Source Identification from In-Vehicle CAN-FD Signaling: What Can We Expect?

Yucheng Liu[1] and Xiangxue Li[1,2,3(✉)]

[1] School of Software Engineering, East China Normal University, Shanghai, China
xxli@cs.ecnu.edu.cn
[2] Shanghai Key Laboratory of Trustworthy Computing, Shanghai, China
[3] Westone Cryptologic Research Center, Beijing, China

Abstract. Controller Area Network (CAN) is significantly deployed in various industrial applications (including current in-vehicle network) due to its high performance and reliability. Controller area network with flexible data rate (CAN-FD) is supposed to be the next generation of in-vehicle network to dispose of CAN limitations of data payload size and bandwidth. The paper explores for the first time Electronic Control Unit (ECU) identification on in-vehicle CAN-FD network from bus signaling and the contributions are four-fold.

- Technically, we discuss the factors that might affect ECU recognition (e.g., CAN-FD controller, CAN-FD transceiver, and voltage regulator) and look into the signal ringing and its intensity where dominant states along with rising edges (from recessive to dominant states) suffice to fingerprint the ECUs. We can thereby design ECU identification scheme on in-vehicle CAN-FD network.
- For a given network topology (in terms of the stub length and the number of ECUs), we execute CAN-FD and CAN separately and one can expect considerable performance for the two kinds of protocols by using any signal characteristics (rising edges, dominant states, falling edges, and recessive states). In particular, the recognition rates by dominant states and rising edges of signals outperform significantly those by any other combinations of signal characteristics.
- As a respond to the possible transition mechanism from CAN to CAN-FD, we also allow a hybrid topology of CAN and CAN-FD, namely, there exist on the network ECUs sending purely CAN frames, ECUs sending purely CAN-FD frames, and ECUs sending both CAN and CAN-FD frames, and our suggestion on dominant states and rising edges shows robustness to source identification as expected. This shows convincing evidence on the universal applicability of our approach to forthcoming real vehicles set up by CAN-FD network.
- The proposed approach can be easily extended to intrusion detection against attacks not only initiated by external devices but also internal devices.

D. Gao et al. (Eds.): ICICS 2021, LNCS 12918, pp. 204–223, 2021.
https://doi.org/10.1007/978-3-030-86890-1_12

We hope our results could be used as a step forward and a guidance on securing the commercialization and batch production of in-vehicle CAN-FD network in the near future.

Keywords: Controller Area Network · CAN-FD · ECU identification

1 Introduction

Controller area network (CAN) is one of the most commonly used bus communication protocols between in-vehicle Electronic Control Units (ECU, similar to ordinary computer, consists of a microcontroller (MCU), some memory (ROM/RAM), input/output interface (I/O), analog-to-digital converter (A/D), and large-scale integrated circuits such as shaping and driving). It was introduced by Robert Bosch GmBH in 1983. All ECUs inside the vehicles are connected to each other through CAN bus. However, CAN protocol lacks security mechanisms, such as authentication and encryption [5]. Indeed, an adversary might easily eavesdrop on the bus, obtain all communication messages between ECUs at will, and then initiate a replay attack [23]. He can even modify the obtained messages which will be further injected into the CAN network in an attempt to control some safety-critical functions. We do see various attacks against CAN network recently [2,10,14,19,20]. In response to these attacks, researchers propose a series of countermeasures, represented by Intrusion Detection System (IDS) and Message Authentication Code (MAC). The latter is not practical however, as the length of the CAN frame data field is up to 8 bytes. And an alternative method is to use truncated MACs [22,25]. This method needs to constantly update the key, which will take up more computing resources. What's more, frequent key updates may cause malfunctions when the vehicle is moving. Fortunately, some seminal works [5,15,21] can not only detect the presence of malicious frames but also identify their sender ECUs. This is really essential for fast forensic, isolation, security patch, etc.

Robert Bosch GmBH recommends CAN-FD (CAN with flexible data) [7] in 2012 to meet the requirements of modern vehicles and dispose of CAN limitations of data payload size and bandwidth. Besides compatibility with CAN, CAN-FD has the following advantages: the maximum length of the CAN-FD data field is 64 bytes; it supports variable rates (namely, a frame can use different transmission rates in different stages) and the maximum rate can reach 5Mbit/s (the maximum rate of CAN is 1Mbit/s); it can refine the load of the existing bus and increase the number of the nodes[1] on the bus.

Unfortunately, CAN-FD itself does not convey security protection either (similar to CAN) and existing attacks on CAN might also be feasible on CAN-FD. Take masquerade attack on CAN network [3] as an example. Initiating a masquerade attack and not being detected by the system, an adversary needs to stop the transmission of targeted ECU and imitate it to inject attack messages.

[1] As a slight abuse of terms, we use hereafter *node* and *ECU* indiscriminately.

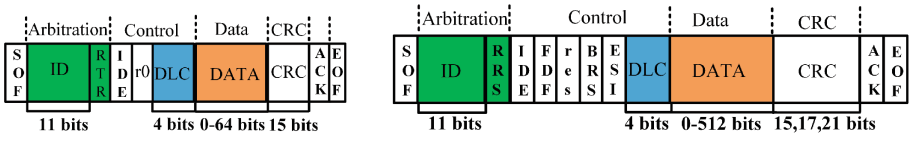
The attack also works on in-vehicle CAN-FD network. Although an intrusion detection system based on topology verification is proposed [26] to detect attacks by using external intruding devices, it can neither detect masquerade attack nor identify the sender of the attack messages. Our proposed mechanism explores for the first time Electronic Control Unit (ECU) identification on in-vehicle CAN-FD network from bus signaling.

2 Background and Related Work

2.1 Controller Area Network

CAN uses differential signals to transmit messages. Namely, the two signals on CAN-H and CAN-L have equal amplitudes relative to 2.5 V (common mode voltage) and opposite polarities. Compared with single-ended signals, differential signals are subtracted less electromagnetic interference [15]. When the ECU sends the recessive bit (1), the voltage on CAN-H and CAN-L is about 2.5 V, so the differential voltage generated is 0 V. For the dominant bit (0), the voltages on CAN-H and CAN-L are 3.5 V and 1.5 V, respectively, and the resulting differential voltage is 2 V.

The nodes inside the CAN network communicate with each other via CAN frames. CAN frames are divided into standard frames and extended frames according to whether they contain extended identifiers. The length of the identifier of the CAN standard frame is 11 bits, and 29 bits for extended frame (including 11 bits identifier and 18 bits extended identifier). At present, most vehicles use CAN standard frames. The composition of standard frames is shown in Fig. 1(a).



