Research on Information Dissemination Security Based on Generative Adversarial Network in Internet of Vehicle Environment

Junting Zhang
(Corresponding author: Junting Zhang)

School of Automobile, Henan College of transportation
Zhengzhou 450000 China
Email: byoungholee@qq.com

(Received July 20, 2023; Revised and Accepted Jan. 17, 2024; First Online Feb. 23, 2024)
The Special Issue on Data Fusion, Deep Learning and Optimization Algorithms for Data Privacy Protection
Special Editor: Prof. Shoulin Yin (Harbin Institute of Technology)

Abstract

Information security transmission is very important in the Internet of Vehicles. In the traditional methods, the remote node storage operation is not cooperative. In addition, the data block with low service evaluation is not effectively processed, and the duplicate redundancy is persistent, resulting in increased bandwidth pressure. Hence, we propose an information dissemination security model based on a generative adversarial network in the Internet vehicle environment. The data is enhanced with generative adversarial networks. Based on the enhanced data, practical features can be extracted to analyze the transmitted data effectively. Experimental results show that the proposed method enhances system security, information transmission security, and network performance compared with other methods.

Keywords: Data Enhancement; Generative Adversarial Network; Information Dissemination Security; Internet of Vehicle

1 Introduction

CACC (cooperative adaptive cruise control) [18,20] based on Internet of vehicle (IoV) means that the vehicle equipped with adaptive cruise control system uses mobile wireless communication technology such as DSRC (dedicated short range communications) [1, 14] to exchange information with roads, people and cloud to solve practical problems such as traffic accidents and traffic congestion. However, with the application and deployment of IoV and vehicle-road collaboration technologies, due to the openness of mobile communication and application information [7, 15, 21], they not only meet the requirements of vehicle information interaction, but also provide opportunities for virus intrusion. The intrusion of malicious code can be accompanied by the transmission of information between CACC vehicles, which can steal user privacy or interfere with the normal running of vehicles, resulting in serious security threats.

At present, the classical method to analyze the information transmission in the environment of IoV is to use simulation method. Yao *et al.* [22] reviewed the simulation methods in detail and analyzed the nonlinear relationship between traffic flow, information flow and vehicle-vehicle communication events. Based on the existing micro-traffic simulation model, Gupta *et al.* [8] built an information transmission simulation framework in the environment of the IoV and simulated the information transmission between vehicles on the expressway. In references [4, 13, 16], traffic flow was simplified as static traffic flow, and by analyzing the statistical distribution of traffic parameters, the constraints faced by information flow propagation were discussed. However, these simulation methods lack strict theoretical model support. Zhou *et al.* [24] considered communication constraints, analyzed the relationship between information propagation and traffic flow mechanics, established an information flow propagation model, and described the dynamic behavior of information flow propagation. Du *et al.* [6] divided the road into multiple cell units and constructed an information-coupled cell transport model (IT-CTM) to capture the flow of information within and between cells. Adaptive CruiseControl (ACC) is a widely used driver assistance system, and it is also the first step to realize automatic driving. ACC technology mainly collects the distance between the vehicle and the forward vehicle through on-board sensors such as radar, and automatically adjusts the speed of the vehicle to maintain a reasonable safety distance with the front vehicle. Collaborative adaptive cruise control technology is the second step to realize autonomous driving, which means that vehicles equipped with adaptive cruise control system realize vehicle-to-vehicle interconnection communication through vehicle-to-vehicle networking technology. That is, the vehicle can not only sense the environment ahead through radar sensors, but also obtain information between the vehicle in front and other vehicles through wireless communication technologies such as DSRC, so that the connected vehicles can cooperate to complete the control and manipulation. Compared with ACC, CACC technology enables vehicles to obtain more information about surrounding vehicles, such as speed, acceleration, etc., and can react in real time according to the running status of other vehicles. Therefore, it can shorten the safety distance between vehicles, reduce the speed fluctuation of the road fleet, improve the operating efficiency of road traffic flow, reduce the incidence of road traffic accidents, and improve the driving comfort.

