2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)

Research on vehicle network intrusion detection technology based on dynamic data set

Bin Qiu
[1] State Key Laboratory of Automotive Safety and Energy Conservation, Tsinghua University, Beijing, 100084;
[2] Equipment Industry Development Center, Ministry of Industry and Information Technology, Beijing, 100846

Ke Chen
China Automotive Technology and Research Center Co., Ltd., Tianjin, China

Kexun He
CATARC Software Testing (Tianjin) Co., Ltd., TATC Tianjin, China
hekexun@catarc.ac.cn

Xiyu Fang
CATARC Automotive Test Center (Tianjin) Co., Ltd, TATC Tianjin, China

2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC) | 978-1-6654-0605-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICFTIC54370.2021.9647072

***Abstract*—a new round of scientific and technological revolution and industrial reform promote the intelligent development of automobile and promote the deep integration of automobile with Internet, big data, communication and other industries. At the same time, it also brings network and data security problems to automobile, which is very easy to cause national security and social security risks. Intelligent vehicle Ethernet intrusion detection can effectively alleviate the security risk of vehicle network, but the complex attack means and vehicle compatibility have not been effectively solved. This research takes the vehicle Ethernet as the research object, constructs the machine learning samples for neural network, applies the self coding network technology combined with the original characteristics to the network intrusion detection algorithm, and studies a self-learning vehicle Ethernet intrusion detection algorithm. Through the application and test of vehicle terminal, the algorithm generated in this study can be used for vehicle terminal with Ethernet communication function, and can effectively resist 34 kinds of network attacks in four categories. This method effectively improves the network security defense capability of vehicle Ethernet, provides technical support for the network security of intelligent vehicles, and can be widely used in mass-produced intelligent vehicles with Ethernet.**

Keywords- cyber security; Intrusion detection; data set.

I. INTRODUCTION

A new round of scientific and technological revolution and industrial reform are promoting the development of intelligent vehicles. In 2020, the sales volume of intelligent networked vehicles in China was 3.032 million, and the market penetration was about 15% [1]. It is estimated that networked vehicles worldwide will account for 86% of the automotive market [2]. While promoting the deep integration of automobile with Internet, big data, communication and other industries, intelligent automobile also brings network and data security problems to automobile. The increase of automobile networking rate makes it gradually become the key target of network attack, with prominent security risk and weak security protection foundation [3]. In 2015, Charlie Miller and Chris valasek successfully invaded a jeep remotely, which can remotely control the air conditioner, wiper and even accelerator and brake of the vehicle, which can seriously affect the personal safety of drivers and passengers, resulting in the recall of 1.4 million vehicles [4]. In the past 10 years, the number of automobile network security incidents has been increasing sharply. On the other hand, according to the dynamic monitoring of the Internet of vehicles of the Ministry of industry and information technology, it has been found that more than 2.8 million malicious attacks have been committed by vehicle enterprises, Internet of vehicles information service providers and other related enterprises and platforms since 2020. The risks of platform vulnerabilities, communication hijacking and privacy disclosure are very serious and the harm is more severe [5], which is very easy to cause national security and social security risks.

The risk of automobile network security stimulates the urgent development of automobile network security. The research on automotive network security at home and abroad has also received great attention and development. Car companies such as Weilai, GAC and great wall began to build protection systems, including the intrusion detection capability of cars. For example, Great Wall deployed IDPs system on its new models. Tencent, 360, Baidu and other Internet companies have also set up Internet of vehicles security laboratories to carry out protection research. For example, 360 automotive laboratory has developed the latest on-board intrusion detection engine hardware, which can effectively monitor the on-board operating system at the process level [6]. In domestic research, cryptographic technology, message authentication and intrusion detection technology are effective means for automotive network security protection [7]. The on-board network includes a variety of networks: can, Lin, FlexRay, most, Ethernet, etc. Among them, Ethernet and can bus networks are the most studied. Scholars at home and abroad have carried out a lot of research on the intrusion detection technology of automotive Ethernet and can bus. The intrusion detection technology of information entropy and machine learning has been used for the intrusion detection of can center [8-9], and achieved good results. ECU fingerprint and can frame fingerprint extraction are also used in the research of on-board network intrusion detection technology, and good recognition effect is achieved [10-11]. In addition, by using Bloom filter to detect the cycle of message identifier and some data fields, it is helpful to detect potential replay and tampering attacks [12]. Chevalier [13], a Japanese scholar, has developed an on-board embedded log analyzer, which is used to detect a

978-1-6654-0605-5/21/$31.00 ©2021 IEEE

2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)

Authorized licensed use limited to: Wuhan University. Downloaded on January 08,2023 at 08:37:12 UTC from IEEE Xplore. Restrictions apply.

small number of payload bytes that may change rapidly. The detection results are better interpretable than machine learning methods. Li Ning [14] studied the connection between the on-board terminal close to the mobile network and the off-board network.

