The Detection Method of Abnormal Messages based on Deep Neural Network on the CAN Bus

Chaoqun Xing[1, a]
[1]College of Information Science and Engineering, Ocean University of Chain, Shandong, Qingdao, 266100,Chnia

Keywords: Vehicle security, CAN bus, Anomaly detection, Deep neural network.

Abstract: In recent years, with the increase of the number of external interfaces and electronic control units (ECUs) in vehicles, some unknown attacks about vehicle have emerged, the safety of vehicles has gradually become a top priority for auto manufacturer and vehicle owners. In this paper, we introduce a new attack model for the CAN bus of the vehicle and propose a deep neural network (DNN) based model to detect attacks. We collect normal messages exchanged between the CAN bus by ECUs in real vehicles to construct normal data sets. According to the existing methods, the ability of the attacker is quantified. Different anomaly data sets are constructed based on the strength of the attacker. Finally, the performance of the model is verified by using the normal and abnormal data sets. The detection rate of the proposed method is far more than the existing methods with good robustness and stability.

1 INTRODUCTION

At present, automatic driving and Internet of vehicle (IoV) technology has gradually become a research hotspot in the automotive industry. The emergence of these two technologies make the vehicle easily keep communications with external devices, breaking the closure of traditional vehicle system. Such as: vehicle to vehicle (V2V) and vehicles to infrastructure (V2I)(Biswas, Tatchikou and Dion, 2006). However, the risk of malicious network attacks on vehicles will also increase with the development of these two technologies. These attacks may seriously affect the safety of drivers and passengers. Many researchers in the field of automotive safety have reported and proved various attacks in the networked vehicle. The actual experiments(Valasek and Miller, 2015) have proved that attacker use the vulnerability to access some ECUs in the vehicle remotely and gain control of them through a series of reverse operations so as to launch an attack on the vehicle; In(Mastakar, 2012), an attacker connected to the vehicle's internal network deceive all ECUs, including safety-critical components, such as brakes and engines; In(Checkoway et al.,2011)(Miller and Valasek, 2013), various attack scenarios in the networked vehicle are shown, such as brake failure, the information on the panel is incorrect, and so on. These attacks not only seriously affect the driving safety, but also cause the automobile manufacturers to face a certain degree of economic loss.

In order to improve the security of networked vehicles, relevant researchers have proposed various defence schemes against attacks on networked vehicles. Hoppe et al(2011) proposed a method to detect attacks by identifying significant attack patterns in vehicular networks (sudden increase or decrease in the number of CAN messages in a certain period of time, disappearance of the CAN ID, etc.). Müter et al(2010) defined eight "anomaly detection sensors" and six weighted "applicable standards", and used statistical models to detect anomalies. However, the above methods only consider some abnormal situations of CAN bus. They cannot monitor CAN bus comprehensively and detect all possible abnormal behaviours. Larson et al(2008) proposed a specification-based attack detection method that detects the presence of an attack by comparing the behaviour of the current specification system with the predefined patterns. The limitation of this method is that the system cannot collect all the specification behaviours in the vehicle.

In this paper, we use a DNN model to detect abnormal messages on the CAN bus. We propose a novel attack method to construct abnormal data sets,

Xing, C.
The Detection Method of Abnormal Messages based on Deep Neural Network on the CAN Bus.
DOI: 10.5220/0008869100850096
In *Proceedings of 5th International Conference on Vehicle, Mechanical and Electrical Engineering (ICVMEE 2019)*, pages 85-96
ISBN: 978-989-758-412-1

which assumes that the attacker gain the control of the engine ECU and generate abnormal messages by arbitrarily modifying the messages sent by the engine ECU. Abnormal messages correspond to different movement states of the vehicle, the attacker injects the message which is inconsistent with the current movement state of the vehicle into the CAN bus to launch the attack. Firstly, we collect engine ECU's messages in different movement states and construct the abnormal datasets by using proposed attack model. Secondly, we use the constructed normal and abnormal datasets to train the DNN model. Finally, new messages are classified using the trained model. The experimental results show that the model has the advantages of high classification accuracy rate, low computational complexity and good compatibility with vehicular network. The contributions of this paper are as follows:

1. Proposing a deep neural network model to detect the attacks for CAN messages.
2. Introducing a novel attack model for the CAN bus of the vehicle and quantifying the attacker's ability. Different abnormal datasets are constructed for different levels of attackers.
3. The performance of the model is verified by experiments.

The structure of the paper is as follows:

The second part mainly introduces the background of the CAN bus; the third part introduces the detection technology for abnormal messages of vehicles; the fourth part is mainly about the system model and the attack model. The fifth part describes our method, including the structure of network model and the analysis of experimental results; the last part is the conclusion.

2 BACKGROUND

2.1 CAN Bus

The Controller Area Network (CAN) bus protocol is a serial communication protocol that supports distributed real-time control with high security(Gmbh, 1991). It is one of the most widely used network communication protocols for vehicles. According to speeds of the data transmission, CAN bus can be divided into two categories, one is the high-speed CAN bus with data rate of 125kbps to 1Mbps, mainly used in nodes with high real-time requirements, such as engine management unit, electronic transmission control, etc. The other is the low-speed CAN bus with a data transmission rate of 5 kbps to 125 kbps. It is used in nodes with low real-time requirements, such as seat adjustment, lighting, and mirror adjustment.