To sum up, most of the studies conducted by experts and scholars on the information dissemination of the Internet of vehicles aimed at the well-intentioned information, without considering the intrusion of malicious virus information. In addition, most of the current information dissemination models do not consider the interaction between different types of traffic flows. Based on this, aiming at the realistic road traffic environment in which CACC vehicles and ordinary vehicles are driving together, this paper constructs a dynamic model of virus transmission in the vehicle-connected environment, aiming to provide a theoretical basis for studying the virus transmission law in the vehicle-connected environment, and provide an effective method for suppressing the transmission of virus information, so as to ensure the safety of CACC vehicle information transmission.

2 Proposed Information Dissemination Security Model

Generative adversarial network (GAN) is a generative model based on zero-sum game idea [12, 19], which has been applied in many fields such as image generation, traffic sign recognition and traffic accident detection. GAN is mainly composed of generator and discriminator. The main function of the generator is to learn the distribution of real data, and the discriminator determines whether the data is real data or the data generated by the generator based on the input data. According to the discriminant results, the generator G and discriminator D continue to optimize and finally reach a Nash equilibrium, that is, the discriminator D cannot determine whether the input data is real or generated data.

The objective function of GAN optimization process is:

$$min_G max_D L(G, D) = A + B. \tag{1}$$

Where $A = E_{x-P_{data(x)}} log(D(x))$, $B = E_{z-P_{z(z)}} log(1-D(G(z)))$, E is the mathematical expectation of the distribution specified in the subscript. $P_{data(x)}$ is the distribution of the real data x. $P_{z(z)}$ is the distribution of generated data. $D(x)$ is the discriminant function of the discriminator. $G(z)$ is the data generated by the generator.

In this paper, both the generator and discriminator of GAN are fully connected neural networks, and the active function is the rectified linear unit (ReLU) [3, 9] function, which can effectively alleviate the gradient disappearance problem. The generator and discriminator are optimized using the Adam optimizer [11], which has the advantages of high computational efficiency and is suitable for unstable objective functions.

The learning rate of generator and discriminator is set to 0.01, and the number of neurons in the hidden layer of generator and discriminator is set to 128. After 10^4 rounds of iterative training, GAN extended data set containing 8058 on-board data and 12060 normal operation data is obtained. In order to test the performance of the GAN proposed in this paper, a normalization method is used to process the extended and original GAN data sets to eliminate the dimensional effects and facilitate the comparison of data distribution between the two types of data sets. The normalization processing formula is as follows:

$$y_i = \frac{x_i - \underset{1\leq j\leq n}{min}}{\underset{1\leq j\leq n}{max} - \underset{1\leq j\leq n}{min}} \tag{2}$$

Where, x_i and y_i are the data before and after normalization processing respectively.

In order to analyze the communication probability of CACC vehicles, this paper first automatically divides the road into several cells of equal length based on cells, and each car occupies one cell [2]. To make the simulated traffic flow more consistent with the real situation, the safe distance model was introduced to further improve the simulation accuracy. The model evolution process includes constant velocity, acceleration, deceleration and position updating. Kim *et al.* [11] believed that if CACC vehicles wanted to complete the communication function, there should be at least one CACC vehicle in a communication range on the road network. However, considering that the transmission of wireless signal between vehicles will be affected by other uncertain factors such as driving speed, channel quality and environmental changes, the wireless signal will be dynamically attenuated during transmission. Therefore, if the CACC vehicle wants to complete the communication function, it must ensure the reliability of signal transmission in the communication range while ensuring that there is at least one CACC vehicle in the communication range. Based on this, this paper proposes a calculation method for the communication probability of CACC vehicles in the networked vehicle environment, and the communication success probability P_{com} is:

$$P_{com} = [1 - (P1 + P2)^m]P_{suc}. \tag{3}$$

Where m is the total number of cells discretized into roads within the communication range; $P1$ is the probability that the cell is not occupied by the vehicle; $P2$ is the probability that the cell is occupied by ordinary vehicles. P_{suc} is the probability that the signal will be reliably received.

$$P1 = 1 - \frac{N(t)}{N}. \tag{4}$$

$$P2 = \frac{(1-q)N(t)}{N}. \tag{5}$$

Where $N(t)$ is the total number of vehicles at time t of simulation; N is the total number of discrete cells of the entire lane. q is the ratio of CACC vehicles on road.

