**ECUPrint - Physical Fingerprinting Electronic Control Units on CAN Buses inside Cars and SAE J1939 Compliant Vehicles**

Lucian Popa, Bogdan Groza, Camil Jichici and Pal-Stefan Murvay

**Abstract**—We fingerprint 54 ECUs from 10 cars, one of them being a heavy-duty vehicle that is compliant to the SAE J1939 standard. These later specifications implemented in commercial vehicles offer concrete sender addresses in every CAN frame, making physical characteristics easier to link to specific ECUs. This is not the case for traffic collected inside passenger cars where the allocation of CAN bus identifiers is non-uniform, without explicit sender and receiver addresses, making ECU identification more challenging. While previous research has shown good separation between ECUs even when single features are used, e.g., skews or maximum voltage level, prior results are based on a small number of cars, while our larger experimental basis proves that single features are likely insufficient to separate between a large number of ECUs. Concretely, for a crisp separation, at least four features seem to be needed, i.e., mean voltage, max voltage, bit time and plateau time, while clock skews or any single voltage feature lead to overlaps. We provide clear experimental bounds on the intra and inter-distances between ECUs in distinct cars. We use clock skews and, more importantly, several voltage features to test our methodology on a larger experimental basis: 10 cars containing 54 ECUs.

One of the most common use cases for physical fingerprinting is the development of intrusion detection systems, we enumerate several works in this direction in the related work section. However, in this work we are mostly interested in the extraction of such fingerprints to uniquely identify the ECUs inside a vehicle for forensics purposes. It is worrisome that, according to recent data from the 2020 FBI crime statistics report[1], car theft has increased in recent years. Moreover, a new concern has emerged in the context of vehicle identity theft: VIN cloning. By this attack, the vehicle identification number (VIN) of an existing car is cloned. A report from the National Crime Prevention Council[2] shows that many stolen vehicles also rely on cloned VINs. Physical fingerprints may help to alleviate such problems as they can be stored in authorized databases and inspected/updated as the vehicle goes to standard procedures such as the annual (or biennial) technical safety inspection which is mandatory all over the world. No less, such fingerprints may be checked during regular traffic inspections, which are even more common for heavy duty vehicles (SAE J1939 compliant). While it is true that these fingerprints may vary over time, it is expected that the periodic collection of such fingerprints will offer better clues on how the fingerprints will vary, giving precious hints for forensics and authentication purposes. Ultimately, our work is concerned in showing the intra and inter-distances between such fingerprints over a large number of ECUs, i.e., 54 from

**Keywords**—CAN bus, clock skews, ECU fingerprinting, J1939

**I. INTRODUCTION AND MOTIVATION**

The CAN bus has more than a decade of reported security vulnerabilities[11, 2, 3]. This doesn’t come as a surprise for a bus designed in the 80s. In fact, its survival for so many decades is a strong argument to support its future use, despite some obvious security shortcomings. Moreover, the recent introduction of the higher speed CAN-FD makes it clear that CAN will be present inside cars for the decades that follow. The larger payload of CAN-FD will make it much easier to accommodate cryptographic material inside frames responding to modern security needs. Physical fingerprinting techniques have a complementary role, e.g., for authentication or forensics, which is not going to be ruled out by cryptographic security. There are many incidents which prove this is so. For example, by corrupting an existing ECU (Electronic Control Unit), cryptographic keys can be extracted, as demonstrated in a recent attack by which diagnostic security keys were extracted from a real-world vehicle[4]. Other research works have shown that it is possible to compromise the central information unit even from remote[5] or have attacked ECUs via over-the-air update protocols[6]. Cryptographic keys may also be extracted by various side-channel attacks, these attacks are commonly reported in embedded systems[7] and have been also demonstrated on in-vehicle controllers[8, 9]. Testing vulnerabilities against such attacks is now a required practice in the automotive industry[10]. Consequently, physical fingerprinting technologies are here to stay as an additional layer of protection besides cryptographic security.

**Potential use cases.** Our work goes mostly in the same vein as the work from[11] which attempts to map the existing ECUs inside the vehicles based on physical characteristics. However, the authors in[11] rely exclusively on clock characteristics which prove to be unsatisfactory in our analysis over a larger experimental basis due to obvious overlaps between ECUs in distinct cars. We use clock skews and, more importantly, several voltage features to test our methodology on a larger experimental basis: 10 cars containing 54 ECUs.

Our work goes mostly in the same vein as the work from[11] which attempts to map the existing ECUs inside the vehicles based on physical characteristics. However, the authors in[11] rely exclusively on clock characteristics which prove to be unsatisfactory in our analysis over a larger experimental basis due to obvious overlaps between ECUs in distinct cars. We use clock skews and, more importantly, several voltage features to test our methodology on a larger experimental basis: 10 cars containing 54 ECUs.

One of the most common use cases for physical fingerprinting is the development of intrusion detection systems, we enumerate several works in this direction in the related work section. However, in this work we are mostly interested in the extraction of such fingerprints to uniquely identify the ECUs inside a vehicle for forensics purposes. It is worrisome that, according to recent data from the 2020 FBI crime statistics report[1], car theft has increased in recent years. Moreover, a new concern has emerged in the context of vehicle identity theft: VIN cloning. By this attack, the vehicle identification number (VIN) of an existing car is cloned. A report from the National Crime Prevention Council[2] shows that many stolen vehicles also rely on cloned VINs. Physical fingerprints may help to alleviate such problems as they can be stored in authorized databases and inspected/updated as the vehicle goes to standard procedures such as the annual (or biennial) technical safety inspection which is mandatory all over the world. No less, such fingerprints may be checked during regular traffic inspections, which are even more common for heavy duty vehicles (SAE J1939 compliant). While it is true that these fingerprints may vary over time, it is expected that the periodic collection of such fingerprints will offer better clues on how the fingerprints will vary, giving precious hints for forensics and authentication purposes. Ultimately, our work is concerned in showing the intra and inter-distances between such fingerprints over a large number of ECUs, i.e., 54 from
Sources for fingerprinting. Figure 1 depicts the physical components that exist inside an ECU which are needed to support its functions, including CAN communication. Automotive ECUs are supplied from the battery voltage \textit{VBAT} and connected to the vehicle ground \textit{GND}. Internal supplies for the microcontroller and CAN transceiver denoted as \textit{VCC-M} and \textit{VCC-T} are provided from the ECU power management module denoted in the figure as \textit{PWR}. CAN frames are transmitted and received using the CAN-TX and CAN-RX serial lines by the CAN Controller usually embedded inside the Microcontroller. These frames are converted to and from differential voltages on the CAN-H (CAN High) and CAN-L (CAN Low) lines by a CAN Transceiver. An external oscillator OSC provides the clock signal \textit{CLK} required by the Microcontroller to manage the internal operations allowing it to perform specific timing related actions, e.g., adjust CAN bit time or transmit periodic CAN frames, etc.

In the light of the above, there are two main physical characteristics that are exposed by ECUs on the CAN bus: i) clock skews which can be extracted from the timing of CAN frames and ii) voltages which can be extracted from the physical signal on the bus. Both these characteristics have been used by previous works to identify ECUs, although to the best of our knowledge no previous work has addressed them both in a single paper for the same or several vehicles. Table I provides a summary of the vehicles used in our experiments and the amount of data that has been collected from each of them. Our study focuses on 10 vehicles from which we sample more than two hundred thousand bits for voltage fingerprinting and collect more than eight million frames for skew computations that lead to the identification of 54 ECUs.

