Real-time Security Warning and ECU Identification for In-vehicle Networks

Hongqian Wei, *Member, IEEE*, Qiang Ai, *Member*, *IEEE*, Wenqiang Zhao, Youtong Zhang

***Abstract*—Vehicle intelligence and networking have manifested the significance of the embedded Controller Area Network (CAN) bus. However, the lack of message encryption and identity authentication leaves Electric Control Units (ECUs) exposed to cyber-attacks. To identify the potential attacks on the CAN, intrusion detection systems are required with consideration of their computational burden and the application in vehicles. Therefore, we propose a lightweight ECU identification scheme. Explicitly, the proposed method records the periodic intervals of frames and calculates accumulated clock offsets with the recursive least square algorithm; meanwhile, the empirical rules are adopted to eliminate the noises. Then, the ECU fingerprints have been formulated with the derived clock skew, clock offsets as well as their expectations. Furthermore, to accurately identify the attackers in the masquerade attacks, a double-verified attacker identification approach is proposed, in which the data dependency and intra-inter class algorithm are respectively utilized for better executability. Finally, we have tested the proposed method with an actual vehicle and the results manifest that the proposed method could identify the abnormal ECUs with an identification accuracy of at least 98% and its execution time is less than 3ms.**

***Index Terms*—intelligent connected vehicles, networks, security, vehicle control**

I. Introduction

WITH the development of vehicle intelligence and electrification, intelligent connected vehicles have taken for more than 86% of the market share [1, 2]. The intelligence of vehicles requires the supports from in-vehicle networks, such as the CAN networks. For instance, the CAN is widely used in real-time control systems, such as powertrain controllers [3]. Nevertheless, unlike the Ethernet which includes the identity authentication and encrypted messages, CAN broadcasts the frames without the relevant information authentication mechanism, which are extremely vulnerable to cyber-attacks [4, 5]. Miller and Valasek have intruded into the in-vehicle CAN by means of remote firmware updates via the brittle gateway, which has disabled the brakes and engine system [6]. Similarly, the Keen Security Laboratory of Tencent has attacked the powertrain of Tesla Model S and T-box systems of Mercedes Benz with the Internet-of-Things system in 2017 and 2019, respectively [7]. A technical report powered by Upstream Security has stated that the cyber-attack events for the automobile have grown by 605% from 2016 to 2020 [8].

To effectively prevent the potential cybersecurity incidents, two approaches in existing studies have been adopted: the security communication mechanism and the Intrusion Detection System (IDS). The security communication mechanism designs a new communication protocol or communication medium to protect the information frames. In detail, message authentication codes [9], encrypted frames [10] and medium protocols [5, 11], are effective approaches. For instance, the Security On-board Communication proposed by the AutoSAR organization guarantees frame freshness and integrality [12]. Although the above measures could prevent frames from being tampered, the communication protocol has some difficulties for the application on embedded systems due to their limitation of communication bandwidth and the implementation expense [13]. In contrast, the IDS may be a feasible approach to deal with anomaly attacks, since it would not change original communication protocols and it also could be installed on high-performance processors like the central gateway [10, 11, 14].

A. Related works

According to the state-of-the-art studies [31, 32], the IDS could be classified into two types: the signature-based methods and the anomaly-based methods. A signature-based methods manage to learn the preset features and identify potential attacks by constructing ECU fingerprints. An anomaly-based method focuses on knowing normal behaviors to identify any deviations caused by potential attacks. In the anomaly-based IDS, the message characteristics could be the frame frequency [14], frame ID sequence [15] and its derived information entropy [16], message context [17] and the remote frame intervals [18]. This type of IDS usually could accurately identify the attacks though it may have to deal with a large amount of data in real time [19]. The signature-based IDS exploits the subtle distinctions among different ECUs to formulate their unique fingerprints. These fingerprints include the clock frequency [20] and clock skew [21, 22], the physical voltage signals [23] as well as the signal duration of data frames

This work was supported in part by the *National Key Research and Development Plan of China* (Grant No. 2021YF3101500), in part by the *National Natural Science Foundation of China* under Grant 52202461, and in part by the *China Postdoctoral Science Foundation* under Grant 2022TQ0032 and 2022M710380. (*Corresponding author: Youtong Zhang.*)

H. Wei, Q. Ai, W. Zhao and Y. Zhang are with the Key Laboratory of Low-Emission Vehicles in Beijing, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: bit_hongqian@126.com; aiqiangbit@qq.com, wenq_zhao@163.com; youtong@bit.edu.cn).

[24]. For instance, the accumulated clock offsets are exploited to formulate a unique ECU fingerprint using the Recursive Least Square (RLS) algorithm [21]. Similar signal characteristics could be extended into the remote frames [18], which utilizes different frame responses to distinguish the abnormal ECUs. The above ECU fingerprints process the frames of CAN, which are realized with the software application and not rely on the extra hardware devices [25, 26]. In addition, a novel attack named cloaking attack was developed by emulating the desired transmission delay to deceive the IDS [35]. Two clock skew-based IDSs, i.e., the state-of-the-art IDS and its adaption to the network time protocol are developed for the validation of the proposed cloaking attack. The results indicated that they could succeed with little prediction errors of less than 6%. This type of attacks is hard to detect due to their concealment and simulation for the desired time delay. Recently, physical analog signals of CAN are further processed to differentiate the ECUs for potential attack detection [27-29]. For instance, the VoltageIDS was proposed, which combines the time-frequency features of analog voltage signals of CAN frames and learns the subtle differences of voltage levels with the intelligent neural networks [27]. On this basis, the frame duration has been sampled and learned with the RLS algorithm to assess whether newly arrived frames are transmitted from a legitimate ECU, and the detection accuracy is more than 95% [24, 29].

B. Motivation and Contributions

Although the previous state-of-the-art studies have effectively addressed the potential attacks by detecting the message frequency or physical signals, their actual executability requires to be validated. The IDS should be lightweight for the low computational burden, and the online application in actual ECUs should be experimentally tested; meanwhile, the possible noises of message timestamps should be processed before the attacker identification. Additionally, the identification accuracy of attackers should be considerably focused, which could help the auto repairment to quickly find the attacked ECUs in the post-processing items [22].