The bidirectional generative adversarial network adopts the adversarial tuple strategy, which can not only generate the same data distribution as the training sample, but also output the hidden space feature representation of the training sample. The bidirectional generation adversarial network can capture the hidden space feature representation of training sample data through multiple iterative adversarial networks. Therefore, by making full use of the image semantic feature abstraction ability of the bidirectional generation network and generating the feature sequence with the image essential feature description ability, the perceptual hash code with stronger image content representation ability can be constructed and the image content forensics performance can be improved. In this study, a perception hash image content forensics algorithm based on BiGAN is proposed. By optimizing and enhancing the bidirectional generative adversarial network structure, the bidirectional generative adversarial network can improve its ability to learn complex data distribution, and generate data distribution consistent with complex training samples (such as natural images). At the same time, the learning performance of the potential features of the sample data is enhanced, the hidden space feature representation of the sample image is output and the perceptual hash code is quantitatively generated, which realizes the image content authentication and copyright protection based on the perceptual hash.

The basic model of the image perception hash generation network based on BiGAN consists of four subnetworks, namely as shown in Figure 1, encoding network E, generation network G, joint discriminant network D and jump-layer network S. Where, the encoding network E implements mapping $E : x \to E(x)$ from the original image data x to the potential feature representation $E(x)$. The input is the normalized training image, and the output is the image hidden space feature encoding. The generation network G maps the preset noise distribution z to a data distribution $G(z)$ that is consistent with the target image sample, $G : z \to G(z)$. Joint discriminant D network to distinguish the input data tuples from encoding or generate network $D : ((x, E(x)), (G(z), z)) \to 0, 1$. Aiming at the problems of low quality of generated image and insufficient ability of output feature code representation in bidirectional generative adversarial network, a jump layer network S is added between encoding network E and generating network G to realize the transmission of different dimension feature information between coding network E and generating network G, and the mean square error (MSE) loss is added to the network optimization loss. The content representation ability of the images generated by the generation network G is enhanced, so that the generation network G can output high-quality images with complex texture distribution. At the same time, based on the counter loss of the joint discrimination network D, the reverse excitation coding network E outputs more representative image hidden space feature coding, and improves the quality of the generated image perceptual hash code.

Where, E represents the coding network, G represents the generation network, D represents the joint discriminant network, S represents the added jump-layer network, RealImage represents the training sample Image, and Generated Image represents the generated image.

The perceptual hash generation algorithm based on BiGAN makes full use of the self-learning ability of the bidirectional generative adversarial network, and generates the image perceptual hash code by encoding the hidden space features output by the coding network E. Through the iterative confrontation between the generation network G, coding network E and joint discriminant network D, as well as the optimization and enhancement of the jump-layer network S and the mean square error loss MSE, the hidden space feature representation capability of the image perception hash code is continuously improved, and the effective balance between the authentication robustness and discrimination sensitivity of the perception hash code is achieved.

In the perception hash generation network based on BiGAN, the role of coding network E is to extract the hidden space feature information from the original image and generate the image perception hash sequence. The encoding network designed in this study consists of 10 layers of convolutional neural network, each layer of convolutional neural network includes three data operation processing: image *Conv2d*, *BatchNorm* and activation (*LeakRelu*). The initial input of the coding network is the normalized training sample image. At the same time, in order to improve the hidden space feature representation ability of the output feature coding, the convolutional layer output of the fifth, sixth and seventh layers of the encoding network E is transmitted to the same dimensional network layer of the generating network G through the jump-layer network S as an input, and the image feature information of different dimensions extracted by the coding network E is linked to the generating network G to improve the visual quality of the generated image and the network convergence speed. The convolutional output of the last layer of the coding network E is used as the image hidden space feature representation sequence to generate the image perception hash code. The detailed parameter information of coding network E in Bigan-based perceptive hash generation network is shown in Table 1. Among them, the parameters in the first column of the config-

