A Method for Detecting Abnormality of CAN Bus in Vehicle

PENG Jing[1], ZHANG Zhihong[2], HE Hong[1]

(1. *School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin Key Laboratory of Complex System Control Theory and Application, Tianjin* 300384, *China.*

2. *Transmission and Transmission Department, Tianjin Radio and Television Station, Tianjin* 300072)

Abstract: With the development of intelligent and networking technology in automobile, the malicious attacks against in-vehicle CAN networks are increasing day by day, and the problem of information safety in automobile is aggravated. In this regard, this paper analyzes the security loopholes and threats which the CAN bus faced, put forward a kind of anomaly detection algorithm for vehicle CAN bus. The method uses support vector machine algorithm to distinguish between normal message and abnormal message, so as to realize the CAN bus anomaly detection. Theoretical and experimental studies show that this method can effectively detect abnormal packets in the CAN bus with a detection rate of over 90%, which can effectively resist malicious attacks such as tampering and cheating on the vehicle CAN bus.

Keywords: Automobile; CAN bus; Information Security; Support Vector Machines; Abnormal Detection

1 Introduction

With the rapid development of automotive electronics, network and information technology, car networking is providing a new path for technological progress in the automotive industry and the intelligent transportation industry. However, the development of networked cars exposes vulnerabilities and provides opportunities for attackers. In 2013 American hackers Miller and Valasek controlled the functions of the steering wheel, throttle and brake of the Ford Maverick and the Toyota Prius through the OBD interface[1]. In 2015, the two hackers took over Jeep Cherokee again, and the car company therefore recalled 1.4 million cars[2]. Tencent Cohen Labs and 360 Company cracked Tesla in 2016 to identify the remote control of Tesla. These cases make car information security issues fully concerned, in November 7, 2016, China released the "Cyber Security Law of the People´s Republic of China", that clearly required depot and car network operators to take the necessary measures to safeguard vehicle network security.

Domestic and foreign scholars have carried on a series of researches on the information security of CAN bus in the car. Woo et al.[3] proposed a lightweight message encryption method using AES-32 algorithm to encrypt CAN messages, but this method has a certain impact on message transmission rate. Yuhe and others from Jilin University proposed to detect the CAN bus anomaly by using information entropy method. This method can detect the abnormal number of a specific message, but can not accurately detect the anomaly of CAN message data bit. Wu Shangshi, Jilin University, put forward a CAN bus dynamic password authentication method, which can improve the authenticity of CAN communication nodes and message integrity, but increased the delay time in the car initialization phase. At this stage, encryption and authentication mechanism is difficult to achieve. This paper presents a mechanism for detecting anomaly on a car gateway. The detection mechanism uses the support vector machine algorithm to detect the data field of a CAN bus message in the vehicle, which can effectively detect any anomaly. This method is conducive to users in a timely manner to resist external malicious attacks.

2 Automobile CAN-Bus

2.1 Automotive CAN bus topology

At present, the interior of an automobile gener-

ally includes tens or even hundreds of ECUs. As shown in Figure 1, different performances of ECUs are connected to different CAN buses and also interconnected to form an in-vehicle network through a gateway. Each ECU is connected to high-speed CAN and low-speed CAN line due to different performance and real-time. Low-speed CAN bus baud rate is generally 250kbps, usually mounted with dashboard, air conditioning, doors, windows, lights and other control unit; high-speed CAN bus baud rate of 500kbps, usually mounted with the engine, airbags and other control unit. The central gateway is the core of the CAN network, which is responsible for the format and rate conversion of different network data and also as the last level of external data entering the automotive interior CAN network.