(a) CAN data frame format (b) CAN-FD data frame format

Fig. 1. CAN/CAN-FD data standard frame format with 11 bit identifier.

CAN is a multi-master control bus and the bus conflicts will occur if two or more ECUs request to send data at the same time. CAN bus can detect and arbitrate these conflicts in real time by CSMA/CD [1] (Carrier Sense Multiple Access/Collision Detection) arbitration method, which supports a lossless bitwise arbitration decision process. For example, if one ECU transmits a dominant bit (0) and another ECU transmits a recessive bit (1), then there is a collision and the ECU transmitting the dominant bit gets priority.

2.2 Comparing CAN-FD with CAN

CAN-FD and CAN differ in the format and the length of the data frame. Compared with CAN frame, CAN-FD adds FDF (Flexible Data Rate Format), BRS (Bit Rate Switch) and ESI (Error State Indicator) fields (see Fig. 1(b)) [7]. Therein, FDF indicates whether the sent frame is a CAN frame or a CAN-FD frame and BRS stands for bit rate conversion. When the bit is a recessive bit (1), the rate is variable, and when the bit is a dominant bit (0), it is transmitted at a constant rate. ESI is an error status indicator: when ESI is a recessive bit (1), it means that the sending node is in a passive error, otherwise it is in an active error state. In addition, according to the role of different bits, CAN specification divides a frame into different fields, as shown in Fig. 1(b). And in the experiment, we set the rate of 2Mbit/s for the data field, and set the rate of 1Mbit/s in the arbitration field, control field and CRC field. The length of the CAN-FD data field is up to 64 bytes, which increases the available load.

Next we say the data rate. The maximum rate of CAN's arbitration field and data field is no more than 1Mbit/s [6]. However, CAN-FD supports variable rates, and the bit rate of its arbitration field and data field might be different. Among them, the arbitration and the ACK stages continue to use CAN2.0 specification (i.e., the highest rate does not exceed 1Mbit/s), and the data field can reach 5Mbit/s through hardware setting, or even higher.

CAN-FD is defined to be compatible with CAN at the physical layer. All CAN-FD controllers can handle a mix of CAN frames and CAN-FD frames. One might use CAN-FD controllers in conjunction with CAN controllers on in-vehicle network. Thus one might see pure CAN frames or both CAN and CAN-FD frames on the bus.

2.3 Related Work

Generally, we have intrusion detection systems (IDS)[2] and cryptographic solutions to strengthen in-vehicle CAN network security. Murvay and Groza [21] pioneered the methodology of studying the differences in CAN signals (sent by ECUs), which are significant for ECU identification. However, they only used the signals corresponding to the CAN frame's identifier field and did not account for the blended signals caused by the collisions between ECUs' simultaneous messages. The limitation was tackled in [5] where 18-bit identifier extension was used to fingerprint ECUs. As vehicles commonly conform to the standard specifications (e.g., ISO, SAE etc.), this scheme was howbeit impractical in real-world applications. Kneib and Huth [15] proposed Scission for in-depth analysis of CAN signals. In particular, Scission can not only detect intrusion messages, but also recognize which ECU sends the intrusion messages. For cryptographic solutions, Lin et al. [14] constructed message authentication code by sending additional messages, and the authors of [22] proposed to use truncated MACs.

For CAN-FD, security experts can pursue stronger security tricks via its higher transmission rates and larger loads. In [26], authors proposed an IDS for

[2] The paper focuses on signaling based IDS.

in-vehicle CAN-FD network based on topology verification. Their method uses variations of the network topology to identify intrusions initiated by external intruding devices (XIDs), but the method cannot detect attacks initiated by attackers using the vulnerability of existing ECUs in the vehicle. Woo et al. [24] proposed a security architecture for in-vehicle CAN-FD according to ISO 26262 standard. However, this method may cause GECU (gate ECU) to generate excessive load as it has to encrypt the data packets by using the targeted ECU's unique key. To relieve pressure on GECU, Agrawal et al. [1] proposed a group-based approach for the communication among different ECUs. However, their method still requires the management of a large number of keys which takes up a large amount of computing resources of the ECUs, making it beyond instant communication.

3 Signaling and Ringing

3.1 ECUs' Voltage Output Behavior

The output voltage of an ECU's regulator varies independently and differently from other ECUs' regulators, as their supply characteristics are different (e.g., different regulators' common-mode rejection ratios) [4]. Given the same power supply (i.e., a 12 V/24 V battery powering all the ECUs), one can get different output voltage of ECU regulators. On the other hand, due to the differences in the internal resistance of the CAN transceiver, the dominant voltages of CAN-H and CAN-L will be different when the dominant bit is sent. When transmitting the recessive bit, both the high and low side transistors are switched off (inside the CAN transceiver) and thus the voltages on CAN-H and CAN-L are basically the same. So the dominant voltage can be used for fingerprint ECU. Similarly, for CAN-FD, the internal components of an ECU mainly include CAN-FD controller, CAN-FD transceiver, and voltage regulator and we have the same rationale of the dominant voltages of (CAN-FD)-H and (CAN-FD)-L on the bus.

3.2 Ringing and Its Intensity

The impedance mismatch occurs at two points on CAN-FD bus [8,11,13], e.g., one at the junction and another at the front of the non-terminal ECUs. The non-terminal ECU causes positive reflection since its impedance can be up to several tens of kΩ, significantly larger than the stub line's characteristic impedance. Conversely, the junction's impedance is lower than the stub line's characteristic impedance, resulting in negative reflection.

From Dominant to Recessive States. Let n denote the number of ECUs connected to the junction through stub lines and ECU1 a transmitter whose signal voltage need reduce by ΔV to transfer from dominant state to recessive state. Since the dominant state's differential voltage value is approximately 2V,

ΔV has a negative polarity. As seen from Fig. 2, a total of $(n+2)$ lines are connected to the junction (i.e.,. n connected stub lines and the two main bus lines). The signal transmitted from ECU1 to the junction follows $(n+1)$ lines in parallel. Thus, the stub lines have the same impedance $\frac{Z_R}{n+1}$, where the Z_R's nominal value is 120Ω. The reflectance (Γ_d) and transmittance (T_d) at the junction are calculated as:

$$\Gamma_d = \frac{\frac{Z_R}{n+1} - Z_R}{\frac{Z_R}{n+1} + Z_R} = -\frac{n}{n+2}, \; T_d = 1 + \Gamma_d = \frac{2}{n+2}. \tag{1}$$

Since Γ_d has a negative polarity, a larger portion of the incident signal is reflected as n increases, and its small part is delivered into other ECUs.

