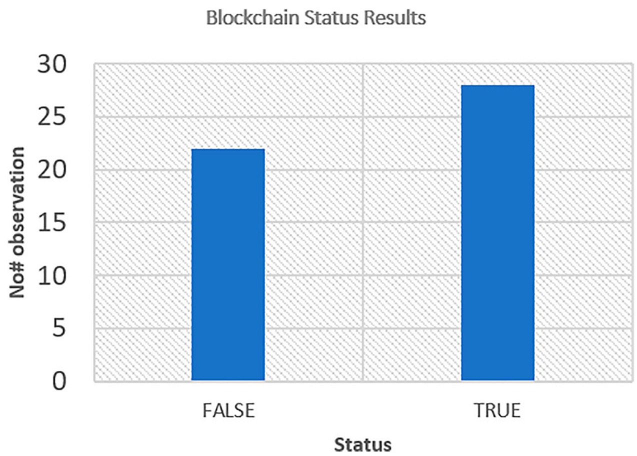

Figure 15: Blockchain status result

4. CONCLUSION

Data in ITS saved with a higher number of dimensions that are very hard to save directly in Blockchain. Data should be decreased in term of dimension before sending it to consensus or storing into blockchain database. Problem with Blockchain is, if data size goes higher and higher, the calculation time for hash increased. To avoid this situation, we conducted comparative study among three techniques namely LDA, PCA and NFM. LDA with dataset1 has 36% misclassification rate. Likewise, PCA gives 32% and NMF gave 35% error rate, respectively. While with dataset2, LDA gave 38% misclassification rate. PCA and NMF gave 41% and 61%, respectively. Our proposed framework performed efficiently with PCA. The data from ITS is reduced to the level where only relevant information can go for the consensus. We do not want all the information passes through PCA, so we introduced the smart contract to initiate the handshakes between the involved parties. Our results from PCA show that accident data is reduced up to the mark and dimensions are reduced which are suitable for Blockchain. We developed a simple blockchain to check if result can be saved and modification can be identified. We tested our blockchain with 22 false values and 28 positive values. False means that blockchain has been modified and positive means no modification has been done. Result shows that blockchain was able to detect any change or modification in blocks. Non-modified data now can be considered as digital proof. Furthermore, private Blockchain made it easy for consensus as an organization can sign this task to the people on whom they trust. Meanwhile, the insurance company can use such information from the Blockchain to get insight for any unfortunate events on the road. Customers can also access these data to present it in the court as data stored in Blockchain is immutable. If data is modified by a hacking attack, our framework will show the block which has been modified and after modified blocks, Blockchain will go to the invalid state. This ensures more security and decreases the time to re-verify data.

ACKNOWLEDGEMENTS

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF2017R1C1B5017464).

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government [grant number 2020R1A2C1012196].

ORCID

Malik Junaid Jami Gul http://orcid.org/0000-0002-2371-4829

REFERENCES

1. L. Figueiredo, I. Jesus, J. a. T. Machado, J. R. Ferreira, and J. L. M. De Carvalho, "Towards the development of intelligent transportation systems," in *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, Oakland, CA, USA, 2001, pp. 1206–1211.

2. S. H. An, B. H. Lee, and D. R. Shin, "A survey of intelligent transportation systems," in *Proceedings – 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*, 2011, pp. 332–337. Available: https://doi.org/10.1109/CICSyN.2011.76

3. P. Papadimitratos, A. La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," *IEEE Commun. Mag.*, Vol. 47, no. 11, pp. 84–95, 2009. doi: 10.1109/MCOM.2009.5307471

4. A. Paul, A. Daniel, A. Ahmad, and S. Rho, "Cooperative cognitive intelligence for internet of vehicles," *IEEE Syst. J.*, Vol. 11, no. 3, pp. 1249–1258, 2017. doi: 10.1109/JSYST.2015.2411856

5. A. Bener, F. M. Abu-Zidan, A. K. Bensiali, A. A. Al-Mulla, and K. S. Jadaan, "Strategy to improve road safety in developing countries," *Saudi Med. J.*, Vol. 24, no. 6, pp. 603–608, 2003.