Therefore, one of the first contributions that we exhibit is to provide comprehensive data on these two types of fingerprints for ECUs inside vehicles, ranging from personal cars up to heavy-duty vehicles that comply with the SAE J1939 standard. The advantage offered by heavy-duty vehicles is that the SAE J1939 standard provides clear information on the source of the messages and fingerprints can be directly linked to a specific sender. In case of passenger cars, such information is largely missing but we show that the CAN frames nicely cluster around specific ECUs and thus the fingerprint of each sender is easy to recognize.

A few words on the relevance of using both timing and voltage characteristics may be in order. Table I presents a brief summary on the weak and strong points for skews and voltages. Skews are easier to collect and they may be preserved when frames are retransmitted by gateways. This depends on the specific implementation, i.e., it will work in case when the retransmission is on-event, similar to the case of computers connected via gateways over large networks as previous works have shown that computers can be fingerprinted from remote [12] despite the fact that packets are running over multi hops in the network. If the gateway buffers the frames and transmits them using interrupts generated by the local timers, then indeed, the skew will not be preserved. Also, the frame arrival time allows the implementation of covert channels based on frame inter-arrival timings [13], [14]. The downside of skews is that they are easier to forge [15]. Moreover, they will not work for on-event transmissions and are affected by processing and arbitration delays. Extracting skews also requires large number of frames for correct estimation. On the other hand, voltage levels are harder to forge, they are feature rich and a single bit may be sufficient to recognize the sender. On the downside, they are harder to collect as they require high sampling rate ADCs and physical access to the bus. Because of this mixed image with pros and cons for both, it seems that using both timing and voltage fingerprints may help and we try to present them in a comparative manner in our work.

Summary of contributions. Briefly, the contributions of our work can be summarized as follows:

1. We collect comprehensive experimental data on skews and voltages from 10 vehicles, ranging from small cars to SUVs and a heavy-duty vehicle, totaling 54 ECUs. Our datasets will be made public and we hope that they can serve for future research works concerned with designing fingerprinting methodologies or intrusion detection systems based on physical characteristics.

2. From the collected data, we extract and analyze five features, partly used in previous research works, which can help to link each frame with a physical ECU and serve as a compact fingerprint for the ECU: clock skews, mean voltage, maximum voltage, bit and plateau time.

3. We present results on skews and the extracted voltage features in a comparative manner, showing concrete data on the intra and inter-distances, which are generally neglected in previous works, with respect to the aforementioned physical features for 54 ECUs.

4. We give a clear quantitative depiction on how these characteristics vary while two of the vehicles are operating for 1 hour.

The rest of the paper is organized as follows. Section II discusses the background on CAN, J1939 and related works while in Section III we show how data collection was performed and present the vehicles from our experiments. In Section IV
we set up the theoretical framework for our analysis. Section V presents the results over the collected experimental data. Finally, in Section VI, we state the conclusions of our work.

II. BACKGROUND AND RELATED WORKS

In this section we provide some brief background on CAN and the J1939 upper layer implementation that is present in one of our vehicles, then we account the related works.

A. Background on CAN and J1939

CAN networks allowing electronic control units or sensors to exchange data at a baudrate of up to 1Mbps using a twisted two-wire cable. Transmissions over the two differential lines, CAN-H and CAN-L, are encoded using dominant (logical '0') and recessive (logical '1') bits and organized into frames with a specific bit structure. Voltage levels for transmission and reception lines between the microcontroller and the CAN transceiver, i.e. CAN-TX and CAN-RX, are nominally 5V TTL for a recessive bit and 0V TTL for a dominant bit. The voltage level is around 2.5V for both differential lines, i.e. CAN-H and CAN-L, when the bus is idle or a recessive bit is transmitted and around 1.5V on CAN-L with 3.5V on CAN-H when a dominant bit is transmitted.

Each frame starts with a dominant bit which represents the start of frame field followed by the arbitration field containing the frame identifier of 11-bit for standard frames and 29-bit for extended frames. Next is the control field which includes the number of bytes contained in the frame data field. The data field holds the actual frame content of up to 8 bytes and is followed by a CRC field used after reception for data integrity check. The last parts of the frame are the acknowledge field and the end-of-frame field. Since two or more nodes may start bit transmission at the same time, all nodes compared each own bit transmission with the actual bus value during the arbitration field. Nodes detecting a lost arbitration will have to wait for the bus to be idle again and restart the transmission. Each receiver must acknowledge the successful frame reception transmitting a dominant bit during the acknowledge field. The standard frame and extended frame structures, i.e. 11-bit ID and 29-bit ID, are shown in Figure 2. The extended frame has a higher arbitration field and structure bit count when compared to the standard frame with 20 additional bits, 18 bits for the identifier, one for the SRR (substitute remote request) bit which is always recessive and one reserved. The IDE (identifier extension) bit is recessive and part of the arbitration field for extended frames while for standard frames it is dominant and part of the control field. There are two reserved bits, i.e., r0 and r1, in the control field for extended frames and only one for the standard frame.

![Frame size and structure illustrated for standard CAN frames (top) and extended CAN frames (bottom)](image)

**J1939 specifics.** Various higher layer protocols were defined over CAN such as ISO-TP and UDS for diagnostics, DeviceNet for industrial automation or SAE J1939 for commercial vehicles. All these protocols make use of the CAN physical and data link layers described earlier and define specific behavior at other OSI layers. According to the standard [16], J1939 uses 29-bit identifiers which are only found in extended frames. The 29-bit J1939 frame identifier contains a frame-priority field of 3 bits used to optimize bus traffic, a 1 bit extended data page field, required to be 0 for J1939, a data page field of 1 bit which defines the page of the Parameter Group Number (PGN), the PDU Format field of 8 bits, the PDU specific field of 8 bits and the 8-bit Source Address which identifies the frame transmitter. Source addresses are unique values and are assigned to network nodes statically or dynamically after initialization according to the J1939 standard.

