Work-in-Progress: Road Context-aware Intrusion Detection System for Autonomous Cars

Tanya Srivastava, Pryanshu Arora
Birla Institute of Technology and Science, Pilani
Pilani, Jhunjhunu, Rajasthan, India 333031

Chundong Wang, Sudipta Chattopadhyay
Singapore University of Technology and Design
8 Somapah Road, Singapore 487372

***Abstract*—The necessity of intrusion detection system (IDS) is concrete for automobiles, and is particularly critical for unmanned, autonomous ones. However, limited work has been done to detect intrusions in an autonomous car while existing IDSs have limitations against strong adversaries. We hence consider the very nature of autonomous car and propose to utilize the *road context* to build a Road context-aware IDS (RAIDS). We hypothesize that given a computer-controlled car, the pattern and data of frames transmitted on the in-vehicle communication network should be relatively regular and obtainable when the car is cruising through continuous road contexts. Accordingly we design RAIDS and implement a preliminary prototype that discerns and identifies anomalous frames fabricated or suspended by adversaries. Evaluation results show that RAIDS effectively detects intrusions that are beyond the capabilities of state-of-the-art IDS.**

***Index Terms*—Road Context-Aware Intrusion Detection System, Autonomous Car, Deep Neural Network**

I. Introduction

A modern automobile needs a protocol, like the Control Area Network (CAN) bus, for the in-vehicle communications among its electrical subsystems, like the engine, steering wheel, and brake, each of which has an electronic control unit (ECU) for monitoring and controlling. Adversaries managed to compromise ECUs and suspend or fabricate frames transmitted on the CAN bus so as to cause a breakdown or traffic accident to a vehicle [1]–[4]. Meanwhile, many technology giants, startups, and academic researchers are developing self-driving, autonomous cars, which, undoubtedly, demand particular care of security and safety. The recent fatal accident of Uber's testing has alerted people to such unmanned vehicles [5]. However, limited work has been done on designing an intrusion detection system (IDS) for the in-vehicle communications of an autonomous car. Whereas, state-of-the-art IDSs even have limitations in the case of manned vehicle. Take the CIDS [2] for example. In accordance with its knowledge of the fingerprints (w.r.t. clock skews) of all ECUs, CIDS detects anomalies when an original ECU stops sending messages or an ECU belonging to adversaries injects messages. Whereas, CIDS shall be oblivious of a compromised ECU sending forged messages. If a strong adversary can manipulate an original ECU to send fake frames, CIDS will malfunction as the fingerprint of the ECU, i.e., the clock skew, seems not anomalous. In other words, such an attack model is beyond the capability of CIDS.

There is a fact that has not been considered for designing IDS in an automotive control system: all frames transmitted on the CAN bus are generated due to the decisions made by the vehicle driver and it is the *road context* that guides a driver to make those decisions. However, since different human drivers have different experiences and habits, they would perform different reactions when facing the same road context, say, a traffic light or heavy rain. It is hence impractical to design an IDS with road context for manned vehicles. *Caveat lector*: unmanned vehicles are orthogonal to manned vehicles concerning the 'driver'. In an autonomous, computer-controlled car, decisions are made by a well-trained machine learning framework upon dynamic road contexts. As a result, the decisions of an autonomous car are more regular and consistent than those of a manned vehicle, and in turn the frames transmitted on the CAN bus should follow an understandable and predictable pattern. In case of an intrusion with forged, delayed, or lost frames on continuous road contexts, a violation of the pattern shall be perceivable.

Motivated by this observation, we have developed a novel IDS for autonomous cars, namely ***R**oad context-**a**ware **IDS*** (RAIDS), to detect abnormal frames from compromised ECUs attached onto the CAN bus. We have built a preliminary prototype of RAIDS with two deep neural networks (DNNs). One DNN extracts and abstracts the road context while the other one is responsible for detecting intrusions. Experimental results with the Udacity data set [6] confirm that RAIDS outperforms an IDS without considering the road context. Though, further optimizations are needed to improve the detection accuracy and reduce the overhead of intrusion detection as RAIDS would be sitting in a real-time autonomous vehicle.

The remainder of this paper is oragnized as follows. In Section 2, we will show an overview of RAIDS, the definition of road context, and how RAIDS leverages the road context to do intrusion detection using two DNNs. In Section 3, we will briefly present the prototype we have developed for RAIDS as well as evaluation results. In Section 4, we conclude this work-in-progress paper and address the work we plan to complete in the near future.

This work was done when Tanya Srivastava and Pryanshu Arora were interns in Singapore University of Technology and Design.

