A new methodology for analyzing vehicle network topologies for critical hacking

Maen Ghadi¹ · Ádám Sali¹ · Zsolt Szalay¹ · Árpád Török¹

Received: 20 May 2020 / Accepted: 4 September 2020 / Published online: 17 September 2020
© The Author(s) 2020

Abstract

This study aims to provide a new approach for describing and measuring the vulnerability of in-vehicle networks regarding cyberattacks. Cyberattacks targeting in-vehicle networks can result in a reasonable threat considering passenger safety. Unlike previous literature, the methodology focuses on a comparatively large sample of vehicle networks (114 objects) by proposing a new framework of statistical techniques for measuring, classifying, and modelling in-vehicle networks concerning the changed vulnerability, instead of dealing with each vehicle network individually. To facilitate understanding of the vulnerability patterns of in-vehicle networks, the dataset has been evaluated through three analytic stages: vulnerability identification, classification, and modeling. The result has helped in ranking vehicles based on their network vulnerability level. The result of the modeling has shown that every additional remote endpoint installation causes a relevant weakening in security. Higher cost vehicles have also appeared to be more vulnerable to cyberattacks, while the increase in the number of segmented network domains has had a positive effect on network security.

Keywords Vehicle network topology · Cybersecurity · Network vulnerability · Hierarchical clustering · Ordinal logistic regression

1 Introduction and literature review

Transport safety and security are sensitive issues that affect all transport users and providers. Self-driving cars could increase safety, prevent deaths, and increase lower-income mobility. But any discussion on the self-driving cars must involve security issues and particularly cybersecurity. Cyberattacks targeting in-vehicle networks can result in a reasonable threat considering passenger safety.

When the field of automotive cybersecurity is investigated, vehicular ad-hoc networks (VANETs) cannot be ignored. It seems to be clear that in-vehicle and inter-vehicle communication networks are essential to implement highly automated and connected transportation systems.

Accordingly, Sharma and Kaul (2018) present a summary related to the different intrusion detection methods to introduce the different approaches especially focusing on VANETs. In the last part of the paper, the benchmarking of the different models has been performed, presenting pros and cons with regard to the investigated models.

Recently, one of the most known cybersecurity issues of automotive communication technologies is the BlueBorne vulnerability. However, there are other important considerations needed to be analyzed related to Bluetooth technology. Cheah et al. (2017) have performed a comprehensive evaluation focusing on the automotive Bluetooth interface. Accordingly, the researchers have proposed a model environment for analyzing Bluetooth interfaces of road vehicles in detail and have developed a specific application to demonstrate the applicability of the developed model.

In general, modeling cyberattacks and defense-related processes have always been an important issue in the field of security. Accordingly, Ten et al. (2010) have performed an overall investigation related to a high-level control security environment, including real-time analysis, intrusion detection, effect analysis, and defense concepts.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12652-020-02522-w>) contains supplementary material, which is available to authorized users.

✉ Maen Ghadi
ghadi.maen@mail.bme.hu

¹ Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Budapest 1111, Hungary

Beyond the general cybersecurity-related considerations of risk evaluation methods and defense strategy identification processes, due to the numerous specific features of the automotive sector, the vulnerability of vehicular systems are difficult to be represented by models of other disciplines. Therefore Ward et al. (2013) have described how a functional safety based cybersecurity method can be developed based on the approach described in ISO 26262 and the related MISRA Safety Analysis Guidelines. Schoitsch et al. (2015) and others have also emphasized the necessity of a combined approach in the automotive industry to integrate safety and cybersecurity (Schoitsch et al. 2015). Moreover, other studies have drawn attention to the relationship of product life cycle and security, especially focusing on over the air updating possibilities in the automotive industry (Boehner 2019). Besides this, the latest solutions are also needed to be adapted in the automotive security sector. Accordingly Fraga-Lamas and Fernández-Caramés (2019) have introduced the application possibilities of blockchain technologies in a road transport environment. Other researches have emphasized the importance of the most up-to-date results in the field of automotive cybersecurity regarding network architectures, common protocols, and novel design concepts (Lo Bello et al. 2019).

In recent years, the number of attacks on in-vehicle networks has been started to grow dynamically. The in-vehicle network is a system for transferring data among different electronic control units (ECUs) of the vehicle via a serial data bus. In other words, in-vehicle networks facilitate information sharing among distributed applications (Leen and Heffernan 2002). Applying a serial data bus reduces the number of wiring, influencing the weight of a vehicle, and enhances its performance. As today's vehicles contain more than 70 ECUs (Henniger et al. 2009), effective communication processes are required among them to reach their full functionality. As vehicles adopt more and more technological applications and implemented connectivity functions to the external world, threats on electronic functionalities are highly rising.

Telematics systems of modern cars provide value-added features such as remote diagnostics, crash response, and stolen car recovery over a long-range wireless link. These telematic systems connect in-vehicle networks with external communication centers via remote endpoints connectors. Furthermore, the introduction of the vehicle to vehicle and vehicle to infrastructure communication in the future developed autonomous cars (Maile and Delgrossi 2009; Ahmed-Zaid et al. 2011) will broaden the possible attack surface.