2.2 The Structure and Transmission Process of Messages on the CAN Bus

Messages on the CAN bus are mainly divided into four types: data frames (standard data frames and extended data frames), remote frames, error frames, and overload frames. Standard data frames are the most common and numerous frame types in the vehicle, Therefore, standard data frames are studied in this paper. Its fields include: Start Field (SOF), Identifier Field (ID), Control Field (Control), Data Field (Data), Cyclic Redundancy Check Field (CRC), Acknowledgement Field (ACK), and End Field (EOF). The identifier field determines the priority of sending messages, it is used to avoid two nodes competing for the CAN bus at the same time; the control field represents the size of the data field; the data field represents the data information carried by the message; the cyclic redundancy check field is used to detect errors in the message; the acknowledgment field confirms whether the node has received a valid CAN message. The structure of standard data frame is shown in Figure 1.

Figure 1. The structure of standard data frame.

Messages on the CAN bus support multiple access, all nodes send and receive messages through the CAN bus in the vehicle's network. At the same time, messages on the CAN bus are broadcast to all nodes connected to the bus. Each node receives messages from other nodes, but only accepts what it needs and ignore the others. As shown in Figure 2, after the messages sent by node 1 are broadcast to nodes 2 and node 3, the two nodes will check the message to see if they need. If so, receive them; otherwise, ignore them.

The research about messages on the CAN bus mainly focuses on the ID and Data fields. In the standard data frame, the ID field has 11 bits, each ID corresponds to an ECU with a specific function in the vehicle; the Data field contains 0-64 bits of high-dimensional data information transmitted in the message, representing different parameter values of the sensor associated with the ECU. In general, the

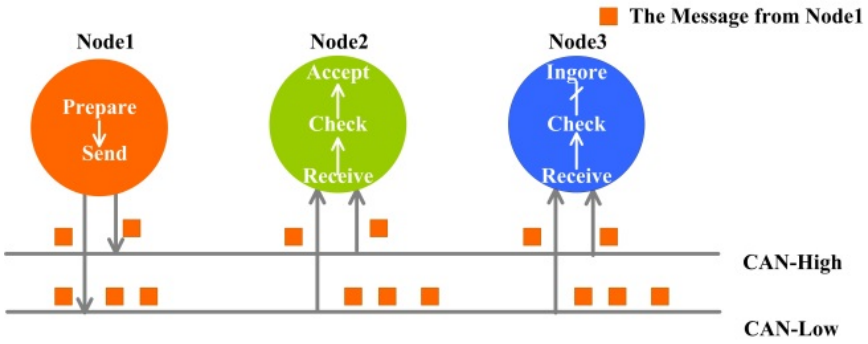

Figure 2. The process of transmitting messages on the CAN bus.

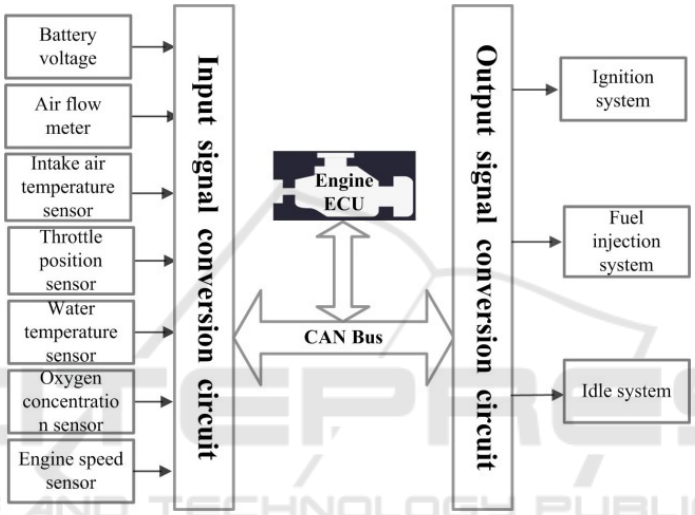

Figure 3. Topology between the engine ECU and various components.

ECU with important functions are often connected with multiple sensors, so there is a fixed ID corresponding to different Data fields in CAN messages. Engine ECU is a typical example, the change of the different bits in the Data field of the message represents the change of the parameters of different sensors in the engine, which is intuitively reflected as the change of the movement state of the vehicle. In this paper, we focus on the Data field in the message of the engine ECU.

2.3 The Topology between the CAN Bus and ECUs

Generally speaking, ECUs are microcontrollers or ARM chips that are used to control, record, or change the state of the vehicle. In modern vehicles, there are usually dozens to dozens of ECUs to control different modules in the vehicle. For example, the engine control module (PCM) receives sensor signals to control fuel supply, air distribution, and adjustment of engine intake pressure through complex calculation. The compensation control coefficient of the engine is also determined according to temperature, load, detonation and combustion conditions.

In addition to the ECU, there are a large number of sensors and actuators in the vehicle. Through their coordination to complete the complex and intelligent operation. The topology between the engine ECU and the various components is shown in Figure 3. The signal of each sensor in the engine is transmitted to the engine ECU connected with the CAN bus through the input signal conversion circuit. The ECU receives the signal and processes it, then sends the corresponding control message to the CAN bus. Finally, the output signal conversion circuit transmits the signal to the corresponding executive unit to control the vehicle.

3 RELATED WORK

3.1 Detection Technology of Abnormal Messages for the CAN Bus

In order to better prevent the malicious attacks on networked vehicles, the detection technology of abnormal messages in vehicles has been widely studied. The technique assumes that the characteristics of abnormal messages sent by an attacker are different from normal messages. The main goal is to detect attackers who attempt to destroy the integrity, confidentiality or availability of vehicle resources. When abnormal messages are detected in vehicles, the detection mechanism of abnormal messages in vehicles will promptly informs the driver by some way to trigger appropriate countermeasures.