There is no intrusion detection device that can adapt to different models and Netcom itself at the on-board end of intelligent Netcom vehicles. There is no technical barrier in the whole market, and the intrusion detection module has not been put into mass production on a large scale. Major automobile manufacturers are speeding up the research and development of intrusion detection technology, but the ever-changing attack methods have brought great trouble to intrusion detection technology and posed a serious security threat. Because the existing detection technology and defense system can not meet the changing challenges, and the artificial intelligence based on machine learning has brought a new opportunity to solve the information security problem of intelligent networked vehicles. Aiming at the attack model of intelligent vehicle, this study studies the lightweight intrusion detection system that can be used in the vehicle end, so as to provide the basis for the network security protection of intelligent vehicle.

II. Intrusion Detection Technology

Intrusion detection technology originates from security audit technology. By monitoring the activities of the system and the behavior of users, it detects the behavior of trying to bypass the protection mechanism, the leap forward of user identity and external intrusion. The model of intrusion detection technology was proposed by Dorothy Denning [15], and the current detection technologies are extended based on this model.

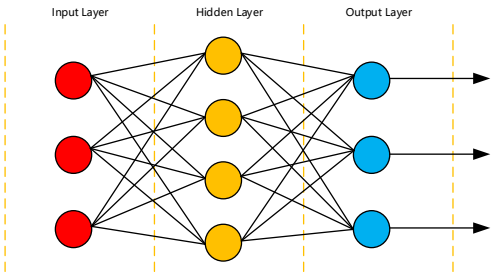

Figure 1 Artificial Intelligence

Thus, intrusion detection is to identify attacks and threats by collecting audit records, network packets, application records or other information in the system. An intrusion detection system is to judge whether there is abnormal behavior in the network or system by collecting network traffic information, user node communication mode, system statistical data, log report and so on, and use it as a basis for alarm or intrusion processing.

According to different analysis methods and technologies, intrusion detection technology is usually divided into feature detection (misuse detection) and anomaly detection:

(1) Signature based detection, also known as misuse detection, assumes that the intrusion activity can be represented by some patterns, and determines whether it is an intrusion activity by detecting whether the subject activity conforms to these patterns. Therefore, feature detection mainly realizes the detection of known attack types. The technical difficulty lies in how to design patterns that can clearly distinguish intrusion activities from normal activities.

(2) Anomaly based detection is the feature extraction and modeling of the subject's normal working model. It can judge whether it is normal activity by detecting whether the active activity is the same as the normal working model. This detection method has high compatibility and can detect unknown attacks. The main disadvantage of this attack method is that it is prone to false positives or false positives. The technical difficulty lies in how to establish the normal working model and how to design the statistical algorithm, so as to avoid treating the normal operation as "intrusion" and ignoring the real "intrusion" behavior.

According to different data sources, intrusion detection can be further divided into host-based intrusion detection and network-based intrusion detection. Host-based intrusion detection mainly uses the audit and tracking log of the operating system as the data source; Network-based intrusion detection uses the original traffic transmitted on the network as the data source.

Whether feature detection or anomaly detection, its purpose is to correctly separate intrusion behavior data from normal data as much as possible. The difference between feature detection and anomaly detection lies in the different classification methods. At present, the widely used classification algorithms can be applied to intrusion detection to distinguish data into normal data and abnormal data. The traditional intrusion detection methods include statistical analysis, predefined rules and so on.

(1) Intrusion detection method based on statistical analysis. For anomaly detection, this method first initializes a system file. During the operation of the system, the anomaly detector constantly compares the current system state with the initial system file. If the deviation exceeds the threshold, it is considered as an intrusion. In some systems, the current system state will be updated to the initial system file, so that the intrusion detection system can adaptively learn the user's behavior pattern. Its disadvantage is that it needs to rely on a large number of known data, and the statistical analysis is not sensitive to the sequence of events, which also makes the system lose the correlation information between events. The false positive and false negative rates are heavily dependent on the setting of the threshold.