Denote Z_{diff} as ECU1's differential input impedance. Now, we have ECU1's front reflectance and transmittance (i.e.,. Γ_s and T_s):

$$\Gamma_s = \frac{Z_{diff} - Z_R}{Z_{diff} + Z_R}, \qquad T_s = 1 + \Gamma_s = \frac{2Z_{diff}}{Z_{diff} + Z_R}. \tag{2}$$

When the signal is at the recessive state, Z_{diff} is much larger than Z_R. Consequently, Γ_s has a positive polarity, and equals approximately one. Thus, ECU1's front end reflection direction is the same as the incident signal's direction, and the incident signal and reflected signals' superposition is about twice the original incident signal.

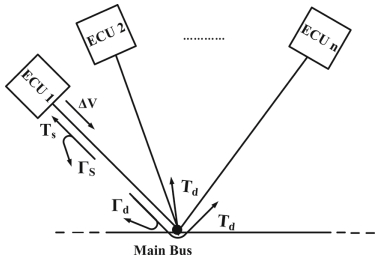


Fig. 2. Reflection and transmission coefficients at junction and non-terminal ECUs.

For a dominant-to-recessive transition, the negative transition signal ΔV is transmitted from ECU1 to the junction, undergoing partial transmission and reflection. The signals are transmitted to other ECUs through the junction and are partially reflected on the other ECUs' front end without changing the direction. At the ECU1's front, the signal returned from the connection is partially transmitted to ECU1. These reflections and transmissions are repeated, resulting in ringing.

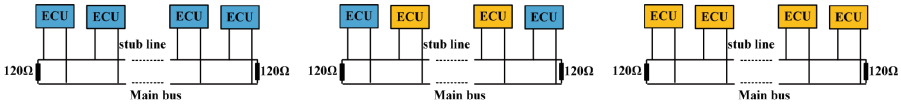
From Recessive to Dominant States. In the transitions from recessive state to dominant state, ECU1's output impedance is very low. In the recessive state,

the electrical energy is released on the network. However, when the signal transfers from recessive to dominant states, ECU1's differential output impedance becomes lowers and starts charging the network. ECU1 generates the signal of 2V, whose polarity is inverted at the junction and reflected onto ECU1. Unlike the dominant-to-recessive transition, the reflection signal is partly received at ECU1 due to the low impedance of ECU1. Since there are no reflections' repetitions, we have small ringing at the recessive-to-dominant state transition.

4 System Model

CAN-FD is designed to transmit large amounts of data at a faster rate and to replace CAN in future design. It has the potential to advance the current state of self-driving automobiles and add additional safety and comfort features in non-automobiles vehicles. As a respond to the possible transition mechanism from CAN to CAN-FD, we allow a hybrid topology of CAN and CAN-FD, namely, there exist on the network ECUs sending purely CAN frames, ECUs sending purely CAN-FD frames, and ECUs sending both CAN and CAN-FD frames.

As shown in Fig. 3, the network consists of two or more CAN nodes, two termination resistors, and bus lines connecting them. A twisted-wire-pair is commonly used for the bus line and its characteristic impedance is defined as **R**. The longest bus line (main bus) is terminated with the termination resistors **R** at both ends for impedance match. CAN nodes are connected to main bus through stub lines. In Fig. 3(a), the ECUs connected to the CAN-FD bus can send both CAN-FD and CAN frames. In Fig. 3(b), blue nodes represent the ECUs that can send both CAN-FD frames and CAN frames, and yellow nodes only send CAN frames. In Fig. 3(c), the ECUs connected to the CAN bus only send CAN frames.



(a) CAN-FD network (b) CAN/CAN-FD hybrid (c) CAN network

Fig. 3. Network topology.

4.1 Threat Models

Without security protection mechanism, the in-vehicle network is vulnerable to various attacks. For example, the bus-off attack [2] can disconnect ECUs from the bus, and the masquerade attack [3] can imitate normal ECUs to inject attack messages. Since one can not determine the sender of any messages, the attacker might use related identifier to impersonate some ECU. This will seriously threaten passengers' safety. We consider two attack modes on in-vehicle CAN and CAN-FD network.

In the attack mode–known ECU, an attacker exploits the vulnerability of existing ECUs inside the vehicle. We mention that modern vehicles generally support wireless connections, such as WiFi, Bluetooth or cellular. Via these interfaces, the attackers might compromise ECUs to achieve various attacks [19, 20]. This type of attacks seems easy to implement and widespread in life (and detailed guidance could even be found freely from some online sites), and our system can detect such attacks accurately and efficiently.

In the second attack mode–unknown ECU, an attacker plugs some extra external device into the bus to send malicious messages. E.g., the device may directly access the bus through the On-Board Diagnostics (OBD)-II port[3].

4.2 Signal Acquisition and Preprocessing

To obtain the differential signal from CAN-FD/CAN bus prototypes, we first link the two probes of an oscilloscope to the (CAN-FD)-H/CAN-H and (CAN-FD)-L/CAN-L lines respectively. Then we use the *difference* function in the software of the oscilloscope to calculate the differential signal (CAN-FD)-H) - (CAN-FD)-L) (or (CAN-H)-(CAN-L)).

Several preprocessing steps are applied to each CAN-FD/CAN signal captured by the oscilloscope. First, all dominant states are extracted from the signals. We set a voltage threshold (=0.9V) and voltage greater than the threshold marks the start of the dominant state. The dominant states are then classified into five sets (denoted as L_1, L_2, L_3, L_4, and L_5) based on the number of contained bits. Let L_i represent all dominant states containing exactly i bits (see Fig. 4). Note that CAN-FD/CAN standards specify that a recessive bit is automatically inserted whenever five consecutive dominant bits appear in a CAN-FD/CAN signal. Thus, no dominant state can contain more than five consecutive dominant bits.

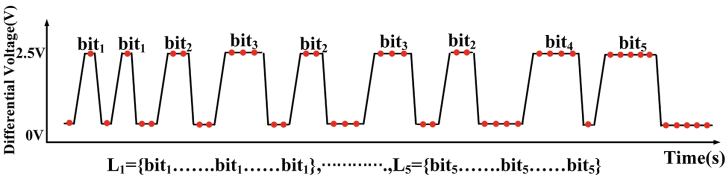


Fig. 4. A CAN-FD/CAN frame is divided into 5 sets.