3.1.1 Principle of Abnormal Messages Detection for the CAN Bus

The simplest detection of abnormal messages in vehicles includes two modules: analysis module and monitoring module. The analysis module is mainly used to process CAN messages in vehicles. The monitoring module is mainly used to extract the characteristics of incoming CAN messages and determine the type of CAN messages according to the training characteristics of existing messages. The analysis module is a feature database, which contains all trained features of incoming messages. Once the monitoring module recognizes a new type, of the abnormal message, the analysis module records and updates the feature database. The general process is shown in Figure 4.

3.1.2 Detection Types of Abnormal Messages for the CAN Bus

Researchers classify anomaly detection into signature-based detection and anomaly-based detection(Loukas et al., 2019)(Singh and Nene, 2013)(Omar, Ngadi and H. Jebur, 2013)(Patcha and Park, 2007). For abnormal messages, the former extracts features from the data set of abnormal messages, generates an intrusion detection system model of abnormal messages and matches a signature for each type of messages. However, It cannot detect if an unknown or a new message is abnormal. It is a blacklist method. The latter uses statistical analysis to learn the basic features of each type of the message to generate the intrusion detection system model of anomalous messages. It focuses on judging anomalous behaviour by comparing with the characteristics of normal behaviour. It is a whitelist method. Due to the continuous development of the Internet of vehicles technology, attacks based on the CAN bus are endless and unpredictable. Therefore, it is difficult to extract the characteristics of all attack events for signature-based detection techniques. In contrast, the anomaly-based detection method can learn the basic characteristics of each event, so it is more suitable for the detection of abnormal messages than the signature-based detection.

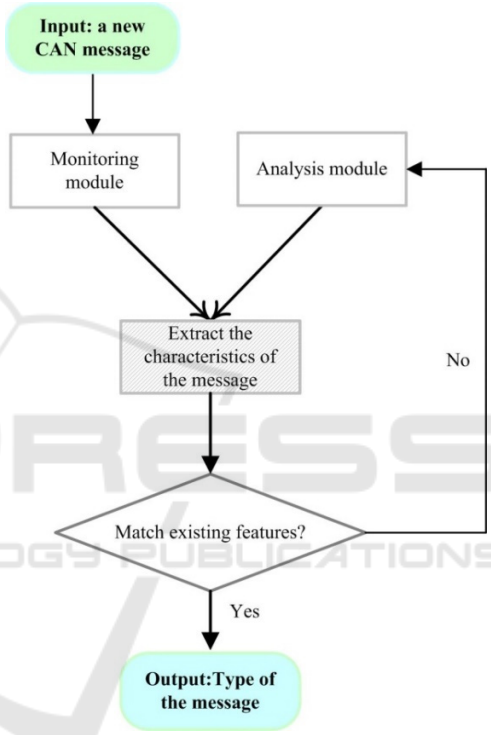

Figure 4. Process of Abnormal Messages Detection for the CAN Bus.

3.2 Existing Methods for Detecting Abnormal Messages in Vehicles

The existing detection methods can be approximately divided into period-based detection and machine learning-based detection. In the category of machine learning, we focus on the detection methods of abnormal messages based on deep learning, which are commonly used at present.

3.2.1 Period-based Anomaly Detection Method

Song et al(2016) is based on the fact that most ECUs in vehicles send CAN messages at a fixed frequency

defined by the manufacturer. A lightweight detection method for abnormal messages in vehicles is proposed. They mainly analyse the arrival time interval of messages on the CAN bus. When new messages appear on the bus, they check the ID value of the message and the arrival time of the last message with the same ID value, and identify anomalies by judging whether the messages appear within the specified time interval. However, there will be high false rate if the threshold for calculating the anomaly is incorrect.

Otsuka and Ishigooka (2018) designed a detection method for delayed decision periods. The proposed system only generates error alarms if it received multiple messages with the same ID field within a maximum period, which can reduce the false alarm rate. However, if the attacker injects a CAN message with the original frequency through the compromised ECU, the system cannot detect the abnormal messages.

Cho and Shin(2016) proposed a clock-based intrusion detection system (CIDS). They modeled the clock characteristics of each ECU and analysed the actual attacker according to the clock fingerprint characteristics. The experimental results show that the false positive rate of CIDS for detecting various types of in-vehicle attacks is 0.055%. One disadvantage of the method is that it is based on clock offset, which is a physical quantity that may change with the change of external environment changes, this method cannot detect the real attacker well.

3.2.2 Deep Learning-based Anomaly Detection Method

Deep learning establishes the ability to recognize and distinguish things by simulating human thinking. It mainly uses various intelligent algorithms to find corresponding features from a large amount of data, and then uses the learned features in the classification and prediction of new samples. Deep learning has a multi-level learning structure, which can perform multiple abstract transformations on the input features and has powerful feature expression capabilities. At present, a large number of deep learning algorithms have been applied to the detection of vehicle's abnormal messages.

In(Kang and Kang, 2016), the authors propose a detection mechanism related to abnormal messages of vehicles based on deep neural network (DNN) model. They first extract feature vectors from CAN messages, then train model parameters using deep belief network (DBN) and traditional stochastic gradient descent method. Finally, anomaly detection is completed by learning the characteristics of data fields (64 bits) in CAN messages. The results show that the detection accuracy based on DNN algorithm is better than that of traditional machine learning method. Kim et al (2017) designed a system that classifies vehicle attacks using DNN model. They used datasets from different attack models as test sets to test the system. The results show that the system classify the normal and attacked scenario correctly.

