In-Vehicle ECU Identification and Intrusion Detection from Electrical Signaling

Xiangxue Li, Yue Bao, and Xintian Hou

1 Introduction

Controller area network (CAN) protocol has strong anti-interference ability and can effectively suppress electromagnetic interference [1]. It relies on differential signals to transmit messages. Differential signals with dominant state (logical 0) and recessive state (logical 1) are transmitted through high (CAN-H) and low (CAN-L) lines. When the signal represents dominant state, CAN-H voltage is approximately 3.5 V, and CAN-L voltage is approximately 1.5 V, which results in a dominant differential voltage of approximately 2.0 V on CAN bus. For recessive state, both CAN-H and CAN-L voltages are approximately 2.5 V, yielding the differential voltage $\approx$0 V [2].

There are growing instances of hacking vehicles due to loose security protection of CAN protocol [3–7]. We have seen various IDSs of in-vehicle CAN networks for decades [8–13]. These suggestions cannot determine which ECU launches the particular attacks. Moreover, a smart attacker might mimic certain characteristics of the target ECU to launch an attack [9, 10, 12, 13]. Fortunately, some seminal work [1, 14–17] can not only detect malicious frames but identify their sender ECUs. The strategy counts on CAN signal unique characteristics, e.g., the hardware and

X. Li (✉)
East China Normal University and Shanghai Key Laboratory of Trustworthy Computing, Shanghai, China

Shanghai Key Laboratory of Privacy-Preserving Computation, MatrixElements Technologies, Shanghai, China
e-mail: xxli@cs.ecnu.edu.cn

Y. Bao · X. Hou
CATARC Software Testing (Tianjin) Co. Ltd, Tianjin, China
e-mail: baoyue@catarc.ac.cn; houxintian@catarc.ac.cn

V. K. Kukkala, S. Pasricha (eds.), *Machine Learning and Optimization Techniques for Automotive Cyber-Physical Systems*, https://doi.org/10.1007/978-3-031-28016-0_15

topology information (delineated by the signal's characteristics so that even if two ECUs send identical message, corresponding signals are divergent).

Signal characteristics are not only affected by vehicle power supply but also related to the hardware characteristics of the sending device itself. It is difficult for an attacker to imitate some particular device's signal characteristics. Thus CAN signals show special functionality in detecting attack messages and identifying sender ECUs. Murvay and Groza [18] 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 [15] where 18-bit identifier extension was used as the ECU's fingerprint.

One more interesting work-Sample was proposed in [17] with low time complexity and the advantage of robustness and recognition rates. Kneib and Huth [1] proposed Scission with in-depth analysis of CAN signals. Scission uses the rising and falling edges of CAN signal to design IDS. However, their method could be affected by CAN topology easily. Once the number of ECUs or the length of stub lines change, the characteristics of rising and falling edges would become different.

2 System Model and Ringing Effect

We can further look into the ringing generation mechanism and recognize the fuzzy discrepancy between transitions *from dominant to recessive state* and those *from recessive to dominant state*. Ringing intensity is related to the number of ECUs and the stub line length of CAN topology [19–21]. When we fix the number of electronic control units, longer stub line results in more intense ringing. The fluctuation of ringing intensity would further tweak falling edges' voltage, which might set off false alarms of IDS (i.e., not triggered by real attacks). We will investigate the factors that enlarge ringing effect and demonstrate the discrepancy between rising edges and falling edges. Our attempt is to design ECU identification scheme and IDS only from the characteristics of dominant states and rising edges (D.R for short).

Figure 1 shows CAN bus topology deployed widely in automotive applications. In particular, Fig. 1a presents linear topology and Fig. 1b depicts start-like topology. A twisted wire is commonly used for CAN bus, and the twisted wire's characteristic

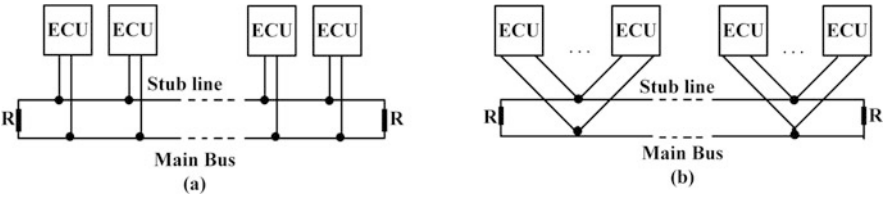

Fig. 1 CAN topology

impedance is marked as **R**. For linear topology (as recommended in the standards [22, 23]), the longest bus is called main bus. Two terminating resistors are arranged at left end and right end, and the resistor resistance is set to **R** to match the bus's characteristic impedance. In the automotive field, terminating resistors are commonly installed in the two farthest CAN nodes (called the terminal CAN nodes) to improve productivity. Other nodes are referred to as non-terminal nodes. ECUs are connected to the main bus using a twisted wire (i.e., stub lines). Stub line is attached to the main bus through a connector (indicated by the black circle in Fig. 1) called junction [21].

2.1 Threat Models

We consider two types of in-vehicle attacks: known-ECUs attack manipulates existing ECUs, and unknown-ECUs attack inserts extra devices to CAN bus.

Automotive manufacturers install ECUs during vehicle production. Attackers rely on additional interfaces to compromise a known ECU to transmit malicious CAN frames. These interfaces include WiFi, Bluetooth, and cellular communication modules. Telematics ECU [6, 24] is a prevalent example, installed widely in modern vehicles to enable supplementary functions. This kind of ECUs are connected to an external network (e.g., a cellular network), providing a target for the attackers.

Instead of exploiting existing ECU's vulnerability, an attacker can connect an unknown ECU to attack CAN network directly. Alternatively, he may plugin a special device to the network via the vehicle's On-Board Diagnostics (OBD)-II port.

2.2 Difference Between ECUs Voltage Outputs

The differences in voltage stabilizing ability of the regulator inside an ECU results in different outputs (V_{OUT}) [14], even for the same power supply (V_{IN}). ECU output voltage variations may stem from the differences in ground voltage and capacitors (denoted C_1,C_2 and C_3 in Fig. 2). Further, industrial typical 5% error tolerance is employed in CAN transceiver resistors, which leads to voltage changes.

2.3 Ringing Effect

The impedance mismatch occurs at two points over the CAN bus (Fig. 3) [20, 21], one at the junction and another at the front of non-terminal ECUs. Non-terminal ECU causes positive reflection as its impedance can be up to several tens of kΩ, significantly larger than the stub line characteristic impedance which is further larger than the junction's impedance, resulting in negative reflection.

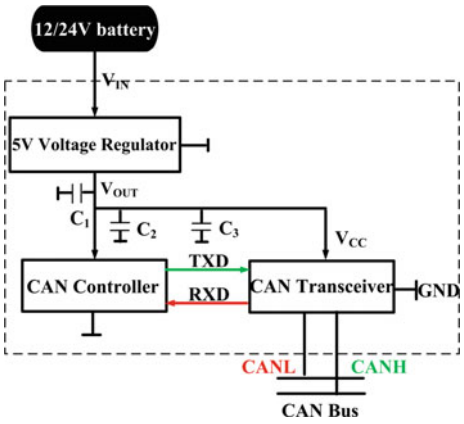

Fig. 2 CAN application schematic

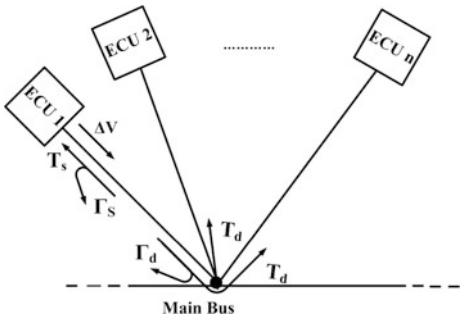

Fig. 3 Reflection

2.3.1 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 would be reduced by ΔV to transfer from dominant state to recessive state. Since the dominant state's value is approximately 2 V, ΔV has a negative polarity. In Fig. 3, a total of $(n+2)$ lines are connected to the junction (i.e., the overall number of 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:

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

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.