4.3 Feature Extraction

We extract the statistical features from the preprocessed electrical CAN-FD/CAN signal. Due to limited computing resources of ECU, we are more

[3] The OBD-II port is near the dashboard interface, and the staff can understand the status of the vehicle in real time through the port.

interested in time domain features of the signal and avoid complex frequency domain conversion. Prior work also discerns the versatility of these features in ECU identification [5]. We extract 8 features for each set (see Table 1) and a total of 40 features for each electrical CAN-FD/CAN signal. As too many features might cause over fitting and computational cost in practice, we use the Relief-F [12] algorithm to weight these features. We thus get a general feature set (see Table 2). In the table, `order` column represents the order of the input features, and `feature` column represents the features selected by the algorithm, e.g., rms(L_5^{40}) means that rms from the set of dominant states of length 5 is selected as the first feature of the input (where 40 represents the order of this feature among all features).

Table 1. Vector x is time domain representation of the data and N its dimension.

Feature	Description
Maximum	$Max = Max(x(i))_{i=1....N}$
Minimum	$Min = Min(x(i))_{i=1....N}$
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x(i)$
Range	$R = Max - Min$
Average Deviation	$adv = \frac{1}{N}\sum_{i=1}^{N} \lvert x(i) - \mu \rvert$
Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x(i) - \mu)^2$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i) - \mu)^2}$
Root Mean Square	$rms = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x(i)^2}$

Table 2. Selected features for classification ordered by their rank

Order	Feature	Order	Feature
1	rms(L_5^{40})	11	max(L_1^{1})
2	adv(L_2^{13})	12	min (L_4^{26})
3	σ^2 (L_4^{30})	13	R(L_3^{20})
4	rms(L_3^{21})	14	rms(L_4^{32})
5	mean (L_1^{3})	15	max(L_4^{25})
6	σ (L_4^{31})	16	adv(L_1^{5})
7	σ^2 (L_3^{22})	17	mean (L_2^{11})
8	σ (L_2^{15})	18	rms(L_2^{16})
9	R(L_4^{28})	19	max(L_3^{17})
10	min(L_3^{18})	20	σ (L_5^{39})

4.4 Identifying ECUs

ECU identification is a multiclass classification problem and we use supervised learning to identify the source of CAN-FD/CAN signal. In particular, logistic regression (LR) is easy to implement with very small amount of calculation, which is very important for limited computing resources of ECU. To show the robustness of our system, we also execute support vector machines (SVM) algorithm with good generalization ability.

For the training phase, we generate fingerprints from multiple CAN-FD/CAN frames of each ECU. The resulting fingerprints are then used together to train the classifiers. For the testing phase, we have two types of tests. The first is to evaluate the model obtained by the training stage (i.e., whether or not it can determine the source of newly received frames), and the second is on intrusion detection.

5 Source Identification and Intrusion Detection

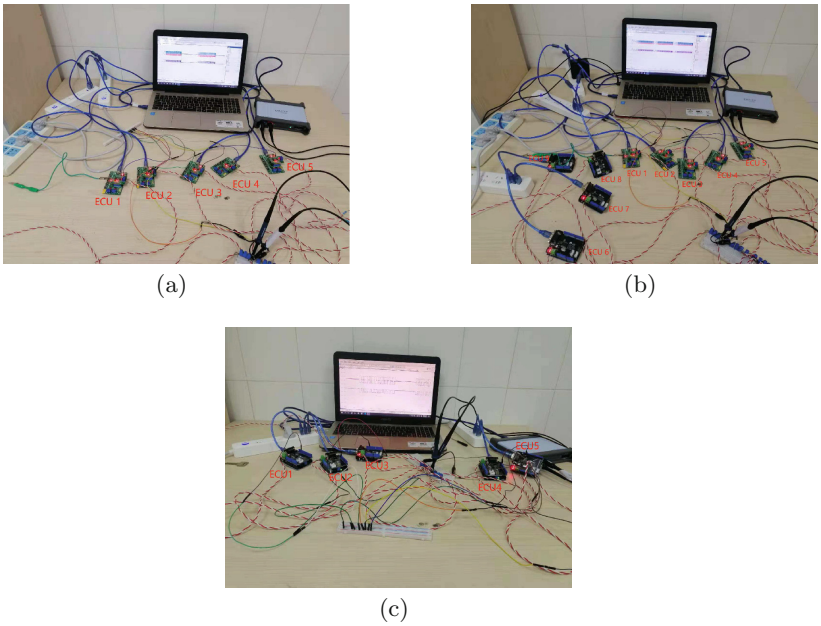
5.1 Experiment Setup

Our system adapts to different bus prototypes (we have three different network prototypes, see Fig. 5). Type A (see Fig. 5(a)) contains five CAN-FD nodes that

can send both CAN-FD and CAN frames. Type B (see Fig. 5(b)) contains five CAN-FD nodes (the same as in Type A) and four extra CAN nodes that send purely CAN frames. Type C (see Fig. 5(c)) contains five CAN nodes that send purely CAN frames. Although the total number of ECUs in real cars might be up to 70 or even larger, in-vehicle networks are physically divided into several subnets, e.g., power-related or comfort-related. As analyzed in Sect. 3, ringing mainly exists between ECUs and junctions. Thus the rationale of fingerprinting ECUs in real cars is the same as that in our experiments. CAN protocol defines low-speed CAN and high-speed CAN. Generally speaking, high-speed CAN connects the ECUs related to the important functions of the vehicles. For example, the ECU that controls the brakes and the ECU that controls acceleration are both on high-speed CAN, and the data transmission speed of high-speed CAN is 500 kbit/s. Our CAN bus prototype takes high-speed CAN network topology.

Each CAN node that sends CAN frames consists of an Arduino UNO board and a CAN shield from Seed Studio. Each CAN shield consists of an MCP2515 controller [16] and an MCP2551 transceiver [17], and the bit rate at which they send data is 500kbit/s. For CAN-FD nodes, each one consists of a STM32F105 shield and a MCP2517FD controller [18]. MCP2517FD is known as compact, cost-effective and efficient CAN-FD controller and uses SPI interface and MCU (Microcontroller Unit) communication. In the experiments, we set the bit rate of MCP2517FD as 1Mbit/s in the arbitration phase, control phase and CRC phase, and 2Mbit/s in the data transmission phase. We mention that using signal characteristics sampled at high bit rate to identify devices is more difficult than at low bit rate. If our method shows effectiveness on the high-speed CAN-FD (and CAN), it would also function well on the low-speed CAN-FD (and CAN, respectively). To maintain the consistency of experimental environments, we require that all the stub lines, oscilloscope, and other components used in the experiments are the same in all three prototypes (except the nodes of different functions).