To this end, this work designs a real-time ECU identification scheme for the in-vehicle CAN, which is composed of three parts: data collection and preprocessing, anomaly detection, and attacker source identification. Explicitly, in the data collection and preprocessing step, the periodic frame intervals in the receivers are recorded and meanwhile the abnormal data caused by the environmental noise are eliminated with the empirical rule algorithm. Then, a typical RLS algorithm is introduced to formulate the clock skew, which provides a significant design reference for the attack detection. Subsequently, a double-verified attacker identification algorithm is presented based on the data dependency and intra-inter class algorithm.

II. CAN AND ATTACKS

A. CAN

CAN was developed by Bosch GmbH [31], and is widely used in real-time systems due to its cost-effective budget and full-duplex communication efficiency. To better understand the CAN, we will introduce the CAN topology where the standard CAN 2.0 protocol is taken as an example.

CAN topology information: This part mainly describes the CAN physical layer and data link layer. In the CAN physical layer, CAN signals are coded with the Non-Return to Zero method and released at a twisted pair of wires, namely CAN-H(high) and CAN-L(low). In the CAN physical protocol, the recessive bit “1” indicates that CAN-H and CAN-L are both zero-biased at 2.5 volts while the dominant bit “0” indicates that CAN-H and CAN-L go around about 3.5 volts and 1.5 volts, respectively. In the CAN communication, the successive dominant/recessive bits fill the data frame.

The CAN protocol defines four types of frames: data frame, remote frame, error frame and overload frame. The standard data frame includes seven fields: Start of Frame (SOF), arbitration field, control field, data field, Cyclic Redundance Check (CRC) field, Acknowledgement (ACK) field, and End of Frame (EOF), as shown in Fig. 1.

Start	Arbitration		Control			Data field	CRC field		ACK field		End
SOF	ID	RTR	IDE	r0	DLC	Data	CRC sequence	CRC	ACK	ACK slot	EOF
1	11	1	1	1	4	0-64	15	1	1	1	7

Fig. 1. Standard data frame of CAN 2.0.

B. Adversary Model

Current common types of attacks in the bus network include spoofing attack, suspension attack, Denial of Service (DoS) attacks, fuzzing attacks and masquerade attack [21]. For the DoS attacks and fuzzing attacks, the attack is highly frequent, which could be easily detected by observing their periods. For the spoofing attack, the adversary can monitor the CAN bus and inject a series of forged messages to the CAN bus, confusing other controllers to execute some wrong instructions. The suspension attack takes advantage of the arbitration mechanism to stop or suspend the target ECU; i.e., an ECU would stop the message transmission when a synchronous message with a higher priority is requesting a transmission at the same time. If it fails for several times, this ECU would go offline. With this arbitration mechanism, the adversary can force the target ECU to lose the arbitration for multiple times, causing the messages from the target ECU invalid.

The masquerade attack would not change the ID sequence and transmission frequency. To execute a masquerade attack, the adversary needs to control two ECUs. As shown in Fig. 2, ECU B is firstly controlled by programming an aggressive firmware via the system vulnerability, and then monitors the communication message of ECU A. By learning the ID and frequency of a message from ECU A, ECU B can make ECU A invalid with the suspension attack; then, ECU B can send the forged message with the same ID and intervals as ECU A, implementing the wrong instructions to control vehicles.

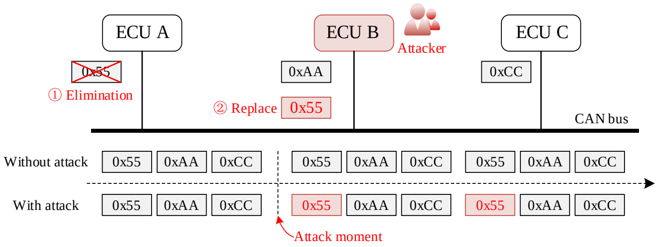


Fig. 2. Masquerade attack.

C. Clock Offsets

Different ECUs present distinct timing deviations due to the clock difference and manufacturing deviation. The timing deviation could be described with the "clock offset" and "clock skew". They are defined as follows.

- **Clock offset**: defined as the timing difference between the reported clock $\mathbb{C}_i$ and the actual clock $\mathbb{C}_t$. In this study, the clock offset is denoted by offsets between two non-true clocks.
- **Clock skew**: defined as the frequency difference between the reported clock and the actual clock, such as $d\mathbb{C}_i/dt - d\mathbb{C}_t/dt$. In this study, the clock skew is defined with two non-true clocks.

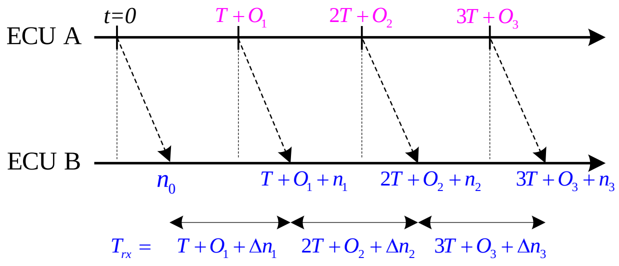


Fig. 3. Clock offset and timing analysis for frame arrivals.

To accurately determine the clock skew, an estimation process is required. As shown in Fig. 3, ECU A broadcasts a frame every T ms and ECU B periodically receives this frame. However, ECU A would send the frame with a small delay (named as clock offset O_i) due to the inconsistence of system clocks. Meanwhile, ECU B receives the frame after the time delay $iT+O_i+n_i$, where n_i denotes the network transmission delay in the medium and its average value approaches to a fixed value. That is, the difference of n_i is nearly same, exactly $E[\Delta n_i]=0$. Thus, the interval between each frame arrival is marked as $T_{rx,i}=iT+O_i+\Delta n_i$; therefore, the expectation value of the timestamp intervals can be expressed as $\mu_{rx}=E[T_{rx,i}]\approx T$. As proved in [33], the average difference between the estimated and measured arrival times is defined in the following equation.

$$\mathrm{E}[\mathbb{N}]=\mathrm{E}\left(i\left(T-\mu_{rx}\right)+O_i+\Delta n_i\right)\approx E\left(O_i\right) \tag{1}$$

Therefore, the *average clock offset* can be estimated, which is indeed different for distinct transmitters. Under such conditions, if the average clock offset could be summed up, the *accumulated clock offsets* could be thereby determined. Furthermore, their slopes can be obtained to represent the *clock skew*, which is unique for the ECU devices.