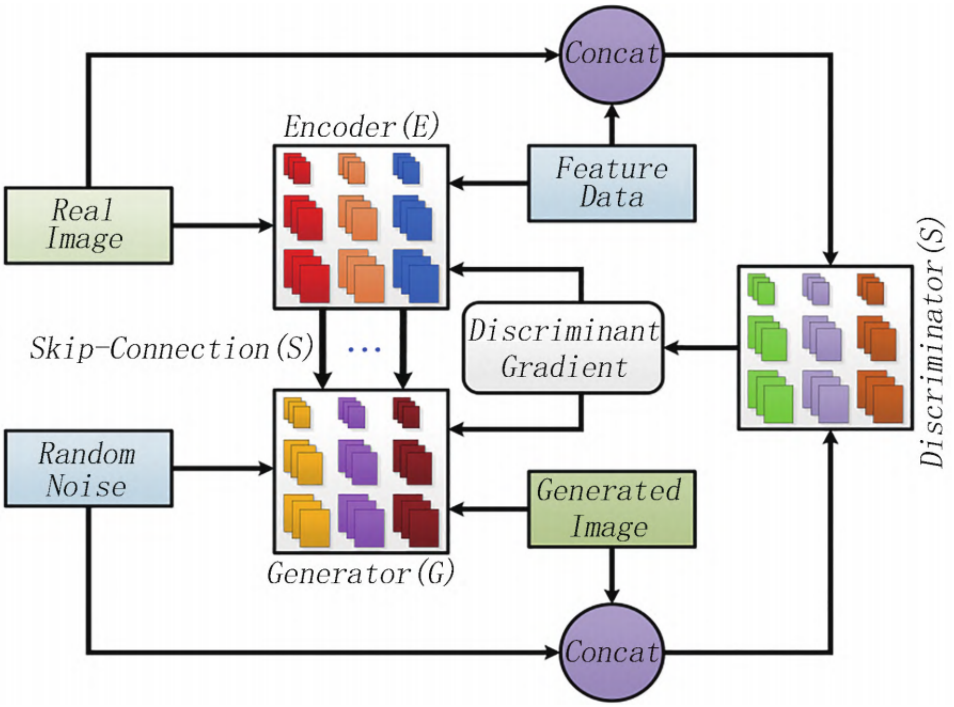


Figure 1: The structure of perceptive hash image forensics algorithm based on BiGAN

uration column represent the number of filters. It can be seen that the coding network E adopts the multiplication method to increase the number of filters, and finally forms an image perception hash code with good hidden space feature representation ability. The parameters in the second column are the size of the sensitivity field of the convolutional operation, the length of the convolutional step, and the number of rows/columns added to the input side. The third column parameter is the slope of LeakRelu activation function adopted. In addition to the last convolution layer, BatchNorm operation is applied after every convolution operation in the experiment to ensure that the training sample data is distributed in the sensitive region of the activation function, so as to avoid the disappearance of gradient and speed up the convergence of the model. In the experiment, a self-learning coding network structure model with 10-layer network structure, including 10 convolutional layers, 9 batch normalization layers and 9 activation layers, is selected to ensure that the network output hidden space feature coding with strong representation ability, considering the influence of the number of convolutional layers on the network operation efficiency and the ability to generate sensing hash code feature representation.

