International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 03 | Mar 2021 www.irjet.net p-ISSN: 2395-0072

Optimized Anomaly Detection System and vulnerability check using black-box attack to vehicle network

Kiran B. Mohan[1], Nisha Mohan P. M[2]

[1]M Tech Student, APJ Abdul Kalam Technological University, Kerala, India
[2]Asst. Professor, Mount Zion College of Engineering, Kadammanitta, Kerala, India

---

Abstract - *ECUs are the critical components played critical role on an automotive system. Attack from the external attackers and abnormal behavior on ECUs can breach the automotive security and it harms the end users, Anomaly Detection System should be able to detect such behaviors and act accordingly without any delay. AI is widely used for ADS in vehicle network. Black-box attack is the popular and efficient attack to ADS, which do not require a knowledge about the deep learning model used in ADS and not required model internal details and model training data details. In this paper, first we propose an efficient anomaly detection system, which used optimized method to adopt Long Short Term Memory(LSTM) deep learning model for the ADS in-vehicle network. We tuned the existing parameters of the LSTM model and optimized the proposed model by using the characteristics of dataset, which from a practical in-vehicle network. Second, we propose a black-box attack to the LST based ADS, which requires only a small test dataset to train a new victim model. We were able to build a victim model that leads to the wrong interpretation within 50 man-hours. It proves that the community should focus on how to protect the system for the future work, not only focus on how to build an efficient ADS.*

Key Words*: *Deep learning, LSTM, Automotive, Anomaly detection

1.INTRODUCTION

Electronic Control Units are used to construct various subsystems in in-vehicle network. Several peripherals have been connected using the same set of wires which enables different controllers to share the same signal of a single sensor. Controller Area Network(CAN)[1] provides address dependability and fault detection. Attackers used the existing security holes to retrieve sensitive information of automotive critical components such as brake ECU, engine ECU etc [2], [3]. Man-in-the-middle attack, spoofing etc are the possible security vulnerabilities are happen. If adversaries compromise the primary interfaces system, these attacks can be successful.

1.1 Anomaly detection

If there has an efficient anomaly detection method with low latency, protection from anomaly attacks can be easily implemented. Marchetti1 and Stabili proposed identifier sequences based method [4]. Even the system has low computational resources, system is still vulnerable as the content of ID can be intercepted and manipulated. Narayanan proposed Hidden Markov model(HMM) to detect anomalous status in the vehicle [5]. This can be integrated into new as well as old cars, But anomaly detection will be accurate only to pre-specified specific anomalies. Support Vector Machine (SVM) , Support Vector Data Description (SVDD)[6] are proposed to learn the normal behaviour and get trained and then detect unexpected behaviour from the learned experience and report as anomalies. Chockalingam et al. proposed to LSTM to detect anomalies[7] without train the model with anomaly dataset.

However, Attackers can still attack these Anomaly Detection Systems using machine learning by white box, Model tampering, Black box attacks. An attacker, who has the full access and knowledge to the whole deep learning internals and its training data, can attack the system called white-box attack. An attacker attacks the system without have the prior knowledge of either the deep learning model internals and model's training data called black-box attack. Tampering of deep learning model is called model tampering attack. Among them black-box attack is the most popular and efficient. Till now in the use case of in-vehicle network gateway, there have been few studies for the vulnerabilities ADS using LSTM model.

In this paper, We develop and evaluate ADS using LSTM model, which can efficiently detect the anomaly, which not required anomaly dataset to train the model. Second, deploy black-box attack to LSTM based ADS, We show that the black-box attack can be easily mounted on ADS based on LSTM using around 50 man-hours. Victim model shows similar performance as tested on the anomaly dataset after it has been trained. Therefore, it outputs an opposite result compared to the original model. This experiment shows that compared to the original model final output would lead to wrong interpretation.

2. RELATED WORKS

Papernot investigates and show that the classes of algorithms introduced previously to craft adversarial samples classified by feed-forward neural networks can be adapted to recurrent neural networks. In a experiment, they show that adversaries can craft adversarial sequences misleading both categorical and sequential recurrent neural networks. Anderson et al. developed adversarial input generators to attack a recurrent neural network (RNN) [8] used to classify the sentiment of IMDb movie reviews as being positive or negative. To this end, they developed LSTM network as well as two baseline models SVM[9] and Naive Bayes and evaluated their accuracy under the attack by two black-box adversaries and a white-box adversary. Their results showed that though LSTM is more robust than other two models, it's still very susceptible to white-box attack with generated adversary still preserving the sentiment of input review, making us question whether LSTMs really learn the sentiment in this task.

