Anomaly Detection of Vehicle CAN Network Based on Message Content

Xiuliang Mo[1,2], Pengyuan Chen[1,2](✉), Jianing Wang[3], and Chundong Wang[1,2]

[1] Key Laboratory of Computer Vision and System, Ministry of Education,
Tianjin University of Technology, Tianjin 300384, China
cpy1001@foxmail.com

[2] Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology,
Ministry of Education, Tianjin University of Technology, Tianjin 300384, China

[3] Sichuan University, Chengdu 610207, Sichuan, China

Abstract. With the rapid advance of intelligent vehicles, auxiliary driving and automatic driving have been paid more attention to. While vehicle security has become increasingly prominent, which is seriously related to the property and personal safety. The attacker can send abnormal information to the controller through internal CAN bus. Because of the particularity of the vehicle CAN network information communication protocol, the encryption authentication technology cannot effectively solve the safety problem of the vehicle network. In the paper, a novel anomaly detection method based on CAN packet content is proposed. The scheme is effective in preventing in-vehicle ECU attacks caused by malicious modifications. Statistical thinking is adopted to analyze the characteristics of normal message content. Then a confidence interval based on normal features is defined for detecting abnormal network messages. Its detection performance has been demonstrated through experiments carried out on real CAN traffic gathered from an unmodified licensed vehicle.

Keywords: Anomaly detection · CAN network · CAN bus · Data frame · Mahalanobis distance

1 Introduction

Cars have become an indispensable tool in our lives. The car network is closely related to our lives. The CAN bus is currently the most widely used in-vehicle bus network technology. Because of its strong real-time communication, and short on-off cycle, the CAN bus has been widely utilized. However, with the increase in the

Our work is supported by the General Project of Tianjin Municipal Science and Technology Commission under Grant (No. 15JCYBJC15600), the Major Project of Tianjin Municipal Science and Technology Commission under Grant (No. 15ZXDSGX00030), and NSFC: The United Foundation of General Technology and Fundamental Research (No. U1536122).

J. Li et al. (Eds.): SPNCE 2019, LNICST 284, pp. 96–104, 2019.
https://doi.org/10.1007/978-3-030-21373-2_9

number of peripheral access interfaces of the networked cars, security problems have become more prominent. These interfaces can be utilized to access the CAN network, attack the CAN network, and forge the CAN message to allow the cars ECU to execute and consume ECU resources. These operations can directly send control instructions to the CAN bus of the car [6]. These hidden dangers are no longer as simple as stealing information and money, but actually threatening our personal safety.

The anomaly detection mechanism proposed in this paper is mainly for the anomaly detection of CAN bus data block. The CAN message primarily consists of ID block and data block. Only detecting CAN ID is unable to find whether the content of the message data block has an abnormality, the data field of CAN bus carries important control commands, sensor information and other key information to control the operation of the on-board system [7]. Thus, It is extremely significant to detect the anomaly of the data field of the CAN bus. The performance of the algorithm for the injection of malicious messages has been evaluated. Moreover, computational requirements of the algorithm are low enough to be compatible with the very limited hardware constraints of microcontrollers used to develop the ECUs (Electronic Control Units) embedded in modern vehicles.

2 Related Work

Kammerer et al. proposed the design of a star-coupled router as the central gateway for all sub-networks in the car to enhance the security of the CAN bus [3]. Each sub-network is connected to the router through the CAN interface system. The router has a routing configuration table. The router uses the information in the routing configuration table to detect and filter the CAN data frames. This method increases the cost and requires higher computing and storage capabilities of the router.

Groza et al. conducted various researches on the safety problems of the onboard CAN bus, proposed a series of light broadcast authentication protocols such as EPSB and Libra-CAN, and verified the availability of these protocols in the scenario where the number of ECU nodes is small [1]. However, in the real vehicle environment, involving multiple ECU nodes, the effectiveness of the algorithm is difficult to verify.