---

**TABLE I**

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Model year</th>
<th>No. ECU’s</th>
<th>No. IDs</th>
<th>Busload</th>
<th>Temperature</th>
<th>Battery voltage</th>
<th>Collected frames (skew)</th>
<th>Collected bits (voltage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honda Civic</td>
<td>2012-2017</td>
<td>6</td>
<td>43</td>
<td>31%</td>
<td>9°C</td>
<td>12.3 V</td>
<td>1,039,512</td>
<td>40,075</td>
</tr>
<tr>
<td>Opel Corsa</td>
<td>2006-2014</td>
<td>4</td>
<td>29</td>
<td>23%</td>
<td>8°C</td>
<td>13.9 V</td>
<td>442,992</td>
<td>9,187</td>
</tr>
<tr>
<td>Hyundai i20</td>
<td>2014-2020</td>
<td>7</td>
<td>40</td>
<td>35%</td>
<td>12°C</td>
<td>14.2 V</td>
<td>616,296</td>
<td>17,767</td>
</tr>
<tr>
<td>John Deere Tract.</td>
<td>2010-2018</td>
<td>3</td>
<td>33</td>
<td>19%</td>
<td>4°C</td>
<td>14.2 V</td>
<td>154,779</td>
<td>4,031</td>
</tr>
<tr>
<td>Dacia Duster</td>
<td>2010-2017</td>
<td>3</td>
<td>12</td>
<td>14%</td>
<td>10°C</td>
<td>14.4 V</td>
<td>247,154</td>
<td>9,086</td>
</tr>
<tr>
<td>Dacia Logan</td>
<td>2012-2019</td>
<td>6</td>
<td>46</td>
<td>14%</td>
<td>10°C</td>
<td>12.6 V</td>
<td>629,662</td>
<td>31,579</td>
</tr>
<tr>
<td>Hyundai ix35</td>
<td>2009-2015</td>
<td>6</td>
<td>26</td>
<td>45%</td>
<td>9°C</td>
<td>13.5 V</td>
<td>847,161</td>
<td>23,104</td>
</tr>
<tr>
<td>Ford Fiesta</td>
<td>2017-2020</td>
<td>6</td>
<td>46</td>
<td>51%</td>
<td>5-7°C</td>
<td>14.9 V</td>
<td>2,243,359</td>
<td>43,861</td>
</tr>
<tr>
<td>Ford Kuga</td>
<td>2013-2019</td>
<td>9</td>
<td>70</td>
<td>65%</td>
<td>9°C</td>
<td>13.7 V</td>
<td>1,233,545</td>
<td>28,024</td>
</tr>
<tr>
<td>Ford Ecosport</td>
<td>2018-2021</td>
<td>4</td>
<td>87</td>
<td>43%</td>
<td>9°C</td>
<td>15.0 V</td>
<td>759,421</td>
<td>22,808</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>54</td>
<td>432</td>
<td></td>
<td></td>
<td></td>
<td>8,213,381</td>
<td>229,510</td>
</tr>
</tbody>
</table>

* the dataset is publicly released to serve for future investigations and can be retrieved from the ECUPrint project on GitHub and the authors institution server

**TABLE II**

<table>
<thead>
<tr>
<th>Skews</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ easy to collect</td>
<td>✓ easier to forge</td>
<td>✓ may not work for on-event frames</td>
</tr>
<tr>
<td>✓ may be preserved through gateways (possible to retrieve from distinct buses)</td>
<td>✓ affected by arbitration and processing delay</td>
<td>✓ many frames for estimation</td>
</tr>
<tr>
<td>✓ single bit/frame is sufficient</td>
<td>✓ harder to collect, may require high sampling rate ADCs</td>
<td>✓ harder to collect, may require high sampling rate ADCs</td>
</tr>
<tr>
<td>✓ feature rich fingerprint</td>
<td>✓ require physical access to the same bus</td>
<td>✓ require many frames for estimation</td>
</tr>
</tbody>
</table>

---

**REFERENCES**


B. Related works

The academic community was quick to react with various solutions for preventing attacks on in-vehicle networks reported in past years [2], [17], due to the inherent vulnerabilities of the CAN protocol. One way intruders can be prevented from transmitting malicious signals on the vehicle bus is to include authentication data inside CAN frames [18] but considering the maximum allowed payload for a CAN frame, i.e., 8 bytes, there is little space to fit both the authentication tag and actual data. The AUTOSAR standard for Secure On-board Communication (SecOC) [19] recommends truncated message authentication codes, i.e., MACs, and random freshness data bits to be included in frames with safety critical information. Another proposal is to use ID-hopping techniques [20], [21] for frame identifiers which makes it very hard or impossible for an attacker to find the correlation between the frame identifier and frame content.

Other lines of work have considered fingerprinting ECUs, removing the need of modifying the CAN frame content, by using physical characteristics, i.e., clock skews derived from cyclic frames or voltage fingerprints. Clock skews have been initially proposed to fingerprint computers [12] and well after that they have been shown to be effective in fingerprinting smartphones [22]. They have also been analyzed as possible features that can be used to detect intrusions in several types of networks such as nodes in wireless sensor networks [23] and access points in wireless networks [24]. Only recently they were suggested for in-vehicle networks to identify electronic control units from the bus [25]. However, clock skews can be faked by the host and their use as fingerprints is vulnerable to cloaking attacks as shown by [15], [26]. Clock skews are still efficient for identifying legitimate ECUs which will not change their skew to evade correct classification [11]. Clock offset variations over temperature were analyzed in [27] where temperature-varied ECU fingerprinting is discussed for source identification and intrusion detection.

Fingerprinting ECUs based on voltages is a research topic that attracted much more interest in recent years. While similar approaches were used in the past for fingerprinting devices in wireless communication [28] and even wired Ethernet [29], the first work to present voltage based physical fingerprinting of CAN nodes by using basic signal processing tools is [30]. The authors in [31] used voltage profiles and ACK voltage thresholds from CAN frames to fingerprint ECUs and detect adversarial ECUs testing their approach both on experimental and real in-vehicle networks. Choi et. al [32] have used an experimental setup based on Arduino boards and real data from two cars, a Hyundai Sonata and a Kia Soul. Single frame based physical fingerprinting has been shown as feasible in [33] on extracted voltage information done with an oscilloscope and a USB data acquisition board.

A different approach is presented in [34] where voltage samples are split in 3 separate groups before being used for fingerprinting. Rising edge, falling edge and the dominant bit level are evaluated using several statistical characteristics: mean, standard deviation, variance, skewness, etc. Two vehicles were used for data collection: a Fiat 500 and a Porsche Panamera, each of them having 6 ECUs on the analyzed bus. An extension of the work was done in [35] with more details related to the outside temperature when the voltage data was collected, the number of frames and other properties for each data set. Edge based identification is studied in a recent work [36] using voltage data collected with a PicoScope 5204 and a resolution of 8 bits down to a sample rate of 2 MS/s. Kneib et. al [37] analyze the impact of voltage sampling method for a previously proposed intrusion detection methods [34], [35]. They emphasize the effects of the sampling selection over the signal quality and perform an evaluation of the performance impact change due to the sampling method on a vehicle concluding that using an average of samples is the best option for the IDS. The same authors propose a low-resource constraint voltage-based intrusion detection method that can be implemented on automotive graded microcontrollers [38].

Average and standard deviation of voltage distribution were considered as features for voltage fingerprinting done in [39] on a CAN bus prototype with nodes communicating at 500 kbps. The bus contained 9 different types of ECUs along with a dedicated node which verified if the transmitters are genuine based on the specified fingerprints proving that their method reduces the false alarm rate, miss detection rate and the authentication latency if compared to the proposal from [31]. A common aspect with all these approaches is that they all fingerprint CAN nodes based on signals transmitted by individual nodes. In contrast, the work in [40] proposes fingerprinting the network layout with the ability to detect the insertion of new nodes based on time domain reflectometry. Fingerprinting ECUs using the bit time was proposed in [41] where the authors define a classification model using statistical measurements of the collected data, e.g., mean, standard deviation, etc. Similarly, the physical characteristics used in [42], i.e., voltage thresholds, rising and falling edge, open the room for detecting spoofing and bus-off attacks. Authors from [43] suggest a voltage based IDS that can be connected on the CAN bus as an independent device without influencing the behavior of the CAN network, e.g., bandwidth, and has an accuracy of more than 97% for detection of malicious voltage signals.

Xiao et. al [44] define a CAN bus authentication scheme using reinforcement learning that is based on the CAN bus physical layer features, i.e., signal voltages and arrival intervals and emphasize the framework evaluation experiments performed on a 500 kbps CAN bus with 18 legitimate ECUs and one adversarial ECU.

Additionally to the voltage fingerprinting proposals, authors from [45] evaluate voltage corruption techniques that would result in masquerade attacks not detected by voltage-based intrusion detection systems. As prevention for the voltage corruption attacks, they propose a defense method that is applied during re-training due to required fingerprint updates.