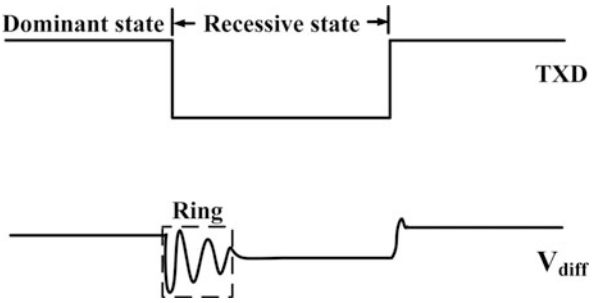

Fig. 4 Ringing for signals from dominant to recessive states

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):

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

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 front end reflection direction is the same as the incident signal direction, and the incident signal and reflected signal superposition is about twice the original incident signal.

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 (Fig. 4).

2.3.2 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 lower and starts charging the network. ECU1 generates the signal of 2 V, 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.

3 Dominant States and Rising Edges for Source Identification

3.1 Signal Measurement and Preprocessing

In order to measure the differential signal on the CAN bus, we connect two channels of an oscilloscope CAN-H and CAN-L, respectively. Each CAN frame's differential signal would be obtained based on the oscilloscope's differential function.

Several preprocessing steps are applied to each CAN signal captured by the oscilloscope. First, all dominant states are extracted from the signals. We set a voltage threshold value as 0.9 V: 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. 5). Note that CAN standard specifies that a recessive bit is automatically inserted whenever five consecutive dominant bits appear in a CAN signal. Thus, no dominant state can contain more than five consecutive dominant bits. By dividing a CAN frame into 5 sets, we have the following gains: (a) redundant features can be eliminated (the dominant states with the same number of dominant bits in a CAN frame have similar characteristics, and these dominant states with similar characteristics are in the same set); and (b) the influence of outliers might be eliminated to make the classification more accurate.

3.2 Feature Extraction

The sets obtained above are subjected to feature extraction, where the measured voltages are discrete values. Feature extraction is essential in ECU identification and needs to be time-efficient. Domain transformations should be avoided if possible. To reflect the characteristics of these discrete values, Table 1 qualifies the features that reflect the characteristics of a group of discrete values from the time domain feature quantities (x is time domain representation of data and N its dimension). Some work also discussed various features for ECU identification [15].

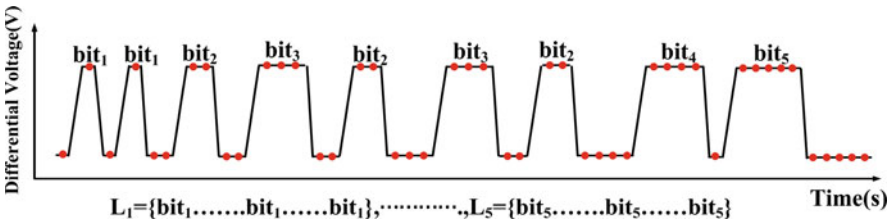

Fig. 5 A CAN frame is divided into 5 sets

Table 1 Features

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 ordered by ranks

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})

Dominant states with the same number of dominant bits indicate analogous characteristics. Thus, CAN frames are divided into five sets (Sect. 3.1). For the unlikely case of empty sets (all characteristics obtain null values), one may replace the missing values with statistical properties, e.g., mean or median. For each set, eight features (Table 1) are extracted, yielding 40 CAN signal features in total.

Relief-F [25] can evaluate the features by calculating a score for each one and selecting the most important ones. We finally opt for 20 feature subsets for each CAN frame (Table 2). The order column represents the sequence number that Relief-F sorts in descending order according to the scores of the features, and these sequence numbers correspond to the dimension of each feature in the input feature set.

3.3 Training and Testing

We view ECU identification from a received CAN frame as a classification problem and use supervised learning to identify the signals' sender. The training set comprises 200 CAN frames for each ECU. After the training phase, a classifier is created that can be used to identify the sender of a CAN frame.

Algorithm 1 ECU identification and IDS

```
1: function TRAINING(S: original CAN signal)
2:     for i=1 to len(S) do
3:         /*Divide the signal S_i into ECU_I*/
4:         ECU_I ← DECODE(S_i)
5:         /*dominant state and rising edge*/
6:         [L_1, L_2, ..., L_5] ← PREPROCESSING (S_i ∈ S)
7:         F_i ← EXTRACTION(L_1, L_2, ..., L_5)
8:         TrainingSet(i)←[F_i: ECU_I]
9:     end for
10:    Classifier←GET_TRAINING_
         ALGORITHM(TrainingSet)
11:    return Classifier
12: end function
13:
14: function TESTING(S: a new CAN signal)
15:    /*Divide the signal S into ECU_I*/
16:    ECU_I ← DECODE(S)
17:    /*dominant state and rising edge*/
18:    [L_1, L_2, ..., L_5] ← PREPROCESSING (S)
19:    F ← EXTRACTION(L_1, L_2, ..., L_5)
20:    [Result, Probability]← IDENTIFICATION
         (F, classifier)
21:    if Probability < threshold then
22:        return Unknown ECU Adversary
23:    else if Result ≠ ECU_I then
24:        return Known ECU Adversary
25:    else
26:        return Normal
27:    end if
28: end function
```

Table 3 Comparison among voltage-based approaches

	Choi et al. [15]	Scission [1]	Simple [17]	Our system
Sampling rate	2.5 GS/s	20 MS/s	50 MS/s	50 MS/s
Identification rate	96.48%	99.85%	99.10%	99.15%
False positive	3.52%	0%	0.899%	0.85%
Signal type	Differential	Differential	Differential	Differential
Domain transformations	Yes	Yes	No	No
Unknown ECU	No	Yes	Yes	Yes

The training phase results in a classifier, which is then used to predict new frames in testing phase. The testing phase includes two tasks. ECU identification tests whether the system correctly identifies frames' source and examines the impact of stub lines' length on the execution ability. Intrusion detection assesses the system's capability of detecting attacks (Sect. 2.1). The system performance on identification and intrusion detection will be discussed in Sect. 4. Algorithm 1 describes the training and testing processes. Table 3 compares some voltage-based proposals.

4 Evaluation

Four CAN bus prototypes (each containing 3/6/9/13 ECUs) are equipped to simulate different CAN networks. All prototypes have the same configuration (except the number of ECUs). Take the prototype with 3 ECUs and 1-m stub lines as example. We first assemble 3 ECUs, each containing an Arduino UNO board and a CAN shield. The shield comprises MCP2515 CAN controller [26] and MCP2551 CAN transceiver [27]. Each ECU is connected through a stub line to the main bus of length 3 m. Two 120Ω resistors are connected to the two ends of the main bus. Use an oscilloscope, one of its probes being connected to the CAN-H line of the main bus and another to CAN-L. Adjust the sampling rate of the oscilloscope to 20 MS/s.