Studnia et al. proposed to design a state based anomaly detection system to determine whether the data frames transmitted on the bus meet the current state of the car (such as stationary, normal driving, and emergency procedures after collision) [4].

Many researchers tried to apply different neural network models to anomaly detection in CAN networks, such as recurrent neural network, deep neural network, and so on. Based on the trained neural network model, a predicted data packet can be output, and then the predicted data packet is compared with the real data packet, and if the error exceeds the acceptable range, an abnormal alarm is performed. Therefore, The sensor can identify malicious attacks on the vehicle [8–11]. However, due to the limited calculation and storage of the ECU, The proposed algorithm is constrained.

Narayanan et al. proposed a Hidden Markov Model to detect anomalous states. Using this model, while a vehicle is in operation, the system is able to detect and issue alerts [2]. Similarly, due to the limited calculation and storage of the ECU, The proposed algorithm is constrained.

The algorithm proposed in this paper mainly monitors the data frame of a CAN message. This design choice has several advantages over the prior art. First, We remark that proposed solution for improving CAN bus security complies with the hardware constraints of a typical automotive ECUs, having very low memory and computational requirements. Secondly, attacks against malicious tampering with CAN message data frame content can be detected in real time.

3 Preview

3.1 Controller Area Network

Automotive electronic components are connected in the car through the CAN network, and electronic components ECUs communicate via CAN messages. CAN network model is shown in Fig. 1. In CAN bus, ECU and ECU controller use data frame to transmit information, and broadcast messages to CAN network with specified ID. CAN data frame can carry 0–8 bytes of data, and the message can use 11-bit standard frame ID or 29-bit extension frame ID. The standard format of the data frame is shown in Fig. 2. It should be noted that the CRC field can only provide some form of protection against random modification of a CAN data frame. Since it is not built upon strong crypto and authentication primitives, attackers can inject data over the CAN bus or introduce arbitrary modifications in legit CAN data frames and easily compute a new valid CRC.

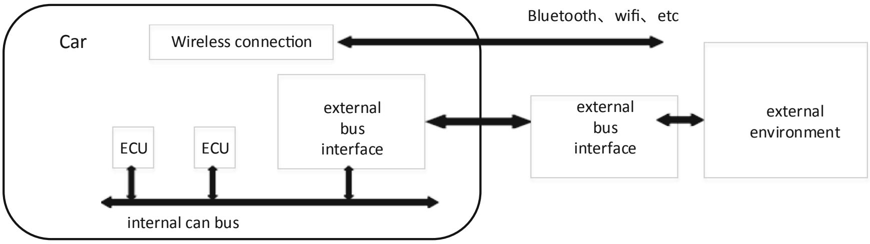

Fig. 1. On-board CAN network model.

3.2 Security Problems in CAN

Previous studies have pointed out the following security vulnerabilities in the CAN protocol.

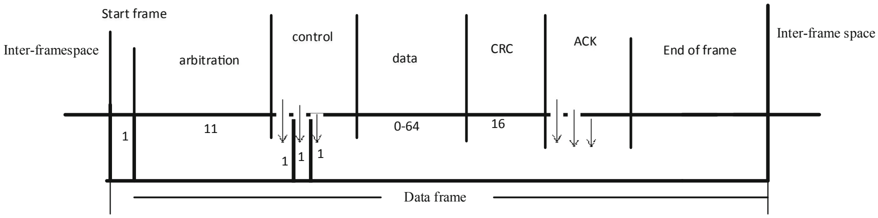

Fig. 2. CAN message standard format.

Lack of security mechanism: Every CAN message can be easily spoofed because it does not contain sender information and does not use any message authentication mechanism other than CRC.

Broadcast nature: Each CAN message is transmitted to all nodes connected to a single CAN bus, which allows an attacker to easily eavesdrop or analyze CAN messages.

Due to these security vulnerabilities in the CAN protocol, it poses a serious threat to today's smart cars.

