Intrusion Detection System on Controller Area Network Based on Autoencoder

Kamronbek Yusupov[1[0009-0003-0034-0945]], Md Rezanur Islam[1[0000-0002-1183-7741]and Insu Oh[2[0000-0002-6545-9125]], Kangbin Yim[2[0000-0002-1361-1455]]

[1] Department of Software Convergence, Soonchunhyang University, Asan, Korea
[2] Department of Information Security Engineering, Soonchunhyang University, Asan, Korea
{yuskamron, arupreza, catalyst32, yim}@sch.ac.kr

Abstract. In modern world, vehicles are becoming an integral part of our lives, so car protection becomes a priority. Inside modern vehicles, a standard Controller Area Network (CAN) communication protocol is used to maintain communication between Electronic Control Units (ECUs). Despite their practical, adaptable CAN architecture, they lack encryption and authentication. Because of this, attackers manage to hack the CAN communication system and gain access to CAN messages exchanged between ECUs. To overcome this problem, we propose an Intrusion Detection System (IDS) based on the Autoencoder unsupervised learning method. The current paper reviews efficient preprocessing techniques and demonstrates the results of our Autoencoder model. Based on the results, proposed unsupervised Autoencoder model, effectively copes with the tasks of detecting and distinguishing abnormal CAN messages, such as DoS and Replay, from normal messages CAN. The results were also demonstrated in the form of plot showing the difference between the attacked and non-attacked data. Given the encouraging results obtained from our experiments, we propose the use of Autoencoder-based IDS.

Keywords: Intrusion Detection System, CAN Bus, Autoencoder.

1 Introduction

With the great growth in development in the field of mechanical engineering and automotive technology, new vehicle capabilities have been discovered such as self-driving cars and connecting the car to Internet networks. The introduction of new technologies into vehicles has opened new platforms and opportunities to attack the vehicle. With the connections of new technologies and the connection of the car to the world, there is a risk of potential threats to the vehicle. Since car protection is often associated with the safety of people's lives, this aspect has become a top priority. The car has minicomputers, that is, Electronic Control Unit (ECU) [1], which are responsible for certain actions in the car such as the transmission, engine, various equipment, and braking systems. In modern cars there are more than 100 of these control units, they are connected using a communication protocol called Controller Area Network (CAN) [2]. This protocol is widely used in this field and can be called a

communication standard due to its tailored structure and ease of use. But since CAN does not have encryption and authentication, this system is considered not secure since this system has some loopholes through which one can penetrate the system and intercept various CAN messages. After which the attacker will be able to control or send different messages to another control unit. Thus, the driver loses control over the vehicle, for example, as the engine stops or brakes fail.

To avoid such cases, an Intrusion Detection System (IDS) was developed, which detects anomalies in CAN messages and neutralizes them from other messages. An effective IDS must demonstrate fast detection because time is of the essence. Also, the IDS must effectively distinguish normal data from abnormal. Because models that have been trained with Supervised learning are likely to make mistakes when detecting intrusions or classifying attacks. But the strength of Autoencoder is the ability to learn without labeling and find the differences between normal and attack messages on its own.

The main goal of the current scientific article is to build an effective unsupervised model that can effectively cope with the tasks of detecting anomalies and finding the difference between normal and attacked data sets. We believe that this approach plays a key role and is important because with this method it is possible to avoid potential threats that may be associated with the car and its driver.

2 Related Work

With the implementation of a large number of ECUs in the automotive industry, the life of drivers and passengers is becoming more convenient and comfortable. However, due to CAN vulnerabilities, the car remains a prime target for an attacker. By exploiting CAN vulnerabilities, it is possible to interfere with the normal operation of the system, such as the injection of malicious messages, command falsification, obstruction of receiving priority messages that are exchanged within the network, which can threaten the driver or passengers. Since this problem is considered a priority, many researchers have demonstrated their experimental results. Zareno et.al presented a two-layer architecture that learns unsupervised [3] and an efficient algorithm that detects anomalies with high accuracy. However, the results of these experiments are relevant until the system architecture is fully verified. Choi et.al proposed an intrusion detection system based on an unsupervised autoencoder [4]. Therefore, as a result, the demonstrated model was able to effectively detect anomalies with an accuracy of up to 91%, while other previously proposed models were able to detect with an accuracy of up to 80%. Hanselmann et.al presented the CANet neural network method which underwent unsupervised learning [5]. This method has demonstrated good results in intrusion and anomaly detection tasks in CAN communication systems. Experimental results showed that the proposed model can detect anomalies with a true negative rate of 0.99. Leslie et.al experimented with an unsupervised learning method for anomaly and malicious traffic detection tasks [6]. They used the E-HAC algorithm, thanks to this algorithm the model effectively copes with the tasks of detecting spoofing attacks in CAN communication systems. Narasimhan et.al presented an algorithm that learns unsupervised to cluster different attacks within regular messages exchanged between

ECUs [7]. However, according to the author, the proposed model can work just as effectively with other data such as computer networks and wireless technology networks. Provotar et.al proposed a model based on unsupervised learning to detect anomalies and different types of intrusions. The proposed model [8] showed encouraging results with an accuracy of 87% and 91%. Novikova et.al presented a successfully developed autoencoder model that has functionality as anomaly detection on a CAN communication system [9]. The demonstrated model was able to detect attacks from 96 cases with an accuracy of 91 cases. An autoencoder-based CANnolo model with LSTM was demonstrated by Longari et.al for anomaly and intrusion detection tasks in CAN communication networks [10]. The proposed model exhibited good performance and accuracy for IDS. Hossain et.al proposed a supervised method based on LSTM [11]. Based on the results of the current experiment, the model effectively copes with the classification of different types of attacks and intrusion detection in CAN communication systems.