The system can also be evaluated on real vehicles, e.g., Nissan Sentra 2016 and Subaru Outback 2011 [17].

4.1 ECU Identification

CAN signals are acquired using the digital storage in the oscilloscope PicoScope 5244D MSO with a sampling rate of 1 GS/s (the oscilloscope captures 1G data points from the signal waveform in one second) and a flexible resolution. Set the sampling rate as 20 MS/s (higher sampling rate increases data volume and hardware costs).

4.1.1 Classification Algorithms

To evaluate the influence of classification algorithms on system performance, two algorithms are employed: Linear Regression (LR) and Support Vector Machine (SVM). For each ECU in the prototype, approximately 200 frames are collected. Table 4 demonstrates that the model accuracy on a simple topology (i.e., only 3 ECUs in the entire network) averages above 99.99% irrespective of the classification algorithms. When the topology becomes complicated (e.g., 13 ECUs, Table 5), the average SVM and LR accuracies are above 98.25%. When the stub line length equals 3 m, the SVM average accuracy is 98.01%, and LR's is 98.11%. In other words, the proposed model accuracy remains high for complex network structures.

4.1.2 CAN Topology

We also explore the effect of changes in stub lines' length on the developed system performance. The system identification rate is tested for the stub line's length = 1, 2, or 3 m. Figure 6 just shows the topology for 1-m stub line.

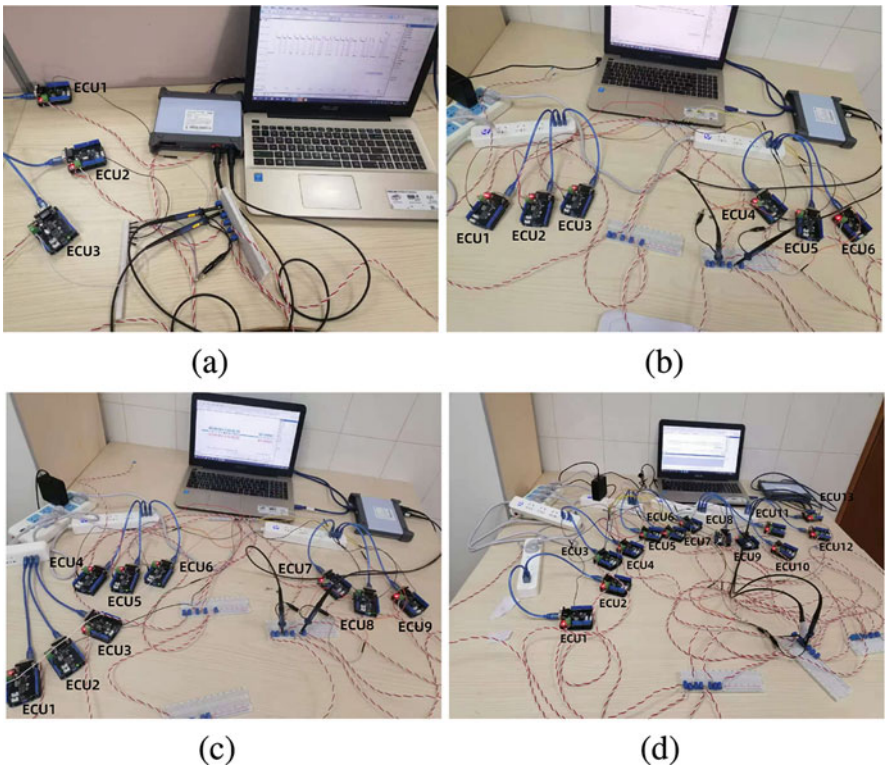

Fig. 6 Topologies of 1-m stub line and 3/6/9/13 ECUs

Same feature extraction and classification algorithms are utilized for each topology (i.e., 3/6/9/13 ECUs). Tables 4 and 5 show that if the number of ECUs is fixed, increasing stub lines' length has no impact on the system recognition accuracy. When using (falling edges and recessive states, F.R for short) and ((dominant states and rising edges) and (falling edges and recessive states), D.R.F.R for short) respectively, recognition rates decrease with the increasing of stub lines' length.

4.1.3 CAN Signal States

A series of comparative experiments are conducted to inspect the influence of CAN signal states on the system using F.R only, or D.R.F.R. As discussed above, the ringing mainly occurs in dominant-to-recessive states transitions, and the more complex the CAN bus topology, the more intense the ringing effect. Table 5 shows the results when the system uses D.R, and the average minimum recognition rates are 98.25% (for 13 ECUs and one-meter stub-line), 98.21% (for 13 ECUs and two-meter stub line), and 98.01% (for 13 ECUs and 3-meter stub line). When the system

Table 4 Recognition rate and recognition rate for 3/6 ECUs

Stub line	Algorithm	Signal state (3 ECUs)			Signal state (6 ECUs)		
		D.R	F.R	D.R.F.R	D.R	F.R	D.R.F.R
1-m	SVM Min	100	71.05	96	99.31	69.75	95.15
	SVM Avg	100	74.42	97.6	99.97	71.23	95.89
	LR Min	99.99	79.61	96	99.41	70.11	94.63
	LR Avg	99.99	80.66	98.4	99.99	72.36	95.99
2-m	SVM Min	99.35	70.09	96.21	98.89	68.11	94.21
	SVM Avg	99.98	72.32	97.32	99.85	69.87	94.32
	LR Min	99.98	79.61	96.12	99.01	69.11	94.21
	LR Avg	99.99	81.32	97.21	99.49	70.55	95.1
3-m	SVM Min	99.98	72.27	95.41	98.89	67.99	92.01
	SVM Avg	99.99	73.43	96.67	99.25	69.43	92.85
	LR Min	99.99	77.1	96.21	99.31	68.01	92.21
	LR Avg	99.99	78.29	97.32	99.55	70.53	93.32

Table 5 Recognition rate and recognition rate for 9/13 ECUs

Stub line	Algorithm	Signal state (9 ECUs)			Signal state (13 ECUs)		
		D.R	F.R	D.R.F.R	D.R	F.R	D.R.F.R
1-m	SVM Min	99.21	58.75	88.51	97.99	59.57	83.98
	SVM Avg	99.51	59.23	89.91	98.45	61.89	84.21
	LR Min	98.89	59.51	89.01	97.76	58.01	83.9
	LR Avg	99.25	60.35	90.11	98.25	62.58	84.11
2-m	SVM Min	98.8	57.35	87.11	97.51	54.25	80.99
	SVM Avg	99.01	58.25	88.26	98.21	56.75	81.21
	LR Min	98.38	57.11	88.11	97.55	54.91	79.21
	LR Avg	98.89	58.55	88.39	98.25	55.35	81.68
3-m	SVM Min	98.59	53.99	86.21	97.35	47.99	75.21
	SVM Avg	98.99	54.43	87.77	98.01	48.43	78.77
	LR Min	98.19	53.01	85.81	97.45	46.01	73.99
	LR Avg	98.71	55.53	86.34	98.11	47.53	77.34

uses F.R, the average minimum recognition rates are 61.89%, 55.35%, and 47.53%, respectively. If the system uses D.R.F.R, the average minimum recognition rates are 84.11%, 81.21%, and 77.34%. This demonstrates that using D.R is not affected by ringing and enables a higher recognition rate than other states.

