R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 891

Effects of the Sampling Technique on Sender Identification Systems for the Controller Area Network

Marcel Kneib[1], Oleg Schell[2]

Abstract: As a result of the ongoing development of vehicle electronics and additional wireless communication interfaces, the possibilities for attacks and their negative consequences are increasing. Once an attacker has obtained access to the internal vehicle communication, in the case of the Controller Area Network (CAN) the attacker is able to forge all messages of the connected Electronic Control Units (ECUs) without a receiving ECU being able to recognize any suspicious behavior. The use of cryptographic methods is only possible to a limited extent due to restricted resources of the ECUs, which is why sender identification systems have been presented which are able to detect these kind of attacks. Presented approaches use different procedures to capture the analog signals on which the detection of attacks respectively the identification of the sender is based. This work shows that the impact on the performance of the sender identification system by the different sampling methods is minimal and therefore the selection of the appropriate technique can be mainly based on the available resources and the communication structure of the corresponding vehicle platform. This is shown on the one hand by the direct analysis of the analog signals captured from a real vehicle as well as by an evaluation of the previously introduced sampling methods using a recently published sender identification system. In addition, an assessment of the procedures based on different parameters shows which method is to be preferred for which application.

Keywords: Automotive Security; Sender Identification; Intrusion Detection

1 Introduction

The connectivity of modern vehicles, as well as the associated amount of interfaces, is constantly increasing. This trend does not only allow additional comfort functionalities and complex driver assistance systems, but also offers additional possibilities to attack a vehicle and its functions [HKD11; LL18]. Evidence that this is not only a theoretical threat was demonstrated by the attack of Miller and Valasek [MV15], as well as the latest research of the Tencent Keen Security Lab [Ca19]. Due to the absence of authenticity in the Controller Area Network (CAN) [Ro91], which is still the most commonly used bus technology in the automotive domain, an Electronic Control Unit (ECU) cannot check whether a received message was sent by a legitimate sender. This enables the forgery of messages, i.e. the execution of impersonation attacks. This problem still exists for its successors, CAN with flexible data rate (CAN-FD) [Ro12] and CAN-XL [CA20]. Unfortunately, the use of cryptographic methods is limited due to the constrained resources of the platforms used in

[1] Robert Bosch GmbH, Mittlerer Pfad 9, 70499 Stuttgart, Germany, marcel.kneib@de.bosch.com
[2] Bosch Engineering GmbH, Robert-Bosch-Allee 1, 74232 Abstatt, Germany, oleg.schell@de.bosch.com

doi:10.18420/inf2020_80

vehicles and the low payload and bandwidth of CAN. As an alternative or in combination with attack detection, methods have been presented in the past which provide sender identification on the basis of analog signals [Kn20]. Due to the static configuration of the internal vehicle communication, such systems allow to verify whether a message was sent by a valid ECU. For identification, however, the signals of CAN messages must first be recorded, for which the considered sender identification approaches suggest different procedures. While some methods capture the entire signal in order to extract the signal characteristics [Ch18; KH18], others concentrate on specific parts [Fo19] or individual points of a frame to determine the sender [KSH20]. The signal recording procedure has a corresponding effect on various properties, such as hardware requirements, cost, complexity and signal quality. In addition, the requirements and architecture of the actual system also have a major influence on the type of recording. This paper presents the different recording approaches and analyzes the associated effects on the relevant properties of sender identification systems for CAN. In addition, the associated performance is analyzed using the example of the recently presented work Edge-based Sender Identification (EASI) [KSH20] utilizing data from a series production vehicle. Furthermore, this work presents the individual application possibilities of the different sampling techniques, so that the reader is able to assess the optimal methodology with corresponding effects and constraints according to the respective requirements.

2 Sampling Approaches

For the CAN communication standard components are used, which can be produced in large quantities and very cost-effectively. These components only provide the connected microcontroller with access to the digital content of the message and not to the analog signals. For this reason, the actual recording of the signals must be independent of the existing hardware and therefore has to be considered and implemented by the respective sender identification approaches.