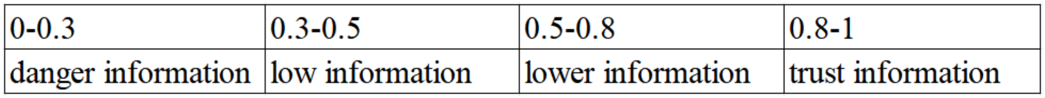
3 Simulation Experiment

Due to the participation of a large number of users, each user may set a different local policy II, which will affect the stability of the P2P network, so the local policy should be given by the designer according to the security prin-

Table 1: Detailed parameter design in encoding network

Layer	Function	value
1	Conv2d, BatchNorm2d, LeakyRelu	32,[3,1,1],0.01
2	Conv2d, BatchNorm2d, LeakyRelu	64,[4,2,1],0.01
3	Conv2d, BatchNorm2d, LeakyRelu	128,[4,2,1],0.01
4	Conv2d, BatchNorm2d, LeakyRelu	256,[5,1,0],0.01
5	Conv2d, BatchNorm2d, LeakyRelu	512,[4,2,0],0.01
6	Conv2d, BatchNorm2d, LeakyRelu	512,[4,1,0],0.01
7	Conv2d, BatchNorm2d, LeakyRelu	512,[4,2,0],0.01
8	Conv2d, BatchNorm2d, LeakyRelu	1024,[4,1,0],0.01
9	Conv2d, BatchNorm2d, LeakyRelu	2048,[1,1,0],0.01
10	Conv2d, BatchNorm2d, LeakyRelu	1024,[1,1,0]

Table 2: Service evaluation analysis



0-0.3	0.3-0.5	0.5-0.8	0.8-1
danger information	low information	lower information	trust information

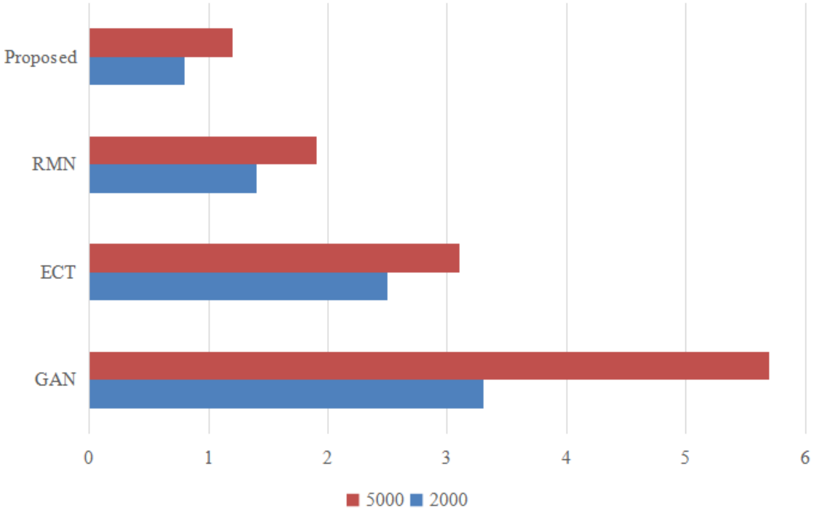


Figure 2: Visual result

ciple of the default value, common users use the default value. In order to explain the value range of local policy II, according to the general safety principle, the analysis diagram is given, the designer can determine the specific value according to the weighted result.

This model takes 4 language variables y, k, w and z to analyze service evaluation, and their weights are all 25%, as shown in Table 2.

When the information is published, it is first divided into several data blocks, and distributed agents store it in neighboring nodes. In this case, subsequent nodes need to search for data blocks when accessing them. Since the local resource list stores a large amount of data block information, the subsequent node only needs to search for one data block, which is a constant less than the total number of data blocks. Its value size is restricted by the amount of information stored in the local resource list, and its search efficiency is comprehensively affected by data block availability, reliability k, optimal search, and data integrity w, etc. Their thresholds are generally set at 0.5, 0.6, 0.5, 0.3. When they are lower than this value, the data service evaluation value is lower than 0.5, then the data is cached or cleared; Otherwise, the data is safely available.

Because the higher the activity of nodes, the more times the nodes are evaluated, the greater the sharing degree of the stored data blocks, the more reliable the comprehensive service evaluation, the effective supervision of data security, and the enhanced stability of the system. You can modify the threshold if you need to search for as much information as possible but the information is not widely shared.

Between any two nodes in the cluster, because of the uniqueness of its coding, there must be different bits of coding, so the probability of establishing a communication link between any two nodes is 100%. The comparison of the communication cost simulation experiment data between the random key model and the information transmission model in the process of key establishment is shown in Table 3 and Figure 2.