3.3 Mahalanobis Distance

The Mahalanobis distance is an effective way to calculate the similarity between samples or between samples and their shared patterns. It is not affected by the dimension and can amplify small changes in the variable. It can also be used to identify outliers. The Mahalanobis distance was proposed by Indian statistician Mahalanobis in 1936. The generic formula for evaluating the Mahalanobis distance between the sample vector and population G can be found in Eq. 1.

$$D_M\left(x\right) = \sqrt{\left(x-\mu\right)^T S^{-1}\left(x-\mu\right)} \tag{1}$$

where μ and S represent the mean vector and covariance matrix of the overall sample G, respectively.

4 System Resign

4.1 Overview of the System

The whole system is generally divided into a model training part and a test part.

Training phase: After the message flows into the data processing module, the data frame is calculated by the method proposed in the paper to obtain the confidence interval of the data frame feature. The feature storage module is responsible for storing these features, which will be used to identify CAN data frame anomalies.

Test phase: The decision module is responsible for determining whether a CAN message has an exception. If it is a normal message, you can also control whether it can enter the message collection module. Otherwise the system will be alert. If the detection system performs well, there is no need to update the CAN message feature database of the data storage module. Otherwise it will need to update its message library. Detection system structure is shown in Fig. 3.

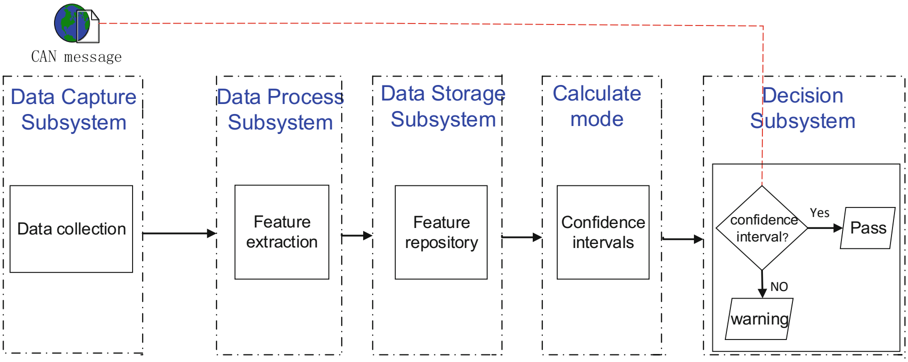

Fig. 3. System architecture.

4.2 Overview of the Proposed Algorithm

Since the CAN data frame can have up to 8 bytes, it uses 4-bit encoding. We assume that the data frame samples of the same CAN ID follow the same distribution. First, preprocess the on-board CAN bus message samples to obtain a standardized on-board CAN bus message samples. Samples of packets with incomplete data bits are self-filled to 8 bits, and all filled with 0. Finally, Normalized message samples are grouped based on the same CAN ID. Suppose the group has M sample vectors $X_1 \sim X_m$, each sample vector has 8 features (every 8 bits as a feature), μ and S represent the mean vector and covariance matrix of the overall sample G, respectively. The Mahalanobis distance is evaluated as shown in formula (1). The algorithm steps are as follows.

Then, we can define a confidence range $\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}} (n-1), \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}} (n-1)\right)$, where S is the variance of samples, n is the number of the samples, and $1-\alpha$ is Confidence level (here take 0.95), $\overline{x}$ is the mean of all d_i values, and $t_{\frac{\alpha}{2}}$ value can be gained from table in [5]. The value of d falls below an empirical threshold $\left(0 < d < \bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}} (n-1)\right)$ or exceeds an empirical threshold $d > \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}} (n-1)$ indicates that the value is abnormal.

It is worth noting that the t distribution is used because the experimental data collected can only be regarded as the sampling of the normal CAN message distribution, so we use the confidence interval of the Mahalanobis distance to perform anomaly detection, and the detection effect will be better.