To simulate the in-vehicle network as realistically as possible, we use twisted pair as the communication cable in all three prototypes. Each ECU is connected to main bus through two twisted pairs (CAN-H and CAN-L) (or (CAN-FD)-H and (CAN-FD)-L). All ECUs are powered by a battery which supplies electric power to each ECU via USB ports. It is required that main bus (twisted pair as well) should be longer than any other stub line on the network (our configuration sets the length of main bus as the sum of those of stub lines). There is a 120 Ω resistor at each of the two ends of main bus. CAN-FD/CAN signals are measured by the oscilloscope PicoScope 5244D MSO with a sampling rate of 25 MS/s and a resolution of 8 bits. Two probes of the oscilloscope are connected to (CAN-FD)-H/CAN-H and (CAN-FD)-L/CAN-L respectively. For each ECU (CAN-FD or CAN node), we use 200 frames as the training set (the size of the training set could be adjusted according to the performance of the model). The machine learning library Scikit and programming software Python3 are used.



(a) (b)

(c)

Fig. 5. Three prototypes of network topology: (a) Type A: CAN-FD nodes, (b) Type B: CAN-FD nodes and CAN nodes, (c): CAN nodes

5.2 Sender Identification

Sender Identification on Pure CAN. For Type C (see Fig. 5(c)), we consider the ringing effect. In particular, we execute SVM and LR by using dominant states and rising edges, recessive states and falling edges, and ((dominant states and rising edges) and (recessive states and falling edges)) respectively. The experimental results are shown in Table 3, Table 4, and Table 5. Each diagonal cell in the same matrix represents the accuracy of the two classification algorithms. As expected, dominant states and rising edges suffice to fingerprint ECUs.

Using Dominant States and Rising Edges. We then evaluate whether our system can correctly classify ECUs for Type A and Type B. Table 6 lists the confusion matrix which allows visualization of the performance of classification algorithms for 5 ECUs that send CAN-FD frames (Type A). We can see that the recognition rate of our system is sufficient to correctly recognize the ECU, and the error rate is very low. Table 7 lists the confusion matrix of 9 ECUs (Type B), of which 5 ECUs send CAN-FD frames, and the remaining 4 ECUs send CAN frames. From the result, we can see that our system can still correctly classify and recognize ECUs even in the case of a hybrid network.

Table 3. Confusion matrix using SVM/LR respectively for Type C and dominant states-rising edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	99.89/ 99.77	0/0	0/0	0.11/ 0.23	0/0
ECU 2	0/0	99.59/ 99.79	0/0	0.41/ 0.21	0/0
ECU 3	0.14/ 0.46	0/0	99.76/ 99.54	0/0	0/0
ECU 4	0/0	0/0	0.2/ 0.02	99.8/ 99.98	0/0
ECU 5	0.2/ 0.08	0/0	0/0	0/0	99.8/ 99.92

Table 4. Confusion matrix using SVM/LR respectively for Type C and recessive states-falling edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	86.52/ 84.66	0/0	5.23/ 6.01	8.25/ 9.33	0/0
ECU 2	0/0	88.21/ 87.11	6.47/ 7.56	0/0	5.32/ 5.33
ECU 3	14.34/ 11.46	0/0	85.66/ 88.54	0/0	0/0
ECU 4	0/0	0/0	15.12/ 14.62	84.88/ 85.38	0/0
ECU 5	4.32/ 5.01	0/0	4.66/ 3.84	5.17/ 6.23	85.85/ 84.92

Table 5. Confusion matrix using SVM/LR respectively for Type C and (dominant states and rising edges)-(recessive states and falling edges).

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	96.12/ 95.34	1.81/ 2.56	0/0	2.07/ 2.1	0/0
ECU 2	4.79/ 5.03	95.21/ 94.97	0/0	0/0	0/0
ECU 3	5.44/ 4.16	0/0	94.56/ 95.84	0/0	0/0
ECU 4	0/0	0/0	4.12/ 5.02	95.88/ 94.98	0/0
ECU 5	2.81/ 2.9	0/0	2.34/ 2.18	0/0	94.85/ 94.92

Table 6. Confusion matrix using SVM/LR respectively for Type A and dominant states-rising edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	99.12/ 99.34	0/0	0/0	0.88/ 0.66	0/0
ECU 2	0/0	99.21/ 99	0/0	0/0	0.79/1
ECU 3	0.24/ 0.46	0/0	99.76/ 99.54	0/0	0/0
ECU 4	0/0	0/0	0.12/ 0.02	99.88/ 99.98	0/0
ECU 5	0.15/ 0.08	0/0	0/0	0/0	99.85/ 99.92

Using Recessive States and Falling Edges. To argue the effectiveness of our recommendation, we also consider the recognition rate if recessive edges and falling edges are used. As depicted in Sect. 3.2, ringing intensity of falling edges of signals is higher than that of rising edges. Thus recognition rate would be affected when the falling edges are used. Table 8 shows the recognition rates 81.54˜86.21% for Type A. Due to space limitation, we write in the Appendix A (Table 12) the confusion matrix using SVM/LR respectively for Type B where we can see really low recognition rates (78.01˜83.89%).