D. Empirical Rules Based on The Gaussian Distribution

The Empirical Rule (ER) algorithm is a statistical approach based on the Gaussian distribution [33]. If a random variable X follows a Gaussian distribution with an average value μ and variance σ^2, marked as $X\sim N\left(\mu,\sigma^2\right)$, its Probability Density Function (PDF) is expressed in Eqn. (2) and features are plotted in Fig. 4.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2}$$

According to the PDF, the three-delta principle is widely utilized in the data process; the probability of a random variable falling within the intervals of $(\mu-\lambda\sigma,\mu+\lambda\sigma)$ conforms to the following rules:

$$\begin{aligned} &P(\mu-\sigma\le x\le\mu+\sigma)\approx 0.6827,\\ &P(\mu-2\sigma\le x\le\mu+2\sigma)\approx 0.9545,\\ &P(\mu-3\sigma\le x\le\mu+3\sigma)\approx 0.9973 \end{aligned} \tag{3}$$

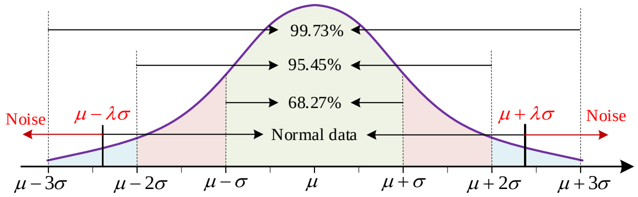


Fig. 4. The illustration of ER for the noisy data.

In the practical application of IDS, the frame interval is influenced by the environmental noise or the message loss. However, this influence cannot be eliminated with the error frame mechanism. To this end, we introduce the ER to filtrate the abnormal noises. For instance, the IDS would preset a threshold λ and establish an interval of $(\mu-\lambda\sigma,\mu+\lambda\sigma)$ to remove the accidental noisy data.

III. Proposed IDS and Attacker Identification

The proposed IDS utilizes the subtle differences in the frame intervals, with which the abnormal attacker sources can be accurately located. The proposed attacker identification scheme includes three steps: the data collection and preprocessing, the intrusion detection, as well as the double-verified attacker identification. The former two steps are similar with existing studies [21] and the noteworthy points lie in the denoising process since the noises in the clock offsets may incur wrong identification of the ECU fingerprints. The last double-verified attacker identification concerns the improvement of the ECU mapping accuracy and the potential attacker identification. In detail, the frame intervals are recorded and further denoised with the ER algorithm; and the clock offsets are extracted and accumulated by the RLS algorithm to establish the clock skew, constituting the rudiment of the ECU fingerprints. Therefore, the masquerade attack is detected by observing the abnormal clock skews. Next, the double-verified attacker identification is realized by combining the mathematical expectation of clock offsets and the intra-inter distance algorithm.

A. Data Collection and Preprocessing

To realize the ECU fingerprints, frame intervals are measured firstly. The IDS devices can be connected to the CAN bus and monitor all frames from the target ECU. Then, frame intervals $T_{r,i}$ of all ECU*ᵢ* are recorded at the receivers. The RLS algorithm can be utilized to estimate the *average clock offset*, which should converge to a non-zero constant.

Nevertheless, we found that the frame intervals come with abnormal noises in the practical application. These noises deviate less than an average frame period, which cannot be detected with the error-frame method [24]. Therefore, we introduce the ER algorithm to eliminate abnormal noises and determine the normal range of the frame intervals. The steps about the data collection and preprocess are expressed.

Step 1. Data collection.

The detection node (IDS) is connected to the CAN bus and monitors the frame intervals from the target ECU. The ER parameters are updated when N sets of data are sampled. Meanwhile, IDS records the frame interval O_i, and calculates the average interval μ_k and the standard deviation σ_k.

Step 2. Noisy data filtration.

The noises are accurately identified with the ER algorithm. If the data satisfy $\left|O_i-\mu_k\right|/\sigma_k<2$, the recorded data are regarded as healthy data; otherwise, the newly recorded data are regarded as noisy data, which would be abandoned.

Step 3. Data storage and parameter update.

1). *Save reasonable data for the subsequent intrusion detection.* The database is updated cyclically using a sliding window based on the first-in-first-out principle. In the sliding window, the first data in the database are removed and the newly reasonable data will be filled to the end.

2). *Update the ER parameters.* According to the fresh database, the mean value μ_k and standard deviation σ_k are updated again.

B. Anomaly Detection Based on The Clock Skew

Since different ECUs present distinct clock offsets, the average clock offsets can be summed up to formulate the *accumulated clock offset*. In detail, the IDS includes two parts: the clock skew formulation and anomaly detection. Since the clock offsets in the same ECU are constant, the accumulated clock offsets are linearly distributed. The accumulated clock offset is identified with the following equation.

$$O_{acc}[i]=s[i]\cdot n[i]+e[i] \tag{4}$$

where $O_{acc}[i]$ denotes the accumulated clock offset at the step i, and $n[i]$ is the frame instance number; s is the regression parameter, representing the estimated clock skew; e denotes the regression error, representing the model residual, which can be utilized to detect the abnormal information. In this model, the RLS algorithm is used to estimate the clock skew.

As aforementioned, if an anomaly intrusion happens, the clock skew would deviate from the reasonable path, resulting in a non-zero model residual. Under such assumptions, the cumulative sum (CUSUM) method can be introduced to observe the sudden change of the model residual. In the CUSUM, two diagnostic variables, denoting the upper and lower boundaries of the accumulated identification error, are defined as follows, respectively.

$$\begin{aligned} L^{+} &\leftarrow \max\left[0, L^{+}+\left(e_i-\mu_e\right)/\sigma_e-\lambda\right] \\ L^{-} &\leftarrow \max\left[0, L^{-}-\left(e_i-\mu_e\right)/\sigma_e-\lambda\right] \end{aligned} \tag{5}$$

where $\mu(e)$ and $\sigma(e)$ denote the average value and standard deviation of the model residuals at the observation window $\mathcal{N}$, and λ is a noisy parameter, representing the maximum acceptable error. Then, the diagnostic variables can be observed in real time and if their values exceed the preset threshold, the IDS can declare that there is an anomaly intrusion.

C. Double-verified Attacker Identification

The traditional attacker identification method, such as the dynamic time warping (DTW), may be inaccurate and time-consuming [22]. Therefore, we propose a new double-verified attack identification approach based on the data dependency of clock offsets and the intra-inter distance algorithm. In the proposed algorithm, there are two steps to validate the attacker sources.