Chundong Wang is the corresponding author (cd_wang@outlook.com).

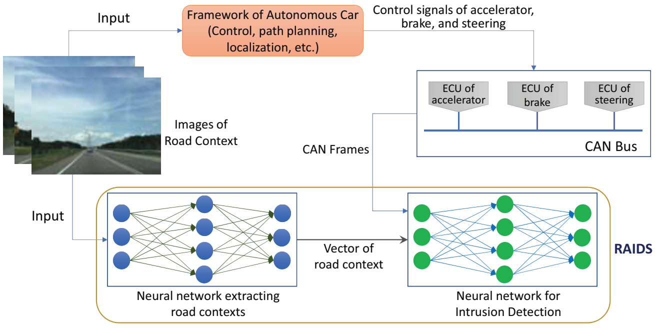
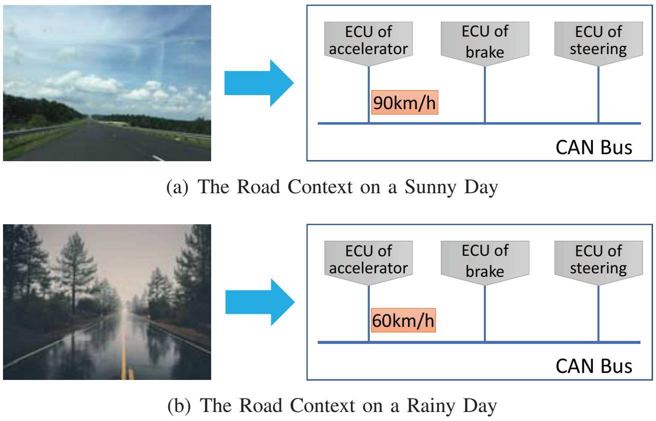

Input
Framework of Autonomous Car (Control, path planning, localization, etc.)
Control signals of accelerator, brake, and steering
ECU of accelerator
ECU of brake
ECU of steering
CAN Bus
Images of Road Context
CAN Frames
Input
Vector of road context
RAIDS
Neural network extracting road contexts
Neural network for Intrusion Detection

Fig. 1. An Illustration of RAIDS's Architecture



(a) The Road Context on a Sunny Day

(b) The Road Context on a Rainy Day

Fig. 2. An Illustration of the Impact of Road Context (Weather) on CAN Frames

II. RAIDS

Overview The essence of RAIDS is to leverage the ongoing road context, which has not been considered by previous work [2]–[4], to validate if CAN frames transmitted on the in-vehicle network are anomalous or not for an autonomous car. If so, RAIDS will report the occurrence of an intrusion; otherwise, RAIDS deems that there is no security threat.

Figure 1 illustrates the architecture of RAIDS and its interaction within an autonomous car. As shown by the upper half of Figure 1, the self-driving framework of autonomous car processes images that contain road contexts to maneuver the vehicle. In the lower half of Figure 1, images are also fed to RAIDS. RAIDS is composed of two DNNs. The first DNN is used to extract and abstract the road context from the same image that autonomous car is processing. The second DNN is responsible for intrusion detection and takes in two inputs. One input is the extracted road context. The other one is CAN frames issued by ECUs installed in the accelerator, brake, steering wheel, etc., as commanded by the self-driving framework regarding the image. Based on the two inputs as well as well-trained weights, the second DNN determines if CAN frames are genuine or not. RAIDS immediately informs the self-driving framework of autonomous car in case that abnormal frames are found, i.e., the incident of an intrusion.

Road Context We define the road context as *all the information an autonomous car is facing when it is cruising*. In other words, the road context includes but not limited to, 1) the traffic light, pedestrians, vehicles, obstacles, bumps, and pits in front of, behind, and beside the autonomous car, 2) the lane lines, cross, turn, joint, and fork of roads, 3) the weather condition, such as rain, fog, cloud, and snow, and 4) the sunrise, sunset, night and tunnel lights. The road context determines control signals made by the self-driving framework. Different road contexts shall cause different signals, which in turn entail different CAN frames transmitted on the in-vehicle network. For example, as illustrated in Figure 2(a) and Figure 2(b), a sunny day suggests that the autonomous car should run at a velocity of 90 km/h, but a rainy or foggy day demands the vehicle to be slower at about 60 km/h for the sake of safety.