Table 7. Confusion matrix using SVM/LR respectively for Type B and dominant states-rising edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5	ECU 6	ECU 7	ECU 8	ECU 9
ECU 1	98.89/ 99.15	0/0	0/0	0/0	0.91/0.7	0.01/0.03	0/0	0/0	0.19/0.12
ECU 2	0/0	98.01/ 99.21	0/0	1.2/0.78	0/0	0/0	0.79/ 0.01	0/0	0/0
ECU 3	0/0	0/0	98.99/ 99.01	0.92/0.89	0/0	0/0	0/0	0/0	0.09/0.1
ECU 4	0/0	0/0	0/0	99.29/ 99.11	0/0	0/0	0.7/0.89	0.01/0	0/0
ECU 5	0/0	0/0	0/0	0/0	98.99/ 99.31	0/0	0/0	0/0	1.01/0.69
ECU 6	1.01/0.9	0/0	0/0	0.01/0.1	0/0	98.98/ 99	0/0	0/0	0/0
ECU 7	1.32/0.98	0/0	0/0	0/0	0.01/0.01	0/0	98.67/ 99.01	0/0	0/0
ECU 8	0/0	0/0	0.9/ 0.96	0.01/0.03	0/0	0/0	0/0	99.09/ 99.01	0/0
ECU 9	1.11/1.8	0/0	0/0	0/0.03	0/0	0/0	0/0	0/0	98.89/ 98.17

Table 8. Confusion matrix using SVM/LR respectively for Type A and recessive states-falling edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	84.12/ 85.34	12/ 13.14	0/0	3.88/ 1.52	0/0
ECU 2	0/0	86.21/ 85	11.79/ 12.78	2/ 2.22	0/0
ECU 3	5.14/ 6.46	4.12/ 4.36	82.76/ 81.54	3.51/ 3.96	4.47/ 3.68
ECU 4	0/0	15.82/ 16.62	0/0	84.18/ 83.38	0/0
ECU 5	0/0	12.32/ 12.01	2.93/ 3.17	0/0	84.75/ 84.82

Table 9. Confusion matrix using SVM/LR respectively for Type A and (dominant states and rising edge)-(falling edges and recessive states).

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5
ECU 1	94.32/ 95.24	3.36/ 3.14	0/0	0/ 0	2.32/ 1.62
ECU 2	0/0	93.21/ 94.21	5.78/ 5.01	0/0	1.01/ 0.78
ECU 3	5.14/ 1.46	0/0	93.76/ 94.54	1.1/ 0.45	0/0
ECU 4	0/0	5.2/ 6.33	0/ 0.09	94.8/ 93.58	0/0
ECU 5	5.05/ 5.15	0.2/ 0.23	0/0	0/0	94.75/ 94.62

Using (Dominant States and Rising Edges) and (Recessive States and Falling Edges). We also compare the execution rates when the system uses (dominant states and rising edges) and (Recessive States and falling Edges). Table 9 and Table 10 show the results of Type A and Type B respectively, both lower than that using dominant states and rising edges.

Table 10. Confusion matrix using SVM/LR respectively for Type B and (dominant states and rising edges)-(recessive states and falling edges).

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5	ECU 6	ECU 7	ECU 8	ECU 9
ECU 1	93.89/ 94.15	5.98/ 5.51	0.13/ 0.34	0/0	0/0	0/0	0/0	0/0	0/0
ECU 2	0/0	92.01/ 93.21	0/0	0/0	6.01/ 5.89	0/0	0/0	1.98/ 0.9	0/0
ECU 3	0/0	0/0	94.01/ 93.1	0/0	5.53/ 6.01	0/0	0/0	0.46/ 0.89	0/0
ECU 4	3.9/4.01	0/0	0/0	95.29/ 95.11	0.81/ 0.88	0/0	0/0	0/0	0/0
ECU 5	0/0	0/0	5.8/6.91	0/0	93.99/ 92.31	0/0	0/0	0/0	0.21/ 0.78
ECU 6	6.01/6.4	0/0	0/0	2.1/1.5	0/0.01	91.98/ 92.09	0/0	0/0	0/0
ECU 7	5.32/4.91	0/0	0/0	0/0	1.08/ 1.01	0/0	93.67/ 94.01	0/0	0/0
ECU 8	0/0	0/0	0.9/ 0.2	0.01/0.03	5.9/6.86	0/0	0/0	93.09/ 92.01	1.01/ 1.82
ECU 9	1.11/1.8	0/0	0/0	1.01/ 0.03	0/0	0/0	0/0	5.1/6.01	93.89/ 92.17

5.3 Detecting Known/Unknown ECUs

The proposed ECU identification scheme is readily extended to intrusion detection system on in-vehicle CAN-FD network and the resulting IDS can not only detect attacks initiated by external devices but also internal devices. The recognition rate can be up to 99%. Due to space limitation, we write the evaluation in the Appendix B and C.

In practice, one can deploy the offline trained models on dedicated ECU which is inserted to the bus. Main function of the exact ECU is to monitor the traffic silently and detect anomaly. In the ECU, a digital signal processor (DSP chip, a microprocessor especially suitable for digital signal processing operations) can be integrated to establish the function of an oscilloscope: collect signals in real time and pass them to the model for detection.

6 Discussions

Sample Rate. We duplicate the experiments at various sample rates to inspect our system's effectiveness, especially for Type B. Note that at different sample rate one will be at different position of sample sizes (which might convey tight relationship with system performance). Fortunately, our approach manifests robustness as expected (due to the contribution of rising edges and dominant states). Table 11 shows the average identification and false positive rates at the sample rates 10˜25MS/s (1000 frames for each ECU).

Comparable Performance Between Type A and Type C. For a given network topology (in terms of the stub length and the number of ECUs), one may note considerable performance for Type A (CAN-FD) and Type C (CAN) by

Table 11. LR Performance at various sample rates.

Sample rate(MS/s)	10	15	20	25
Identification rate	97.11	98.95	99.01	99.15
False positive rate	2.89	1.05	0.99	0.85

using any signal characteristics (rising edges, dominant states, falling edges, and recessive states). In fact, Type C could obtain generally a tiny little better recognition rate than Type A. On the one hand, CAN-FD supports data size up to 512 bits, a drastically larger number than that (64 bits) in CAN specification, thus the cumulative effect of ringing for Type A might be more powerful than for Type C. On the other hand, CAN-FD provides variable transmission rate and our experiments specify the bit rate 2 Mbit/s for the data field of CAN-FD frames and 1Mbit/s for other fields (e.g., arbitration field, control field and CRC field), whereas CAN frame (Type C) regulates the rate of 500kbit/s. Namely, our experiments have the bit width 2000ns in a CAN frame, and 1000ns in the non-data field of and 500ns in the data field of a CAN-FD frame. Now, it is more likely for Type A (than Type C) that ringing of recessive states functions unceasing (even though the bit itself was already completed on the network)[4] and thus involves the coming dominant states before it attenuates to be unnoticeable.

Applicability to CAN-FD Network in Real Vehicles. The controllers used in our evaluation conform to ISO11898-1:2015 and support CAN-FD [18]. We also take into account the possible transition mechanism from CAN to CAN-FD (i.e., Type A and Type B). Our results show expressive evidence on the universal applicability of our approach to forthcoming real vehicles set up by CAN-FD network. We do hope our results could be used as a step forward and a guidance on securing the commercialization and batch production of in-vehicle CAN-FD network in the near future.