The first verification initially determines the possible ECU sources according to the distribution of clock offsets. At first, the average clock offsets would be determined with the RLS algorithm. Then, the correlation coefficients between the sampled average clock offsets and the existing ones are evaluated with the following expression:

$$\rho\left(O_i,O_{ref,i}\right)=\frac{N\left(\sum_{i=1}^{N}O_iO_{ref,i}\right)-\sum_{i=1}^{N}O_i\sum_{i=1}^{N}O_{ref,i}}{\sqrt{\left\{N\sum_{i=1}^{N}O_i^2-\left(\sum_{i=1}^{N}O_i\right)^2\right\}\left\{N\sum_{i=1}^{N}O_{ref,i}^2-\left(\sum_{i=1}^{N}O_{ref,i}\right)^2\right\}}} \tag{6}$$

where $O_{ref,i}$ denotes the reference offsets in the existing ECUs, which can be updated with the historical data. N represents the frame instance number in the sliding window. If the above correlation coefficient is approaching to 1, the current ECU can be initially determined as the attacker, which constitutes the first validation in the proposed approach.

The second verification judges the forged ECU with the intra-inter distance algorithm [34]. To improve the identification accuracy, more elements including the average clock offsets O_i, their mathematical expectations $E\left(O_k\right)$, and the estimated clock skew s_i, are selected as the identification parameters. Since these parameters are not binary, we utilize the Euclidean distance as a metric. Euclidean distance between two values can be defined as $d(\alpha,\beta)=\sqrt{(\alpha-\beta)^2}$. Therefore, the intra and inter distances $\Xi^{\kappa}_{\text{intra}}$ and $\Xi^{\kappa}_{\text{inter}}$ based on the identification parameters such as the skews and clock offsets are defined in Eqn. (7):

$$\Xi^{\kappa}_{\text{intra}}(i)=\left\{d\left(\phi(id'),\phi(id'')\right),\ \forall id',id''\in \text{ECU}_i, id'\neq id''\right\}$$
$$\Xi^{\kappa}_{\text{inter}}(i,j)=\left\{d\left(\phi(id'),\phi(id'')\right),\ \forall id'\in \text{ECU}_i, \forall id''\in \text{ECU}_j\right\} \tag{7}$$

where the parameter $\phi(id)$ denotes the resulting pair, which should run over all identification parameters. According to the intra-inter distance algorithm, the inter distances represent the distances among the messages transmitted by a distinct ECUs while the intra distances represent the distance among the messages from the same ECU. In the practical application, we measure the identification parameters in the newly received messages and run the intra-inter distance over these parameters. Hence, the attackers are compared with the existing ECU fingerprints.

Step 1. Preparation of the ECU identification parameters. To realize the attacker identification, the identification parameter database needs to be built offline, which is the reference.

a). The IDS samples every arriving frame interval $T_{r,i}$ for the target ECUs at each sliding window with the fixed scale N=20. Then, the aforementioned data preprocess step would work to eliminate the noisy data, thereby forming the database $\mathbb{N}$.

b). The average clock offsets O_i are obtained by running the RLS algorithm. Thus, the average clock offsets in the healthy communication would be selected as the reference $O_{ref,i}$, which is subsequently utilized to calculate the correlation coefficients in the first validation of the attacker identification. Accordingly, other required identification parameters, such as $E(O_k)$, and the estimated clock skew s_i, could also be determined.

c). With the sliding window forwarding, the RLS parameters and noise preprocessing parameters are updated. Then, the database $\mathbb{N}$ would be updated simultaneously.

d). By running multiple rounds of the above calculation, the corresponding identification parameters are established. Then, the offline training, such as the intra-inter distance is further executed to formulate the unique ECU fingerprints.

Step 2. First validation of the attacker source identification based on the clock-offset correlation.

a). The IDS samples the newly arriving frame intervals and calculates the average clock offset O_i with the RLS algorithm.

b). The correlation coefficients between the newly arrived offsets and the references in the target ECUs are calculated, with which the most relevant ECU can be further determined by computing over all target ECUs. If the correlation coefficients are higher than 0.8, the possible counterfeit ECUs are initially determined.

Step 3. Secondary validation of the attacker identification based on the intra-inter distance algorithm.

a). The system extracts the identification parameters according to the sampled frame intervals.

b). The system establishes the testing matrix with the intra-inter distance algorithm in the light of the identification parameters.

c). Next, we evaluate the testing matrix compared with the existing ECU fingerprints and thereby determine the counterfeit ECU.

d). Finally, the attacker source is located and the abnormal ECU is determined.

IV. Performance Evaluation

To evaluate the effectiveness of the proposed attacker identification method, an experimental test is implemented based on a mass-produced vehicle.

A. Experimental Setup

Real vehicle: In this article, a *Chery* electric vehicle is used for our experiments in an isolated and secure environment, as shown in Fig. 5. The test device is wired within one CAN bus via the On-Board Diagnostic (OBD-II) port in the actual vehicle. As for the test nodes, typical controllers including a vehicle control unit (ECU1), motor control unit (ECU2) and two battery management systems (ECU3 and ECU4) are employed. These devices are electrically disconnected from the actual power components for the sake of security in the practical experiments. The CAN baud-rate is set at 500Kbps and communication period among the four ECUs ranges from 10ms to 500ms.

Execution environment: The proposed method is programmed using the embedded devices (*S32K148, NXP*), which are running at the main frequency of 48MHz and have a 2MB flash memory. The embedded devices can record the timestamps and frame intervals for the actual intrusion detection analysis. Therefore, it could evaluate the validity of the IDS in the actual in-vehicle environment.

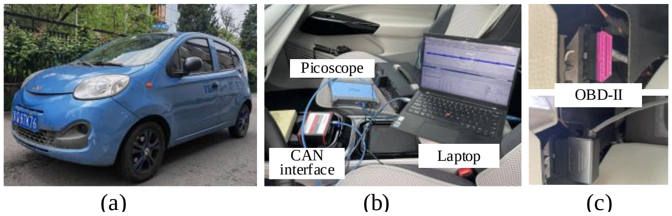


Fig. 5. Experimental environment. (a) Chery vehicle, (b) Actual test environment, (c) OBD-II ports.

Attack introduction: Considering that the masquerade attack is more sophisticated than other attacks, like suspension attacks and spoofing attacks, we have selected the masquerade attacks to test the proposed method. Since the masquerade attack would not change the transmission periods and ID message, and only change the sender source, we make ECU4 invalid and replace it with a new attacker ECU3. For instance, a frame massage with ID 0x401 originally generated from ECU4 would be transmitted by ECU3 when the attack instance is introduced.