4.1.4 On Real Vehicles

We can also check the method on real vehicles, Nissan Sentra 2016 and Subaru Outback 2011 and [17]. There will be 11 rounds of CAN signal, collected from these two vehicles. Table 6 shows that, for Nissan Sentra, using F.R yields the lowest

Table 6 Minimum/average recognition rate in Nissan Sentra and Subaru Outback

Vehicle	Algorithm	Signal state		
		D.R	F.R	D.R.F.R
Nissan Sentra	SVM Min	98.85	44.4	96.01
	SVM Avg	99.15	46.42	96.87
	LR Min	98.34	58.61	94.01
	LR Avg	99.08	60.20	95.41
Subaru Outback	SVM Min	98.37	62.05	91.80
	SVM Avg	99.10	63.88	93.05
	LR Min	98.05	77.23	88.52
	LR Avg	99.09	80.13	91.47

Table 7 IDS for known ECUs (support vector machine and logistic regression)

Vehicle	True	Predicted (SVM)		Predicted (LR)	
		No attack	Yes	No attack	Yes
Prototype	No attack	98.11	1.89	97.98	2.02
	Yes	2.15	97.85	2.59	97.41
Nissan Sentra	No attack	99.12	0.88	99.16	0.84
	Yes	1.89	98.11	1.79	98.21
Subaru Outback	No attack	98.99	1.01	99.01	0.99
	Yes	1.69	98.31	1.75	98.25

average accuracy of 46.42%. When D.R.F.R is used, the lowest average accuracy is 95.41%. For the Subaru outback, the lowest average accuracy equals 99.09% for D.R, 63.88% when using F.R, and 91.47% when using D.R.F.R.

4.2 Intrusion Detection

4.2.1 Known ECUs

We assume that the system has the knowledge: which identifiers are used, which ECUs are allowed to use them. If the ECU selected by the model as source is not allowed to send frames with the identifier of the received frame, an attack will be assumed. It is not allowed that multiple ECUs use same identifier.

We consider the most complex topology (i.e., 13 ECUs and 3-m stub lines) as a prototypical setup. 11 out of the 13 ECUs are seen as legitimate, and the remaining two as attackers. More than 1400 frames are collected, 500 of which are valid, and more than 900 counterfeit. Table 7 show the detection rate 97.85%.

Similar tests are conducted on real data of Nissan Sentra and Subaru Outback. The compromised ECUs are simulated on the real vehicles by adding two additional ECUs. These ECUs consist of an Arduino board and a CAN shield. Adding these ECUs differs from the situation of unknown ECU attack. Namely, unknown ECUs' electrical signals are not trained by the algorithm. In contrast, the two

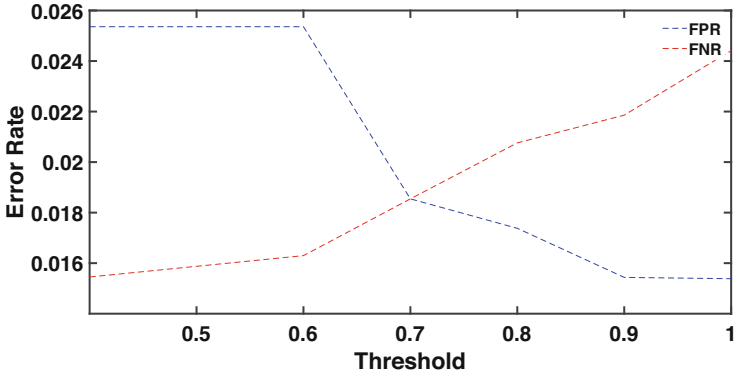

Fig. 7 FN and FP rates at varying thresholds (Subaru outback)

additional ECUs are based on the five original ECUs, and the CAN electrical signals' characteristics of the seven ECUs can be extracted and retrained. In the Nissan Sentra case, the additional ECUs imitate ECU A and ECU B, and 400 counterfeit frames are collected for each one. For Subaru Outback, the ECUs are used to fake ECU I and ECU J, and generate two sets of 400 counterfeit frames. The detection results for Nissan Sentra and Subaru Outback are shown in Table 7.

4.2.2 Unknown ECUs

Unknown ECUs' identification is related to novelty detection, i.e., the identification of new or unknown data not used in a machine learning algorithm's training [28]. We use the threshold trick (Fig. 7) and the instances with probabilities lower than the threshold are classified into unknown class.

We set the number of ECUs as 13, and the stub line's length as 3 m. To evaluate whether the system is capable of detecting unknown ECUs, the network is configured using 12 ECUs, and the 13th ECU is removed. Then monitor the resulting network and collect approximately 500 frames from each ECU for feature extraction. Now a new model can be trained (without the knowledge of the 13th ECU in the signals). Once the model training completes, ECU #13 is re-inserted to the network. Then, 3290 frames are acquired from the network with all 13 ECUs. The appropriate threshold is obtained by calculating the false positive (FP) and false negative (FN) rates (Fig. 8). The threshold 0.83 yields approximately equal values of FP and FN. And the system's identification rate is 97.89%.

For Nissan Sentra and Subaru Outback, 400 normal frames from each vehicle are selected. For Nissan Sentra, ECU M is added with the message ID {1201}, and 200 corresponding frames are collected. Overall, 600 frames are obtained from Nissan Sentra. Again, FN and FP are used to calculate the appropriate threshold (resulting in threshold value = 0.8). Figure 9 shows the results. The system achieves 98.54%

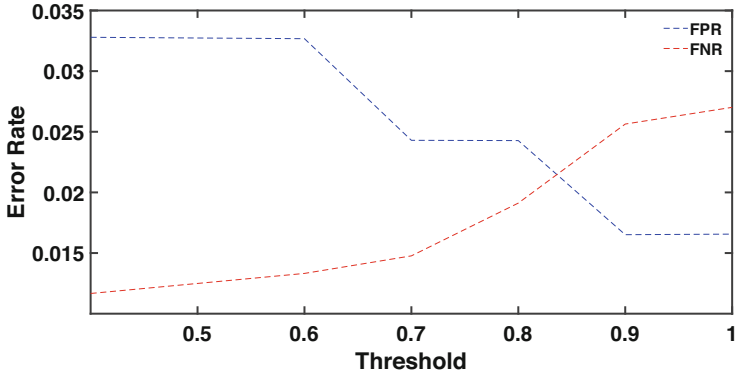

Fig. 8 FN and FP rates at varying thresholds (Prototype)

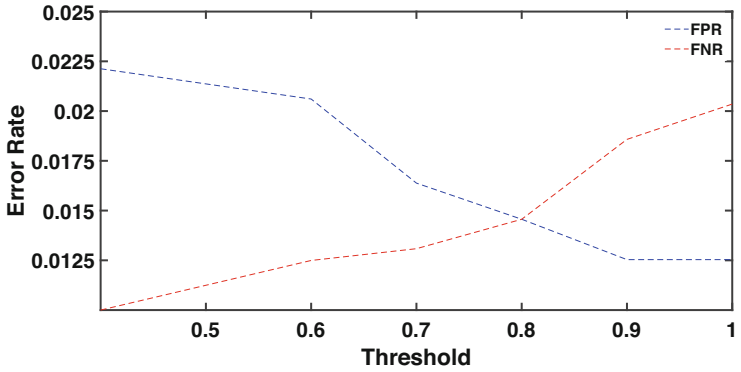

Fig. 9 FN and FP rates at varying thresholds (Nissan Sentra)

identification rate. Similarly, for Subaru Outback, ECU N sending e messages with IDs{537, 538} is inserted. Six hundred frames are collected and utilized to calculate the appropriate threshold (i.e., 0.7, see Fig. 7), and the system's accuracy is 98.15%.