Beyond performing security tests and improving the efficiency of the defense system, security can already be considered during the design process. Tsigkanos et al. (2015) have introduced a novel approach allowing security software engineers to represent security requirements and the

topology of the operational environment at the same time. A similar approach has been followed by Mirkovic (2010) and others through developing adequate facilities, tools, and processes to ensure resources for researches in cybersecurity, also emphasizing the importance of network topology. The importance of in-vehicle communication network topologies has already been investigated by researchers (Miller and Valasek 2014; Lin and Sangiovanni-Vincentelli 2012), however detailed and comprehensive analysis of the different topologies have not been performed in the automotive industry so far.

Generally, the attack possibilities will be strongly influenced by the different remote endpoints, the topology of the in-vehicle networks, and the applied safety programs. Lee and Kim (2011) have shown that the topology structure has a direct impact on the robustness of a network under attack. Hegde et al. (2013) have applied a simulation software (i.e., CANoe) to study and compare the effect of applying different in-vehicle network topologies on the network performance. Furthermore, researchers have demonstrated how vulnerability in remote endpoints can be exploited to access in-vehicle networks, penetrate ECUs, and then control or sabotage vehicle maneuver (Checkoway and McCoy 2011; Miller and Valasek 2013).

The growing demands for the new functionality of modern vehicles, as well as the possible accompanying cyberattacks, define a complex evaluation environment that cannot be handled with traditional techniques. Rather than focusing on individual attacks, this paper provides a comprehensive analysis of a relatively large sample of in-vehicle networks. The analysis has based on the development of a complex statistical methodology for identifying, classifying, modeling, and evaluating in-vehicle network topologies in regard to potential cyber incidents.

2 Data description

In this study, we aimed to test the degree of exposure of vehicular network topologies to cyberattacks based on the data of 114 vehicles. The topology database was generated based on multiple resources. The most helpful resource was the VAG group's Self Study Program manuals (V. Group 2019) that gave a detailed description of the vehicles' networks. Other service manuals and wiring diagrams were also used from Chevrolet (Chevrolet 2009), Mercedes (Mercedes 2019), Mitsubishi (Minato 2019), and Toyota (Toyota 2019). The rest of the materials were collected from various training, research, and presentation materials (Suzuki 2012; Köhl et al. 2003; Burns 1996) available online or downloaded from official databases.

The used data has included information about vehicle characteristics (i.e., manufacturer, model, year of

production, manufacturer's suggested retail price (MSRP), engine type and capacity), some relevant details related to the internal network (i.e. the number of segmented network domains, ECUs, remote endpoints), and the processed information on network topology security (i.e., accesses protected by gateways). The MSRP was collected from online (i.e., Autotrader.com and Kbb.com), which represents the original value of the vehicle on release. A significant pillar of the developed evaluation framework is the identification of the derived value of Appendix 1 that gives a comprehensive characterization of vulnerability based on some of the relevant variables used in this study. Besides this, it has to be emphasized that collecting such data has meant a considerable challenge, and remarkable efforts, resources, and time have been needed to be allocated to search for the required details.

3 Methodology

Communication among the electronic components of the vehicle is essential for providing its main functions. Different possible combinations of network types, ECUs, sensors, and other electronic devices within a vehicle can result in different network topologies. The topology of the internal vehicle networks usually vary by manufacturers (Miller and Valasek 2014) and is designed to balance among efficiency, convenience, and cost with modern technological solutions. Due to the continuous development of vehicle technologies and functionalities, automotive network architecture is getting more complex and susceptible to new ways of penetration from attackers (András et al. 2015).

This study aims to provide a new approach for describing and measuring the vulnerability of in-vehicle networks concerning malicious interventions. The methodology focuses on a relatively large sample size (i.e., difficult to analyze every object alone) by proposing a new framework of statistical techniques for measuring, classifying, and modeling in-vehicle networks according to the changed vulnerability.

The developed methodology was the result of a unique research process considering the existing cybersecurity standards applied in the automotive industry and in the IT sector. We did take a deeper look into the SAE recommendations (SAE J 2016; Szalay et al. 2017) for Cybersecurity Best Practices for Modern Vehicles, which pointed out that the methods of Critical Security Controls for Effective Cyber Defense (CIS CSC) can be directly applied in numerous industrial segments. However, in the case of the automotive sector, the proposed approaches should be adapted according to the specific characteristics of the sector.

In the first step we investigate the layered approach recommended by the Cybersecurity Framework of National Institute of Standards and Technology. The layered approach

describes the key points of designing and building a comprehensive and systematic cybersecurity protection system for vehicles. Following this, our classification model has numerous connection points with the mentioned layered approach. The technique of layered cybersecurity protection applies a risk-based identification and protection process focusing on the safety-critical modules of the vehicle. Since the external availability and protection level of safety-critical elements is a considerably emphatic part of our model, the output of the risk-based identification process is critical from the viewpoint of our evaluation method.