B. ECU Clock Skews

To validate whether the clock offsets can be optimized with the RLS algorithm, we conducted an experiment on the real vehicle to extract the accumulated clock offsets. In this experiment, four ECUs are used and two types of ID messages are recorded. ECU1 sends messages with IDs 0x541 and 0x542 at a period T=10ms; ECU2 transmits frame messages with IDs 0x469 and 0x46A at a period of 100ms; ECU3 sends messages with IDs 0x403 and 0x460 at a period of 100ms; ECU4 sends messages with IDs 0x411 and 0x401 at a period of 50ms. Then, we run the RLS algorithm over the four ECUs to evaluate the accumulated clock offsets, which are shown in Fig. 6.

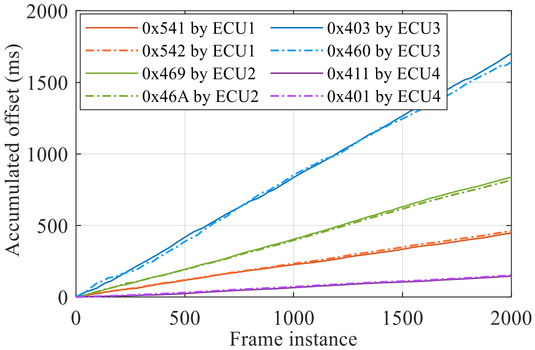


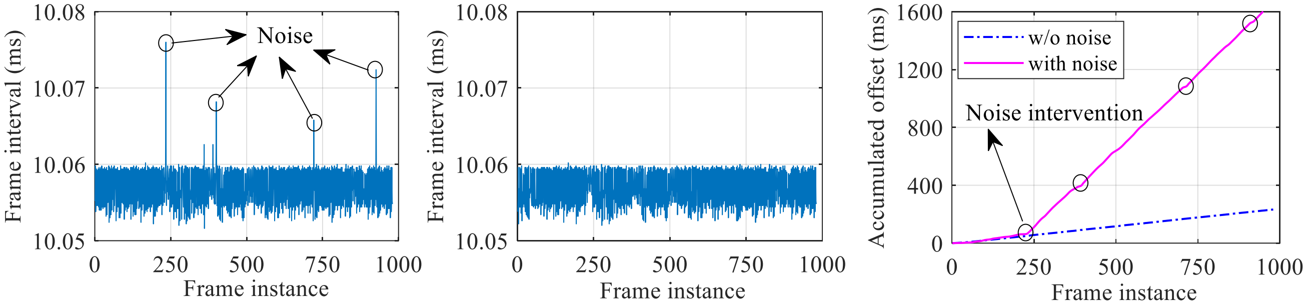
Fig. 6. Clock skews for target ECUs in the real vehicle.

According to Fig. 6, the accumulated clock offsets in the same ECU share a similar tendency regardless of the ID sequence, and different ECUs present distinct clock skews, which indicates that the clock skew features are only related to the hardware configuration. Specifically, the clock skew (slope of the accumulated offset curve) of ECU3 is 0.85×10^{-3} ; ECU4 has the smallest clock skew with 1.06×10^{-4}, which indicates that the frame periods have little effects on the ECU clock offsets. Therefore, the deviation of the accumulated clock offsets could be assessed when the ECU is tampered.

In the experiments, we found that the original frame intervals come with some high-frequency noises, which would affect the identification accuracy of the average clock offsets. Therefore, we introduce the ER algorithm to eliminate the intervention of noises. In this test, frame intervals of ECU1 are recorded before and after the data preprocess are shown in Fig. 7. The original frame intervals without the de-noising process come with some noises at frames 246, 395, 734 and 891. Since the amplitude of this noise is less than a standard frame period, the IDS system cannot mark them as an error frame. However, the processed frame intervals with the proposed de-noising algorithm in Fig. 7(b) present good distribution characteristics and there is little abnormal noise. Fig. 7(c) shows the accumulated clock offsets under the presence and absence of noise. Due to the abnormal noise intervention at frame 246, the accumulated clock offset has deviated from its original slope. Thus, the estimated clock skew with the noise intervention would be recognized as an anomaly attack. On the contrary, the estimated clock skew with the noise preprocess would follow the right tendency.

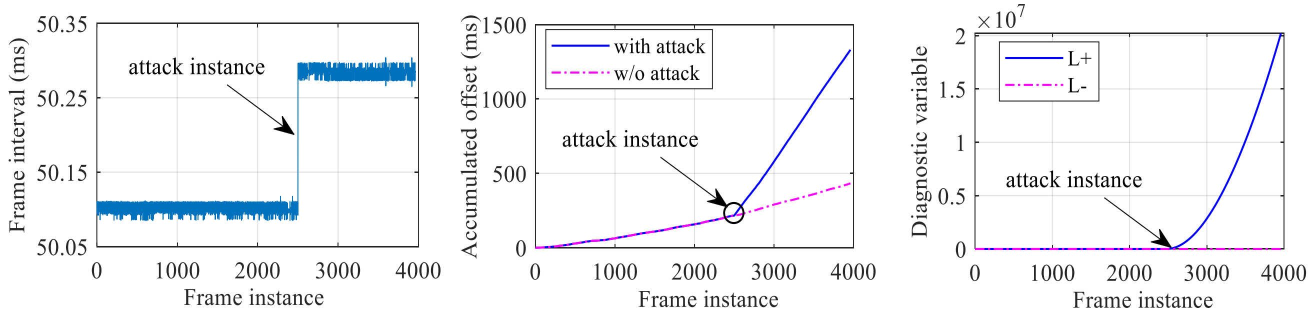
C. Evaluation of the Anomaly Detection

The above results of the accumulated clock offsets prove that the model residual in the RLS algorithm approaches to zero under healthy conditions. If an attack occurs, the clock skew would contort, causing non-zero model residuals. Hence, the diagnostic variables defined in (5) can visualize the sudden increase or decrease under the abnormal attack conditions. We introduce the masquerade attack on the ECU4 at frame number 2500. Before the attack instance, healthy messages with ID 0x401 were generated by the ECU4; however, when the attack occurs, the healthy messages are overwritten and replaced by the new messages with the same ID 0x401 transmitted by the attacker ECU3.