4.3 Discussions

4.3.1 Environmental Factors

The voltage signal is really sensitive to environmental factors, such as temperature change. To pursue robustness against environmental factors, we adopt a method of threshold-based online model update. When the recognition rate of the model is lower than a threshold, the IDS composes an update batch with already classified fingerprints from all ECUs and thus does not require additional computing capacity.

We use the data Nissan sentra (on 02/01/2019 and 02/18/2019 [17]) to evaluate the system: select the frames sent by ECU A and ECU B from these two data sets, and then perform preprocessing and feature extraction.

We first see whether robust sender identification can be kept up over the entire training data without performing update operations. We extract approximately 2500 normal frames from the 02/01/2019 data set, the first 200 frames per ECU of the set are used for the initial training and the remaining 1000 frames of the data set for test, and this leads to the average recognition accuracy 99.31%. Then we select 1200 frames from the 02/18/2019 data set, and all the frames are classified using the already trained classifiers. The classification accuracy is 95.23%.

Next, we introduce automatic update mechanism to improve the recognition rate. The following metrics are used: recognition rate, false positive rate, false negative rate, and F-Score. Recognition rate represents the source of how many frames the model can correctly identify. False positive rate refers to the case that an unknown ECU is incorrectly classified as valid. False negative rate refers to the case that an valid ECU is classified as unknown. F-Score represents the comprehensive classification ability used to evaluate the model. We update the model online according to F-Score. When the F-Score is lower than the threshold (0.9, as demonstrated in experiments), the model will be automatically updated: the 02/01/2019 data is used to train the model and the 02/18/2019 data is used to verify the average recognition rate of the updated model. Now we manage the average recognition rate 99.12%.

4.3.2 Sample Rate

We duplicate the experiments at various sample rates to inspect system effectiveness, especially in a complex network environment (i.e., 13 ECUs and 3-m stub lines). Note that at different sample rate one will be at different position of sample sizes (which might convey tight relationship with system performance). The approach manifests robustness as expected (due to the contribution of rising edges and dominant states). Table 8 shows the average identification and false positive rates at the sample rates 2∼20 MS/s. The experiments allow each ECU to use 1000 frames.

Table 8 Performance at various sample rates for Linear Regression

Sample rate (MS/s)	2	5	10	15	20
Identification rate	97.11	97.85	98.11	98.15	98.21
False positive rate	2.89	2.15	1.89	1.85	1.79

4.3.3 Limitation and Battery/ECU Aging

The method can detect compromised ECUs by monitoring CAN bus. 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. We mention that this is an open problem in signaling-based ECU identification schemes [1, 14–16] and our focus of the work is on the connection between signal ringing and ECU identification.

Generally, the service life of car battery is of $3 \sim 5$ years and its real usage duration is also related to the driver's driving habits. Therefore, the aging of the car's battery might affect the characteristics of the electrical signal sent by each ECU, and one would see different impact level for different position of the ECU in the CAN network [29]. On the other hand, ECU has a relatively long service life and the aging process is really slow. One may thus not consider the impact of aging on electrical signals.

5 Source Identification on In-Vehicle CAN-FD Networks

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 section discusses ECU identification on CAN-FD network from bus signaling. If a model shows robustness to source identification, then we get convincing evidence on its applicability to forthcoming real vehicles set up by CAN-FD network. ECU identification can be easily extended to intrusion detection against attacks not only initiated by external devices but also internal devices.

5.1 CAN-FD

Robert Bosch GmBH recommends CAN-FD [30] to dispose of CAN limitations of data payload size and bandwidth. Besides its compatibility with CAN, CAN-FD has the advantages: the maximum length of the 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).

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 [13] 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. The attack also works on in-vehicle CAN-FD network. We should explore ECU identification on in-vehicle CAN-FD network.

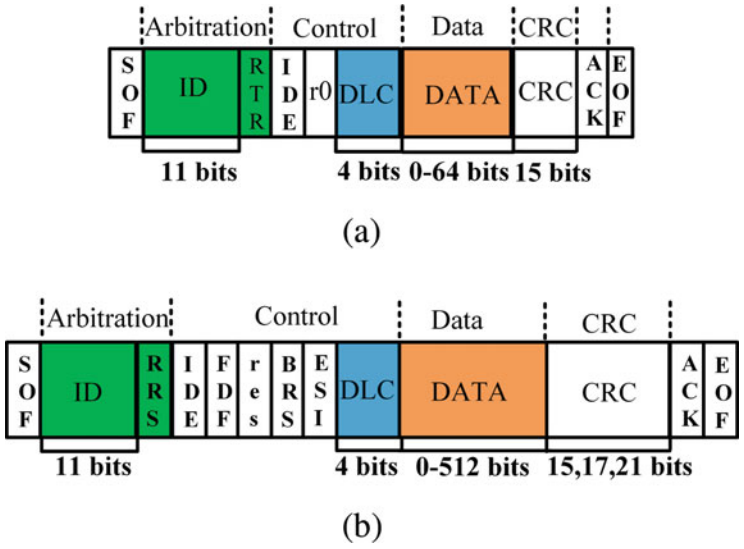

Fig. 10 CAN/CAN-FD frames with 11-bit identifier. (**a**) CAN data frame format. (**b**) CAN-FD data frame format

Comparing CAN-FD with CAN 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.

CAN-FD and CAN differ in the format and the length of the data frame (Fig. 10). Compared with CAN frame, CAN-FD adds FDF (Flexible Data Rate Format), BRS (Bit Rate Switch) and ESI (Error State Indicator) fields (see Fig. 10b) [30]. 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 active error) state. A CAN-FD frame is divided into different fields (Fig. 10b). For example, we can set the rate of 2Mbit/s for the data field and 1Mbit/s for the arbitration field, control field and CRC field. The length of the CAN-FD data field is up to 64 bytes, increasing available load.

The maximum rate of CAN arbitration field and data field is no more than 1Mbit/s [3]. However, CAN-FD supports variable rates, and the bit rate of its arbitration field and data field might be different. 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 Security For CAN-FD, security experts can pursue stronger security tricks via its higher transmission rates and larger loads. In [31], an IDS was proposed

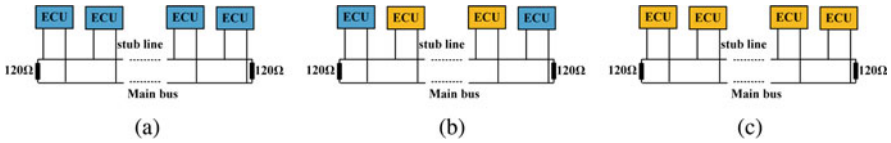

Fig. 11 Network topology. (**a**) CAN-FD network. (**b**) CAN/CAN-FD hybrid. (**c**) CAN network

for in-vehicle CAN-FD network based on topology verification. It uses variations of network topology to identify intrusions by external intruding devices (XIDs), but it cannot detect attacks via the vulnerabilities of existing ECUs. Woo et al. [32] proposed a security architecture for in-vehicle CAN-FD according to ISO 26262. This method may cause GECU (gate ECU) to generate excessive load as it has to encrypt data packets using the targeted ECU's unique key. To relieve pressure on GECU, Agrawal et al. [33] proposed a group-based approach for the communication among different ECUs. However, it should manage a large number of keys which requires a large amount of computing resources of the ECUs, making it beyond instant communication.