Since in principle every ECU can send a message at any time, it must be ensured that parallel transmission and thus corruption of the currently sent message does not occur. For this purpose, an ECU first checks whether the bus is free and then begins to send its message. In simplified form, the message consists of an unique message identifier, which also defines the priority of the message, and the associated content. During the transmission of the identifier it can happen that other ECUs start the transmission. The sending ECUs check whether their currently transmitted signal corresponds to the signal currently on the bus, and if not, the transmission of the respective ECU is stopped. The characteristics of the analog signals transmitted in this situation cannot be used for recognition as they contain characteristics of several ECUs. Therefore, all approaches focus on the segment succeeding the identifier.

As introduced in [KSH19], each frame consists of several symbols which represent the transmitted bits on the bus. As there are different kinds of symbols, the most approaches [Ch18; Fo19; KH18] first group those symbols according to the voltage transitions. There are four

different transition groups g, the rising ($g = 1$) and falling ($g = 2$) edges, and the stable high ($g = 3$) and low ($g = 0$) levels. The k-th symbol of a frame m sent by ECU e is defined by

$$S_k^{g,(e,m)} = (x_1, ..., x_l) \tag{1}$$

where, x_i, $i \in \{1, .., l\}$ are the individual voltage values of the symbol. Hence, elements of this l-tuple can be accessed according to the projection of the set theory with $S_k^{g,(e,m)}[i] = x_i$. Each symbol group, defined by

$$G^{g,(e,m)} = \bigcup_{k=1}^{K} S_k^{g,(e,m)}, \tag{2}$$

can contain a different amount K of symbols. Since the subsequent calculation of the characteristics from these symbols is computationally expensive, some approaches [Ch18; KH18] initially calculate an *average symbol* per group according to Equation (3), respectively use the averaged symbol directly as characteristic [Fo19].

$$\overline{S}^{g,(e,m)} = (\frac{1}{K}\sum_{k=1}^{K} S_k^{g,(e,m)}[1], \ldots, \frac{1}{K}\sum_{k=1}^{K} S_k^{g,(e,m)}[l]) \tag{3}$$

Another possibility is to utilize only a *single symbol* for the calculation of the characteristics. Without loss of generality, the first symbol of a group is considered for the further calculations, defined by

$$\mathring{S}^{g,(e,m)} = S_1^{g,(e,m)}. \tag{4}$$

The third variant, the *composite symbol* [KSH20], assembles the symbol from individual sample points of several symbols in a group. Based on the available number of samples per symbol L and the number of samples to be used per symbol P, the number of required symbols $K = \lceil \frac{L}{P} \rceil$, $P \leq L$ is given. For $L = 20$, $P = 2$, $K = 10$ according to

$$\tilde{S}^{g,(e,m)} = \bigcup_{p=1}^{P}\bigcup_{k=1}^{K} S_k^{g,(e,m)}[K * (p-1) + k], \tag{5}$$

the resulting sample points are illustrated in Fig. 1 for $g = 1$.

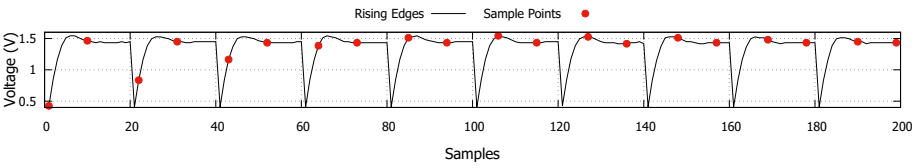

Fig. 1: Considered sampling points of the rising edges for the composite symbol.

3 Signal Analysis

3.1 Data set

For the initial analysis of the effect on the signal quality, a data set is reused which has already been utilized for the evaluation of sender identification approaches [KH18; KSH20]. The signals were recorded from a Fiat 500 which has six internal ECUs, each using up to seven different identifiers. In order to increase the number of ECUs, two additional Raspberry Pis were connected, each equipped with a CAN shield. One Raspberry Pi was connected to the bus in the trunk, while the other Pi was attached to the on-board diagnostics port together with a PicoScope 5204 at a sampling rate of 500 MS/s and a resolution of 8 bit. Since this data set is only slightly affected by changing environmental conditions, it allows the effects of the different sampling approaches to be analyzed as accurately as possible.