Environmental Factors. The electrical characteristics of CAN signals may remain unchanged for several months [21]. However, in actual vehicles, changes in the internal temperature of the vehicle will affect the characteristics of electrical signals. A typical example is that the voltage output may deviate from 0.012 V to 0.026 V [15] when we start the vehicle from a cooled turn-off engine to a warmed-up engine. This situation may also exist for the CAN-FD network. Howbeit, the length of CAN FD frame is greater than 512 bits, and the number of dominant states contained would be much likely greater than that in CAN frame. We might thus expect an acceptable impact of temperature changes on signal characteristics (and further on the system). Precise assessment is left as one of the future work.

[4] It is already reported [8,9] that for CAN-FD protocol, high-speed data phase and low-speed arbitration phase challenge the same ringing surrounds (as ringing does not depend on the transmission rate), and ring of some recessive bit might not converge until criterion and interfere with the next dominant bit.

Battery/ECU Aging. Generally speaking, the service life of car battery is of 3 to 5 years and its real usage duration is also related to the driver's driving habits. Battery aging might affect the characteristics of the electrical signals. The same problem exists on CAN network. For now, however, we can not track the impact of battery aging on the system by simulating CAN-FD nodes and car battery as there is no CAN-FD vehicle for real driving. We hope we can explore the interesting topic in the coming future. On the other hand, ECU has a relatively long service life and the aging process is really slow. It is thus rational not to consider the impact of ECU aging on electrical signals.

Limitation of the Model. Our method can detect compromised ECUs by monitoring CAN bus. It can determine whether particular frame on the bus originates from some ECU that is allowed to commit the corresponding identifier. If not, the system will issue a warning. Otherwise said, an attack will be detected once a known ECU professes some message identifier affiliated with another normal ECUs. However, if a known ECU abuses its own identifier (that is permitted under normal circumstances) to launch some attack, our system cannot recognize the attack.

Acknowledgement. The work was supported by Shanghai Municipal Education Commission (2021-01-07-00-08-E00101), the National Natural Science Foundation of China (Grant No. 61971192), and the National Cryptography Development Fund (Grant No. MMJJ20180106).

A Source Identification on Type B and Recessive States-Falling Edges

As depicted in Sect. 3.2, ringing intensity of falling edges of signals is higher than that of rising edges. Thus recognition rate would be affected when the falling edges are used. Table 12 show the results for Type B (and Table 8 for Type A) and we can see really low recognition rates.

Table 12. Confusion matrix using SVM/LR respectively for Type B and recessive states-falling edges.

	ECU 1	ECU 2	ECU 3	ECU 4	ECU 5	ECU 6	ECU 7	ECU 8	ECU 9
ECU 1	79.89/ 78.15	15.98/ 16.51	3.12/ 4.32	0/0	0/0	0/0	0/0	0/0	1.01/ 1.02
ECU 2	0/0	80.01/ 79.21	0/0	0/0	16.01/ 17.99	0/0	0/0	3.78/ 2.8	0/0
ECU 3	0/0	0/0	78.01/ 79.1	0/0	18.53/ 17.01	0/0	0/0	3.73/ 3.89	0/0
ECU 4	16.01/ 15.99	0/0	0.01/ 0.19	80.29/ 80.11	3.6/ 3.71	0/0	0/0	0/0	0/0
ECU 5	0/0	0/0	16.48/ 15.91	0/0	78.99/ 79.31	0/0	1.32/ 1.01	0/0	3.21/ 3.77
ECU 6	15.01/ 14.98	0/0	0/0	3.1/ 3.25	0.91/ 0.76	80.98/ 81.01	0/0	0/0	0/0
ECU 7	15.32/ 15.91	0/0	0/0	0/0	1.01/ 1.1	0/0	83.67/ 82.99	0/0	0/0
ECU 8	0/0	0/0	15.91/ 14.99	2.01/ 2.18	5.9/ 6.86	0/0	0/0	80.09/ 81.01	1.99/ 1.82
ECU 9	14.11/ 15.01	0/0	0/0	1.01/ 1.99	0/0	0/0	0/0	0.99/ 0.83	83.89/ 2.17

Table 13. Confusion matrix of the IDS using SVM

Support vector machines			
Prototype	True	Predicted	
		No Attack	Yes
CAN-FD	No Attack	99.38	0.62
	Yes	1.5	98.5
CAN-FD And CAN	No Attack	99.01	0.99
	Yes	1.18	98.82
CAN	No Attack	99.58	0.52
	Yes	0.99	99.01

Table 14. Confusion matrix of the IDS using LR

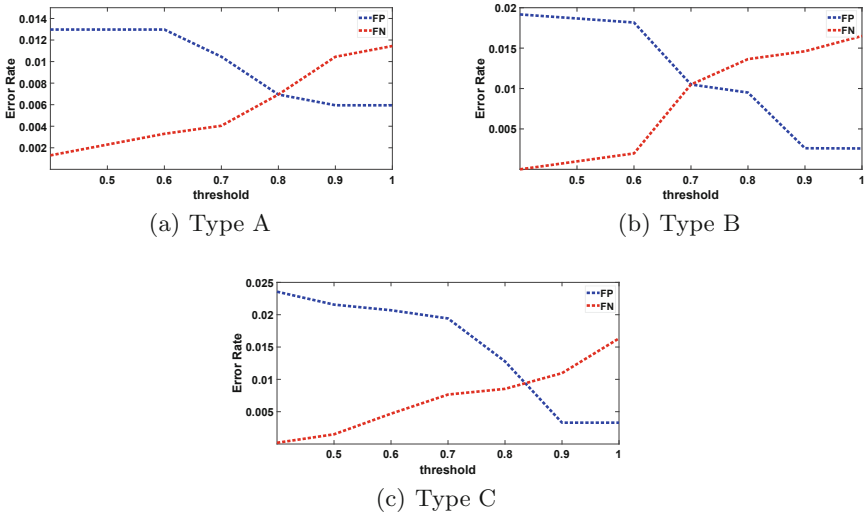
Logistic regression			
Prototype	True	Predicted	
		No Attack	Yes
CAN-FD	No Attack	99.85	0.42
	Yes	1.88	98.12
CAN-FD And CAN	No Attack	99.11	0.89
	Yes	1.89	98.11
CAN	No Attack	99.44	0.56
	Yes	0.89	99.11

B Detecting Known ECUs

For Type C (Fig. 5(c)), we assume that ECU 1 is normal and the attackers can use other ECUs to send messages with the same identifier as ECU 1. We collect a total of 500 frames, of which 300 are used as attack frames and the rest as normal frames. As shown in Table 13, we achieve a detection rate of 99.01%. For Type A (Fig. 5(a)), we use the same assumptions and operations as for Type C

and achieve a detection rate of 98.5% (see Table 13). For Type B (see Fig. 5(b)), we regard ECU 7, ECU 8 and ECU 9 as attackers (equipped with the ability of sending both CAN and CAN-FD frames). We collect 1000 frames, of which 600 are used as attack frames and the rest are normal. Table 14 shows the results with comparable performance to Type A and Type C.