(a) Original frame intervals (b) Frame intervals after preprocessing (c) Accumulated clock

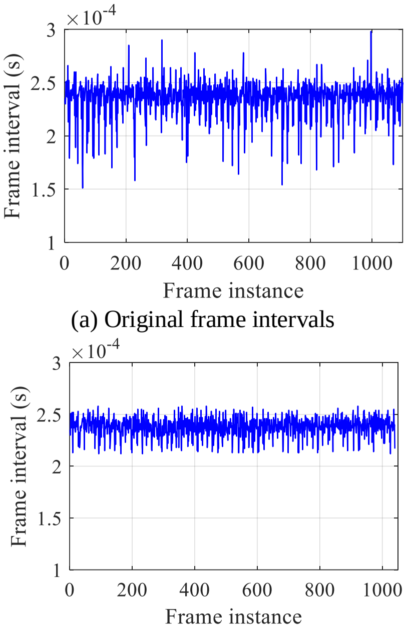
Fig. 7. Comparison results of the noise preprocessing on the ECU1.



(a) Frame intervals (b) Accumulated clock offset (c) Diagnostic variables

Fig. 8. Intrusion detection results for the masquerade attack on the ECU4.

Fig. 8 shows the detection results of the masquerade attacks. From Fig. 8(a), frame intervals increased from 50.103ms to 50.268ms after masquerade attacks, manifesting that the difference of hardware configuration surely affect the periodic frame intervals. When the attack occurs at frame 2500, the slope of the accumulated clock offset begins to tilt. This abnormal curve also leads to a sudden increase of the diagnostic variable L+, exceeding the preset threshold 10. The mentioned anomaly would be regarded as the anomaly intrusion behavior. Normally, when the clock skew of the attacker ECU is greater than that of the original ECU, the diagnostic variable L+ would suddenly increase; on the contrary, when the clock skew is smaller than that of the original ECU, the diagnostic variable L- would suddenly increase.



(a) Original frame intervals

(b) Frame intervals after the ER data process

Fig. 9 Results of the validation of ER data processes

To validate the effectiveness of the ER data processing method, we select the datasets from [21] to make another test. The relevant results are shown in Fig. 9. In the dataset, the frame intervals of periodic messages are recorded as shown in Fig. 9(a), in which the average frame interval is about 245ms. However, some frame intervals have obviously exceeded 300ms and some frame intervals are less than 200ms, both of which may result in the wrong clock skew. In contrast, the frame intervals after the ER data process behaves more regularly and the frame intervals are mostly constrained within in 230-250ms, which implies that the three-delta principle could be effective for the test dataset in the actual ECUs.

D. Identification of Attacker Sources

The effectiveness of the proposed double-verified attacker identification algorithm is evaluated. The masquerade attack is mounted at the number 250 where the healthy frame originally transmitted by the ECU4 would be replaced by forged frames with the same ID 0x401 from the attacker ECU3. In this test, the correlation coefficients of the four ECUs in the first validation step are illustrated in Fig. 10.

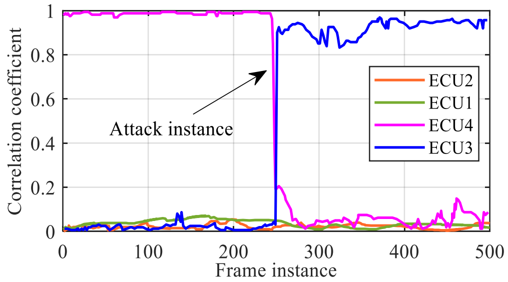


Fig. 10. The correlation coefficients of the four ECUs in the first validation of the attacker identification.

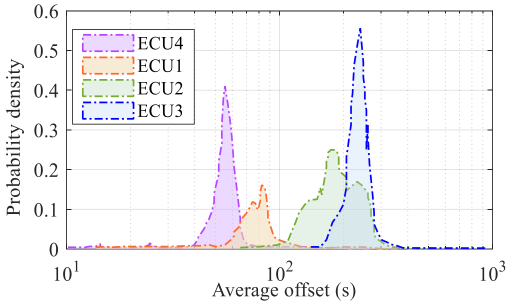


Fig. 11. The probability distribution of clock offsets in the four ECUs.

In Fig. 10, the correlation coefficient of ECU4 approaches to 1 before the attack instance while those of the other ECUs are lower than 0.1, and after the attack is introduced, the correlation coefficient of ECU4 has rapidly decreased to less than 0.1 while that of the anormal ECU3 is about 0.9, which shows that the corresponding frame has been transmitted by the abnormal ECU rather than the healthy ECU. In the experiment, most attacker sources can be well detected with a high identification accuracy of more than 90%. Thus, if the identification accuracy is not considerably considered, the first validation is sufficient.

To achieve more accurate identification and higher robustness of the related algorithms, the double-verified algorithm is introduced. Fig. 11 shows the probability distribution of the average clock offsets in the four ECUs. Different ECUs present distinct probability density curves regardless of the maximum probability density or the average offsets. Therefore, it is reasonable to distinguish the frame offsets according to their mathematical expectation values. Thus, we select the expectation of the average clock offset as an identification parameter in the second validation.

Next, we compared the proposed method with the existing dynamic time warping (DTW) method [22]. The attack test has been implemented in the embedded systems for more than 10000 times. The confusion matrix about the identification of the proposed method and the DTW algorithm are listed in Tables I and II. From Table I, the identification accuracy of all ECUs is more than 98%, which manifests that the proposed double-verified attacker identification algorithm can well distinguish the ECUs and precisely locate the masquerade attack. The maximum mismatching rate among all ECUs is only 0.272% for ECU2. Table II shows the identification accuracy of the DTW method. The identification accuracy of the traditional DTW method is lower than 95%. The maximum mismatching rate is 5.233% for ECU4.