(2) Intrusion detection method based on predefined rules. The knowledge about intrusion is transformed into rule base. Taking the expert system as an example, its rules are if then structure. The former is the condition of intrusion, the latter is the response measures taken after discovering intrusion, and the condition of intrusion can be the characteristics of intrusion behavior. This method has a high detection rate for known attacks or intrusions, but it can not detect when it is not in the predetermined rules. In addition, the rule base often needs to be updated dynamically, which will make maintenance more

Authorized licensed use limited to: Wuhan University. Downloaded on January 08,2023 at 08:37:12 UTC from IEEE Xplore. Restrictions apply.

difficult. When changing rules, we also need to consider the impact on other rules in the rule base.

III. METHOD

Aiming at the information security problems of intelligent networked vehicles, especially the communication security and application security problems, this paper analyzes the real-time sea volume heterogeneous data generated by key parts of intelligent vehicles (including central gateway, intelligent cockpit and on-board computing platform) in on-board Ethernet through machine learning technology. Through the real-time analysis of massive heterogeneous data, the key features are extracted, and the machine learning algorithm model is used to analyze whether its behavior is attack behavior or normal behavior.

A. Data set construction

In order to adapt the machine learning algorithm to the real vehicle environment and complete the requirements of integration verification between the project and the real vehicle, it is also necessary to collect and sort out the data in the real vehicle, focusing on network data, log data and can data. In the ECU gateway, the nfqueue rule target is defined through Netfilter and iptables mechanisms, and the network data packet is extracted from the kernel and passed to the user process. In this way, the network data collection can be realized.

B. data pre-processing

The process of preprocessing data in the real environment is mainly divided into four steps: first, obtain the data in KDD CUP99 data format from the data source, read it into memory, and modify the protocol_ The type feature and flag feature are uniquely hot coded (processing disordered discrete features). At this time, the dimension of the data will rise, and then the processed data features will be normalized for some large value ranges. At this time, after the dimensionality reduction operation of principal component analysis, the data can be sent to the prediction model for prediction.

The original data set refers to the KDD cup 99 data set. Each sample in the constructed network traffic data set is a connection record, which is composed of 41 features.

The program flow first reads the parameters required by the algorithm from the file, then constructs the execution sequence of independent heat coding, normalization and principal component algorithm, finally sends the data to be processed into the algorithm processing sequence one by one, and finally obtains the data that can be sent into the prediction model for processing.

Read data from the data source to the corresponding data structure in memory. Data sources can be in the form of text files, strings in source code, and files in specific formats (such as CSV format, text files divided by specific separators). The content of data includes two categories: one is preprocessing parameters, including independent heat coding parameters, data normalization parameters and dimensionality reduction parameters of principal component analysis; The other is the data to be preprocessed, which is the characteristic parameters of the data packets transmitted in the network (such as protocol name, etc.). The module handles different data formats in different data sources into a standardized form, and does not care about and distinguish the content of data.

The preprocessing algorithm includes three algorithms: one hot encoding, normalization and principal component analysis (PCA). The original data are processed by these three algorithm modules in turn to obtain the data format that can be sent to the prediction model, which is stored in the DataTable for sending to the prediction model.

With the important characteristics of data set_ Type (feature No. 2) as an example. protocol_ There are three types of feature values: ICMP, TCP and UDP. However, the machine learning model has no way to deal with such symbol features, so the symbol features need to be digitized. The specific method of independent heat coding is to expand the feature dimension to the number of values, and then set the corresponding value location value to 1.

IV. VEHICLE-BUS INTRUSION DETECTION ALGORITHM

(a) Intrusion detection and defense method based on signal

This project puts forward the method of intrusion detection for message and network from the perspective of signal. On board can network is a signal-based network, in which almost every message communicated is assigned one or more fixed signals, and these signals have their own value range and practical significance. Therefore, it can judge whether the message is legal by checking whether the value of the signal contained in the message is in the normal value range, It can also check whether network intrusion occurs by monitoring whether the relationship between the values of multiple signals in the network is abnormal.