6. K. Goniewicz, M. Goniewicz, W. Pawłowski, and P. Fiedor, "Road accident rates: strategies and programmes for improving road traffic safety," *Eur. J. Trauma Emerg. Surg.*, Vol. 42, no. 4, pp. 433–438, 2016. doi: 10.1007/s00068-015-0544-6

7. G. Pauer, "Development potentials and strategic objectives of intelligent transport systems improving road safety," *Transp. Telecommun. J.*, Vol. 18, no. 1, pp. 15–24, 2017. doi: 10.1515/ttj-2017-0002

8. F. Zhu, Z. Li, S. Chen, and G. Xiong, "Parallel transportation management and control system and its applications in building smart cities," *IEEE. Trans. Intell. Transp. Syst.*, Vol. 17, no. 6, pp. 1576–1585, 2016. doi: 10.1109/TITS.2015.2506156

9. H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-Khah, and P. Siano, "Iot-based smart cities: A survey," in *EEEIC 2016 – International Conference on Environment and Electrical Engineering*, 2016, pp. 1–6. Available: https://doi.org/10.1109/EEEIC.2016.7555867

10. "Global smart transport market size 2020 Statista." [Online]. Available: https://www.statista.com/statistics/279583/size-of-the-global-smart-transport-market/ [Accessed January 24, 2020].

11. B. Ran, P. J. Jin, D. Boyce, T. Z. Qiu, and Y. Cheng, "Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-generation transportation modeling," *J. Intell. Transp. Syst.*, Vol. 16, no. 4, pp. 226–242, 2012. doi: 10.1080/15472450.2012.710158

12. "Scorecard – INRIX." [Online]. Available: https://inrix.com/scorecard/ [Accessed January 24, 2020].

13. "WHO World Health Organization." [Online]. Available: http://gamapserver.who.int/gho/interactive_charts/road_safety/road_traffic_deaths/atlas.html [Accessed January 24, 2020].

14. "Traffic Violations in USA Kaggle." [Online]. Available: https://www.kaggle.com/felix4guti/traffic-violations-in-usa [Accessed January 24, 2020].

15. H. Abdi, and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, Vol. 2, no. 4, pp. 433–459, 2010. doi: 10.1002/wics.101

16. Y. Aït-Sahalia, and D. Xiu, "Principal component analysis of high-frequency data," *J. Am. Stat. Assoc.*, Vol. 114, no. 525, pp. 287–303, 2019. doi: 10.1080/01621459.2017.1401542

17. K. Sun, *et al.*, "Classification of soundscapes of urban public open spaces," *Landsc. Urban Plan.*, Vol. 189, pp. 139–155, May 2019. doi: 10.1016/j.landurbplan.2019.04.016

18. A. J. Milewska, D. Jankowska, D. Citko, T. Więsak, B. Acacio, and R. Milewski, "The use of principal component analysis and logistic regression in prediction of infertility treatment outcome," *Stud. Log. Gramm. Rhetor*, Vol. 39, no. 1, pp. 7–23, 2014. doi: 10.2478/slgr-2014-0043

19. A. Daniel, A. Paul, A. Ahmad, and S. Rho, "Cooperative intelligence of vehicles for intelligent transportation systems (ITS)," *Wireless Pers. Commun.*, Vol. 87, no. 2, pp. 461–484, 2016. doi: 10.1007/s11277-015-3078-7

20. J. Ye, "Least squares linear discriminant analysis," in *ACM International Conference Proceeding Series*, 2007. Available: https://doi.org/10.1145/1273496.1273633

21. S. Lei, B. Zhang, Y. Wang, B. Dong, X. Li, and F. Xiao, "Object recognition using non-negative matrix factorization with sparseness constraint and neural network," *Information (Switzerland)*, Vol. 10, no. 2, p. 37, 2019. Available: https://doi.org/10.3390/info10020037

22. B. Ran, P. J. Jin, D. Boyce, T. Z. Qiu, and Y. Cheng, "Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-Generation transportation modeling," *J. Intell. Transp. Syst.*, Vol. 16, no. 4, pp. 226–242, Nov. 2012. doi: 10.1080/15472450.2012.710158

23. V. C. K. Cheung, K. Devarajan, G. Severini, A. Turolla, and P. Bonato, "Decomposing time series data by a non-negative matrix factorization algorithm with temporally constrained coefficients," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015. Available: https://doi.org/10.1109/EMBC.2015.7319146

24. "2012 Vehicle Collisions Investigated by State Police – Data.gov." [Online]. Available: https://catalog.data.gov/dataset/2012-vehicle-collisions-investigated-by-state-police [Accessed January 24, 2020].