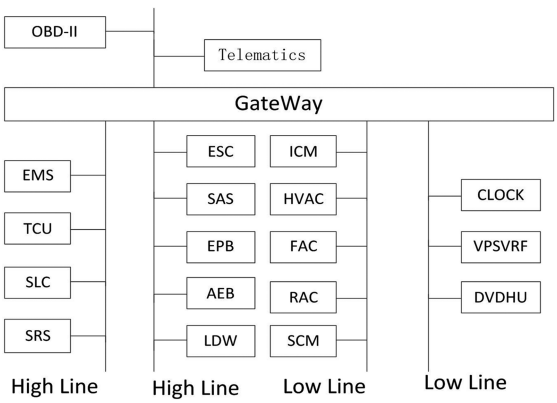


Fig. 1 CAN network topology

2.2 Automotive CAN bus communication mechanism

2.2.1 Electrical characteristics

The physical medium of CAN bus in the car is twisted-pair, the two lines are represented as CAN_H and CAN_L respectively. The signal in the twisted pair is transmitted as a differential voltage:

$$V = V_H - V_L \quad (1)$$

In static state, CAN_H and CAN_L are 2.5V or so, at this moment V = 0V, represents the logical state is "1", it is the recessive state.

In dynamic, CAN_H is about 3.5V, CAN_L is about 1.5V, then V = 2V, this state represents the logic "0", which is dominant. The CAN message consists of such "0" and "1".

2.2.2 CAN message structure

Inside the car, the CAN bus transmits messages in the form of broadcast. As shown in Figure 2, CAN messages have two formats, a standard frame format and an extended frame format. The basic structure of the two frames is the same, including the arbitration field, the control field, the data field, the CRC field and the frame end field. Only the length of the arbitration field is different. The length of the standard frame arbitration field is 11 bits and the length of the extended frame arbitration field is 29 bits. The extended frame and the standard frame each have 0-8 data fields and contain the specific instruction information of the message. The anomaly detection mechanism in this paper mainly detects the data field of CAN message.

2.2.3 Arbitration mechanism

Car CAN bus uses CSMA / CD (Carrier Sense / Collision Avoidance) arbitration mechanism. When multiple nodes on the bus send packets at the same time, the packets with the highest priority are sent first. Each CAN message has a unique ID, "0" has a higher priority than "1", so the smaller the message ID, the higher the priority.

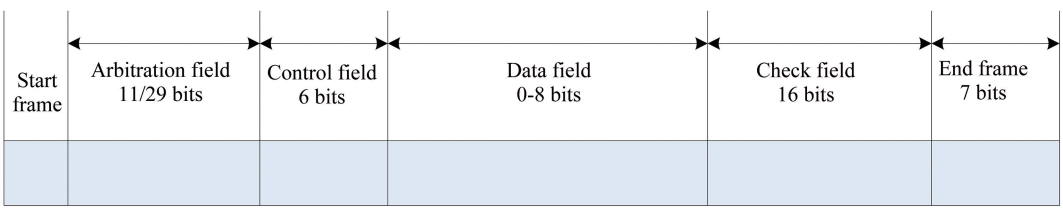


Fig. 2 CAN message structure

The ECU node on the vehicle CAN bus receives the message of the CAN bus in real time. When it is

identified that the message ID matches its own ID, the node receives the message and performs corresponding actions. When the IDs do not match, the node discards the message.

2.3 CAN bus security vulnerability

At the beginning of the design of the CAN bus, the car was a relatively closed system, thus, CAN bus information security issues were not considered, it made the car CAN bus now exposed serious security vulnerabilities.

(1) the broadcast characteristics of the message

Since the CAN bus message is broadcasted, any node on the CAN network can receive the message when the message is sent. This allows hackers to use malicious techniques to simulate the ECU node to monitor the CAN network information, and reverse analysis of information or send forged data to achieve the malicious control of other nodes on the CAN bus.

(2) Information is not secure

CAN message on the bus, without any digital signature or message authentication code protection, resulting in the inability to guarantee the integrity, availability, authenticity and non-repudiation of the information.

(3) Vulnerable denial of service

Malicious attackers use the arbitration mechanism of the CAN bus to continuously send packets with high priority arbitration IDs to the bus so that other ECUs can not use the bus, causing the ECU to fail.

(4) CAN bus interface without safety protection

The interface of OBD clearly stipulates that the pins of CAN network in the car can be directly connected to the car network via the CAN pin. The attacker can replay, flood and tamper with the car CAN network.