3. EXISTING METHOD

Backbone for ADS in-vehicle network topology is Flex-Ray bus in-vehicle network communication, which connects to several gateways. Gateways supports different functionalities, such as multimedia control, power train, x-by-wire, body/comfort control and charge control, etc.

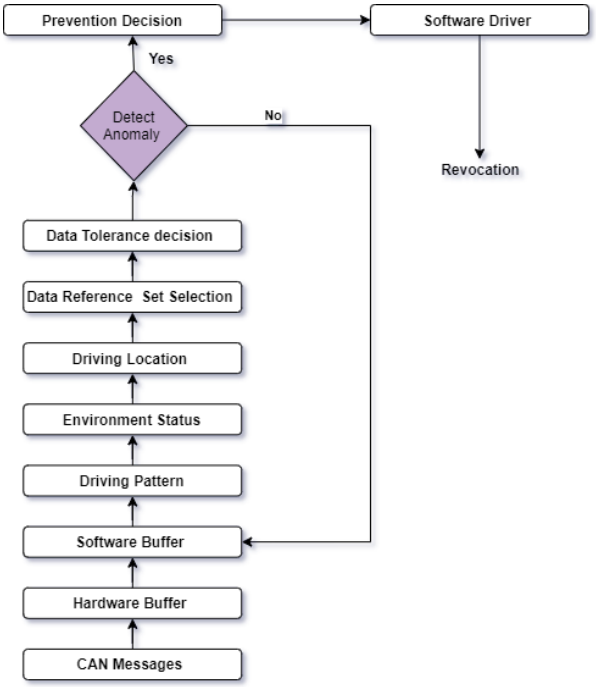

Fig -1: Existing anomaly detection and prevention system workflow

Each gateway will connect to one subsystem in-vehicle network: Flex-Ray, CAN, and MOST. On-Board Diagnostic (OBD-II) port, which is a diagnostics port used to access these networks. Through OBD-II, all the sensitive information transmitted within this vehicular network can be easily accessed. Workflow of an existing anomaly detection and prevention system shows by Fig 1. The existing solutions rely on the value of the data to detect the anomaly in the payload (data) content of the message. The existing solutions compare the incoming data with its references, such as data reference set selection, driving pattern, driving location, environment status and data tolerance decision. These becomes non real-time prevention, caused by long decision (huge time delay) of the anomaly.

Due to the exponential decay in the gradient of the loss function with time, it is difficult to train the model to recognize long-term temporal dependencies for standard RNN. On the other hand, Input, output, and forget gates used by LSTM to have better control over the gradient flow. Maintaining information in memory for long periods to allow LSTM to learn longer-term dependencies. Moreover, data messages need not be decoded as LSTM does not need to know the modeled system's domain. The LSTM based anomaly detector can be used on different vehicles without much modification.

4. PROPOSED OPTIMISED ADS USING LSTM AND ADVERSARY ATTACKS

A. LSTM based Anomaly Detection System

Few works only has been done on LSTM based ADS[10]. In this paper, we proposed, Based on the characteristics of the dataset, which is from practical CAN in-vehicle network, together with tuning existing parameters of LSTM, Optimized the maximum square error. The data type (raw or "converted" physical) and the sample rate are the characteristics of the dataset. The data type and sample rate are tuned together with existing parameters, e.g., length of the sequence and maximum square error.

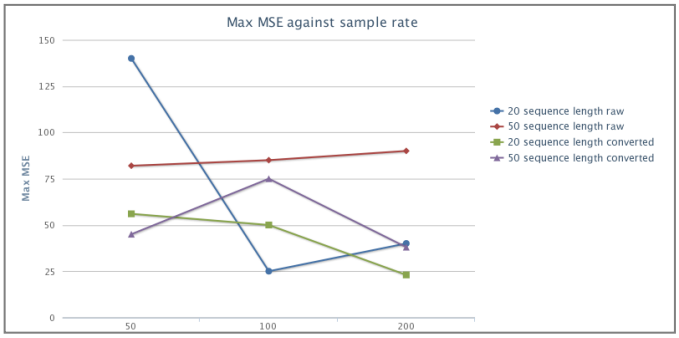

Fig -2: Max MSE that occurs for the specific sample rate, against the sample rate

An experiment is carried out to find the most optimized parameters, such as sample rate of collected velocity dataset, sequence length for LSTM model and dataset type (raw data or physical data), verify with the targeted maximum mean square error(MSE). How the dataset

matches the LSTM model can be evaluated by Mean square error. We collected the velocity of the vehicle from the CAN bus and convert them into the physical value under the highway conditions, in which there have 260 samples for 1 second. With less sudden braking the driving behavior on the highway will be relatively smooth. The velocity dataset collected ranges from 0 Km/h to 160 Km/h with around 1,000 seconds. Fig. 2 shows that the dataset with 200 samples per second, 20 sequence length, and physical type achieves the smallest value, in which "value converted" is the physical value of the dataset. Fig. 3 shows the results of ADS using LSTM model with the optimized parameters, in which squared error is less than 25. It is hard to detect whether the model is the original model or the victim model from the results, If we use the victim model to substitute the original model, because they have similar input/output interface and compatible performance.