25. C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer (Long. Beach. Calif).*, Vol. 50, no. 7, pp. 80–84, 2017. doi: 10.1109/MC.2017.201

26. G. Kambourakis, C. Kolias, and A. Stavrou, "The Mirai botnet and the IoT Zombie Armies," in *Proceedings – IEEE Military Communications Conference MILCOM*, Oct. 2017, Vol. 2017, pp. 267–272.

27. A. Zand, G. Modelo-Howard, A. Tongaonkar, S. J. Lee, C. Kruegel, and G. Vigna, "Demystifying DDoS as a service," *IEEE Commun. Mag.*, Vol. 55, pp. 14–21, 2017. doi: 10.1109/MCOM.2017.1600980

28. J. J. Santanna, and A. Sperotto, "Characterizing and mitigating the DDoS-as-a-service phenomenon," in *Proceedings of the Procedia CIRP*, A. Sperotto, G. Doyen, S. Latré, M. Charalambides, B. Stiller, Eds., Berlin: Springer Berlin Heidelberg, 2014, pp. 74–78.

29. A. a. Cardenas, P. K. Manadhata, and S. P. Rajan, "Big data analytics for security," *IEEE Secur. Priv.*, Vol. 11, pp. 74–76, 2013. doi: 10.1109/MSP.2013.138

30. I. Lee, and K. Lee, "The internet of things (IoT): Applications, investments, and challenges for enterprises," *Bus.*

Horiz., Vol. 58, pp. 431–440, 2015. doi: 10.1016/j.bushor.2015.03.008

31. M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, “Big IoT data analytics: Architecture, opportunities, and open research challenges,” *IEEE Access*, Vol. 5, pp. 5247–5261, 2017. doi: 10.1109/ACCESS.2017.2689040

32. D. Mourtzis, E. Vlachou, and N. Milas, “Industrial big data as a result of IoT adoption in manufacturing,” *Procedia CIRP*, Vol. 55, pp. 290–295, 2016. Proceedings of the Monitoring and Securing Virtualized Networks and Services. doi: 10.1016/j.procir.2016.07.038

33. E. Ahmed, I. Yaqoob, I. A. T. Hashem, I. Khan, A. I. A. Ahmed, M. Imran, and A. V. Vasilakos, “The role of big data analytics in internet of things,” *Comput. Netw.*, Vol. 129, pp. 459–471, 2017. doi: 10.1016/j.comnet.2017.06.013

34. J. Pacheco, and S. Hariri, “IoT security framework for smart cyber infrastructures,” in *Proceedings of the Proceedings – IEEE 1st International Workshops on Foundations and Applications of Self-Systems, FAS-W 2016*, Augsburg, Germany, 2016, pp. 242–247.

35. S. Tweneboah-Koduah, K. E. Skouby, and R. Tadayoni, “Cyber security threats to IoT applications and service domains,” *Wireless Pers. Commun.*, Vol. 95, pp. 169–185, 2017. doi: 10.1007/s11277-017-4434-6

36. M. Frustaci, P. Pace, G. Aloi, and G. Fortino, “Evaluating critical security issues of the IoT world: Present and future challenges,” *IEEE Internet Things J.*, Vol. 5, no. 4, pp. 2483–2495, 2017.

37. “Symantec Internet Security Threat Report – ISTR.” Symantec [Online] 2017.

38. “2011–2013 NYC Traffic Volume Counts Kaggle.” [Online]. Available: https://www.kaggle.com/hanriver0618/2011-2013-nyc-traffic-volume-counts [Accessed: Jan. 24, 2020].