3.2 Metrics

In order to allow an approach-independent evaluation of the signal quality, the metrics *intra- and inter-distance* as well as their combination, the *inter-intra-distance* are used. The distance between two symbols $\mathcal{S}$ and $\mathcal{S}'$ regarding the considered sampling approaches, is defined by

$$Symbol\text{-}Distance(\mathcal{S}, \mathcal{S}', g, (e,m), (e',m')) = \frac{1}{L}\sum_{l=1}^{L}\left|1 - \frac{\mathcal{S}^{g,(e,m)}[l]}{\mathcal{S}'^{g,(e',m')}[l]}\right|. \tag{6}$$

The intra-distance, calculated by Equation (6) with $\mathcal{S} = \mathcal{S}', e = e'$ and $m \neq m'$, is used to evaluate the deviations of the symbol between all frames of a single ECU. In order to additionally analyze the symbol differences between the ECUs, the inter-distance is utilized, which is calculated by Equation (6) with $\mathcal{S} = \mathcal{S}'$ and $e \neq e'$ for all ECUs and their associated frames. Finally, the inter-intra-distance, i.e. the difference between the inter- and intra-distance, is used as the metric to assess the distance between the ECUs, taking into account the magnitude of the deviations of the ECUs with respect to the considered sampling approach.

Furthermore, the statement must be evaluated that the differences of the variations of the sampling approaches are negligible, since they are in the range of the natural variation of the symbols within a frame [KSH20]. Therefore, first the natural deviation of the symbols in the data set within a frame is calculated with $\mathcal{S} = S_k^{g,(e,m)}$ respectively $\mathcal{S}' = S_{k'}^{g,(e,m)}$ for $k \neq k'$. Following, for the comparison of the deviation and thus the verification of the statement, $\mathcal{S}'$ is replaced by the symbols created by the sampling approaches.

Tab. 1: Effect of the sampling technique on the signal quality.

	Intra-Distance	Inter-Distance	Inter-Intra-Distance
Average Symbol	0.3763 %	6.9193 %	6.5430 %
Single Symbol	0.6886 %	6.9351 %	6.2466 %
Composite Symbol	0.6886 %	6.9351 %	6.2466 %

3.3 Analysis

In Tab. 1 the calculated distances for the different sampling techniques are shown. The symbols $g = 1$, i.e. the rising edges, were used for the calculation, since these symbols contain the most important characteristics for distinguishability [KH18; KSH20]. For the average symbol it can be seen that it has the highest inter-intra-distance, mainly due to the lower intra-distance. This indicates that the use of the average symbol allows the best overall differentiation among all ECUs. However, it can also be seen that the single and composite symbols have no noticeable differences and, with an inter-intra-distance that is less than 0.3 % lower, the distinguishability is only minimally reduced.

Tab. 2: Effect of the sampling technique on the intra-frame deviation.

Data set	Average Symbol	Single Symbol	Composite Symbol
0.6425 %	0.4983 %	0.6230 %	0.6101 %

The deviations of the symbols within a single frame for the data set and the considered sampling techniques are shown in Tab. 2. Basically, it can be noticed that the data set shows the biggest and the average symbols the lowest deviations and the single and composite symbol again are close to each other and also close to the data set. All in all, the results confirm the claim that the differences due to the sampling approaches are negligible.

4 Sender Identification System Evaluation

The previous analysis is based on the signal itself respectively on the calculated distances. However, since the sender identification approaches use much more complex characteristics for classification, this chapter analyzes the effect of the different sampling methods on a real system. For this purpose, the Edge-based Sender Identification (EASI) [KSH20], which also uses only a single rising edge for identification, is considered. For the evaluation, the system uses the same configuration and data set used in the original work for the analysis of the behavior of the characteristics to environmental factors as well as the effect of electrical consumers. The utilized vehicle is the same as mentioned in Sect. 3.1, but without having the additional Raspberry Pis connected. The metrics considered for the evaluation of the approximately 55 000 frames are the *true positive* and *true negative rate*, the *identification rate* and the *confidence* of the system. A high true positive rate indicates the system's ability to detect forged frames, the true negative rate allows an assessment of the amount of wrong

alarms, the identification rate analyzes the general performance of sender identification and the confidence gives an indication on how well the learned model fits to the current situation.