Table 3: Comparison of communication overhead with different nodes/s

Model	2000	5000
GAN [25]	3.3	5.7
ECT [26]	2.5	3.1
RMN [5]	1.4	1.9
Proposed	0.8	1.2

As can be seen from Table 1, when the number of nodes is relatively small, the communication cost of the transmission model is less than that of the random key model, because the key path establishment between nodes in the transmission model only needs one communication. In order to improve the security performance, the random key model does not create paths at one time, but generates new keys through mutual random numbers. In the first communication between nodes in the cluster [10, 17, 23], two more times of communication will be carried out. Relative to the improvement of its security performance, a small increase in the communication overhead between nodes in the cluster is acceptable. When the number of nodes increases sharply, the communication cost of the random key model is less than that of the transmission model, because with the increase of the number of nodes, the intercluster communication greatly increases, and the probability of the direct first path establishment of the random key model in the intercluster communication is increased.

4 Conclusions

Based on the dual key building algorithm based on GAN model, a new random key building algorithm based on dual coding is proposed. Theoretical analysis and simulation results show that the new algorithm can effectively improve the probability and security performance of direct dual key establishment between any two nodes, and also save the communication cost in the communication between nodes. Therefore, it can be considered that this is a better performance IoV key establishment algorithm.

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] E. Abuhdima *et al.*, "Impact of Dust and Sand on 5G Communications for Connected Vehicles Applications," *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 229-239, 2022.

[2] I. Ahmed, E. Balestrieri, P. Daponte and F. Lamonaca, "A Method Based on Ellipse Fitting for Automatic Morphometric Parameter Measurements of Fish Blood Cells," in *2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Jeju, Korea, Republic of*, pp. 1-6, 2023, doi: 10.1109/MeMeA57477.2023.10171881.

[3] T. Devi and N. Deepa, "A novel intervention method for aspect-based emotion Using Exponential Linear Unit (ELU) activation function in a Deep Neural Network," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India*, pp. 1671-1675, 2021, doi: 10.1109/ICICCS51141.2021.9432223.

[4] Z. Ding, J. Xiang, "Overview of intelligent vehicle infrastructure cooperative simulation technology for IoV and automatic driving," *World Electric Vehicle Journal*, vol. 12, no. 4, pp. 222, 2021.

[5] Z. Dou, J. Tian, Q. Yang, L. Yang, "Design and analysis of cooperative broadcast scheme based on reliability in mesh network," *Mobile Information Systems*, vol. 2021, pp. 1-18, 2021.

[6] L. Du, S. Gong, L. Wang, X. Y. Li, "Information-traffic coupled cell transmission model for information spreading dynamics over vehicular ad hoc network on road segments," *Transportation Research Part C: Emerging Technologies*, vol. 73, pp. 30-48, 2016.

[7] T. M. Ghazal, R. A. Said, N. Taleb, "Internet of vehicles and autonomous systems with AI for medical things," *Soft Computing*, pp. 1-13, 2021.

[8] M. Gupta, R. B. Patel, S. Jain, H. Garg, B. Sharma, "Lightweight branched blockchain security framework for Internet of Vehicles," *Transactions on Emerging Telecommunications Technologies*, 2022. https://doi.org/10.1002/ett.4520.

[9] I. Jahan, M. F. Ahmed, M. O. Ali, Y. M. Jang, "Self-gated rectified linear unit for performance improvement of deep neural networks," *ICT Express*, vol. 9, pp. 3, pp. 320-325, 2023.

[10] S. U. Jan, I. A. Abbasi and M. A. Alqarni, "LMAS-SHS: A Lightweight Mutual Authentication Scheme for Smart Home Surveillance," *IEEE Access*, vol. 10, pp. 52791-52803, 2022.

[11] D. J. Kim, H. I. Kim, S. H. Lee, C. C. Chung, "Adaptive Feedforward Compensator Based on Approximated Causal Transfer Function for CACC with Communication Delay," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 15192-15197, 2020.