Ringing on CAN-FD Bus For CAN-FD, 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. As in Sect. 2.3, ringing might exist on CAN-FD bus [19, 34, 35].

5.2 System Model

CAN-FD is designed to transmit large amounts of data at a faster rate and to replace CAN in future design. For 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. In Fig. 11a, the ECUs can send both CAN-FD and CAN frames. In Fig. 11b, blue nodes represent the ECUs that can send both CAN-FD frames and CAN frames, and yellow nodes only send CAN frames. In Fig. 11c, the ECUs only send CAN frames.

Signal Acquisition and Preprocessing To obtain differential signals from CAN-FD/CAN bus prototypes, we first link two probes of an oscilloscope to (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. As in Sect. 3.1 (and Fig. 5), the trick of five sets L_1, L_2, L_3, L_4, and L_5 is used as well (Fig. 12).

Feature Extraction Statistical features could be extracted from the preprocessed electrical CAN-FD/CAN signals. We use the features in Table 1 as well for each set

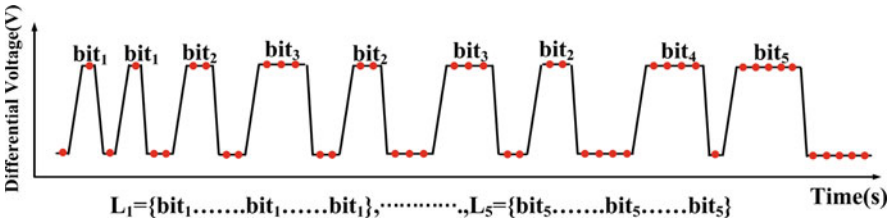

Fig. 12 A CAN-FD/CAN frame is divided into 5 sets

and a total of 40 features for each electrical CAN-FD/CAN signal. Relief-F [25] is also used to weight these features and the feature set in Table 2 can thus be obtained.

Identifying ECUs We use supervised learning, logistic regression (LR) and SVM, to identify the source of CAN-FD/CAN signal. The training phase generates 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 trained model (i.e., whether or not it can determine the source of newly received frames), and the second is on intrusion detection.

5.3 Source Identification and Intrusion Detection

5.3.1 Experiment Setup

The system adapts to different bus prototypes (Fig. 13). Type A (Fig. 13a) contains five CAN-FD nodes that can send both CAN-FD and CAN frames. Type B (Fig. 13b) contains five CAN-FD nodes (the same as in Type A) and four extra CAN nodes that send purely CAN frames. Type C (Fig. 13c) contains five CAN nodes. 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 ringing mainly exists between ECUs and junctions, 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. 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 500kbit/s. Our CAN bus prototype takes high-speed CAN network topology.

Each CAN node consists of an Arduino UNO board and a CAN shield from Seed Studio. Each CAN shield consists of an MCP2515 controller [26] and an MCP2551 transceiver [27], and the bit rate is 500kbit/s. For CAN-FD nodes, each one consists of a STM32F105 shield and a MCP2517FD controller [36]. MCP2517FD is known as compact, cost-effective and efficient CAN-FD controller and uses SPI interface

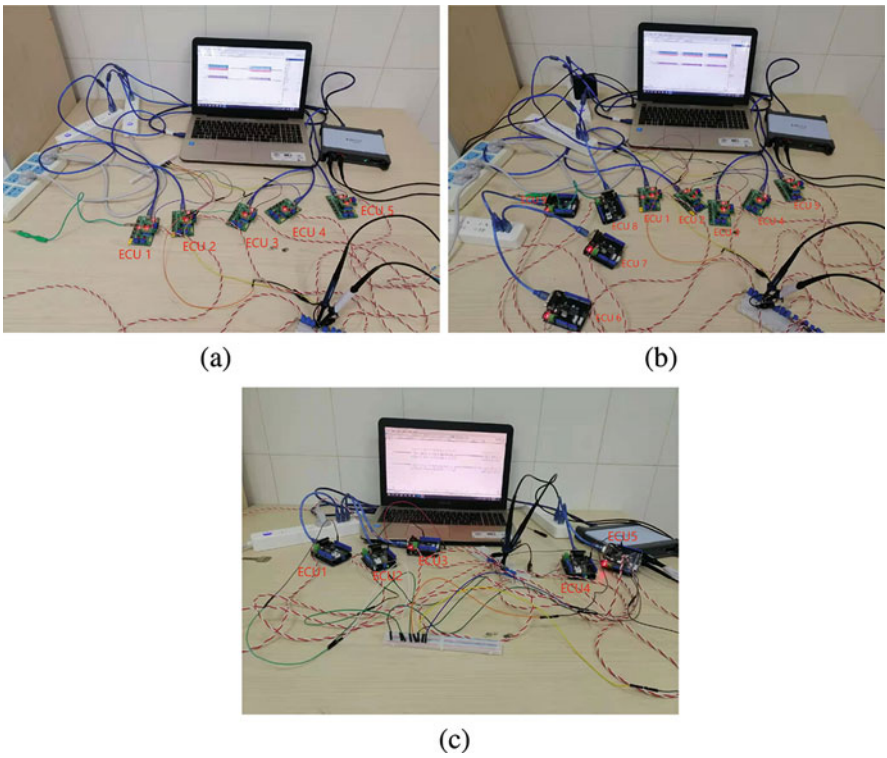

Fig. 13 Three prototypes. (**a**) Type A: CAN-FD nodes, (**b**) Type B: CAN-FD nodes and CAN nodes, (**c**) : CAN nodes

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).

All ECUs are powered by a battery which supplies electric power to each ECU via USB ports. 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 ohm 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 training set (its size could be adjusted according to the performance of the model).

5.3.2 Sender Identification

5.3.2.1 Sender Identification on Pure CAN

For Type C (Fig. 13c), we consider ringing effect. We execute SVM and LR by using D.R, F.R, and D.R.F.R. The results are shown in Tables 9, 10, and 11. Each diagonal cell represents the accuracy of the two classification algorithms. As expected, D.R suffice to fingerprint ECUs.

5.3.2.2 Using Dominant States and Rising Edges (D.R)

We then evaluate whether the system can correctly classify ECUs for Type A and Type B. Table 12 lists the confusion matrix for 5 ECUs that send CAN-FD frames (Type A). The recognition rate of the system is sufficient to correctly recognize ECUs, and the error rate is very low. Table 13 lists the confusion matrix of 9 ECUs

Table 9 SVM/LR for Type C and D.R

	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 10 SVM/LR for Type C and F.R

	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 11 SVM/LR for Type C, D.R.F.R

	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 12 SVM/LR for Type A and D.R

	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

(Type B), of which 5 ECUs send CAN-FD frames, and the remaining 4 ECUs send CAN frames. One may see the system can still correctly classify and recognize ECUs in hybrid network.