III. VEHICLES AND DATA COLLECTION

In this section we present the vehicles that were subject to our experiments, then we describe the tools that were used for data collection.
A. The vehicles in our experiments

For evaluating the physical characteristics of automotive control units, we considered 10 different vehicles, 9 cars and 1 tractor, for which we monitored the in-vehicle CAN communication. The cars used fall in three different body configurations: hatchback (Hyundai i20, Ford Fiesta, Opel Corsa, Dacia Logan), sedan (Honda Civic) and SUV (Dacia Duster, Hyundai ix35, Ford Kuga, Ford Ecosport) with manufacturing dates between 2006 and 2021 as already shown in Table 1. The diversity of cars was chosen by considering that Ford vehicles are designed in the United States of America, Hyundai and Honda originate from the Asian Market while Opel and Dacia have the design facilities in Europe. Also, the tractor employed is a contemporary machinery produced by John Deere, a famed worldwide manufacturer of agriculture equipment.

The graphical depiction of vehicles used for collecting data for our experiments and their CAN bus networks is shown in Figure 2. The topologies illustrated in the figure are derived from the identification of the ECUs based on physical characteristics as we will later detail. This does not represent the exact wiring of the ECUs inside the car of which we are unaware and which is not within the scope of our work. We are specifically interested in the number of ECUs and their fingerprints.

For some of these vehicles, we were however able to identify the role of specific ECUs, as outlined in Figure 3, based on service manuals available on-line as discussed next. We identified the ECUs using the electrical wiring diagrams for the Ford Fiesta module communication networks posted at [46]. The gateway module (GWM), sync module (APIM), headlamp control module (HCM), powertrain control module (PCM), body control module (BCM) and parking aid module (PAM) are the 6 ECUs from the Fiesta OBD network. For Ford Kuga, the communication network diagram posted on [47] includes nodes communicating on the CAN bus connected to the OBD-II port, also named data link connector, i.e., DLC. In this case, we extracted data from 9 control units as follows. The keyless vehicle module (KVM), instrument cluster module (IC), powertrain control module (PCM), fuel additive system module (FUEL), ABS module (ABS), yawrate sensor (YAW), headlamp leveling module (HLM), all-wheel drive control unit (AWD) and electrohydraulic power steering module (EPS) are the Ford Kuga nodes. A Dacia Forum for UK owners [48] hosts the electrical wiring diagram for the Dacia Duster which includes the network diagrams. We used it to extract the ECU names connected on the OBD-CAN bus from the SSESP diagram. In this case there are only 3 ECUs, the ABS control unit (ABS), the injection system control unit (INJ) and the front/rear torque distribution control unit (FRTD). We determined that the John Deere tractor has 3 ECUs, the body control module (BCM), transmission control unit (TCM) and engine control module (ECM) based on the content of the frames specified in the J1939 standard.

For the rest of the vehicles, we do not know the specific role of the ECUs and we simply enumerate the ECUs in Figure 3. For Honda Civic we identified 6 distinct control units and for Opel Corsa 4 ECUs. On the Hyundai i20 we have determined that there are 7 nodes transmitting through the OBD II port. We determined that there are 6 ECUs connected on the Dacia Logan bus, 6 ECUs in the Hyundai ix35 diagnostic network and 4 ECUs in the Ford Ecosport OBD-CAN bus.

B. Tools for data collection

In order to enable the CAN data collection from the cars and tractor, we employed two devices: a CANcaseXL and a PicoScope 5000 Series. The CANcaseXL belongs the XL Family devices produced by Vector, the most widely used networking tools provider in the automotive domain. The CAN communication networks can be interfaced through CANcaseXL using a large variety of software tools developed by Vector, e.g., CANoe, CANalyzer, CANape. We built a custom application using the XL Driver Library which facilitates easy access to several protocol specific functions, e.g., receive CAN frames with a specific baud rate, and it also supports interfacing with the CANcaseXL.

We accessed CAN data from the cars and tractor using the diagnostic port connected to the internal CAN networks having the data collection setup as shown in Figure 4(i). The cars use communication based on the standard CAN specifications while the tractor uses SAE J1939 standard. Therefore, the J1939 OBD port has a specific shape and layout with 9 pins as described in [49]. The OBD port and the pins used to collect the data from the Ford Fiesta are shown in Figure 4(ii) while Figure 4(iii) depicts the specific J1939 9-PIN OBD port inside the tractor. Both pictures outline the CAN-H, CAN-L, GND pins which are of interest for collecting the skew and voltage data.

The clock skew data contains the frame identifiers and associated timestamps. In order to perform data collection for skews we built the XL Driver Library based application, configured with a baudrate of 250 kbps for the tractor and 500 kbps for the vehicles, to receive the available CAN frames through the diagnostic port. We connected the CAN cables directly to the diagnostic port CAN pins and interfaced them via a DB9 female connector to the CANcaseXL. Having the setup ready, we proceeded to log the CAN traffic over periods of 5 to 10 minutes for each vehicle while it was already turned on. Note that the CAN IDs that were collected are part of normal functioning and do not include the initialization phase, i.e., startup processes, in which some specific actions may be performed, e.g. address claiming procedures for J1939.

For voltage data collection, the probes of a 5000 Series Picoscope were connected to the CAN-H and CAN-L lines of the diagnostic port using a specific connector with accessible pins. Several files were extracted from each car and the tractor in order to capture the voltage data for all frames transmitted on the CAN bus, i.e., for each frame identifier, considering the frame identifiers extracted from the timestamp data collection step. We configured the voltage capture settings using the PicoScope 6 version 6.14.23.5207 software tool. Since the monitored lines have the maximum accepted voltage of 4.5V for CAN-H the voltage range configured on the PicoScope for both channels was of ± 5V. In order to collect an adequate amount of data required for fingerprinting transmitters, the
device when using two channels. Frames are transmitted by nodes from the tractor with a bit rate of 250 kbps according to SAE J1939-11, so, we were able to capture voltage data of up to 9-10 separate frames in a window of 5 milliseconds assuming a 500 microseconds average duration for one frame. For the cars, nodes exchange data using a bit rate of 500 kbps following the recommendation of SAE J2284-3 so it would take up to around 260 microseconds on average for a frame to be transmitted, allowing us to lower the capture window to 2 milliseconds. The window size and baud rate are the only differences between the tractor and the car data collection setup.

While the Picoscope tool is indeed a more expensive laboratory device, similar hardware components such as the Xilinx XC6SLX25 FPGA used in the Picoscope 5000 series or a high performance ADCs such as the 2 channel, 500 MSPS ADC08D502 from Texas Instruments are both below 100 USD. So the practical deployment of a tool with the capabilities required in our work should not be extremely expensive.

IV. FRAMEWORK FOR ANALYSIS

In this section we present the theoretical framework for the two fingerprinting methodologies: clock skews and voltage levels.