Table I
IDENTIFICATION ACCURACY OF THE ATTACKER SOURCES FOR THE PROPOSED DOUBLE-VERIFIED ALGORITHM

ECU item	ECU1	ECU2	ECU3	ECU4
0x401 by ECU1	**99.721**	0	0	0.231
0x401 by ECU2	0	**98.672**	0.02	0.191
0x401 by ECU3	0	0.012	**99.122**	0
0x401 by ECU4	0.06	0.272	0.041	**98.223**

Table II
IDENTIFICATION ACCURACY OF ATTACKER SOURCES FOR DTW ALGORITHM

ECU item	ECU1	ECU2	ECU3	ECU4
0x401 by ECU1	**92.433**	0	0	5.233
0x401 by ECU2	0.254	**93.654**	0.32	2.222
0x401 by ECU3	0.432	0.494	**94.223**	0.032
0x401 by ECU4	5.087	3.232	0.011	**91.326**

E. Evaluation of the Real-time Executability

Since the actual embedded system in the vehicle control unit requires high real-time performance, the executability of the proposed method is assessed. We choose the DTW method [22] as a comparison group. The relevant test results are depicted in Table III and almost 10000 tests were performed, in which the average running time, standard deviation and maximum deviation are comprehensively analyzed. According to the comparison results, the average running time of the DTW algorithm is 20.456ms and the maximal time expense is 25.809ms, which is a huge computational burden for embedded systems. The proposed double-verified attacker identification algorithm only consumes about 2.852ms and the maximum consuming time is only 3.737ms, which has reduced the execution time by 86%. The results indicate that the proposed algorithm is more appropriate for the actual application. Notably, the first step validation takes about 0.534ms, denoting that the attacker identification based on the correlation coefficients is more proper. If a high detection accuracy is not considered, one-step verification (denoted by the first validation method) is also an effective and efficient approach.

In general, the double-verified method proposed in this work is superior to the DTW method in terms of the identification accuracy and execution time. The test results on the four ECUs in the experiment have illustrated that the double-verified method has improved the identification accuracy by about 5% as shown in Table I and Table II. The DTW method only analyzes the clock skews, which may misidentify the ECUs. Instead, the proposed method could take the clock skew, clock offsets as well as the mathematical expectations into training process for a better identification accuracy. As for the execution time, the DTW utilizes the time-consuming dynamic programming as the solution-seeking method, which would compare all sampled data and find similar one. In contrast, the intra-inter distance has trained the models offline and utilized them online, which is much more efficient.

In general, the proposed double-verified IDS differs with the existing DTW method [22] and the clock skew method [21]. In detail, compared with the clock skew method, the subsequent ECU source identification and multiple-step validations in the proposed method are supplied, in which multiple elements such as the clock offset, skew as well as the expectation could be utilized for the higher identification accuracy. Compared with the DTW method, the proposed method presents higher performance in the executability and the identification accuracy. Meanwhile, the proposed method has been experimentally tested in embedded systems.

Table III
THE REAL-TIME PERFORMANCE EVALUATION RESULTS.

Methods	Avg. time (ms)	Std. deviation (ms)	Max. deviation (ms)
DTW	20.456	2.542	5.353
First step validation	0.534	0.104	0.204
Proposed method	2.852	0.368	0.885

V. CONCLUSION

This article proposed a real-time ECU fingerprinting and attacker identification approach. It can formulate the unique ECU fingerprints and eliminate abnormal noises, thereby constructing a much more robust IDS. More importantly, rather than applying a complicated global optimization on the counterfeit ECU identification using the high-performant processors, the proposed double-verified attacker identification algorithm can be directly programmed into the embedded systems to locate the attacker sources with a higher identification accuracy and better executability. Experimental results manifest that the proposed attacker identification approach has improved the identification accuracy, reaching a value higher than 98% while the actual execution time has been decreased to less than 3ms. Therefore, the proposed ECU fingerprinting and attacker identification approach has significantly enhanced the in-vehicle networks security with the higher practicability and adaptiveness.

Although the proposed method could effectively address the potential masquerade attacks according to the ECU mapping results, limitations still exist. For instance, the proposed method is hard to address the possible birthday paradox problem under which the clock skews of two ECUs are similar. As for this problem, more fingerprints are required, such as ECU physical features including their high-level voltages, rising-falling edges, duration time of high-voltage levels, plateau time, etc. Additionally, this method is also hard to deal with non-periodic messages, which should be further addressed in the future research.

REFERENCES

[1] B. Mao, F. Tang, Z. M. Fadlullah, and N. Kato, "An Intelligent Route Computation Approach Based On Real-Time Deep Learning Strategy for Software Defined Communication Systems," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1554-1565, 2019, Art no. 1.

[2] Y. Cui, L. Du, H. Wang, D. Wu, and R. Wang, "Reinforcement Learning for Joint Optimization of Communication and Computation in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13062-13072, 2021, Art no. 2.

[3] X. Zhu, H. Zhang, D. Cao, and Z. Fang, "Robust Control of Integrated Motor-Transmission Powertrain System over Controller Area Network for Automotive Applications," *Mechanical Systems and Signal Processing*, vol. 58, pp. 15-28, 2015.

[4] S. Woo, H. J. Jo, and D. H. Lee, "A Practical Wireless Attack on the Connected Car and Security Protocol for In-Vehicle CAN," *IEEE Transactions on intelligent transportation systems*, vol. 16, no. 2, pp. 993-1006, 2014.

[5] C.-W. Lin and A. Sangiovanni-Vincentelli, "Cyber-Security for the Controller Area Network (CAN) Communication Protocol," in *2012 International Conference on Cyber Security*, 2012: IEEE, pp. 1-7.

[6] C. Miller and C. Valasek, "Remote Exploitation of an Unaltered Passenger Vehicle," *Black Hat USA*, vol. 2015, no. S 91, 2015.

[7] S. Nie, L. Liu, and Y. Du, "Free-Fall: Hacking Tesla from Wireless to CAN Bus," *Briefing, Black Hat USA*, vol. 25, pp. 1-16, 2017.

[8] "Upstream Security's 2020 Global Automotive Cybersecurity Report," Available: https://upstream.auto/upstream-security-global-automotive-cybersecurity-report-2020/, 2020.

[9] C. Patsakis, K. Dellios, and M. Bouroche, "Towards a Distributed Secure In-Vehicle Communication Architecture for Modern Vehicles," *Computers & security*, vol. 40, pp. 60-74, 2014.

[10] W. A. Farag, "CANTrack: Enhancing Automotive CAN Bus Security Using Intuitive Encryption Algorithms," in *2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO)*, 2017: IEEE, pp. 1-5.

[11] A.-I. Radu and F. D. Garcia, "LEIA: A Lightweight Authentication Protocol for CAN," in *European Symposium on Research in Computer Security*, 2016: Springer, pp. 283-300.

[12] G. Bella, P. Biondi, G. Costantino, and I. Matteucci, "CINNAMON: A Module for AUTOSAR Secure Onboard Communication," in *2020 16th European Dependable Computing Conference (EDCC)*, 2020: IEEE, pp. 103-110.

[13] T. Hoppe, S. Kiltz, and J. Dittmann, "Security Threats to Automotive CAN Networks–Practical Examples and Selected Short-Term Countermeasures," in *International Conference on Computer Safety, Reliability, and Security*, 2008: Springer, pp. 235-248.