3 Abnormal detection method

The 8 data bits of the CAN bus message represent different meanings and carry important control information. Hackers can modify the important data bits of the ID to maliciously control certain functions of the car, so CAN message data bit anomaly detection is very important. CAN message data anomaly detection can be seen as a two-category machine learning problems, divided into normal data packets and abnormal data packets. In this paper, an algorithm of support vector machine (SVM) is used as an anomaly detection of in-vehicle CAN message data.

3.1 Support Vector Machines

Support Vector Machine (SVM) is a kind of machine learning method based on statistical learning theory, which can get a smaller error classifier through limited training samples. Suppose the sample is (Xi, Yi), i = 1,2, ... n, Xi is the input sample, Yi is the classification identification number and Yi (-1, +1) is defined. The classification surface equation can be expressed as:

$$wx_i + b = 0 \tag{2}$$

Where Xi is the input vector, w is the weight vector, its value can be adjusted, and b is the offset. In order to make the separation larger between training samples, the problem is transformed into the quadratic programming problem:

$$\min \frac{1}{2} \| w^2 \| \tag{3}$$

$$y_i(wx_i + b) \geq 1, i = 1,2,\ldots n \tag{4}$$

Then establishing Lagrange function to solve the optimal:

$$J(w,\beta) = f(w) + \sum_{i=1}^{m} \beta_i h_i(w) \tag{5}$$

Where βi Lagrange multiplier, seeking partial derivative equation:

$$\begin{cases} \dfrac{\partial J}{\partial w} = 0 \\ \dfrac{\partial J}{\partial \beta_i} = 0 \end{cases} \tag{6}$$

We can find the value of w ,β, minimize the objective function f (w), the constraints transform into:

$$\begin{aligned} hi(w) &= 0, i = 1,2,\ldots,m \\ gi(w) &\leq 0, i = 1,2,\ldots,k \end{aligned} \tag{7}$$

Defining the Lagrange Functions:

$$J(w,\beta) = f(w) + \sum_{i=1}^{m} \beta_i h_i(w) + \sum_{i=1}^{k} a_i g_i(w) \quad (8)$$

Finding the maximization objective function:

$$Q(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j d_i d_j x_i^T x_j \quad (9)$$

To get the decision-making surface:

$$\sum_{i=1}^{l} a_i^* d_i x_i^T x + b* = 0 \quad (10)$$

The a_i^* is the optimal solution to the problem.

Because CAN message data is not one-dimensional, but 1-8 multi-dimensional data, each dimension represents a feature of the data. The kernel function is used to process multi-feature data, and the data is mapped from n-dimensional space to a high-dimensional space by the inner product computing kernel function. Through the kernel function, the inseparable function in n-dimensional space becomes separable in the high-dimensional space. In this paper, RBF warp-based kernel function analysis:

$$K(x,x_i) = \exp\left\{ - \frac{|x - x_i|^2}{\sigma^2} \right\} \quad (11)$$

RBF classifier assigns a support vector to the center of each radial basis function. The weight of the function is automatically generated by the algorithm, and the value of σ determines the classification surface. The classification function of support vector machine is similar to a neural network, and the input vector is mapped to a high dimensional feature space through prior selection. The output function is a linear combination of intermediate nodes. The SVM architecture is as shown in the figure 3.

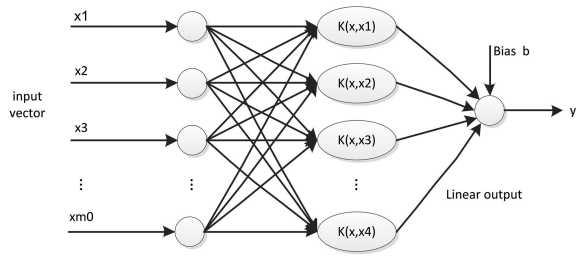


Fig. 3 SVM system structure

3.2 CAN bus detection with Support vector machine

The method of support vector machine is used to count the data characteristics of normal CAN message and to determine the functional relationship between intrusion behavior features and normal features. By analyzing the CAN bus data, it detects if there is any anomaly for data. The dichotomous problem analysis method is used to distinguish the normal message from the abnormal message in the vehicle bus. Every bit of CAN bus message data bit may carry key information, and an exception should be detected for each bit. The final decision function of SVM is decided by a few support vectors. Therefore, considering each bit of data bits as a feature, the message data bits are divided into 8 features, and if the data bits of the message are less than 8 bits, 00 is used instead.