Nair et al(1993) established an warning mechanism to detect abnormal messages. In the method, CAN messages is used to generate the transfer probability and emission probability when the vehicle is attacked. They matched the probability with the hidden Markov model and analysed the data related to the time series to generate the test model. An alert is issued when the monitored message is abnormal. For example, when the speed of the vehicle changes abnormally, the system will detect an obvious abnormality in the messages related to the speed, then the system will generate an alarm. However, an obvious defect of the method is that the system will not generate an alarm when multiple physical quantities in the vehicle change simultaneously.

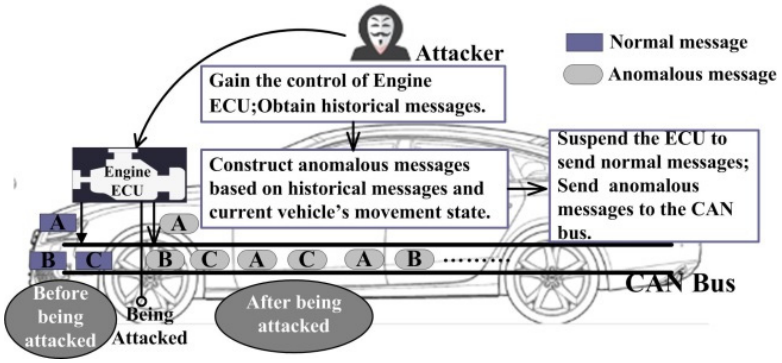

Figure 5. The flow of attacks in the vehicle.

Taylor et al(2016) proposed an anomaly detector based on recurrent neural network (RNN) to detect such as temperature. When the external environment abnormal messages on the CAN bus. The detector works by learning and predicting the data fields in the next message sent by each ECU on the bus. However, the method can only detect the abnormal messages with the specific ID field.

Rieke et al (2017) proposed to detect abnormal behaviour in CAN message sequence together with the implementation of event stream processing in vehicles. They filtered out normal events and analysed a small number of abnormal behaviour caused by abnormal messages. These abnormal behaviours may cause damage to components in the actuator of the vehicle. The disadvantage of this method is that the model takes a long time to train when the number of events in the vehicle is large. In addition, because the data field of the CAN message is not taken into account, the method cannot detect a circumstance that the data of CAN message is incorrect but the sequence is correct.

In (Narayanan, Mittal and Joshi, 2017), a multi-layer context extraction mechanism is proposed. The authors design a rule-based detection method to obtain context information related to vehicles. They collect messages on the CAN bus to generate SWRL semantic rules and use rules to build context-related vehicle information. However, similar to a specification-based approach of machine learning is not used to construct the SWRL URLs.

4 PROPOSED MODEL

We focus on the process of destroying vehicles by attackers, the method of obtaining normal datasets in engine ECU of real vehicles, and the way of constructing abnormal datasets by attackers of different levels.

4.1 System Model

Generally speaking, messages sent by ECU contain information of multiple sensors. In this paper, the experimental vehicle is the 2017 Toyota Vois, we mainly analyse the engine ECU's messages which contain vital information such engine speed closely related to vehicle's movement state.

In the model, the main attack is to modify the content of the message sent by the engine ECU. Different levels of attackers modify the content of the message in different degrees. The attack process is as follows: Firstly, the attacker gains the control of engine ECU in some way and obtains historical messages sent by engine ECU. Secondly, he specifically modifies the data field of the engine ECU's messages according to the current vehicle's movement state and historical messages to generate a series of abnormal ECU's messages. Finally, he disables the current engine ECU and injects abnormal messages into the CAN bus through the comprised ECU to launch the attack, forcing the movement state of the vehicle to change. The attack process is shown in Figure 5.

Due to the high similarity between the two types of messages. we can't detect anomalies by a linear model. Therefore, we use the deep neural network model to quickly and accurately detect abnormal messages in different levels to ensure driving safety.

4.2 Attack Model

At present, there are about two ways for an attacker to enter the network of the vehicle to gain control of the ECU. One is to inject malicious messages into the CAN bus with some special commands through the OBD interface inserted into the vehicle by external devices such as laptop; another attacker uses the remote control unit or entertainment system to interact with the external network. They hijack messages that communicate on the bus and forge them to get the control of the ECU. Compared with the first attack mode, the second is more flexible and efficient. In this paper, the attackers mainly use the second attack mode.

In the attack model, attackers use the comprised ECU to get messages of the engine ECU in the stationary vehicle, injecting messages into the moving state of the vehicle through the comprised engine ECU to force the vehicle to receive wrong commands, thus forming an attack on the vehicle. Here, we have three different levels of attackers: weak attackers, medium attackers and strong attackers.

4.3 Data Set

Data sets are mainly divided into normal and abnormal data sets, the normal data sets refer to the data collected during normal movement state of the vehicle; the abnormal data sets are mainly constructed by attackers of different levels who modify normal messages according to their abilities.

For normal data sets, we collect normal messages from engine ECU of the vehicle in stationary and driving state. We connect the USBCAN-I Pro tool to the OBD interface of the vehicle and use ECANTools universal test software to receive messages of the CAN bus. During the experiment, we first start the vehicle and keep it in a stationary state, collecting data for 30 minutes. Then we drive the vehicle at a constant speed of 30 km/h for 30 minutes to collect the data in a moving state. After filtering, messages associated with engine ECU in different states is obtained. The ID and Data fields of the normal messages are shown in Table 1. Because there are many kinds of data fields in messages, we only intercept part of the content of the data fields here.