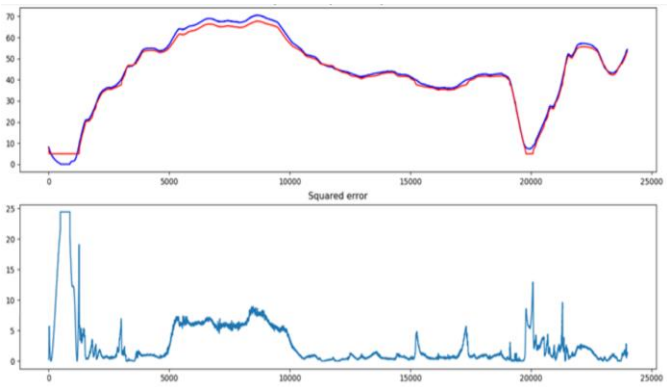

Fig -3: Original(in blue) and Predicted (in red) curves using 200 sampled per second

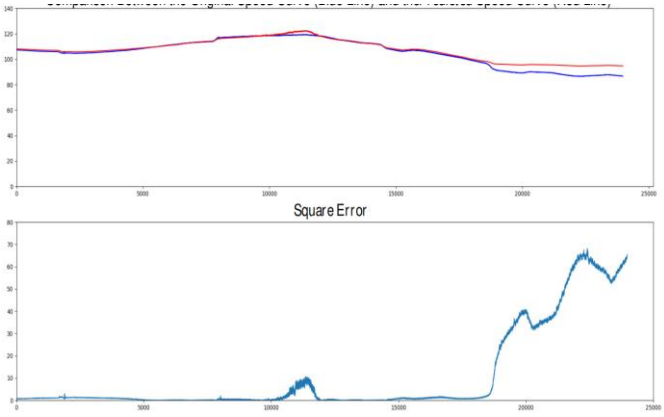

Fig -4: Comparison between original(Blue) and predicted(Red) curves of original model test on non-anomaly dataset

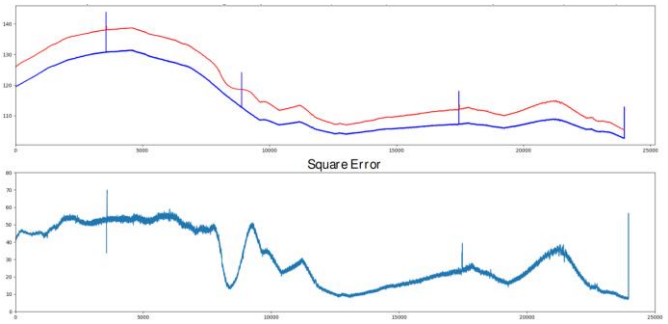

Fig -5: Comparison between original(Blue) and predicted(Red) curves of victim model test on anomaly dataset

B. Black-box attack towards LSTM based Anomaly Detection System

The approach tends to use the power of free query to train a victim model for the existing black-box attacking strategies. The attack strategy consists of training a local model to substitute for the target DNN by using inputs synthetically generated by an adversary and labels assigned by the target DNN. White-box attack techniques can be done to the victim model, and the crafted adversarial examples can be used to attack the targeted DNN. The primary advantage of training a victim model is its total transparency to the adversary, hence essential attack procedures for DNNs, such as back-propagation for gradient computation, can be implemented on the victim model to craft adversarial examples. Since the victim model is a representative for the target DNN in terms of classification rules, performing adversarial attacks to the victim model is expected to be similar to attacking the corresponding targeted DNN. In other words, an adversarial example crafted from a victim model can be highly transferable to the targeted DNN given with the ability to query the targeted DNN.

The steps of the Black-box attack are as follows:
• Hackers obtain the testing dataset with anomaly.
• Using anomaly dataset to train a similar AI model, with the same input and output with the original AI model.
• Replacing the original model with the trained the victim model by hackers.
• Output of this trained victim model will lead to wrong decisions.