The CAN bus message is collected. The maximum data length of the message is 8 bytes, and the message is divided into 8 features to detect the abnormality of each data bit of the different ID message. Figure 4 shows the support vector machine detection model. The processed normal messages and abnormal messages are input to the system for training, and an anomaly detection indicator is obtained for judging whether the on board bus is abnormal.

The data packets in the vehicle are classified according to ID, and the data bits of each type of ID packet are trained to find the optimal classification solution.

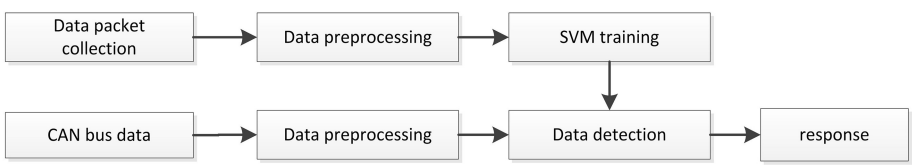


Fig. 4 Support vector machine detection model

Algorithm steps are as follows:

1) collect normal data samples and abnormal data samples, add classification identification number;

2) The training samples are input into the sup-

port vector machine system to get the best classification surface standard;

3) According to the threshold standard, the data in the bus is detected, the normal data and abnormal data are classified;

4) Repeat steps 1-3 for the results of the classification to update the classification threshold criterion to reach the optimal classification test.

4 Experimental simulation

In order to acquire the actual data of a CAN bus in a domestic car, the V-spy is connected to the OBD diagnostic interface in the car, and the collected data is used as an experimental sample. The process of collection is shown in Figure 5.

Fig. 5 data collection

As shown in Figure 6 (a), a total of 3680 packets with the ID 0x101 are collected. The sample data is large enough for testing experiments. Filter the packet data to get all the data that ID is 0x101, as shown in Figure 6 (b).

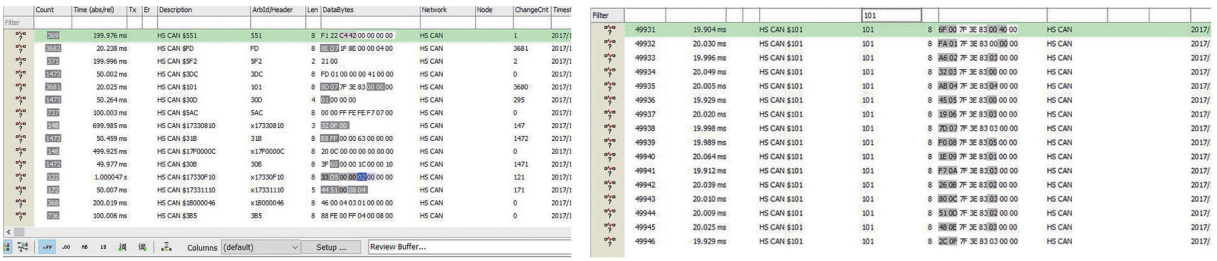


Fig. 6 CAN message

The data of ID 0x101 is summarized. Because of the security relationship, the data of the middle part is processed:

Table 1 normal message with ID 0x101

	ID	Data
1	0x101	6F 00……00 40 00
2	0x101	FA 01……00 00 00
3	0x101	A6 02……03 00 00
4	0x101	F0 08……05 00 00
5	0x101	1E 09……01 00 00
6	0x101	F7 0A……03 00 00
7	0x101	A7 02……00 40 00
…	…	…

Since the above data comes from the real data in the car, there is no abnormal data. Generate abnormal packets in a random and random manner:

Table 2 formal message with ID 0x101

	ID	Data
1	0x101	12 03……43 40 34
2	0x101	2A 01……60 80 40
3	0x101	A6 62……A3 70 89
4	0x101	60 78……05 45 55
5	0x101	1E 09……E1 80 00
6	0x101	47 67……03 34 05
7	0x101	A7 02……90 40 0A
…	…	…

Select the normal ID data 400 as a training sample, to be testing standards. Take 200 abnormal samples and 200 normal samples randomly mixed, as a test case input support vector machine detection system, the following test results:

Table 3 Abnormal detection results

The number of packets	Number of abnormal packets	Number of detected abnormal packets	Detection rate
400	200	189	94.50%

The 20 different ID data were tested, the detection rate were obtained, the summary results shown in Figure 7. It can be seen that the detection results of different ID packets are different, but the detection rates are all above 90%, indicating that the detection method has higher accuracy and can be used to detect abnormalities of data bits of CAN bus messages.