[14] A. Taylor, N. Japkowicz, and S. Leblanc, "Frequency-Based Anomaly Detection for The Automotive CAN Bus," in *2015 World Congress on Industrial Control Systems Security (WCICSS)*, 2015: IEEE, pp. 45-49.

[15] M. Marchetti and D. Stabili, "Anomaly Detection of CAN Bus Messages through Analysis of ID Sequences," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017: IEEE, pp. 1577-1583.

[16] M. Müter and N. Asaj, "Entropy-Based Anomaly Detection for In-Vehicle Networks," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, 2011: IEEE, pp. 1110-1115.

[17] F. Xu and M. Warkentin, "Integrating Elaboration Likelihood Model and Herd Theory in Information Security Message Persuasiveness," *Computers & Security*, vol. 98, p. 102009, 2020.

[18] H. Lee, S. H. Jeong, and H. K. Kim, "OTIDS: A Novel Intrusion Detection System for In-Vehicle Network by Using Remote Frame," in *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, 2017: IEEE, pp. 57-5709.

[19] S.-F. Lokman, A. T. Othman, and M.-H. Abu-Bakar, "Intrusion Detection System for Automotive Controller Area Network (CAN) Bus System: A Review," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1-17, 2019.

[20] S. Halder, M. Conti, and S. K. Das, "COIDS: A Clock Offset Based Intrusion Detection System for Controller Area Networks," in *Proceedings of the 21st International Conference on Distributed Computing and Networking*, 2020, pp. 1-10.

[21] K.-T. Cho and K. G. Shin, "Fingerprinting Electronic Control Units for Vehicle Intrusion Detection," in *25th USENIX Security Symposium (USENIX Security 16)*, 2016, pp. 911-927.

[22] Y. Zhao, Y. Xun, and J. Liu, "ClockIDS: A Real-time Vehicle Intrusion Detection System Based on Clock Skews," *IEEE Internet of Things Journal*, 2022.

[23] W. Choi, H. J. Jo, S. Woo, J. Y. Chun, J. Park, and D. H. Lee, "Identifying ECUs Using Inimitable Characteristics of Signals in Controller Area Networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 4757-4770, 2018.

[24] J. Zhou *et al.*, "A Model-Based Method for Enabling Source Mapping and Intrusion Detection on Proprietary CAN Bus," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[25] S. U. Sagong, X. Ying, A. Clark, L. Bushnell, and R. Poovendran, "Cloaking The Clock: Emulating Clock Skew in Controller Area Networks," in *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, 2018: IEEE, pp. 32-42.

[26] J. Du and Y.-C. Wu, "Distributed Clock Skew and Offset Estimation in Wireless Sensor Networks: Asynchronous Algorithm and Convergence Analysis," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5908-5917, 2013.

[27] W. Choi, K. Joo, H. J. Jo, M. C. Park, and D. H. Lee, "VoltageIDS: Low-Level Communication Characteristics for Automotive Intrusion Detection System," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 8, pp. 2114-2129, 2018.

[28] R. Bhatia, V. Kumar, K. Serag, Z. B. Celik, M. Payer, and D. Xu, "Evading Voltage-Based Intrusion Detection on Automotive CAN," in *Network and Distributed System Security Symposium (NDSS)*, 2021.

[29] M. Kneib and C. Huth, "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*, 2018, pp. 787-800.

[30] S. Jana and S. K. Kasera, "On Fast and Accurate Detection of Unauthorized Wireless Access Points Using Clock Skews," *IEEE transactions on Mobile Computing*, vol. 9, no. 3, pp. 449-462, 2009.

[31] Aldweesh, Arwa, Abdelouahid Derhab, and Ahmed Z. Emam. "Deep Learning Approaches for Anomaly-Based Intrusion Detection Systems: A Survey, taxonomy, and open issues." *Knowledge-Based Systems* 189 (2020): 105124.

[32] Debar, Hervé, Marc Dacier, and Andreas Wespi. "Towards A Taxonomy of Intrusion-Detection Systems." *Computer networks* 31.8 (1999): 805-822.

[33] Goodman, Nathaniel R. "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)." The Annals of mathematical statistics 34, no. 1 (1963): 152-177.

[34] Michael, Mark, and Wen-Chun Lin. "Experimental Study of Information Measure and Inter-Intra Class Distance Ratios on Feature Selection and Orderings." *IEEE Transactions on Systems, Man, and Cybernetics* 2 (1973): 172-181.

Hongqian Wei received the B.S. and Ph.D. degrees in power machinery and engineering from the Beijing Institute of Technology, Beijing, China, in 2016 and 2022, respectively. From 2019 to 2021, He was a visiting scholar with the School of Engineering in Cardiff University, Wales, UK, sponsored by the China Scholarship Council. He is currently a postdoctoral research associate with the School of Mechanical Engineering, Beijing Institute of Technology. His research interests include the vehicle dynamics control and in-vehicle network security for intelligent vehicles. His research has been supported by the National Natural Science Foundation of China and China Postdoctoral Science Foundation.

Qiang Ai was born in Tangshan, Hebei, China, in 1994. He received the B.S. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China. His research interests include permanent magnet synchronous motors, inverters, functional safety, and new energy vehicles.

Wenqiang Zhao was born in Chifeng, China. He received the M.S degree in Power engineering and Engineering Thermophysics from Beijing Institute of Technology in 2019. He used to work on energy management of hybrid electric vehicles in Benz Corporation. Now he is working toward the Doctoral degree in vehicle control and electric drive with the School of Mechanical Engineering at Beijing Institute of Technology. His current research interests include the design of motor controller and vehicle controller.

Youtong Zhang was born in Jilin City, Jilin, China, in 1965. He received the B.S. and M.S. degrees in engine engineering from Jilin University, Changchun, in 1990, and the Ph.D. degree from the College of Mechatronics Engineering, Beijing Institute of Technology, Beijing, China, in 1995. He is currently a Professor with the College of Mechatronics Engineering, Beijing Institute of Technology. He is also the Director of the Laboratory of Low Emission Vehicle, Beijing Institute of Technology. He authored three books and has more than 200 articles. His research interests include the engine electronic control, electrical drive of new energy vehicles, in-vehicle network safety for intelligent vehicles and intelligent farm machinery system. His research has been supported by the National Key Research and Development Program of China.