Through the analysis and test of two kinds of data in different states, we find that the message ID related to engine ECU is 0x2C1. The first and second bytes in message data field represent engine intake pressure sensor (red part in the table), the 3rd and 4th bytes represent engine intake temperature sensor (green part in the table), the engine speed sensor is represented by the fifth and seventh bytes (blue part of the table).

As for the anomalous data set, according to the description in the attack model, the capabilities of the three different levels of attackers are as follows: Weak attackers do not know the actual meaning represented by each byte in the data field of the engine ECU's message, they only modify a single byte of the data field in the message at random, which has less impact on the vehicle. Medium attackers and strong attackers know the physical meaning of each byte in the data field of the engine ECU's message. They launch the attack by replacing the byte of the corresponding data field of messages in the moving state with the partial byte of the data field in the message of the engine ECU in the stationary state. It may cause serious malfunction of vehicles. Specifically, the medium attacker can only modify the bytes of data fields related to a single sensor in a message, while a strong attacker can modify the bytes of data fields related to multiple sensors in a message. Abnormal message related to the engine ECU of the vehicle are shown in Table 2, the modifications are marked by horizontal lines.

5 OUR METHOD

5.1 The Network Structure

Deep Neural Network (DNN) is an artificial neural network (ANN) with two or more hidden layers. It has good effect in solving classification problems with high data dimension and large amount of data. At the same time, DNN has been widely used in computer vision, image processing, speech recognition and other fields because of its remarkable classification performance(Krizhevsky, Sutskever and Hinton, 2012)(Hinton et al., 2012).

Table 1. Normal Messages related to the engine ECU of the vehicle.

ID	Vehicle's movement state	Data
0x2C1	Stationary	8 1 91 0 9A CC 0 CB
		8 2 4 1 50 D 0 F9
		8 5 4B 0 8A EA 0 97
	Driving	0 1 9F 6 6C BF B A7
		8 1 0 0 AF C8 0 4B
		0 1 83 FF 30 CC 6 50
		0 1 22 4 38 C9 6 F9
		8 1 4 0 2 C8 0 A2

Table 2. Abnormal Messages related to the engine ECU of the vehicle.

ID	Source of Messages	Data
0x2C1	Weak attacker	**9** 1 0 0 AF C8 0 4B
		0 1 83 FF 30 CC 6 **CB**
		0 1 22 4 38 C9 **7** F9
	Medium attacker	8 **2 4** 0 AF C8 0 4B
		0 1 83 **1 50** CC 6 50
		0 1 22 4 38 **EA** 6 **97**
	Strong attacker	8 **5 4B** 0 AF **CC** 0 **CB**
		0 **2 4** **0 8A** CC 6 50
		0 1 22 **0 9A** **CC** 6 **CB**

DNN provides an effective way to simulate the non-linear relationship between input and output. Therefore, we input the non-linear features in the CAN message data field into the DNN model to deal with the problem of abnormal CAN messages detection in vehicles. The network model used in this paper is shown in Figure 6. Each node in the middle hidden layer calculates the output with the activation function of the input value. In the network model, the input is a set of K samples $\{(d^{(1)}, y^{(1)}), \ldots\ldots, (d^{(K)}, y^{(K)})\}$ where $d = \langle d_0, d_1, d_2, \ldots\ldots d_{63} \rangle$, which contains 64-bit data field feature vectors in CAN bus message, y is the type of the corresponding message.

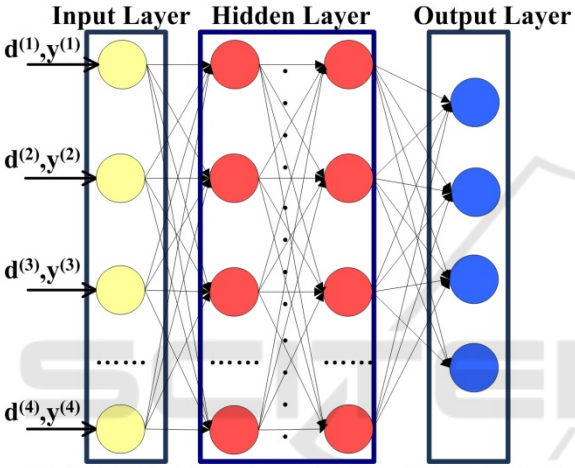

Figure 6. The Deep neural network model.

5.1.1 Activation Function

The activation function of the model is ReLU function, which is a kind of non-linear activation function. It can alleviate the problem of gradient disappearance to a certain extent. At the same time, it can directly train the deep neural network by supervised learning without relying on unsupervised layer-by-layer pre-training, thus speeding up the training speed. In addition, the derivation of function is very simple. The derivative form of the function is very simple. Compared with other linear activation functions, it can represent the complex classification boundary. Functional formulas and images are shown in (1) and Figure 7 respectively.

$$f(x) = \max(0, x) \tag{1}$$

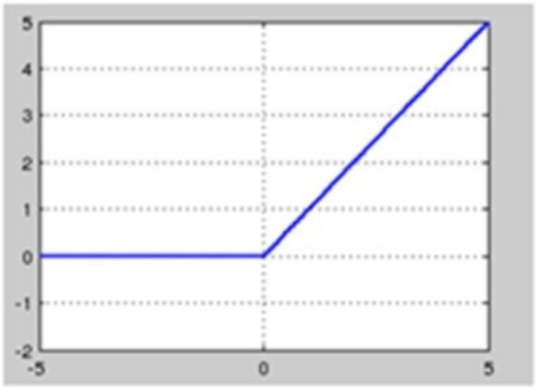

Figure 7. Rule function.