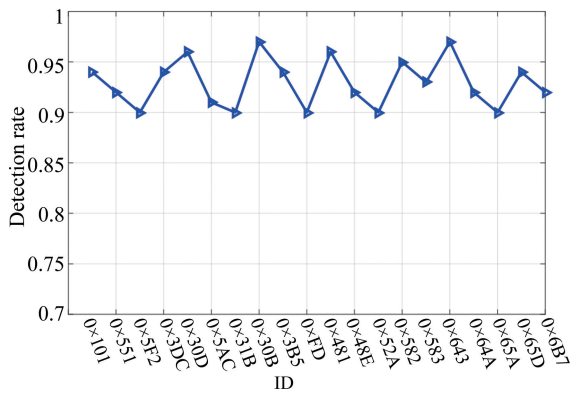


Fig. 7 Different ID messages test results

5 Conclusion

In this paper, according to the characteristic of CAN message, the data part of the message is classified by the feature, and a CAN bus anomaly detection model based on SVM is proposed. By collecting the normal message on the vehicle CAN bus and inputting the normal message and the abnormal message into the SVM detection system, the system learns the difference between the normal message and the abnormal message through the machine learning method to obtain the detection standard value. This value can be used to detect whether any abnormal data occurs in real time. Experiments show that this method can effectively detect malicious attacks such as data tampering on the CAN bus in the vehicle.

References

[1] Miller C and Valasek C(2013). *Adventures in automotive net-works and control units*. Las Vegas

[2] Craig Smith (2016). *The car hacker´s handbook: a guide for the penetration tester*. An Francisco, pp.63-84

[3] S. Woo, H. J. Jo, and D. H. Lee(2015), "*A Practical Wireless Attack on the Connected Car and Security Protocol for In-Vehicle CAN*," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 5, pp. 993 - 1006.

[4] Groza B, Murvay P S(2012). *Broadcast Authentication in a Low Speed Controller Area Network*. E-Business and Telecommunications. Springer Berlin Heidelberg, pp.330-344.

[5] Hoppe T, Kiltz S, Dittmann J(2008). *Security threats to automotive CAN networks—Practical examples and selected short-term countermeasures*. Reliability Engineering & System Safety, 96(1).pp.11-25.

[6] Schweppe H and Roudier Y(2012). *Security and privacy for in-vehicle networks*. Vehicular Communications, Sensing, and Computing (VCSC), 2012 IEEE 1st International Workshop on. IEEE, pp. 12-17.

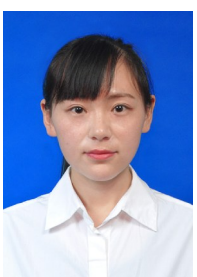
[7] Miller C and Valasek C(2014). *A survey of remote automotive attack surfaces*. Black Hat USA.

[8] Miller C and Valasek C(2015). *Remote exploitation of an unaltered passenger vehicle*. Black Hat USA.

[9] Wolf M, Weimerskirch A, Paar C(2004). *Security in automotive bus systems*. Proceedings of the Workshop on Embedded Security in Cars .

[10] Mundhenk P, Steinhorst S, Lukasiewycz M(2015). *Security analysis of automotive architectures using probabilistic model checking*.pp.1-6.

Authors' Biographies



Peng Jing, female, born in 1989, master, Her main research direction is intelligent network of automotive information security
E-mail: pjzy52817@126.com

He Hong, female, born in 1960, professor, M.Sc. Tutor. Her main research direction is detection technology and automation devices, electronic information processing and mobile robot navigation.
E-mail: heho604300@126.com