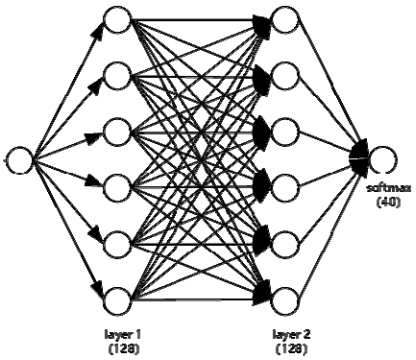

Figure 2 Intrusion Detection Method

① Signal value range check

It is a common method of intrusion detection to judge the legitimacy of the message by checking whether the value of the data in the message is legal. Many studies also propose to check the value range of communication data for the can network, but they just check each byte, which is not accurate, And can only rely on a large number of normal data to count the value range of each byte. The on-board can network is a signal-based network, which specifies the starting position,

Authorized licensed use limited to: Wuhan University. Downloaded on January 08,2023 at 08:37:12 UTC from IEEE Xplore. Restrictions apply.

length, range and other attributes of each signal on the message in its communication matrix. Therefore, the values of each signal in the communication message can be extracted according to these information, and then check whether the values of these signals are within the legal range, that is, check whether the data in the message conforms to the provisions of the communication matrix. Through this method to judge whether the message is legal or not, it can effectively defend against partial tampering and injection attacks.

② Signal value relationship check

Each signal in the on-board can network has its unique practical significance, such as a series of signals representing engine speed, vehicle speed, door status, etc. these signals do not exist alone in the on-board system. Some signals may have a unique relationship, such as the linear relationship between engine speed and vehicle speed during normal driving, For another example, when the wiper washing liquid is sprayed, the wiper will also start. Therefore, network intrusion can be detected by monitoring whether the relationship between these signals is abnormal. According to the type of relationship, the signal relationship can be divided into two types: the relationship before and after the same signal value (signal change rate) and the relationship between signals. For the relationship between signals, this project makes an in-depth study on the relationship between two signals and the complex relationship composed of three or more signals.

Because many signal values often represent actual physical parameters, the change of some signal values will tend to a smooth curve or straight line, such as the signal representing vehicle speed. Due to the limitation of vehicle acceleration, the growth of signal value cannot be infinite, that is, the growth rate of signal value will have an upper limit, so the values of these signals can be monitored, And judge whether the network is abnormal by calculating whether the change rate exceeds its upper limit.

Taking the relationship between two signals as an example, this paper illustrates the application of the relationship between signals in intrusion detection and defense. For signal a and signal B, it is assumed that they have the following relationship:

When the value of signal a is in the range of [A1, A2], signal a and signal B satisfy the relationship: F (a, b) < 0, f (a, b) is a binary expression containing two signal values of a and B.

The relationship here is only an example and can be matched more flexibly: the preconditions can be set according to the actual relationship between the two signals, or they can be set so that there is no need to check the preconditions; The relation can be not only an inequality, but also an equation, but also a value range.

For the above relationship, the value of update signal can be collected and recorded in real time. When a meets the conditions within the range of [A1, A2], the real-time detection of signal B will be started, that is, every time a message with signal B is received, the value of signal B will be read, and whether the value of signal B meets the relationship f (a, b) < 0 will be checked. If not, an exception will be reported.

For the positions of signals a and B, there can be the following two cases:

A and B are on the same message: at this time, there is no need to record the values of the two signals. When receiving the message, directly read the values of the two signals to check whether the relationship is satisfied. At this time, the legitimacy of the message can even be judged by the inspection results.

A and B are on different messages: at this time, the value of signal a needs to be recorded and updated in real time to check the value of signal B. when the relationship is not satisfied, an exception is detected.

(b) Implementation technology of intrusion detection and defense based on signal

① Signal value extraction method

In order to use the signal value for intrusion detection and defense, the signal value in the message needs to be extracted first. Therefore, the extraction method of the signal in each message needs to be recorded, and the relevant attributes of each signal need to be known: start bit, length, format (symbol type, size end). As shown in the following table, the signal analysis method of a message is recorded, including the number of signals to be extracted, the identifier of each signal and various attribute values of each signal. The SID of the signal is the signal identification, which is used to distinguish different signals in the same message. Canid + Sid can determine a unique signal. The attribute information of the signal can be obtained by parsing the DBC file of the corresponding network, and the SID needs to be numbered by itself, which can be numbered and identified according to the sequence of the signal.

Table 1 signal extraction method

name	**lenth**	**value**	**description**
CID	4Byte	-	CANID
SNUM	1Byte		Number of signals to be extracted
SID1	1Byte	[0-63]	Identifier for signal 1
Format1	1Byte	-	Signal 1 format
Start1	1Byte	[0-63]	The start bit of signal 1
Length1	1Byte	[1-63]	The length of signal 1, in bit
…	…	…	

② Signal value range check

Identify a unique signal through canid + Sid and set its value range. The range can be expressed in a variety of ways: < = max, > = min, [min, Max], or a value list, or a prohibited value list, etc. These range information will be recorded in the

Authorized licensed use limited to: Wuhan University. Downloaded on January 08,2023 at 08:37:12 UTC from IEEE Xplore. Restrictions apply.