5.1.2 Loss Function

The loss function of the model is a cross entropy loss function, which is often used in the classification problem of unbalanced distribution of positive and negative samples. It is mainly used to determine the closeness of the actual output to the expected output and characterize the distance between the actual output (probability) and the expected output (probability). Generally speaking, the smaller the value of cross-entropy, the two probability distributions. In general, the smaller the cross entropy, the closer the two probability distributions are. The cross entropy loss function is:

$$L = -\frac{1}{n}\sum_x [y \ln a + (1-y)\ln(1-a)] \tag{2}$$

$$a = \sigma(z) = \sigma(\sum w_j x_j + b) \tag{3}$$

Where y is the expected output, a is the actual output of the neural network, x is the sample, and n is the total number of samples. The derivative of the weight and the bias vector is:

$$\begin{aligned} \frac{\partial L}{\partial W_j} &= -\frac{1}{n}\sum_x \left(\frac{y}{\sigma(z)} - \frac{1-y}{1-\sigma(z)}\right)\frac{\partial \sigma}{\partial W_j} \\ &= -\frac{1}{n}\sum_x \left(\frac{y}{\sigma(z)} - \frac{1-y}{1-\sigma(z)}\right)\sigma'(z)x_j \\ &= \frac{1}{n}\sum_x \frac{\sigma'(z)x_j}{\sigma(z)(1-\sigma(z))}(\sigma(z)-y) \\ &= \frac{1}{n}\sum_x x_j(\sigma(z)-y) \end{aligned} \tag{4}$$

$$\frac{\partial L}{\partial b} = \frac{1}{n}\sum_{x}(\sigma(z) - y) \quad (5)$$

It can be seen from the above formula that the update speed of the weight is completely affected by $\sigma(z) - y$, that is, the error. If the error is large, the update speed is fast; otherwise, the speed is slow. Using the cross entropy loss function can overcome the problem that the weight of the model is updated too slowly.

5.1.3 Classifier

The last layer of the model is the softmax classifier. The classifier is a multi-class regression model developed on the basis of logistic regression, which is suitable for classification problems where the value of a class label is greater than 2. When an input is given, each output gets a value between 0 and 1, which represents the probability that the input belongs to the category, finally, the category corresponding to the input is acquired according to the maximum probability principle.

5.2 Experiment Setup

The experiment was performed on an Ubuntu system equipped with a 3.4GHz Intel CPU, using kears in the high-level neural network database and the open source TensorFlow framework as the back-end, which involved the feature extraction, data processing and training, classification.

5.2.1 Model Parameter

In the experiment, we used a total of 25,000 messages to train network model, Including 20,000 normal messages and 5,000 abnormal messages from different levels of attackers. We divide messages into the training set, the verification set, the test set with a ratio of 6:3:1, so as to avoid over-fitting in the training process. We used the method of training and testing at the same time to classify the data in each set. Finally, all the data is used as a test set to test the overall performance of the trained model.

The input layer of the model consists of 64 neurons, corresponding to the 64-bit features of the data fields in the CAN message; the hidden layer has 3 layers, each layer has 64 neurons; the output layer contains 4 neurons with softmax function. The type of messages corresponding to the output label are as follows: 0- normal, 1-- weak attacker, 2-- medium attacker, 3--strong attacker. In each training of the model, the batch size is defined as 128, the learning rate is 0.001.

5.2.2 Experimental Result

We measure the performance of the model by using classification accuracy rate, which is the percentage of the number of messages correctly classified as a percentage of the total number of messages. We specifically analysed the classification accuracy rate of each type of messages and the classification accuracy rate of all messages of the model. The experimental results are shown in Figure 8. In order to enhance the reliability of the experimental results, we use the k-Nearest Neighbour (KNN) as baseline to classify the same datasets. The experimental results are shown in Figure 9.

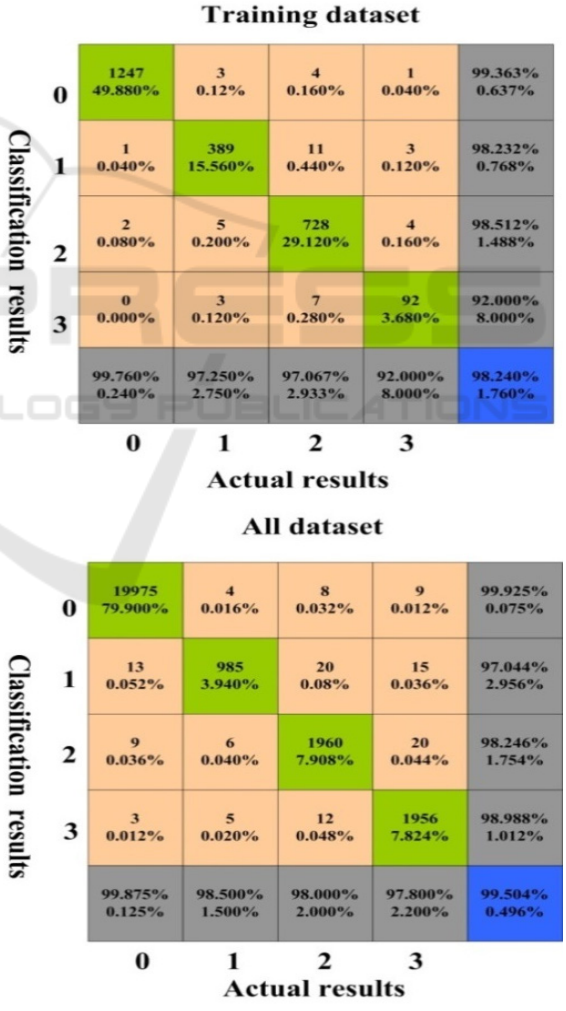

Figure 8. Classification accuracy rate under the DNN.