Algorithm 1. Calculating Mahalanobis distance from data frames

```
Input: CAN bus messages $M = \{m_1, m_2, m_3, ..., m_n\}$
Output: Mahalanobis distance between samples and population mean vector $D_M = \{d_{M_1}, d_{M_2}, \cdots, d_{M_n}\}$
for each data frame do
    Group by ID
    for every 8 bit do
        Convert into decimal
    end for
    Calculate Mahalanobis distance
end for
```

5 Experiments

5.1 Experiment Setup

The data set used in this experiment is a real CAN bus data from an unmodified licensed vehicle, which contains 150000 CAN messages from the CAN bus with a speed of 250 kbit/s.

The data in its original state is a collection of text files containing comma separated values with ID, DLC fields (Indicates data length) and data fields. We divide the raw data files into different files by ID. For each different ID file, we assign 80% of the file data to the training model and the 20% file data is split into normal and simulated anomaly among the testing packets. Each byte in data field is an attribute of the message. The data frame is composed of 8 attributes (n0, n1, n2, n3, n4, n5, n6, n7) to represent the corresponding meaning in the communication protocol. In actual situations, the received CAN message needs to be analyzed based on the specific communication protocol. We convert the bus message from hexadecimal to decimal, which does not lose its attribute characteristics. The experimental data comes from the real vehicle environment and there are no abnormal messages. Therefore, we use the odd-order random method to generate abnormal messages. Each of these attribute value is obtained by a random assignment of 0–255. We inject these abnormal data frame into the message which CAN ID is legal to get the abnormal test data. Finally, these abnormal data are randomly mixed into the normal data as test samples.

5.2 Anomaly Detection

Select 500 normal CAN message with ID 0x10d as training samples. Take 200 abnormal samples randomly mixed with 300 normal samples as test cases and input to system for detection. The test results are shown in Table 1. It should be noted that for the attack of maliciously injecting data packets, the proposed detection algorithm has no false alarm, hence the false positive rate is 0.

Table 2 shows the range of Mahalanobis distances between the sample data frames of each ID and their overall samples. Figure 4 shows the detection rate of the different kind of CAN ID message. From the result, we can know that the anomaly

Table 1. Test results

Messages	Abnormal packets	Abnormal packets detected	Detection rate (%)
500	200	185	92.5

detection for the data field of the CAN message has different detection effect on different CAN ID. We find that the detection rate of an abnormal message whose data frame contains many 0 or changes less is high. In contrast, the detection rate of an abnormal message whose data field changes greatly is low.

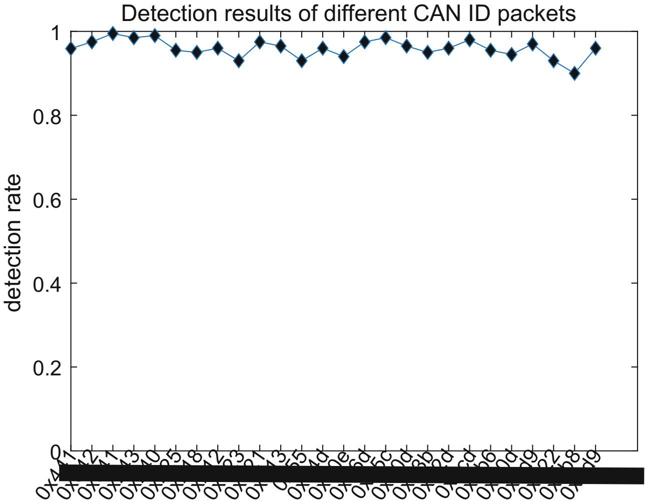

Fig. 4. Detection results of different CAN ID packets.

Table 2. Mahalanobis distance interval for different ID samples.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13
min_{D_M}	1.117	1.534	0.884	0.868	0.903	0.860	1.621	1.410	0.873	0.865	0.913	1.029	0.861
max_{D_M}	5.298	3.938	5.620	4.725	5.738	5.634	5.534	5.050	5.507	5.532	4.158	4.956	6.021

5.3 Time and Memory Cost

This section gives an accurate examination of the computational complexity and memory requirements of the proposed algorithm to demonstrate its low computational requirements, making it suitable for common ECUs in modern vehicles.