Extracting Road Context To extract the road context contained in an image, we leverage a DNN to figure out the *features* of the image when making it go through multiple layers of DNN [7] [8]. We note that the self-driving framework of autonomous car proceeds further with features belonging to each image so as to generate control signals, which later would be converted to CAN frames. As a result, we can utilize the features of an image as the abstraction of road context included in that image. To obtain features of an image, we process the image via a convolutional neural network (CNN). The features generated by the CNN are abstracted in a vector. This CNN of RAIDS works in parallel with the DNN of self-driving framework but is much simpler. The reason is twofold. First, the self-driving DNN does not terminate with the features of an image but must do further computations to produce control signals for the car. Second, the feature vector yielded by RAIDS's CNN can be coarse-grained as long as it sufficiently supports detecting intrusions at the next stage of RAIDS.

Supervised Learning with Road Context for IDS The second stage of RAIDS is mainly a DNN of supervised learning, which, by referring to the learned pattern between historical road contexts and CAN frames, determines whether the frames being transmitted on the CAN bus are anomalous or not. RAIDS must ensure that the CAN frames under detection and the features extracted are resulted from the same image. Moreover, leveraging the genuine relationship learned between road contexts and CAN frames, RAIDS shall not only detect the occurrence of intrusions but can also identify the ECU(s) being compromised for the purpose of mitigation.

III. Prototyping and Testing

Prototype A preliminary prototype of RAIDS has been implemented. The CNN used for extracting road context is Resnet 18 that gives a vector of 1,000 features per image [7]. An LSTM network is employed as the second stage to examine a sequence of images against relevant CAN frames [8]. A '0' output value means no intrusion is detected while a '1' output alerts the self-driving framework of a possible intrusion. We have also implemented an IDS using a DNN without considering the road context.

Attack Model We consider an attack model that exceeds the capability of CIDS. We assume the existence of strong adversaries who can manipulate original ECUs of a car to forge harmful frames and drop normal frames. Without loss of generality, we suppose compromising ECUs of steering wheel, accelerator, and brake.

Testing With the data set provided by Udacity self-driving course [6], we have trained and tested both IDSs. Experimental results show that RAIDS manages to detect intrusions with manipulated ECUs. In particular, RAIDS attains an accuracy rate of 79 percent on average while for the IDS without road context it is about 74 percent. Though, due to the inevitable loss of road context in image processing, the current prototype of RAIDS achieves suboptimal performance. We are working on the reduction of loss in extracting road context so as to further promote RAIDS's accuracy.

IV. Conclusion and Future Work

In this paper, we present a novel approach, i.e., RAIDS, which leverages the road context to detect intrusions on the in-vehicle network of an autonomous car. RAIDS is a two-stage framework with two collaborating DNNs. Experiments with a preliminary prototype of RAIDS confirm that it is effective against strong adversaries and achieves higher accuracy than an IDS without road context. In the near future, besides targeting the aforementioned improvement of detection accuracy, we shall reduce the overhead of RAIDS so that it can be fitted in a real-world autonomous vehicle.

Acknowledgment

This work is partially supported by Singapore University of Technology and Design under the research grant SRIS17123.

References

[1] Stephen Checkoway, Damon McCoy, Brian Kantor, Danny Anderson, Hovav Shacham, Stefan Savage, Karl Koscher, Alexei Czeskis, Franziska Roesner, and Tadayoshi Kohno. Comprehensive experimental analyses of automotive attack surfaces. In *Proceedings of the 20th USENIX Conference on Security*, USENIX Security '11, pages 77–92, Berkeley, CA, USA, 2011. USENIX Association.

[2] Kyong-Tak Cho and Kang G. Shin. Fingerprinting electronic control units for vehicle intrusion detection. In *Proceedings of the 25th USENIX Security Symposium*, USENIX Security '16, pages 911–927, Austin, TX, 2016. USENIX Association.

[3] Je-Won Kang Min-Joo Kang. Intrusion detection system using deep neural network for in-vehicle network security. *PloS one*, 11(6):e0155781, 2016.

[4] Armin Wasicek, Mert D. Pes, Andre Weimerskirch, Yelizaveta Burakova, and Karan Singh. Context-aware intrusion detection in automotive control systems, June 2017. 5th Embedded Security in Cars (ESCar '17).

[5] Daisuke Wakabayashi. Self-driving uber car kills pedestrian in arizona, where robots roam, March 2018. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html.

[6] Udacity Inc. The udacity open source self-driving car project, April 2018. https://github.com/udacity/self-driving-car.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[8] Yuchi Tian, Kexin Pei, Suman Jana, and Baishakhi Ray. DeepTest: Automated testing of deep-neural-network-driven autonomous cars. In *Proceedings of the 40th International Conference on Software Engineering*, ICSE '18, pages 303–314, New York, NY, USA, 2018. ACM.