A. Clock skews

To set up a theoretical framework, we rely on the well established formalism from [50]. The clock of the system is defined as a piecewise continuous function $C : \mathbb{R} \rightarrow \mathbb{R}$ that is
twice differentiable. With respect to this function, we will use the offset of each device clock, which is the difference between the reported time and the true time, e.g., \( C_{\text{offset}} = C^{A}(t) - t \). Subsequently, the first derivative of the clock function \( C^{A}(t) \) represents the clock frequency and its second derivative \( C'^{A}(t) \) is the clock drift. The clock skew between two clocks can be then defined as the difference in the first derivative, i.e., \( C^{A,B}_{\text{skew}} = C'^{A}(t) - C'^{B}(t) \). Further, the drift \( C^{drift} \) of a clock relative to another is the difference between the second derivative of the clocks, i.e., \( C^{A,B}_{\text{drift}} = C''^{A}(t) - C''^{B}(t) \).

In our practical scenario, we consider that the true clock is the frame cycle time multiplied with the frame number, i.e., \( t = i \times \delta_{\text{id}} \) where \( \delta_{\text{id}} \) is the cycle time of the frame carrying a specific ID. The clock skew can be estimated from cyclic frames carrying a specific ID as: \( C_{\text{skew}}(id) \approx \frac{t_j - t_i}{(j - i)} \times \delta_{\text{id}} \).

Here \( t_i \) and \( t_j \) are the timestamps of the i-th and j-th frame respectively. We note that the work in [50] proposes the use of more complex linear approximation algorithms to extract the clock skews from network packets. The fact that CAN bus frames have fixed cycle time makes the previous simple formula to work reasonably well in estimating the skews if one simply takes the mean or median over the arrival time of sufficient frames. Of course, more demanding algorithms will lead to more accurate results.

To serve as a graphical example, Figure 5 shows the cycle time (i), clock offsets (ii) and the convergence of the skew (iii) computed with the above approximation for two IDs from a trace of 30 seconds. The left side of the figure is for an ID with a cycle time of 50ms and the right side for an ID with a cycle of 100ms. In our setup the true time is the expected arrival time of each frame based on its pre-defined cycle time and the current clock on the tool which records the data (a CANcase as earlier discussed in the data collection section). Note that while the cycle time is very distinctive for the two IDs, the slope of the clock offset is identical, i.e., the same skew equal to 1.00009, demonstrating that these IDs originate from the same ECU. The clock offset is significantly affected between the 50th and 150th frame for the ID on the right side of the figure, which is likely due to its lower priority. Computing the skew in this specific portion may lead to erroneous results. Such events are rare, but they show that clock skews may occasionally be unreliable and more frames have to be considered to get a correct image. As shown in (iii), for the first ID the value of the skew converges after just a dozen frames due to its more stable arrival time which is likely due to its higher priority. However, for the second ID the skew is wrongly estimated to be between 1.00010 and 1.00020 during frames 50–150. This suggests that for correct computation of the skew more than 100 frames may be needed. As a partial conclusion, skews are good for fingerprinting ECUs if there are sufficient frames but not very good for detecting intrusions in real-time since the arrival time of individual packets may be deceiving.

B. Voltage features

As already stated, another approach for fingerprinting ECUs is based on unique characteristics of the physical signals generated by each node on the CAN bus. These unique features can be observed by analyzing minute variations in the voltage levels of the CAN High and CAN Low lines. There are several causes for the uniqueness of signal characteristics. First, there are minute, uncontrollable, differences in the manufacturing process of CAN transceivers which will generate unique characteristics in their signaling behavior. Second, the shape of the transmitted signal is influenced by the transmission medium itself and all the other nodes located along the transmission path [51]. In the current analysis we focus on four features that can be extracted from the voltage levels: the mean voltage, maximum voltage, bit time and plateau time for isolated dominant bits, i.e., between two recessive bits. The last two characteristics depend also on the clock of the node. We now formally define these characteristics.

We use the following voltage features, which are going to be formalized next, and are already considered in related works: i) the mean voltage level in [34], [36], [52], ii) max voltage level in [32], [34], [36], [52] and iii) bit time in [41], [53], [54], [55]. In addition to these, we also note that the plateau time of the bit (not used in previous works) also provides good indications on the sender ECU, so we introduce this as an additional metric to compare it with existing metrics. Nonetheless, in the experimental analysis, by combining these four metrics the identification results are significantly improved. While machine learning techniques may give a much better separation between distinct ECUs, the experimental analysis that follows will show that these characteristics seem sufficient to distinguish between ECUs and thus they may serve as a compact fingerprint for ECUs inside the car.

By mean voltage we refer to the mean voltage during the dominant state of the bus, i.e., as the bit approaches
the nominal value of ∼2V. Other papers, have also included the rising and falling edges in computing the mean voltage, however, we only refer to the mean value as long as the bit stays in the dominant voltage range. For each frame carrying a specific ID, having the collected samples for an isolated dominant bit which is a transition from recessive to dominant and back, i.e., s1, s2, ..., sf, we define the mean and maximum value based on a fixed range τ as follows:

\[
\begin{align*}
V_{\text{mean}}(id) &= \text{mean}\left\{s_i : i = \ell / 2 - \tau, \ell / 2 + \tau \right\} \\
V_{\text{max}}(id) &= \max \left\{s_i : i = 1, \ell / 2 - \tau \right\}
\end{align*}
\]

The value of τ has to be chosen based on the duration of a bit and the number of samples per second such that the middle portion lies on the plateau. For example, in our experiments with a 500 kbps CAN and 500 MS/s we have \(\ell = 2000\) and one sample each 2ns for which we set the value of \(\tau = 150\) which fitted the plateau for all bits inside all cars from the experiments.

Let indexes \(\alpha \leq \ell / 2, \beta > \ell / 2\) and \(\sigma\) be the sampling time of the signal \(s_1, s_2, ..., s_f\), then we define the bit time and plateau time as follows:

\[
\begin{align*}
T_{\text{bit}}(id) &= \min_{\alpha, \beta} \left\{ (\beta - \alpha) \sigma : |s_\alpha| \leq \epsilon, |s_\beta| \leq \epsilon \right\} \\
T_{\text{plat}}(id) &= \max_{\alpha, \beta} \left\{ (\beta - \alpha) \sigma : |s_\alpha - V_{\text{mean}}(id)| \leq \epsilon \right\}
\end{align*}
\]

For the datasets that we collected from all cars, the sampling time \(\sigma\) is 2ns. The threshold \(\epsilon\) was selected based on empirical analysis on the intra/inter-distances at 20mV.

The previous definitions were concerned with isolated dominant bits that occurred in almost all of the IDs which we collected. If this is not the case, our approach can be easily scaled as follows. According to the CAN standard, each 5 consecutive bits of the same value are to be followed by a bit of distinct polarity (due to the CAN bit stuffing mechanism). The plateau time can be immediately extracted from multiple bits simply by dividing the entire plateau time with the number of bits (that is 5 in the worst case). Bit time can be derived from the overall time of multiple bits time by subtracting the plateau for all but 1 bit (that is the plateau time of 4 bits in the worst case). The mean and max voltage are unchanged when defined over multiple bits.

Figure 6 provides a graphical depiction of the voltage levels collected for IDs originating from distinct ECUs. The first ID is from Ford Ecosport and the second one is from Hyundai ix35. The first ID (i) shows a clear spike during the rising time with the extracted features \(V_{\text{mean}} = 2.195\,\text{mV}, V_{\text{max}} = 2.236\,\text{mV}, T_{\text{bit}} = 2.686\,\mu\text{s}, T_{\text{plat}} = 1.478\,\mu\text{s}\). The second ID (ii) has a flatter plateau with the extracted features \(V_{\text{mean}} = 2.191\,\text{mV}, V_{\text{max}} = 2.195\,\text{mV}, T_{\text{bit}} = 2.640\,\mu\text{s}, T_{\text{plat}} = 1.417\,\mu\text{s}\). Note that the figure presents the number of samples on the X-axis and each sampling point corresponds to 2ns, thus, the ∼1300 samples for the bit time lead to around 2.6 \(\mu\text{s}\).