Tab. 3: Effect of the sampling technique on the sender identification performance.

	Average Symbol	**Single Symbol**	**Composite Symbol**
True Positive Rate	99.82 %	99.16 %	99.59 %
True Negative Rate	100 %	100 %	100 %
Identification Rate	99.98 %	99.91 %	99.98 %
Confidence	99.81 %	99.57 %	99.87 %

In Tab. 3, it can be noticed that the use of the average symbol achieves the best results considering the real sender identification system. While no false alarms have occurred in any of the analyses, the usage of the single symbol leads to a slight decrease of the true positive rate. Assuming that an attack requires three messages which are not detected by the system, the probability of a successful attack is increased from 5.8^{-9} to 5.9^{-7} by using the single symbol instead of the average symbol. Accordingly, even by using EASI with the most lightweight configuration, the single symbol, a high probability of detecting potential attacks is achieved and thus still provides a high increase in security. Overall, as already determined during the direct signal analysis in Sect. 3.3, no significant differences can be observed for the different sampling techniques.

5 Assessment

Tab. 4: Assessment of the signal acquisition approaches.

	Performance	**Additional Hardware Requirements**	**Resource Requirements**	**Multi-Channel Capability**	**Complexity**	**Timing Restrictions**
Average Symbol	+	-	○	-	○	○
Single Symbol	-	-	+	+	+	+
Composite Symbol	○	○	+	-	-	-

An overview of the assessment is presented in Tab. 4, where the approaches are compared relatively in terms of the individual aspects. While the use of the average symbol provides the best results in the previous analyses, it is also expected to have the highest resource consumption. In particular, as with the single symbol, an external analog-to-digital converter (ADC) is required to record the entire signal or symbol at the appropriate sampling rate. For instance, a required sampling rate of 20 megasamples/second is assumed for the classic CAN, but due to the higher requirements respectively the shorter symbol duration of CAN-FD, higher sampling rates will be necessary. The acquisition of a composite symbol offers some advantages, as under certain assumptions regarding the used microcontroller it is possible to utilize the internal ADC for the acquisition. However, this requires a particularly fast comparator [KSH20] to be able to detect the individual level changes fast enough and the observed frames must have a certain amount of corresponding symbol transitions. In principle, both the acquisition of single and composite symbols require the least resources in

terms of calculation and storage, because in the case of the average symbol, it is necessary to store the entire signal in order to process it before the signal characteristics can be extracted. Provided that the computing capability of the implementing ECU is sufficient, however, this disadvantage can be compensated by calculating the running average. With the single and composite symbol this is omitted as the symbols can be used directly. Depending on the communication architecture of the vehicle under consideration, it may be necessary to analyze several CANs in parallel. For example, the networks of many vehicles are nowadays separated by domain or functions, which contributes to an increased security [RFS18]. In the case that multiple channels have to be observed, the usage of the single symbol is especially advantageous, as the sampling unit is only occupied for the time span of the symbol. For a single or a small number of bus segments, the composite symbol approach shows the lowest cost, but the complexity of time-critical sampling should not be underestimated. Reaching a high sender identification performance and a high robustness against fluctuations and signal changes potentially caused by environmental conditions and electrical consumers [KSH19], will allow to prevent attacks by disturbing ongoing transmissions of forged frames [Fo19; KH18]. If the possibility of preventing an attack is intended, the decision whether a forged frame is present must be made in a correspondingly short time. The type of signal acquisition has a considerable influence in this respect, since a certain number of symbols of the same type must have been transmitted for the generation of the composite symbol. A certain number of symbols of the same type are also used for the average symbol, whereby the amount can also be defined variably. For example, a time span can be defined after which all captured symbols are used for the calculation of the average symbol. In case only a single symbol is acquired during this time, this corresponds at least to the single symbol whose direct usage causes the least negative effect on the required processing time.