C Detecting Unknown ECUs



(a) Type A

(b) Type B

(c) Type C

Fig. 6. Error rates at varying thresholds.

For unknown ECUs, we adopt a threshold-based method to extend our model. For Type A, we first remove ECU 5 and obtain about 500 frames from the remaining ECUs. These data are used to train a new model. Then we plug ECU 5 back to the network and sample a total of 600 frames now. The obtained model is used to classify the newly collected data and Fig. 6(a) shows the False Positive (FP) and False Negative (FN) rates for different threshold values. The recognition rate can be up to 99.36% at threshold = 0.8. For Type B, we remove ECU 8, use the remaining ECUs to train a new model, and then plug ECU 8 back to the network. We collect now a total of 1,000 data which will be classified by the obtained model. Figure 6(b) shows FP and FN vs threshold, and the recognition rate is 99% at threshold = 0.7. For Type C, we use similar method and Fig. 6(c) shows FP and FN vs threshold. We see the 99.1% recognition rate at threshold = 0.83.

References

1. Agrawal, M., Huang, T., Zhou, J., Chang, D.: CAN-FD-Sec: improving security of CAN-FD protocol. In: Hamid, B., Gallina, B., Shabtai, A., Elovici, Y., Garcia-Alfaro, J. (eds.) CSITS/ISSA -2018. LNCS, vol. 11552, pp. 77–93. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16874-2_6
2. Cho, K., Shin, K.G.: Error handling of in-vehicle networks makes them vulnerable. In: Proceedings of ACM CCS, pp. 1044–1055 (2016)
3. Cho, K., Shin, K.G.: Fingerprinting electronic control units for vehicle intrusion detection. In: 25th USENIX Security Symposium, pp. 911–927 (2016)
4. Cho, K., Shin, K.G.: Viden: attacker identification on in-vehicle networks. In: Proceedings of 2017 ACM CCS, pp. 1109–1123. ACM (2017)
5. Choi, W., Jo, H.J., et al.: Identifying ECUs using inimitable characteristics of signals in controller area networks. IEEE Trans. Veh. Technol. **67**(6), 4757–4770 (2018)
6. GmbH, R.B.: CAN Specifcation Version 2.0 (1991)
7. GmbH, R.B.: CAN with Flexible Data-Rate (2012)
8. H. Mori, Y.S., et al.: Novel ringing suppression circuit to increase the number of connectable ECUs in a linear passive star CAN. In: International Symposium on Electromagnetic Compatibility - EMC EUROPE, pp. 1–6 (2012)
9. Islinger, T., Mori, Y.: Ringing suppression in can fd networks. CAN Newsletter (2016)
10. Karl, K., Alexei, C., et al.: Experimental security analysis of a modern automobile. In: IEEE Symposium on Security and Privacy, pp. 447–462 (2010)
11. Kim, G., Lim, H.: Ringing suppression in a controller area network with flexible data rate using impedance switching and a limiter. IEEE Trans. Veh. Technol. **68**(11), 10679–10686 (2019)
12. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Bergadano, F., De Raedt, L. (eds.) ECML 1994. LNCS, vol. 784, pp. 171–182. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-57868-4_57
13. Lim, H., Kim, G., et al.: Quantitative analysis of ringing in a controller area network with flexible data rate for reliable physical layer designs. IEEE Trans. Veh. Technol. **68**(9), 8906–8915 (2019)
14. Lin, C., Sangiovanni-Vincentelli, A.L.: Cyber-security for the controller area network (CAN) communication protocol. In: 2012 ASE International Conference on Cyber Security, pp. 1–7. IEEE Computer Society (2012)
15. Marcel, K., Christopher, H.: Scission: signal characteristic-based sender identification and intrusion detection in automotive networks. In: Proceedings of the 2018 ACM Conference on Computer and Communications Security, pp. 787–800 (2018)
16. Microchip-Corporation: Stand-Alone CAN Controller With SPI Interface (2005)
17. Microchip-Corporation: MCP2551 High-Speed CAN Transceiver (2007)
18. Microchip-Corporation: Externa CAN FD Controller with SPI Infertface (2017)
19. Miller, C., Valasek, C.: Adventures in automotive networks and control units. Def Con **21**(260–264), 15–31 (2013)
20. Miller, C., Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015(S 91) (2015)
21. Pal-Stefan, M., Bogdan, G.: Source identification using signal characteristics in controller area networks. IEEE Signal Process. Lett. **21**(4), 395–399 (2014)
22. Schweppe, H., Roudier, Y., et al.: Car2x communication: securing the last meter-a cost-effective approach for ensuring trust in car2x applications using in-vehicle symmetric cryptography. In: 2011 IEEE VTC Fall, pp. 1–5 (2011)

23. Tobias, H., Jana, D.: Sniffing/replay attacks on can buses: A simulated attack on the electric window lift classified using an adapted cert taxonomy. In: Proceedings of the 2nd workshop on embedded systems security (WESS), pp. 1–6 (2007)
24. Woo, S., Jo, Hyo Jin, A.O.: A practical security architecture for in-vehicle CAN-FD. IEEE Trans. Intell. Transp. Syst. **17**(8), 2248–2261 (2016)
25. Woo, S., Jo, H.J., et al.: A practical wireless attack on the connected car and security protocol for in-vehicle CAN. IEEE Trans. Intell. Transp. Syst. **16**(2), 993–1006 (2015)
26. Yu, T., Wang, X.: Topology verification enabled intrusion detection for in-vehicle CAN-FD networks. IEEE Commun. Lett. **24**(1), 227–230 (2020)