As a partial conclusion, voltage levels offer much more features that can be extracted and they are also well suitable for analysis with more demanding machine learning algorithms. In this work we stay to the previously stated characteristics which are easy to collect and offer enough data to distinguish between the ECUs as shown later.

C. Distinguishability based on skews and voltage features

Having defined the skews and voltage features, we now measure the distinguishability of the fingerprints based on the intra and inter-distances between the fingerprints for each frame carrying a specific ID. Since the skews and voltages are real valued and not binary, we will use the Euclidean distance as a metric, i.e., \(d(u, v) = \sqrt{(u - v)^2}\) represents the Euclidean distance of two values which can be extended to any number of features. The intra and inter-distances based on skews, i.e., \(D_{\omega}^{\alpha}\), voltage features, i.e., \(D_{\omega}^{\text{mean}}, D_{\omega}^{\max}\), bit time, i.e., \(D_{\omega}^{\text{bit}}\), and plateau time, i.e., \(D_{\omega}^{\text{plat}}\), where \(\omega \in \{\text{inter}, \text{intra}\}\), are defined as follows:

\[
D_{\text{intra}}^{\alpha}(i) = \left\{d(\varphi(id'), \varphi(id'')) : \forall id', id'' \in \text{ECU}_i, id' \neq id'' \right\}
\]

\[
D_{\text{inter}}^{\alpha}(i, j) = \left\{d(\varphi(id'), \varphi(id'')) : \forall id' \in \text{ECU}_i, \forall id'' \in \text{ECU}_j \right\}
\]

Where the pair \((\alpha, \varphi)\) runs over the five fingerprints, i.e., \((\alpha, \varphi) \in \{(\text{skew}, C_{\text{skew}}), (\text{mean}, V_{\text{mean}}), (\max, V_{\text{max}}), (\text{bit}, T_{\text{bit}}), (\text{plat}, T_{\text{plat}})\}\) and \(i, j = 1, n\) run over all the \(n\) ECUs. That is, the intra-distances account for the distances between distinct IDs originating from the same ECU while the inter-distances account for the distance between IDs sent by distinct ECUs.

V. EXPERIMENTAL RESULTS

This section presents experimental results based on the previous methodologies and data collected inside the cars and the commercial vehicle.

A. ECU separation based on clock skews and voltage features

We now discuss the separation between ECUs in each vehicle based on clock skews and voltage features. The results that follow indicate that neither skews, nor single voltage features like mean and max voltage, are sufficient for a fine
grain separation. A specific issue with skews is that frames (IDs) coming from the same ECU may have distinct skews if they are replayed by a gateway ECU. For voltage levels, the mean and max voltage may at times be very close even for distinct ECUs. These situations are outlined next in the discussions according to the case.

1) **Honda Civic**: we determined that there are 6 ECUs and 43 IDs on the network. The Honda Civic provides a crisp separation between 6 ECUs. Figure 7 (i) shows the separation of the ECUs based on skews suggesting 6 ECUs right from the beginning. Regarding voltages, Figure 8 (i) shows the separation of the ECUs based on the mean-max voltage features. While a separation between 5 ECUs is clear, the ID from ECU3 is quite close to the IDs from ECU2 and ECU4 because of similar mean and max voltage features. However, the skew suggests this is a distinct ECU and it is indeed so as the voltage patterns from Figure 9 (i) and (ii) prove that ID 18E from ECU3 has a distinct pattern compared with ID 091 that is originating from ECU4. Figure 10 which provides separation based on the plateau and bit time...
(i) confirms that there are 6 clusters corresponding to 6 ECUs.

2) Opel Corsa: we determined that there are 4 ECUs and 29 IDs. The separation between the ECUs was one of the easiest from all cars. This was not only due to the small number of ECUs but also due to the very stable ID arrival timings and voltage features. Figure 8(ii) shows how the IDs separate along the mean-max voltages and leaves little doubts on the separation.

3) Hyundai i20: we determined that there are 7 ECUs and 40 IDs on the network. The separation between the 7 ECUs was effortless with no overlaps in terms of skews and voltages similar to the case of the Honda Civic. We depict only the mean-max voltage clustering of the IDs against the 6 ECUs in Figure 8(iii). Further details are given in the next section related to the inter and intra-distances.

4) John Deere: we determined 3 ECUs and 33 IDs to be present on the network. As stated, the address oriented nature of J1939 upper layer specifications provide excellent information for separating between ECUs as the sender addresses are already known. The resulting classification based on skews is presented in Figure 7(ii) which clearly shows the IDs clustering around three physical ECUs. This supports the findings from our analysis based on the J1939 specifications which, according to sender addresses indicates the presence of 3 ECUs. We have seen that the clock skews determined for 3 IDs out of the 21 IDs sent by the BCM, i.e., ECU1, are slightly different from the rest. However, the mean-max voltage level separation from Figure 8(iv) confirms that these originate from the same ECU but since the skew is slightly different for those IDs, we suspect that a 4th ECU may be present behind the BCM which acts like a gateway.

5) Dacia Duster: has 3 ECUs and 12 IDs. This was the simplest topology from our cars (only 3 ECUs similarly to the John Deere tractor) and with the smallest number of IDs. Figure 8(v) shows how the IDs separate along the mean-max voltages.

6) Dacia Logan: we determined 6 ECUs and 46 IDs. We expected a similar simplicity to the Dacia Duster but this was not the case. The fingerprint lead to the conclusion that this model has 6 ECUs. Interestingly, this car has the ECUs with the closest skews, with ECU2 and ECU3 differing with only 1ppm, i.e., 0.999973 vs. 0.999974. The voltages confirm that these ECUs are distinct with a large difference in the mean voltage levels of about 70mV. Figure 8(vi) shows how the IDs separate along the mean-max voltages. Figure 10(ii) shows the graphical separation based on the bit and plateau timings which confirms the presence of 6 distinct ECUs.

7) Hyundai ix35: we determined 6 ECUs and 26 IDs. The number of IDs is somewhat low for 6 ECUs which made us suspect that there may be a gateway behind the OBD port which filters and re-transmits incoming IDs from internal buses. The voltage features shown in Figure 8(vii) suggest however the presence of 6 distinct ECUs.

8) Ford Fiesta: we determined 6 ECUs and 46 IDs. The Ford Fiesta creates a few problems when attempting to separate between some of the IDs using skews. The main problem is that the timing of the IDs is not very stable. By taking different portions of the trace the computed skew was slightly different. Figure 11 gives such an example. ID 073(i) has a constant arrival time resulting in the same skew regardless of the portion of the trace. For IDs 364(ii) and 360(iii) there is a significant variation in the cycle time and the offsets change during the second half of the plots. This means that the skew that can be computed from the first 3000 frames is distinct from the one computed from the first 6000 frames. In this case the skew provides an unreliable separation between the ECUs. Fortunately, in this case, the mean-max ECU voltage separation works well as suggested in Figure 8(viii). Figure 10(iii) shows the bit and plateau time separation with 6 clusters corresponding to the 6 distinct ECUs.