6 Conclusion

Basically, no significant differences in performance with respect to detection and identification rates could be determined by the different sampling methods. For this insight, on one hand the signals were analyzed directly and on the other hand the performance effects of the sampling method on a sender identification system were investigated. The small difference in performance enables the selection of the method based on the available resources and the underlying communication architecture. The average symbol, for example, not only shows slightly higher performance but also has the highest resource usage, while the recording of a single or composite symbol has advantages depending on the specific application. For single CAN buses the composite symbol is the most cost effective option, while the recording of a single symbol with an external ADC is advantageous for monitoring several CAN segments.

References

[Ca19] Cai, Z.; Wang, A.; Zhang, W.; Gruffke, M.; Schweppe, H.: 0-days & Mitigations: Roadways to Exploit and Secure Connected BMW Cars. Black Hat USA/, 2019.

[CA20] CAN in Automation: CAN XL is knocking at the door./, Jan. 2020, URL: https://www.can-cia.org/news/cia-in-action/view/can-xl-is-knocking-at-the-door/2020/1/3/, visited on: 01/03/2020.

[Ch18] Choi, W.; Joo, K.; Jo, H. J.; Park, M. C.; Lee, D. H.: VoltageIDS: Low-Level Communication Characteristics for Automotive Intrusion Detection System. IEEE Transactions on Information Forensics and Security 13/8, pp. 2114–2129, Aug. 2018, ISSN: 1556-6013.

[Fo19] Foruhandeh, M.; Man, Y.; Gerdes, R.; Li, M.; Chantem, T.: SIMPLE: Single-Frame Based Physical Layer Identification for Intrusion Detection and Prevention on in-Vehicle Networks. In: Proceedings of the 35th Annual Computer Security Applications Conference. ACSAC '19, Association for Computing Machinery, San Juan, Puerto Rico, pp. 229–244, 2019.

[HKD11] Hoppe, T.; Kiltz, S.; Dittmann, J.: Security threats to automotive CAN networks—Practical examples and selected short-term countermeasures. Reliability Engineering & System Safety 96/1, pp. 11–25, 2011, ISSN: 0951-8320.

[KH18] Kneib, M.; Huth, C.: Scission: Signal Characteristic-Based Sender Identification and Intrusion Detection in Automotive Networks. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, ACM, New York, NY, USA, pp. 787–800, 2018, ISBN: 978-1-4503-5693-0.

[Kn20] Kneib, M.: A Survey on Sender Identification Methodologies for the Controller Area Network. In (Reinhardt, D.; Langweg, H.; Witt, B. C.; Fischer, M., eds.): SICHERHEIT 2020. Gesellschaft für Informatik e.V., Bonn, pp. 91–103, 2020.

[KSH19] Kneib, M.; Schell, O.; Huth, C.: On the Robustness of Signal Characteristic-Based Sender Identification. 2019.

[KSH20] Kneib, M.; Schell, O.; Huth, C.: EASI: Edge-Based Sender Identification on Resource-Constrained Platforms for Automotive Networks. In: Proceedings of the 27th Network and Distributed System Security Symposium. 2020.

[LL18] Luo, Q.; Liu, J.: Wireless Telematics Systems in Emerging Intelligent and Connected Vehicles: Threats and Solutions. IEEE Wireless Communications 25/6, pp. 113–119, 2018.

[MV15] Miller, C.; Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015/, p. 91, 2015.

[RFS18] Ring, M.; Frkat, D.; Schmiedecker, M.: Cybersecurity Evaluation of Automotive E/E Architectures. 2. ACM Computer Science in Cars Symposium/, 2018.

[Ro12] Robert Bosch GmbH: CAN with Flexible Data-Rate Specification. 2012.

[Ro91] Robert Bosch GmbH: CAN Specification. 1991.