3 Dataset Preparation and Pre-Processing Methods for Autoencoder

The dataset that was used in our experiment was collected from a real car KIA SOUL [12]. Our first set of data included normal CAN messages and attacked data such as DoS, Replay. The number of messages in the given sets was not the same, so we had to take a certain number in all the data set. Thus, we only extracted 150,000 messages from each dataset to train the model. 50 thousand messages each for testing the model. Our dataset consisted of 14 bars that showed important information about messages in the CAN communication system inside the car. The dataset includes a "Time_Ofset" which shows the timestamp of each event in the system. "CAN ID" is a column that indicates the unique identifiers for each CAN message. Using CAN ID, the system determines which message was sent from which de-vice and sets priority. The "Time_gap" column shows the time interval of each CAN message. Using this column in deep learning, you can build a model that can effectively classify attacked data and normal data. "Data Length Code (DLC)" This column shows information about the number of bytes in each CAN message. After that, to determine the sequence of bytes in the data set, there is a "Payload" column. In our dataset there is one whole Payload that demonstrates all the sequences of bytes and columns that demonstrate each byte separately. The last column is the Label which informs us whether the message is "R - Regular" or "T - Attack".

In any scientific work, the key stage is the data processing part. To create an effective model, we need to determine the correct method for processing data. So, we made a big focus on pre-processing methods. After receiving the data set, we had to sort it and clean it to avoid further problems that might arise during training. Since there were 14 columns in our data set, for our experiment we had to add one column there that will be useful to us in further work. The column that we added this conversion of CAN ID to categorical data Scaled_CAN_ID, we did so to avoid retraining the model with large numbers. Various researchers have conducted experiments using the entire

data set to train a model, but in the current experiment we showed that for effective anomaly detection you do not need to use all the columns of the data set - we only needed two of them. We only used columns such as Scaled_CAN_ID and Time_Gap, because using these two columns you can determine whether it is an attack or an Attack free. After preparing the data set in the format we needed, we used Mean Normalization [13] and Data Standardization [14] to change the entire value and time interval in the range of numbers from 0 to 1. Thanks to this method, the entire average value was subtracted and then divided between the range between values. This way we got the minimum and maximum values. After that we used the permutation method, this method helps us avoid overtraining the model. The second method we used was time series input [15], we set the time steps with values equal to 10 steps. This method works in such a way that the model learns with 10 CAN messages and then begins to predict the next message. After prediction, the sliding window goes down one step and again selects 10 messages, and this function works until the end of the data set. In this way, the model is trained with all messages within the data set. Ultimately, we collected 100 thousand CAN messages from each data set for training and 50 thousand of each for testing. Finally, the entire dataset was formatted into a 3D format and each input data was converted into array NumPy.

4 Intrusion Detection System Based on Autoencoder

Our main goal was to create a deep learning model that learns without the help of a teacher, that is, without the help of labeling data, for IDS intrusion detection on a CAN communication system. To do this, we selected 3 types of data sets from two sources, which included a set of normal CAN messages, Denial of Service (DoS) and Replay attack. After preparing the necessary data preprocessing, we built the unsupervised model architecture we needed using Autoencoder. In the current experiment, we built an LSTM-based autoencoder with two layers. For the experiment, a dataset with a sequence of 10 and two features was used to train the model. To build the model, we used the Tensorflow 2.13.0 library. To build an effective model, we used the Dropout and Normalization methods to obtain stability and avoid overfitting the model. We chose the RMSprop optimizer with a learning rate of 0.0001. For the loss function, we used Mean Squared Error (MSE) which will help us identify the difference between the input versions that were before and the versions that were restored after the decoder. To save only the most effective model, we used the early_stoping method with values of 10 epochs. The early_stopping method works until it finds the most effective model, after which it gives a chance of 10 epochs to identify a model more effective than the previous one; if it does not find it, it saves what it has. We trained our model with Normal dataset and tested with the Attack dataset DoS and Replay. The results showed that when we consider the aspect of "Time Gap," a clear difference becomes apparent. In data the reconstruction error is zero indicating very little dissimilarity. However, in the case of a DoS attack as shown in Fig.1 the reconstruction error deviates significantly. It usually starts at 0.5. Whereas some numbers may overlap in the context of "CAN ID Sequence," abnormalities, such as assaults,

often emerge beyond the 0.015 threshold whereas the bulk of values tend to cluster around the 0 mark in normal data. For detecting replay attack in the context of CAN ID sequence, while a fraction of values may align with attack data, a substantial portion becomes discernible beyond the 0.015 threshold as shown in Fig.1. Conversely, in the time gap sequence, the reconstruction error tends to initiate at approximately 0.022. Consequently, the histogram plot allows us to readily establish a discernible threshold point for our IDS, aiding in effective anomaly detection and system security.