[12] P. Li, A. A. Laghari, M. Rashid, J. Gao, T. R. Gadekallu, A. R. Javed, S. Yin, "A Deep Multimodal Adversarial Cycle-Consistent Network for Smart Enterprise System," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 693-702, 2023.

[13] J. Liu, L. Zhang, C. Li, J. Bai, H. Lv and Z. Lv, "Blockchain-Based Secure Communication of Intelligent Transportation Digital Twins System," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22630-22640, 2022.

[14] Q. Pan, J. Wu, J. Nebhen, A. K. Bashir, Y. Su and J. Li, "Artificial Intelligence-Based Energy Efficient Communication System for Intelligent Reflecting Surface-Driven VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19714-19726, 2022.

[15] Y. Ren, F. Zhu, J. Wang, P. K. Sharma and U. Ghosh, "Novel Vote Scheme for Decision-Making Feedback Based on Blockchain in Internet of Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1639-1648, 2022.

[16] P. Sharma and H. Liu, "A Machine-Learning-Based Data-Centric Misbehavior Detection Model for Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4991-4999, 2021.

[17] Y. Sun, S. Yin, J. Liu, L. Teng, "A Certificateless Group Authenticated Key Agreement Protocol Based on Dynamic Binary Tree," *International Journal of Network Security*, vol. 21, no. 5, pp. 843-849, 2019.

[18] T. Tapli and M. Akar, "Cooperative Adaptive Cruise Control Algorithms for Vehicular Platoons Based

on Distributed Model Predictive Control," in *2020 IEEE 16th International Workshop on Advanced Motion Control (AMC), Kristiansand, Norway*, pp. 305-310, 2020, doi: 10.1109/AMC44022.2020.9244429.

[19] L. Wang, S. Yin, H. Alyami, *et al.*, "A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images," *Geoscience Data Journal*, 2022. https://doi.org/10.1002/gdj3.162.

[20] P. Wang, C. Jiang, X. Deng, L. Wang, H. Deng and Z. He, "A multi-mode cooperative adaptive cruise switching control model for connected vehicles considering abnormal communication," in *2017 6th Data Driven Control and Learning Systems (DDCLS), Chongqing, China*, pp. 739-744, 2017, doi: 10.1109/DDCLS.2017.8068165.

[21] L. Yang, A. Moubayed and A. Shami, "MTH-IDS: A Multitiered Hybrid Intrusion Detection System for Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 616-632, 2022.

[22] Z. Yao, Y. Wu, Y. Wang, B. Zhao, Y. Jiang, "Analysis of the impact of maximum platoon size of CAVs on mixed traffic flow: An analytical and simulation method," *Transportation Research Part C: Emerging Technologies*, vol. 147, pp. 103989, 2023.

[23] S. Yin, H. Li, L. Teng, A. A. Laghari, V. V. Estrela, "Attribute-based Multiparty Searchable encryption model for Privacy Protection of Text Data," *Multimedia Tools and Applications*, 2023. ttps://doi.org/10.1007/s11042-023-16818-4.

[24] L. Zhou, T. Ruan, K. Ma, C. Dong, H. Wang, "Impact of CAV platoon management on traffic flow considering degradation of control mode," *Physica A: Statistical Mechanics and Its Applications*, vol. 581, pp. 126193, 2021.

[25] J. Zhang and Y. Zhao, "Research on Intrusion Detection Method Based on Generative Adversarial Network," in *2021 International Conference on Big Data Analysis and Computer Science (BDACS), Kunming, China*, pp. 264-268, 2021, doi: 10.1109/BDACS53596.2021.00065.

[26] Y. X. Zhang, L. Zou, "Research on information dissemination on social networks based on edge-based compartmental theory," *International Journal of Modern Physics B*, vol. 35, no. 24, pp. 2150249, 2021.

Biography

Junting Zhang biography. Junting Zhang is with School of Automobile, Henan College of transportation. Research interests are Information security, Computing networks, Internet of vehicles security.