5.3.2.3 Using Falling Edges and Recessive States (F.R)

We also consider the recognition rate if F.R are used. As ringing intensity of falling edges of signals is higher than that of rising edges, recognition rate would be affected when falling edges are used. Table 14 shows the results for Type B and Table 15 shows the recognition rates 81.54∼86.21% for Type A. We can see really low recognition rates.

5.3.2.4 Using (Dominant States and Rising Edges) and (Falling Edges and Recessive States) (D.R.F.R)

We also compare the execution rates when the system uses D.R.F.R. Tables 16 and 17 show the results of Type A and Type B respectively, both lower than that using D.R.

5.3.3 Detecting Known ECUs

Now we evaluate whether our system can recognize malicious frames sent by an attacker using known ECUs. For Type C (Fig. 13c), we assume that ECU 1 is normal and an attacker 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. Table 18 shows a detection rate 99.01%. For Type A (Fig. 13a), we use the same assumptions and operations as for Type C and achieve a detection rate of 98.5% (Table 18). For Type B (Fig. 13b), we regard ECU 7, ECU 8 and ECU 9 as attackers (capable 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 18 shows the results with comparable performance to Type A and Type C.

Table 13 Confusion matrix using SVM/LR respectively for Type B and D.R

	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 14 Confusion matrix using SVM/LR respectively for Type B and F.R

	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	83.89/82.17

Table 15 SVM/LR for Type A and F.R

	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 16 SVM/LR for Type A, D.R.F.R

	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

5.3.4 Detecting Unknown ECUs

We adopt a threshold-based method. For Type A, we first remove ECU 5 and obtain about 500 frames from the remaining ECUs to train a 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 newly collected data and Fig. 14 shows False Positive (FP) and False Negative (FN) rates. 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 1000 data which will be classified by the obtained model. Figure 15 shows recognition rate 99% at 0.7. Type C uses similar method and Fig. 16 shows 99.1% recognition rate at 0.83.

5.4 Discussions

Sample Rate The experiments could be reproduced at various sample rates, especially for Type B. At different sample rate one will be at different position of sample sizes (which might be closely related to system performance). Table 19 shows the average identification and false positive rates at the sample rates 10 ∼25 MS/s (1000 frames for each ECU).

Comparable Performance Between Type A and Type C For same topology, one may note considerable performance for Type A (CAN-FD) and Type C (CAN) by 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. First, CAN-FD supports data size up to 512 bits, drastically larger than 64 bits in CAN specification, thus the cumulative effect of ringing for Type A might be more powerful than for Type C. Second, CAN-FD provides variable

Table 17 SVM/LR for Type B and D.R.F.R

	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

Table 18 IDS using Support Vector Machines/Logistic Regression

Prototype	True	Predicted (SVM)		Predicted (LR)	
		No attack	Yes	No attack	Yes
CAN-FD	No attack	99.38	0.62	99.85	0.42
	Yes	1.5	98.5	1.88	98.12
CAN-FD&CAN	No attack	99.01	0.99	99.11	0.89
	Yes	1.18	98.82	1.89	98.11
CAN	No attack	99.58	0.52	99.44	0.56
	Yes	0.99	99.01	0.89	99.11

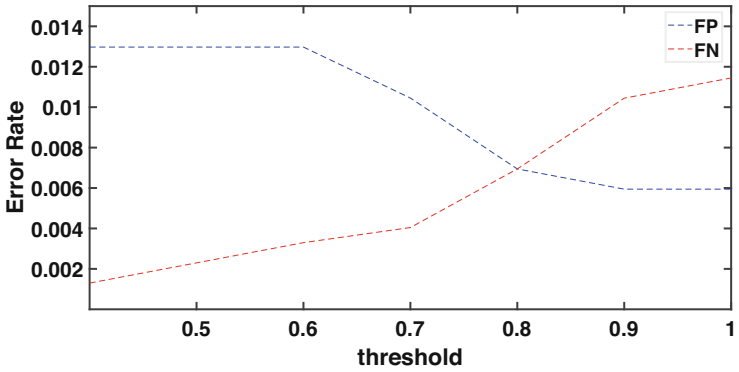

Fig. 14 Error rates at varying thresholds (Type A)

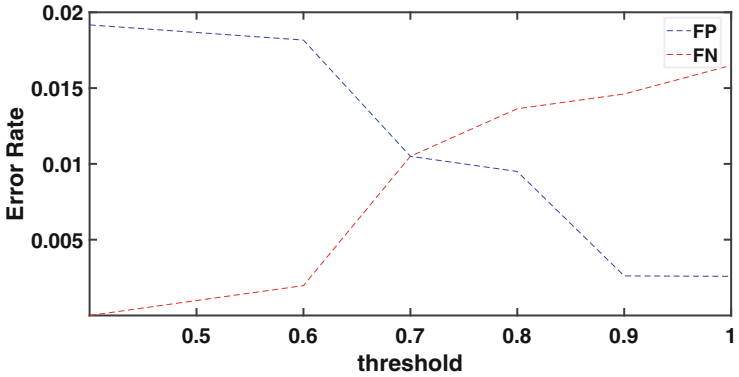

Fig. 15 Error rates at varying thresholds (Type B)

transmission rate and the experiments specify 2Mbit/s for data field of CAN-FD frames and 1Mbit/s for other fields (e.g., arbitration, control and CRC), whereas Type C regulates 500kbit/s. Namely, we have the bit width 2000 ns in a CAN frame, and 1000 ns in non-data field of and 500 ns in data field of a CAN-FD frame. Now, it is more likely for Type A (than Type C) that ringing of recessive states functions

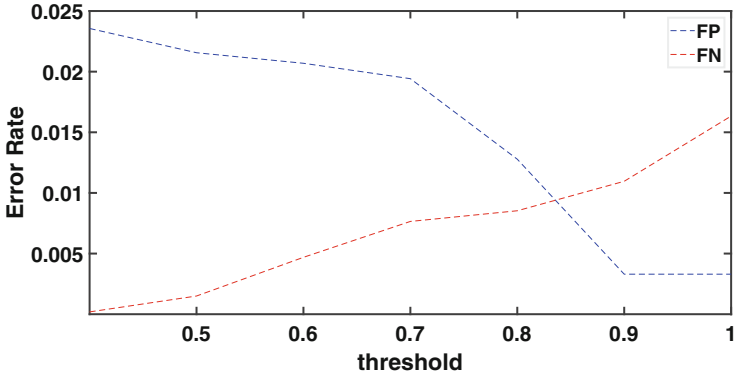

Fig. 16 Error rates at varying thresholds (Type C)

Table 19 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

unceasing (even though the bit itself was already completed on the network)[1] and thus involves the coming dominant states before it attenuates to be unnoticeable.

Applicability to CAN-FD Network in Real Vehicles The controllers used herein conform to ISO11898-1:2015 and support CAN-FD [36]. Possible transition mechanism from CAN to CAN-FD (i.e., Type A and Type B) is also considered. The results show expressive evidence on the applicability to forthcoming real vehicles set up by CAN-FD network. These 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 In real vehicles, the changes of internal temperature will affect the characteristics of electrical signals. A typical example is that the voltage output may deviate from 0.012 to 0.026 V [1] when we start the vehicle from a cooled turn-off engine to warmed-up. This may also exist for CAN-FD network. Howbeit, CAN-FD frames are longer 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).