It can be seen from the confusion matrix that when a small part of the data is used as the test set,

the classification accuracy rates of the two models are: 98.240%, 87.160%; when all the data are used as the test set, the accuracy rates are: 99.504%, 93.660%. Therefore, no matter part of the data or all of the data are used as the test set, the classification accuracy based on the deep neural network model is much higher than the traditional machine learning algorithm.

5.3 Analysis of Results

5.3.1 Impact of Different Levels of Attackers on the Performance of Models

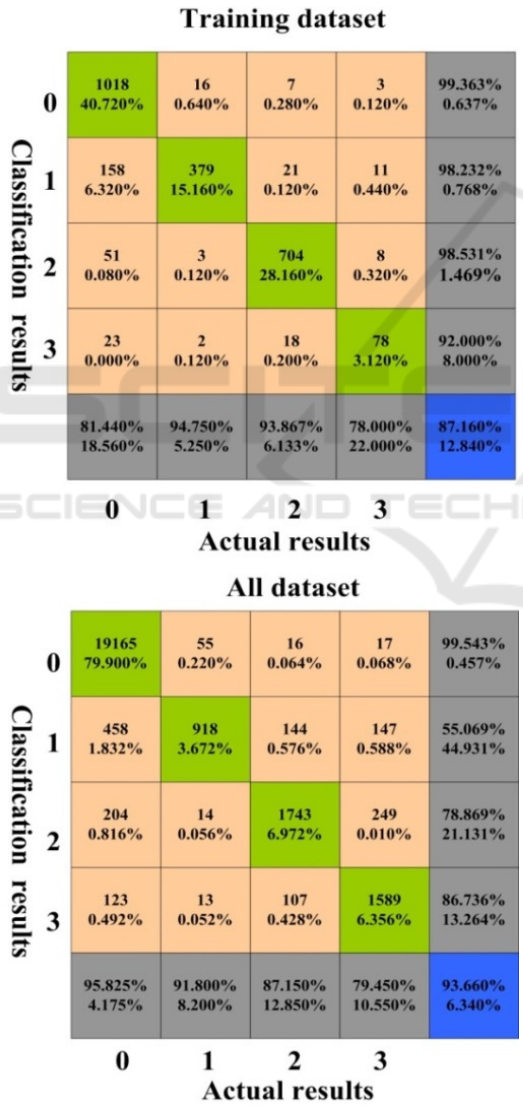

Figure 9. Classification accuracy rate under the KNN.

According to the confusion matrix, the accuracy rate of the recognition of messages is different, that is, the classification accuracy rate of the model changes with the level of the attacker. For the abnormal messages sent by different levels of attackers, the overall classification accuracy rate of the two models is shown in Table 3.

Table 3. Classification accuracy of two network models for different types of messages.

Type of messages	Classification accuracy rate of the model	
	DNN	KNN
Normal	99.875%	96.825%
Weak attacker	98.500%	91.800%
Medium attacker	98.000%	87.150%
Strong attacker	97.800%	79.450%

As can be seen from the table, the recognition accuracy rate of both models declines with the enhancement of the attacker's level. This is because weak attackers may only modify constant bytes that are independent of sensor parameters. It is relatively easy for the model to recognize messages whose features vary greatly; Medium attackers can modify the bytes associated with a certain sensor's parameter. The modified message is more similar to the original message, models are not easy to recognize such messages; Strong attackers can modify more bytes than medium attackers. The characteristics of modified messages are more similar to normal messages, these messages are difficult to recognize by models.

5.3.2 Impact of the Number of Hidden Layers on the Performance of the Model

We try to modify the number of hidden layers of the neural network model to find a network model with the best performance. Figure 10 describes the classification accuracy rate of each type of messages and all messages under different hidden layers.

It can be seen from the line graph that the classification accuracy rate of the model is high when the hidden layer is 4, regardless of the certain type of message or all messages. At this time, the performance of the deep neural network model is optimal.

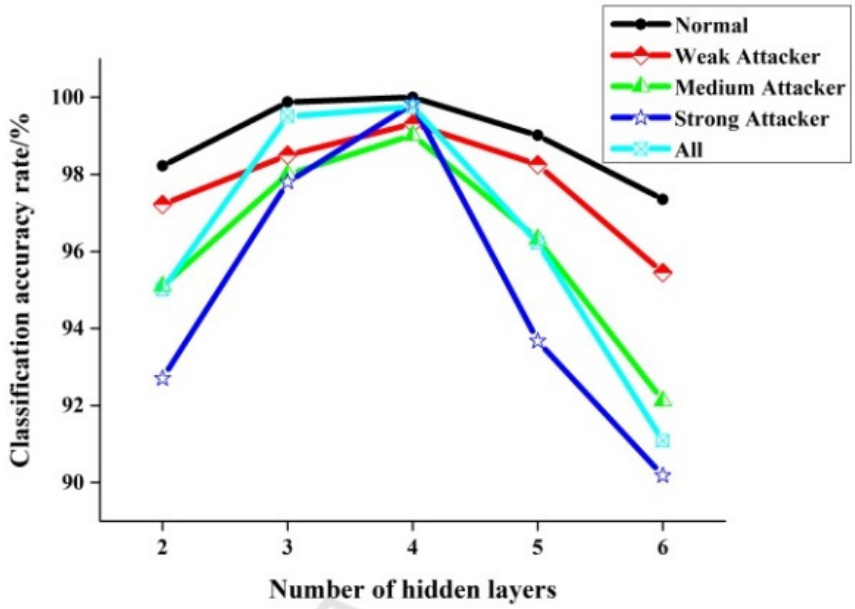

Figure 10. Classification accuracy rate of the model under different hidden layers.