9) Ford Kuga: we determined 9 ECUs and 70 IDs. We identified IDs that originate from the same ECU but have very distinctive skews again suggesting the presence of a gateway. The upper part of Figure 12 shows IDs 1D0 and 208, the first having a cycle time of 20ms and the second a cycle time of 25ms. This results in a skew of 0.999956 for the first and 1.002180 for the latter. We again suspect that the second one, i.e. ID 208, is redirected from another bus as the inter-arrival time, i.e., the cycle plot, shows significantly more noise. However, the mean-max voltage separation from Figure 8(ix) allows us to identify separate clusters corresponding to 9 ECUs. Figure 11(iv) contains the 9 groups of bit and plateau time which leads to the same 9 ECUs identified by the other two voltage features.

10) Ford Ecosport: we determined 4 ECUs and 87 IDs on the network. We noticed similar problems when computing the skews on Ford Ecosport as we had for Ford Kuga. The
Ecosport shows distinct skews for IDs coming from the same ECUs. The lower part of Figure 12 shows such differences for IDs 091 having a skew of 0.999610 and ID 045 having a skew of 0.999929 both having a cycle time of 20ms although they both originate from ECU4. Due to the clear differences in the resulted skews we suspect that ECU4 acts as a gateway and consequently some of the IDs come with a distinct skew. Similar problems with the skews occur at ECU1 while all the IDs from ECU2 and ECU3 separate well from the rest. Voltage separation is almost perfect between the 4 ECUs in the Ford Ecosport as it can be seen in Figure 8 (x).

**B. Overview of intra and inter-distances**

We now summarize the results on intra and inter-distances. Figure 13 shows the skews and voltage inter and intra-distances plotted as heatmaps. In the plots, we use a threshold of 10ppms for the skews, 25mV for the mean, max voltage separation and 10ns for the bit time and plateau time. The inter-distances are much cleaner for ECUs inside the same vehicles but there are far more collision in the inter-distances between ECUs in distinct cars. We discuss these in detail in what follows.

Skews provide a good separation for ECUs in the same vehicle as shown in Figure 13 (i). There are some exceptions considering the high variations in the skews extracted from the Ford Kuga and Ford Ecosport, i.e., there are more clusters based on skews compared to the number of ECUs that were separated through voltage features. Inter-distances are reasonable but insufficient as several overlaps exists, specifically for Dacia Duster. John Deere, Hyundai i20, Hyundai ix35 and Honda Civic. For example, there is a notable collision between ECU1 and ECU2 in the Dacia Duster and there are several collisions in the Ford Kuga or Ford Ecosport. Smaller collisions also occur between ECU1 and ECU2 in the Honda Civic or between ECU5 and ECU7 for Hyundai i20.

The mean and max voltage fingerprint from Figure 13 (iii), (iv) provide a crisper separation between the ECUs but there are still overlaps for intra-distances in the Ford Kuga, John Deere and Honda Civic. For Honda Civic there is an overlap between ECU3, ECU4 and ECU5 as for John Deere the overlap is between ECU1 and ECU2. Still, there is a notable collision between ECU4 in the Hyundai i20 and ECU6, ECU7 in the Ford Kuga. Smaller collisions also occur between ECU2 in the Opel Corsa and ECU3 in the Honda Civic or ECU4 in the Opel Corsa and ECU3 in the Ford Ecosport. The voltage patterns may still be distinctive and a finer grain analysis with machine learning algorithms will very likely distinguish between the ECUs. Such an analysis is out of scope for the current work but we leave our dataset public to be accessible for future works.

As expected since the bit time and plateau time depend both on voltages and clocks they yield a cleaner separation. The bit time heatmap from Figure 13 (v) is cleaner than the plateau time heatmap from Figure 13 (vi) with regards to inter-distances. Still, the bit time for several ECUs from the Honda Civic is close to the bit time for ECU1 or ECU4 from the Opel Corsa. Similarly, the bit times are close for some ECUs from the Hyundai i20, Dacia Logan, Hyundai ix35, Ford Fiesta and Ford Ecosport. As for inter-distances, there are negligible collisions for the bit time measured for 2 ECUs in Honda Civic, John Deere and Dacia Logan. Overlaps between the bit time for more ECUs are visible for the Hyundai i20, Hyundai ix35, Ford Fiesta, Ford Kuga and Ford Ecosport but there are no bit time overlaps for ECUs from Opel Corsa, Dacia Duster. The plateau time provides a clearer separation when it comes to ECUs from the same vehicle but there are more overlaps between ECUs from different vehicles as shown in Figure 13 (vi). The plateau time is also close for pairs of ECUs from John Deere, Dacia Logan, Ford Fiesta, Ford Kuga and the Ford Ecosport. The overlaps from inter-distance perspective are clearly visible between ECU2 from Honda Civic and ECU5 from Hyundai i20 or ECU1 from Opel Corsa and ECU4 from Hyundai i20. Same for ECU3 from Dacia Logan, ECU4 from Hyundai ix35, ECU2 from Ford Fiesta and ECU5 from Ford Kuga.

Finally, when merging multiple features, i.e., the mean, max, bit and plateau time, the overlaps almost fully disappear as Figure 13 (ii) shows. There is a minor overlap between the IDs from the Ford Fiesta and three other cars, Hyundai i20, Hyundai ix35 and Dacia Logan and some overlaps between IDs from the Ford Kuga and the Ford Ecosport ECUs, but the overlaps are minor. This suggests that at least 4 voltage

---

**Fig. 11.** Cycle time (left) and offset (right) for 3 IDs in the Ford Fiesta

**Fig. 12.** Cycle time and offsets for IDs 1D0 and 208 in Ford Kuga and IDs 045 and 091 in Ford Ecosport
features have to be used for a reliable separation.

C. Impact of environmental variations

We now put to test two cars from our experiments, i.e., the Honda Civic and the Ford Fiesta, against environmental variation effects on skews and voltages. For this purpose, we conducted tests by measuring the evolution of skews and voltages before and after one hour of normal car operation. It is known and well emphasized by previous research works that physical parameters will exhibit variations during car operation. However, there are no details in related works on how these values actually change. A reason for which we proceed to further investigations. Briefly, our investigations show that variations are non-uniform and not necessarily increasing or decreasing as the car operates. This suggests that predictions will be hard or impossible to make. The use-case presented by us in the introduction will likely call for the same environment, e.g., an authorized garage, and thus environment changes should be minimal. Clearly, a more comprehensive study will be needed in order to determine how such variations occur over a larger time interval, e.g., one or several years. Such a large interval was out of reach for us in the current research communication. If insufficient over long time intervals, the proposed fingerprinting methodology will at least allow one to construct a quick mapping of the ECUs inside the car based on data collected from the bus, in a similar vein to the work in [11].

Figure 14 shows the evolution of the skews before and after 1 hour runtime for the Honda Civic (left) and the Ford Fiesta (right). In both cars the values are shifting to the right, meaning either that clocks on ECUs run faster or that the CANCase clock (used in the measurements) actually runs slower. Since the shifts are unequal, we suspect that the controllers are mainly responsible for the shift in skews and thus the oscillator begin to run slightly faster as the car is running (this may also be due to the battery charging as the car is running).

Figure 15 shows the shifts in mean voltage levels after 1 hour of runtime. For the Ford Fiesta (left) the voltage is shifting to the right showing higher voltages after 1 hour of runtime. For the Honda Civic however, the shifts are not
always positive, notably, $ECU_6$ exhibits negative variations. The variations are in the 10-20mV range for the Ford Fiesta but for the Honda Civic they are much higher commonly around 40mV and topping for $ECU_3$ at 59mV. This is generally above our separation threshold of 20mV.