[1] It is reported [34, 37] that for CAN-FD, high-speed data phase and low-speed arbitration phase challenge the same ringing surrounds (as ringing does not depend on transmission rate), and ring of some recessive bit might not converge until criterion and interfere with the next dominant bit.

Battery/ECU Aging Battery aging might affect the characteristics of the electrical signals. 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. This interesting topic might be explored 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.

6 Conclusion

The chapter introduces in-vehicle ECU identification by using CAN electrical Signaling. This can be viewed as side-channel information exploit on CAN networks. In designing the identification algorithms, signal characteristics of different phases in the signals has different impacts on the algorithm accuracy. The problem of source identification is also important on in-vehicle CAN-FD networks. ECU identification algorithms can be trivially extended to in-vehicle IDS systems.

Acknowledgments The author is supported by the National Natural Science Foundation of China (61971192), Shanghai Municipal Education Commission (2021-01-07-00-08-E00101), and Shanghai Trusted Industry Internet Software Collaborative Innovation Center.

References

1. Kneib, M., Huth, C.: Scission: signal characteristic-based sender identification and intrusion detection in automotive networks. In: Proceedings of the 2018 ACM SIGSAC conference on computer and communications security, pp. 787–800 (2018)
2. ISO 11898-2. Road vehicles - Controller area network (CAN) - Part2: High-speed medium access unit. ISO Standard-11898, International Standards Organisation (ISO) (Dec. 2016).
3. Robert Bosch GmbH. CAN specification version 2.0, Robert Bosch GmbH, Stuttgart, Germany, 1991. Available: http://www.bosch.com (1991)
4. Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S.: Experimental security analysis of a modern automobile. In: IEEE symposium on security and privacy (2010)
5. Miller, C., Valasek, C.: Adventures in automotive networks and control units. Def Con **21**, 15 (2013)
6. Miller, C., Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat U S A **2015**, 91 (2015)
7. Tencent Keen Security Lab. Experimental security assessment of Mercedes-Benz cars. https://keenlab.tencent.com/en/whitepapers/Mercedes_Benz_Security_Research_Report_Final.pdf
8. Kang, M., Kang, J.: A novel intrusion detection method using deep neural network for in-vehicle network security. In: IEEE 83rd vehicular technology conference (VTC Spring), pp. 1–5 (2016)
9. Muter, M., Asaj, N.: Entropy-based anomaly detection for in-vehicle networks. In: Intelligent vehicles symposium (IV). IEEE (2011)
10. Song, H.M., Kim, H.R., Kim, H.K.: Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network. In: 2016 International conference on information networking, pp. 63–68 (2016)

11. Taylor, A., Leblanc, S., Japkowicz, N.: Anomaly detection in automobile control network data with long short-term memory networks. In: DSAA 2016, pp. 130–139 (2016)
12. Guo, F., Wang, Z., Du, S., Li, H., Zhu, H., Pei, Q., Cao, Z., Zhao, J.: Detecting vehicle anomaly in the edge via sensor consistency and frequency characteristic. IEEE Trans. Veh. Technol. **68**(6), 5618–5628 (2019)
13. Cho, K.-T., Shin, K.G.: Fingerprinting electronic control units for vehicle intrusion detection. In: Proc. of the 25th USENIX security symposium, Aug. (2016)
14. Cho, K., Shin, K.G.: Viden: attacker identification on in-vehicle networks. In: Proceedings of 2017 ACM CCS, pp. 1109–1123 (2017)
15. Choi, W., Jo, H.J., Woo, S., Chun, J.Y., Park, J., Lee, D.H.: Identifying ECUs using inimitable characteristics of signals in controller area networks. IEEE Trans. Veh. Technol. **67**(6), 4757–4770 (2018)
16. Choi, W., Joo, K., Jo, H.J., Park, M.C., Lee, D.H.: VoltageIDS: low-level communication characteristics for automotive intrusion detection system. IEEE Trans. Inf. Forens. Secur. **13**, 2114 (2018)
17. Foruhandeh, M., Man, Y., Gerdes, R., Li, M., Chantem, T.: Simple: single-frame based physical layer identification for intrusion detection and prevention on in-vehicle networks. In: 35th Annual computer security applications conference, pp. 229–244 (2019)
18. Murvay, P.S., Groza, B.: Source identification using signal characteristics in controller area networks. IEEE Signal Process. Lett. **21**(4), 395–399 (2014)
19. 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)
20. Lim, H., Kim, G., Kim, S., Kim, D.: 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)
21. Mori, H., Suzuki, Y., Maeda, N., Obata, H., Kishigami, T.: 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, Rome, pp. 1–6 (2012)
22. High-Speed CAN (HSC) for vehicle applications at 500 kbps, SAE J2284-3, SAE International, Warrendale, PA, USA (2002)
23. Studnia, I., Nicomette, V., Alata, E., Deswarte, Y., Kaniche, M., Laarouchi, Y.: Survey on security threats and protection mechanisms in embedded automotive networks. In: 2013 43rd Annual IEEE/IFIP conference on dependable systems and networks workshop, pp. 1–12 (2013)
24. Checkoway, S., McCoy, D., et al.: Comprehensive experimental analyses of automotive attack surfaces. In: 20th USENIX security symposium. USENIX Association (2011)
25. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Machine learning: ECML-94, pp. 171–182. Springer, Berlin Heidelberg (1994)
26. Microchip-Corporation: Stand-Alone CAN Controller with SPI Interface (2005). Microchip MCP2515. https://www.mouser.com/datasheet/2/268/MCP2515-Stand-Alone-CAN-Controller-with-SPI-200018-708845.pdf
27. Microchip-Corporation: MCP2551 High-Speed CAN Transceiver (2007). http://ww1.microchip.com/downloads/en/devicedoc/21667e.pdf
28. Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw. **12**(2), 181–201 (2001)
29. Kneib, M., Schell, O., Huth, C.: On the robustness of signal characteristic-based sender identification. CoRR, vol. abs/1911.09881 (2019)
30. Robert Bosch GmbH. CAN with flexible data-rate (2012). https://www.can-cia.org/fileadmin/resources/documents/proceedings/2012_hartwich.pdf
31. Yu, T., Wang, X.: Topology verification enabled intrusion detection for in-vehicle CAN-FD networks. IEEE Commun. Lett. **24**(1), 227–230 (2019)
32. Woo, S., Jo, H.J., et al.: A practical security architecture for in-vehicle CAN-FD. IEEE Trans. Intell. Transp. Syst. **17**(8), 2248–2261 (2016)
33. Agrawal, M., Huang, T., et al.: CAN-FD-Sec: improving security of CAN-FD protocol. In: ESORICS 2018, Lecture notes in computer science 11552, pp. 77–93 (2018)

34. Mori, H., Suzuki, Y., 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)
35. 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)
36. Microchip-Corporation. External CAN FD Controller with SPI Interface MCP2517FD (2017). http://ww1.microchip.com/downloads/en/DeviceDoc/MCP2517FD-External-CAN-FD-Controller-with-SPI-Interface-20005688B.pdf
37. Islinger, T., Mori, Y.: Ringing suppression in CAN FD networks. CAN Newsl. Jan, pp. 12–16 (2016)