Computational Complexity: Since the message transmitted on the CAN bus is binary, Sect. 4 discusses the real-time detection algorithm that requires the conversion of binary data to decimal. The time complexity required for this step is O(N), N is the number of digits required. When calculating the Mahalanobis distance between a CAN message data frame of a specific ID and its overall mean value, the required time complexity can be regarded as a constant, that is, the calculation can be completed in a linear time. The computational complexity of the final real-time detection is evaluated as O(N).

Memory Requirements: The proposed detection system requires the use of an index data structure to store a confidence interval for a specific ID with respect to the Mahalanobis distance. The size of the structure is evaluated as $N * L$, where N is the number of unique ID on the internal network (in this paper, 27 different CAN ID are found), and L is the number of bits required to store the confidence interval. The confidence interval data is FLOAT type. The memory of the FLOAT type consists of three parts: the sign bit (1 bit), the exponent part (8 bits) and the mantissa part (23 bits). So for the detection system, in the feature storage module, $2 * 32 * 27/8 = 216$ bytes is required, in the calculation module, $(8 * 8 + 8) * 32/8 = 288$ bytes is required. $8 * 8 * 32$ is used to store the inverse of the covariance matrix of a particular ID, and $8 * 32$ is used to store the mean vector. Common low-end ECUs are generally composed by microcontrollers with 1 computational core, with few hundreds of Kilo Bytes of RAM. And its operations can be carried out by a common microcontroller equipped with a single core. Hence, the proposed live-detection algorithm can be implemented on common low-end ECUs.

6 Conclusion

A novel anomaly detection algorithm proposed in this paper detects the abnormality of the CAN message data frame, which can effectively prevent the attacker from maliciously modifying the content and launching an attack on the ECU in the vehicle. In this paper, when collecting vehicle bus messages, the collection time is not long, high speed and other conditions are not considered.

Future work includes not only the integration of our algorithms with detection methods based on replay attacks and traffic attacks, but also the study of the approximate location of tracking malicious ECUs.

References

1. Groza, B., Murvay, S., van Herrewege, A., Verbauwhede, I.: LiBrA-CAN: a lightweight broadcast authentication protocol for controller area networks. In: Pieprzyk, J., Sadeghi, A.-R., Manulis, M. (eds.) CANS 2012. LNCS, vol. 7712, pp. 185–200. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35404-5_15
2. Narayanan, S.N., Mittal, S., Joshi, A.: Using data analytics to detect anomalous states in vehicles. arXiv preprint arXiv:1512.08048 (2015)
3. Kammerer, R., Frömel, B., Wasicek, A.: Enhancing security in CAN systems using a star coupling router. In: 2012 7th IEEE International Symposium on Industrial Embedded Systems (SIES). IEEE (2012)

4. Studnia, I., et al.: Security of embedded automotive networks: state of the art and a research proposal. In: SAFECOMP 2013-Workshop CARS (2nd Workshop on Critical Automotive applications: Robustness & Safety) of the 32nd International Conference on Computer Safety, Reliability and Security (2013)
5. Tang, D.: Probability Theory and Mathematical Statistics. Tianjin University Press, Tianjin (2009)
6. Miller, C., Valasek, C.: Adventures in automotive networks and control units. Def Con **21**, 260–264 (2013)
7. Taylor, A.: Anomaly-based detection of malicious activity in in-vehicle networks. Université d'Ottawa/University of Ottawa (2017)
8. Kang, M.-J., Kang, J.-W.: Intrusion detection system using deep neural network for in-vehicle network security. PloS One **11**(6), e0155781 (2016)
9. Kang, M.-J., Kang, J.-W.: A novel intrusion detection method using deep neural network for in-vehicle network security. In: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE (2016)
10. Wang, C., et al.: A distributed anomaly detection system for in-vehicle network using HTM. IEEE Access **6**, 9091–9098 (2018)
11. Taylor, A., Leblanc, S., Japkowicz, N.: Anomaly detection in automobile control network data with long short-term memory networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE (2016)