6 CONCLUSIONS

In this paper, a new attack model for the CAN bus of the vehicle is proposed and the attack is accurately detected by deep neural network (DNN). According to the existing methods, the ability of attackers is quantified and different anomaly data sets are constructed. Finally, Normal messages of engine ECU and abnormal messages from different attackers verify that the deep neural network model has high classification accuracy for each type of messages. Which proves that deep neural network has a good effect in the detection of abnormal messages on the CAN bus.

REFERENCES

Biswas, S., Tatchikou, R. and Dion, F. (2006) 'Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety', *IEEE Communications Magazine*, 44(1), pp. 74–82. doi: 10.1109/MCOM.2006.1580935.

Checkoway, S. *et al.* (2011) '<Cars-Usenixsec2011.Pdf>'. doi: 10.1109/TITS.2014.2342271.

Cho, K. and Shin, K. G. (2016) 'Fingerprinting Electronic Control Units for Vehicle Intrusion Detection This paper is included in the Proceedings of the Fingerprinting Electronic Control Units for Vehicle Intrusion Detection'.

Gmbh, R. B. (1991) 'CAN Specification'.

Hinton, G. *et al.* (2012) 'Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups', *IEEE Signal Processing Magazine*. IEEE, 29(6), pp. 82–97. doi: 10.1109/MSP.2012.2205597.

Hoppe, T., Kiltz, S. and Dittmann, J. (2011) 'Security threats to automotive CAN networksPractical examples and selected short-term countermeasures', *Reliability Engineering and System Safety*, 96(1), pp. 11–25. doi: 10.1016/j.ress.2010.06.026.

Kang, M. J. and Kang, J. W. (2016) 'Intrusion detection system using deep neural network for in-vehicle network security', *PLoS ONE*, 11(6), pp. 1–17. doi: 10.1371/journal.pone.0155781.

Kim, J. *et al.* (2017) 'Method of intrusion detection using deep neural network', *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*. IEEE, pp. 313–316. doi: 10.1109/BIGCOMP.2017.7881684.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', in *ImageNet Classification with Deep Convolutional Neural Networks*.

Larson, U. E., Nilsson, D. K. and Jonsson, E. (2008) 'An approach to specification-based attack detection for in-vehicle networks', *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 220–225. doi: 10.1109/IVS.2008.4621263.

Loukas, G. *et al.* (2019) 'A taxonomy and survey of cyber-physical intrusion detection approaches for vehicles', *Ad Hoc Networks*, 84, pp. 124–147. doi: 10.1016/j.adhoc.2018.10.002.

Mastakar, G. (2012) 'Experimental Security Analysis of a Modern Automobile', pp. 1–16. Available at:

http://users.cis.fiu.edu/~carbunar/teaching/cis5374/slides/autosec.g.mastakar.pptx.

Miller, C. and Valasek, C. (2013) 'Adventures in Automotive Networks and Control Units', *Def Con 21*, p.99.

Müter, M., Groll, A. and Freiling, F. C. (2010) 'A structured approach to anomaly detection for in-vehicle networks', *2010 6th International Conference on Information Assurance and Security, IAS 2010*, pp. 92–98. doi: 10.1109/ISIAS.2010.5604050.

Narayanan, S. N., Mittal, S. and Joshi, A. (1993) 'OBD SecureAlert : An Anomaly Detection System for Vehicles'.

Narayanan, S. N., Mittal, S. and Joshi, A. (2017) 'Using semantic technologies to mine vehicular context for security', *37th IEEE Sarnoff Symposium, Sarnoff 2016*. IEEE,pp.124–129.doi: 10.1109/SARNOF.2016.7846740.

Omar, S., Ngadi, A. and H. Jebur, H. (2013) 'Machine Learning Techniques for Anomaly Detection: An Overview', *International Journal of Computer Applications*, 79(2), pp. 33–41. doi: 10.5120/13715-1478.

Otsuka, S. and Ishigooka, T. (2018) 'CAN Security : Cost-Effective Intrusion Detection for Real-Time Control Systems Overview of In-Vehicle Networks'. doi: 10.4271/2014-01-0340.Copyright.

Patcha, A. and Park, J. M. (2007) 'An overview of anomaly detection techniques: Existing solutions and latest technological trends', *Computer Networks*, 51(12),pp.3448–3470.doi: 10.1016/j.comnet.2007.02.001.

Rieke, R. *et al.* (2017) 'Behavior Analysis for Safety and Security in Automotive Systems', *Proceedings - 2017 25th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2017*. IEEE, pp. 381–385. doi: 10.1109/PDP.2017.67.

Singh, J. and Nene, M. J. (2013) 'A Survey On Machine Learning Techniques For Intrusion Detection Systems', *International Journal of Advanced Research in Computer and Communication Engineering*, 2(11), pp. 4349–4355.

Song, H. M., Kim, H. R. and Kim, H. K. (2016) 'Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network', *International Conference on Information Networking*, 2016–March,pp.63–68.doi: 10.1109/ICOIN.2016.7427089.

Taylor, A., Leblanc, S. and Japkowicz, N. (2016) 'Anomaly detection in automobile control network data with long short-term memory networks', *Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016*, pp. 130–139. doi: 10.1109/DSAA.2016.20.

Valasek, C. and Miller, C. (2015) 'Remote Exploitation of an Unaltered Passenger Vehicle', *IO Active*, 2015, pp. 1–91. doi: 10.1088/2041-8205/762/2/L23.