DBC file. The value range information of the signal can be obtained by analyzing the DBC file of the corresponding network, More accurate value range can also be set manually.

③ Signal relationship check

The threshold value of signal change rate, i.e. the change threshold value of unit time, should be set before checking the signal change rate. The time can use standard time such as MS or s, or the clock unit provided by the actual environment. Periodic messages can also use cycle as the unit, so it will be more convenient to record and calculate. In addition to collecting the value of the recorded signal, it is also necessary to collect the arrival time of the message to calculate the change rate of the signal in two adjacent messages (messages with the same ID), and then judge whether the network is abnormal by checking whether the value exceeds the set threshold.

Check the relationship between signals, respectively identify the two signals with relationship through canid + SID, and set the preconditions for the relationship check and the specific relationship formula F. The test case in this paper adopts the linear relationship of bivariate once, that is, the relationship is Ma + Nb + L < = 0 (where m, N and L are known integers). In addition, in order to prevent false positives caused by packet loss, for the signal relationship check in different messages, it is also necessary to record the arrival time of the signal and set an effective time of the signal value.

V. CONCLUSION

Aiming at the information security problem of intelligent networked vehicle on-board Ethernet, this research applies the massive heterogeneous data generated by the key parts of intelligent networked vehicle in real time, extracts the key features through the real-time analysis of the massive heterogeneous data, designs the network intrusion detection algorithm based on the deep neural network, and applies the intrusion detection algorithm based on the existing on-board bus on-board terminal, The verification test is carried out through the built bench. The intrusion detection method formed in this study can be widely used in key parts with Ethernet communication, such as central gateway, intelligent cockpit and on-board computing platform, fill the gap of relevant domestic technologies, provide strong support for vehicle active defense, and improve the information security defense level of intelligent vehicles.

REFERENCES

[1] Huang Peng, et al. Research on data security of intelligent networked vehicle [R]. China Automobile Industry Association.

[2] https://www.capgemini.com/wp-content/uploads/2019/08/Connected-Vehicle-Trend-Radar.pdf.

[3] Herzliya, Upstream security's 2021 global automotive cybersecurity report [EB/OL]. https://upstream.auto/2021report, 2020-12-05/2021-08-25.

[4] Miller C, Valasek C. Remote exploitation of an unaltered passenger vehicle[J]. Black Hat USA, 2015, 2015(S 91).

[5] Zou Bosong, et al. Intelligent and Connected Vehicle Safety Penetration White paper2020[R]. China Software Evaluation Center, 2020 ..

[6] Lu Z, Qu G, Liu Z. A survey on recent advances in vehicular network security, trust, and privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(2): 760-776.

[7] Wu Wufei, Li Renfa, Zeng Gang, et al. Overview of Research on Network Security of Intelligent Connected Vehicles[J]. Journal on Communications, 2020, 41(6): 161-174.

[8] Yu He. Research on Information Security of Networked Vehicles and CAN Bus Anomaly Detection Technology [D]. Jilin University, 2016.

[9] Tian Daxin, Yan Huiwen, Wang Congyu. Research on Vehicle Network Security in Internet of Vehicles[J]. Mobile Communications, 2019, 43(11): 40-46.

[10] Cho K T, Shin K G. Fingerprinting electronic control units for vehicle intrusion detection[C]//25th {USENIX} Security Symposium ({USENIX} Security 16). 2016: 911-927.

[11] Kneib M, Huth C. Scission: Signal characteristic-based sender identification and intrusion detection in automotive networks[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018: 787-800.

[12] Groza, B, Pal-Stefan Murvay. Efficient Intrusion Detection With Bloom Filtering in Controller Area Networks[J]. IEEE Transactions on Information Forensics and Security 2019,14(4): 1037-1051.

[13] Chevalier Y, Rieke R, Fenzl F, et al. ECU-Secure: characteristic functions for in-vehicle intrusion detection[C]//International Symposium on Intelligent and Distributed Computing. Springer, Cham, 2019: 495-504.

[14] Li Ningning. Research and implementation of vehicle network intrusion detection technology based on machine learning [D]. University of Electronic Science and Technology of China, 2019.

[15] Denning D E. An intrusion-detection model[J]. IEEE Transactions on software engineering, 1987 (2): 222-232.

Authorized licensed use limited to: Wuhan University. Downloaded on January 08,2023 at 08:37:12 UTC from IEEE Xplore. Restrictions apply.