39. “Vehicle – Dataset – DataHub – Frictionless Data.” [Online]. Available: https://datahub.io/machine-learning/vehicle [Accessed: Jan. 24, 2020].

40. R. Abdul, A. Paul, J. Gul M., W.-H. Hong, and H. Seo, “Exploiting small world problems in a SIoT environment,” *Energies*, Vol. 11, no. 8, p. 2089, 2018. doi: 10.3390/en11082089

41. M. M. Rathore, F. Saeed, A. Rehman, A. Paul, and A. Daniel, “Intrusion detection using decision tree model in high-speed environment,” in *Proceedings of the 2018 International Conference on Soft-computing and Network Security (ICSNS)*, IEEE, 2018, pp. 1–4.

42. A. Rehman, M. Mazhar Rathore, A. Paul, F. Saeed, and R. W. Ahmad, “Vehicular traffic optimisation and even distribution using ant colony in smart city environment,” *IET Intell. Transp. Syst.*, Vol. 12, pp. 594–601, 2018. doi: 10.1049/iet-its.2017.0308

43. M. J. Gul, P. Samin, and R. Riaz, “Live devices: Insider threat to resources,” *Int. J. Comput. Sci. Inf. Technol.*, Vol. 5, pp. 1131–1134, 2014.

44. I.-C. Lin, and T.-C. Liao, “A survey of blockchain security issues and challenges,” *Int. J. Netw. Secur.*, Vol. 1919, pp. 653–659, 2017. doi: 10.6633/IJNS.201709.19(5).01

45. A. Gervais, G. O. Karame, V. Capkun, and S. Capkun, “Is bitcoin a decentralized currency?” *IEEE Secur. Priv.*, Vol. 12, no. 3, pp. 54–60, 2014.

Authors

Anandkumar Balasubramaniam received his Bachelors in Electronics and Communication Engineering Under Anna University, Chennai, India in 2012. Currently, he is pursuing his Masters Combined Ph.D. program under the supervision of Prof. Anand Paul at Kyungpook National University, Daegu, South Korea. His research areas include Intelligent Transportation Systems, Internet of Vehicles, Smart City and Smart Vehicles.

Malik Junaid Jami Gul received his Master's degree in computer science and Information Technology, Pakistan in 2015. Currently, he is pursuing his Ph.D. with Dr Anand Paul at Kyungpook National University, Daegu, South Korea. His research interests include Big Data Analytics, Internet of Things, Smart Systems, Network Traffic Analysis and Monitoring, Smart City, Operating system security, Intrusion Detection, and Computer and Network Security.

Varun G Menon (SM'19) is currently Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has completed his Ph. D in Computer Science and Engineering from Sathyabama University, India in 2017. His research interests include sensors, IoT, fog computing, brain computer interface and underwater acoustic sensor networks. He is a Distinguished Speaker of ACM. He is currently Associate Editor for IET Quantum Communications Journal, Editorial Member of IEEE Technology Policy and Ethics and a guest editor for *IEEE Sensors Journal*, *IEEE Internet of Things Magazine*, *IEEE Transactions on Industry Informatics*, *Computer Communications Journal and the Journal of Supercomputing*.
varunmenon@ieee.org

Anand Paul received the Ph.D. degree in Electrical Engineering from the National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently working as an Associate Professor in the School of Computer Science and Engineering, Kyungpook National University, South Korea. He is a delegate representing South Korea for M2M focus group and for MPEG. His research interests include Algorithm and Architecture Reconfigurable Embedded Computing. He is IEEE Senior member and has guest edited various international journals and he is also part of Editorial Team for *Journal of Platform Technology*, ACM Applied Computing review and Cyber–Physical Systems. He serves as a reviewer for various IEEE /IET/Springer and Elsevier journals. He is the track chair for Smart human computer interaction in ACM SAC 2015, 2014. He was the recipient of the Outstanding International Student Scholarship award in 2004–2010, the Best Paper Award in National Computer Symposium and in 2009, and International Conference on Soft computing and Network Security, India in 2015.
Corresponding author. Email: paul.editor@gmail.com