The black-box attack strategy consist of training a local model to substitute for the target DNN by using inputs synthetically generated by an adversary and labels assigned by the target DNN as shown in Fig 4. we provide the result of original dataset tested on the non-anomaly dataset. It shows the result of the original dataset tested on non-anomaly dataset. 68.057 is the maximum mean square error of the original model. To show how good the model can estimate the future speed curve, square error should be as small as possible. Square values larger than 69.057

considered as anomalies. To verify the efficiency of the proposed black-box attack to our adopted ADS using LSTM model, we collected a smaller dataset with anomalies around 11.7 MB with 24000 data points. We split the dataset into a new training dataset and a new testing dataset with the ratio 3:1. The trained victim model tested on the anomaly dataset as shown in Fig. 5. 69.923 is the maximum square error for victim model, which approved that compared to original with a 68.057 square error, victim model has the similar performance.

Verified the hacking without enabling the functions of GPU and CUDA in Intel Core i54200U CPU 1.90 GHz with 4GB memory system. With the 24000 data points of 11.7MB size of anomaly data, we spend only 50 man hours to successfully mount a black-box attack on LSTM based Anomaly Detection System. We substitute the original model with victim model, which has been trained. Then, we perform tests on both models using same non-anomaly testing dataset. 89.699 and 523.428 are the maximum square error of original model and victim model respectively. It also approves that the victim model cannot detect the anomalies with square error smaller than 528.428 and bigger than 89.699, which will lead to wrong detection results.

5. CONCLUSIONS

In this paper, We are able to demonstrated efficient application of LSTM model for the detection of anomalies in-vehicle network, Also proved that no need to train the model with anomaly dataset. Based on the characteristics of the dataset we optimized the maximum square error of LSTM together with tuning the sequence length of LSTM. Trained a new victim model with the small size of testing dataset and deployed practical black-box attack to this adopted model. We can easily substitute the original model with our trained model due to the compatible input/output interface. Compared to original model final output would lead to the wrong interpretation, as shown in the experimental results. Our experiment provides the testing framework for the proposing protections from black-box attacks in future work.

REFERENCES

[1] Bosch, "CAN specification, version 2.0," 1991.

[2] T. Hoppe, S. Kiltz, and J. Dittmann.

[3] M. Wolf, A. Weimerskirch, and C. Paar, "Security in automotive bus systems," in Workshop on Embedded Security in Cars, 2004.

[4] M. Marchetti and D. Stabili, "Anomaly detection of CAN bus messages through analysis of ID sequences," in IEEE Intelligent Vehicles SYMPOSIUM. IEEE, 2017, pp. 1577–1583.

[5] S. N. Narayanan, S. Mittal, and A. Joshi, "Using data analytics to detect anomalous states in vehicles," CoRR, vol. abs/1512.08048, 2015. [Online]. Available: http://arxiv.org/abs/1512.08048.

[6] A. Theissler, "Anomaly detection in recordings from in-vehicle networks," in First International Workshop on BIG DATA APPLICATIONS AND PRINCIPLES, Spain, 2014, p. 11–25.

[7] V. Chockalingam, I. Larson, D. Lin, and S. Nofzinger, "Detecting attacks on the CAN protocol with machine learning," 2017. [Online]. Available: http://www-personal.umich.edu/~valli/assets/files/CAN\ AD.pdf

[8] N. Papernot, P. McDaniel, A. Swami, and R. Harang, "Crafting Adversarial Input Sequences for Recurrent Neural Networks," CoRR, 2016.

[9] M. Anderson, A. Bartolo, and P. Tandon, "Crafting Adversarial Attacks on Recurrent Neural Networks," CoPR, 2017. [Online]. Available: https://www.semanticscholar.org/paper/Crafting-Adversarial-Attacks-on-Recurrent-Neural-Anderson-Bartolo/539f7b09681c614bd980b25f3054a79c240d504a

[10] A. Taylor, S. Leblanc, and N. Japkowicz, "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 2016. IEEE, 2016, pp. 130–139

BIOGRAPHIES

Kiran B. Mohan received the B.Tech degree in Computer Science and Engineering from APJ Abdul Kalam Technological University, Kerala, India in 2012. He is currently pursuing M.Tech degree in Computer Science and Engineering from APJ Abdul Kalam Technological University, Kerala, India at Mount Zion College of Engineering, Kadammanitta, Kerala, India. His primary research interests are in Artificial Intelligence (Machine Learning oriented programming), Automotive and embedded system.

Nisha Mohan P.M. received the M.Tech degree in Communication and Networking from MS University, Tirunelveli, India in 2013. She is currently working as Assistant Professor in the Department of Computer science and Engineering at Mount Zion College of Engineering, Kadammanitta, Kerala, India. Her primary research interests are in Cloud Computing, Image Processing, Cyber Security and Artificial Intelligence (Machine Learning oriented programming).