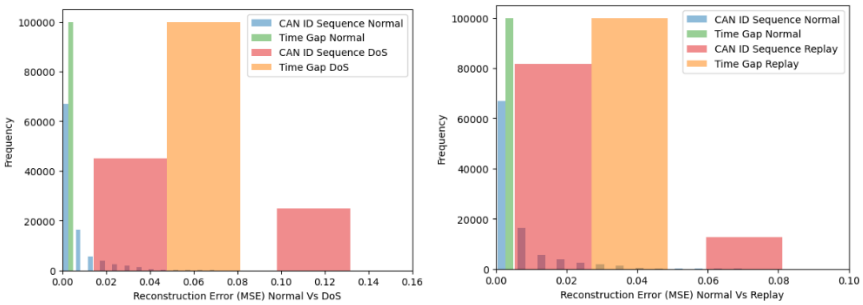

Fig. 1. MSE between Normal, DoS and Replay

5 Conclusion

In the current research, we conducted an experiment to create an effective intrusion detection system based on Autoencoder, which works using the unsupervised learning method. The results show that our proposed model is indeed effective in detecting anomalies in the CAN communication system. Our model was able to find the differences between Attack Free and Attack messages without labeling. We think the results of the current scientific work can become basic knowledge on this topic and will help researchers in future studies. In our future research, we want to use different data sets to more accurately prove our idea that unsupervised models are just as effective as models trained with teachers. After obtaining the results we need, we want to compare the Supervised Learning and Unsupervised Learning models for an intrusion detection system, which of these models is effective for vehicle safety. We hope that our results will inspire researchers to dive deeper into this topic.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2021R1A4A2001810) & This work was supported by Institute for Information & communications Technology Planning&Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2022-0-01197, Convergence security core talent training business (SoonChunHyangUniversity))

References

1. Sim, A. X. A., & Sitohang, B. (2014, November). OBD-II standard car engine diagnostic software development. In 2014 International Conference on Data and Software Engineering (ICODSE) (pp. 1-5). IEEE.
2. HPL, S. C. (2002). Introduction to the controller area network (CAN). Application Report SLOA101, 1-17.
3. Zanero, S., & Savaresi, S. M. (2004, March). Unsupervised learning techniques for an intrusion detection system. In Proceedings of the 2004 ACM symposium on Applied computing (pp. 412-419).
4. Choi, H., Kim, M., Lee, G., & Kim, W. (2019). Unsupervised learning approach for network intrusion detection system using autoencoders. The Journal of Supercomputing, 75, 5597-5621.
5. Hanselmann, M., Strauss, T., Dormann, K., & Ulmer, H. (2020). CANet: An unsupervised intrusion detection system for high dimensional CAN bus data. Ieee Access, 8, 58194-58205.
6. Leslie, N. (2021, March). An unsupervised learning approach for in-vehicle network intrusion detection. In 2021 55th Annual Conference on Information Sciences and Systems (CISS) (pp. 1-4). IEEE.
7. Narasimhan, H., Vinayakumar, R., & Mohammad, N. (2021). Unsupervised deep learning approach for in-vehicle intrusion detection system. IEEE Consumer Electronics Magazine.
8. Provotar, O. I., Linder, Y. M., & Veres, M. M. (2019, December). Unsupervised anomaly detection in time series using lstm-based autoencoders. In 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT) (pp. 513-517). IEEE.
9. Novikova, E., Le, V., Yutin, M., Weber, M., & Anderson, C. (2020). Autoencoder anomaly detection on large CAN bus data. Proceedings of DLP-KDD.
10. Longari, S., Valcarcel, D. H. N., Zago, M., Carminati, M., & Zanero, S. (2020). CANnolo: An anomaly detection system based on LSTM autoencoders for controller area network. IEEE Transactions on Network and Service Management, 18(2), 1913-1924.
11. Hossain, M. D., Inoue, H., Ochiai, H., Fall, D., & Kadobayashi, Y. (2020). LSTM-based intrusion detection system for in-vehicle can bus communications. IEEE Access, 8, 185489-185502.
12. Dataset: HRCL homepage, https://ocslab.hksecurity.net/
13. Wiesler, S., Richard, A., Schlüter, R., & Ney, H. (2014, May). Mean-normalized stochastic gradient for large-scale deep learning. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 180-184). IEEE.
14. Gal, M. S., & Rubinfeld, D. L. (2019). Data standardization. NYUL Rev., 94, 737.
15. Cherdo, Y., Miramond, B., Pegatoquet, A., & Vallauri, A. (2023). Unsupervised Anomaly Detection for Cars CAN Sensors Time Series Using Small Recurrent and Convolutional Neural Networks. Sensors, 23(11), 5013.