The increase in both clock speeds and voltage levels makes us suspect that as the car runs the battery charges and an increased voltage supply results in faster clocks and higher voltages on the bus. This rule however does not apply to one ECU in the Honda Civic which has a decreased voltage. These variations, due to environmental changes, indicate that physical fingerprints need to be continuously updated to avoid misclassification. This has been already pointed out by previous works, but no details were given regarding these variations. Since our investigations show that these variations are not necessarily monotonic, predicting them seems infeasible in the general case. Indeed, in the case of intrusion detection systems re-training and fingerprint updates proved to be necessary as already suggested in works like [31], [34].

VI. Conclusion

Clock skews are much easier to collect, but various abnormalities due to bus loads or computational load on the host controller make skews a much less accurate and less stable fingerprinting mechanism. They also require a significant amount of frames for correct estimation (dozens to hundreds) and deviations are possible for legitimate frames which make clock skews a poor indicator when a small number of samples is available. On the other hand, clock skews may work to identify ECUs behind gateways which cannot be separated based on voltages. Voltage features are more difficult to collect, in our analysis we used a digital oscilloscope while regular in-vehicle controllers usually don’t have this kind of signal acquisition capabilities. But they offer a much more accurate classification even based on single bits. Regarding voltage features, it seems that bit time and plateau time are a better feature to classify senders, however they are insufficient in a large pool of ECUs. In our analysis, grouping four features, i.e., mean voltage, max voltage, bit time and plateau time, seemed to yield a far better classification. Since our dataset is publicly available, we welcome future research works to try various other statistical tests or more demanding machine learning classifiers on the collected data.

REFERENCES


Lucian Popa started his PhD studies in 2018 at Politehnica University of Timisoara (UPT). He graduated his B.Sc in 2015 and his M.Sc studies in 2017 at the same university. He has a background of 4 years as a software developer and later system engineer in the automotive industry as former employee of Autoliv (2014 - 2018) and current employee of Veoneer (2018 - present). His research interests are in automotive security with focus on the security of in-vehicle buses.

Bogdan Groza is Professor at Politehnica University of Timisoara (UPT). He received his Dipl.Ing. and Ph.D. degree from UPT in 2004 and 2008 respectively. In 2016 he successfully defended his habilitation thesis having as core subject the design of cryptographic security for automotive embedded devices and networks. He has been actively involved inside UPT with the development of laboratories by Continental Automotive and Vector Informatik. Besides regular participation in national and international research projects in information security, he lead the CSEAMAN project (2015-2017) and the PRESENCE project (2018-2020), two research programs dedicated to automotive security funded by the Romanian Authority for Scientific Research and Innovation.

Camil Jichici is a PhD student at Politehnica University of Timisoara (UPT) since 2018 and works as a young researcher in the PRESENCE project. He received the Dipl.Ing. degree in 2016 and M.C. degree in 2018, both from UPT. His research interests are on the security of in-vehicle components and networks. He is also working as a software integrator in the automotive industry for Continental Corporation in Timisoara since 2014.

Pal-Stefan Murvay is Lecturer at Politehnica University of Timisoara (UPT). He graduated his B.Sc and M.Sc studies in 2008 and 2010 respectively and received his Ph.D. degree in 2014, all from UPT. He has a 10-year background as a software developer in the automotive industry. He worked as a postdoctoral researcher in the CSEAMAN project and is currently a senior researcher in the PRESENCE project. He also leads the SEVEN project related to automotive and industrial systems security. His current research interests are in the area of automotive security.

Due to space constraints, in this appendix we provide numerical data for the skews and voltages in four of the vehicles from our experiments (Dacia Logan, John Deere, Ford Fiesta and Honda Civic). Numerical data for the rest of the vehicles will be included as a supplemental material along with the datasets. All tables contain specific IDs grouped around each ECU. This association is based on the computed physical fingerprints (skews and voltages) to the best we could ascertain based on the methodology in this work. Since our dataset is public, we invite future research works to examine this association in more detail and possibly come with amendments.

The hexdecimal value of the frame identifier is presented the ID column and the cycle time of the frame in milliseconds, generally taken as a median, is shown in the Cycle column. The following columns contain data for the clock skew, $\Delta_{\text{skew}}$, the extracted mean voltage, $V_{\text{mean}}$, and maximum voltage, $V_{\text{max}}$, measured in Volts, and the identified bit time, $T_{\text{bit}}$, and plateau time, $T_{\text{plat}}$, measured in microseconds, from data collected while the vehicle was running, after it was started.

Data for the Dacia Logan is presented in Table III. Table IV contains the information for the John Deere tractor but omitting IDs 0CFFFF21, 18EAFF21, 18FEE500, 18FFFA21, 18FBB21, 1CEBFF00, 18FFFA21, 1CECF00 which are on-event frames and IDs which are used for multi-frame transmission. All omitted IDs for the John Deere tractor can be easily classified on the voltage levels.

Tables V and VI for the Honda Civic and Ford Fiesta have additional columns, $\Delta$, which contain the deviations compared to the column before for each value obtained from data collected after 60 minutes of driving. In Table V we illustrate the cyclic IDs with the associated information from Ford Fiesta. The following IDs: 455, 720, 727, 728, 72F, 7A5, 7AD, have been omitted since they are on-event (non-cyclic). The last of them in particular, occurred only when the car was started and was not seen again afterwards. As skews cannot be computed for on-event IDs, we remove them from the analysis in order to make the plots in Figure 13 comparable between skews and voltage features. If needed, these IDs can be classified based on the voltage features alone. Table VI contains the measurement data for the 43 IDs from Honda Civic.
### TABLE III

<table>
<thead>
<tr>
<th>No.</th>
<th>ECU ID</th>
<th>Cycle</th>
<th>$\text{V}_{\text{mean}}$</th>
<th>$\text{V}_{\text{max}}$</th>
<th>$\text{V}_{\text{Fmax}}$</th>
<th>$\text{T}_{\text{plat}}$</th>
<th>$\text{bit}_{\text{max}}$</th>
<th>$\text{skew}$</th>
<th>$\Delta$ $\text{V}_{\text{Fmax}}$</th>
<th>$\Delta$ $\text{T}_{\text{plat}}$</th>
<th>$\Delta$ $\text{bit}_{\text{max}}$</th>
<th>$\Delta$ skew</th>
<th>$\Delta$ ∆</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ECU1</td>
<td>186</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE IV

<table>
<thead>
<tr>
<th>No.</th>
<th>ECU ID</th>
<th>Cycle</th>
<th>$\text{V}_{\text{mean}}$</th>
<th>$\text{V}_{\text{max}}$</th>
<th>$\text{V}_{\text{Fmax}}$</th>
<th>$\text{T}_{\text{plat}}$</th>
<th>$\text{bit}_{\text{max}}$</th>
<th>$\text{skew}$</th>
<th>$\Delta$ $\text{V}_{\text{Fmax}}$</th>
<th>$\Delta$ $\text{T}_{\text{plat}}$</th>
<th>$\Delta$ $\text{bit}_{\text{max}}$</th>
<th>$\Delta$ skew</th>
<th>$\Delta$ ∆</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ECU1</td>
<td>186</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ECU2</td>
<td>187</td>
<td>0.999974 1.940 1.941 2.734 1.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>