If we study the CIS CSC methodology further, we find that the cybersecurity gap assessment process should also be an important part of a comprehensive security analysis. Beyond, the operative assessment, the documentation of the performed methods should also be performed in a very detailed way. This consists (SAE J 2016; Szalay et al. 2017):

- A risk assessment,
- The penetration test results,
- And the organizational decisions.

Our methodology can contribute to the 3 main requirements for performing and documenting a risk assessment process. The SAE CBPMBV describes these requirement as follows.

A risk assessment study has to include at least the evaluation of the internal vehicle networks, of the external wireless networks, and all the possible interfaces that can be applied to reach an ECU externally. This characteristic of the topology is considered in our method by the identification of the critical ECU-s and unprotected ECU-s (SAE J 2016; Szalay et al. 2017).

The external and internal availability of vehicle ECUs should be restricted to the most important functions of the given vehicle module and the used connection points has to be secured to protect the access of malicious perpetrators. In accordance with this methodology, we considered these possible attack surfaces as the remote endpoints of the network. The number of remote endpoints and unprotected remote endpoints was included in rating and documentation process.

Logical and physical segmentation methods should be applied to isolate controllers, safety-critical modules and network domains from external connection points to prevent unauthorized access. According to the recommendation, we also considered the aspects of network segmentation as a vital security feature. The number of segmented network domains and the amount of unprotected network domains were the third factor of our evaluation process.

Accordingly, in light of the investigated standard automotive cybersecurity methods, the proposed novel cyber security evaluation model meets the expectations of the current recommendation of the Society of Automotive Engineers.

3.1 In-vehicle network vulnerability

Modern vehicles can contain more than 70 ECUs (Hegde et al. 2013). Each ECU is responsible for one or more specific vehicle functions (SAE 2016; Szalay et al. 2017). For full functionality, the ECUs need to communicate with each other and with the outside world (Zöldy and Zsom-bók 2019). Unprotected ECUs might tempt attackers and consequently posing a threat to vehicles and their passengers. The threat level is generally influenced by the characteristic of the vehicle network topology and its external endpoints.

Safety-critical attacks against today's vehicles require two general stages. The first stage includes the injection of malicious data from outside using either remote or onboard access endpoints. In most of the cases, remote attack possibilities can cause even more risk for vehicle safety (Valasek and Miller 2015), since sending some malicious codes via wireless, internet, or Bluetooth can be more convenient for the attackers. The second stage requires to gain control over some ECUs or sabotage their functionality. The degree of vulnerability is strongly influenced by the function of the hacked ECUs. For example, some ECUs are responsible for passengers' convenience, such as switching the radio or fasten the seatbelt. These ECUs are installed within a convenient or infotainment internal network. In contrast, other ECUs are directly responsible for various physical aspects of the vehicle, such as monitoring the steering wheel angle or controlling the ABS system and can be found within a powertrain network domain.

Cyberattacks can result in a serious threat to passenger safety. Vehicle manufacturers aim to monitor and protect all possible internal and external accesses in the designed network to reduce all possible ways of malicious intervention. Access points are frequently protected by gateway technology. The gateway cannot be installed for every access point due to its cost and the complexity of the network architectures. However this causes relevant security gaps for perpetrators to attack in-vehicle networks.

According to our expectations, the developed analytical framework can reveal the relevancy of the considered three critical aspects that might directly influence vehicle security and safety. The mentioned three aspects analyzed with special attention are the security effect of remote endpoints, the impact of segmented network domains, and the location of critical ECUs. Therefore, the methodology applied for measuring the vulnerability of in-vehicle networks is based on the ratio of unprotected components in case of those three aspects (i.e. remote endpoints, network domain, and ECUs) for every investigated certain network topology separately. Critical ECUs were selected based on their automotive safety integrity (ASIL) level. ASIL framework

is a key component of the safety standard ISO 26262 since it can be used to measure the risk of a specific system component. The more complex the system, the higher the risk of systematic failures and random hardware failures. There are four ASIL values, named A–D. ASIL A is the minimum level of risk. And ASIL D is the maximum. So, ASIL D has stricter compliance requirements than ASIL A. Accordingly, during our research, in the case of critical ECUs, we focused on components characterized by ASIL D.

3.2 Hierarchical clustering analysis

Regarding the security aspects of in-vehicle network topology, previous literature has only concentrated on small sample size and rather analyze every vehicle individually. In the case of larger sample size, it is not very easy to deal with every single vehicle individually. Hence the data should be treated as a group. When dealing with a large dataset, data mining techniques are typically used to extract useful information and patterns from such data and convert it into an understandable structure.

Cluster analysis is a well-known data mining techniques, aiming to classify data into homogeneous groups so it can be easily understood and applied in further operations. The selection of the best clustering method is based on the objectives and field of study, and there is no such thing as a universally "best clustering method" (Hennig and Liao 2013; Luxburg et al. 2012).

Clustering is a qualitative method that aims to classify a set of data, regardless of size, into small homogeneous groups. However, the hierarchical clustering is claimed to perform the best in case of a relatively small sample size. According to Abbas (2008) and Verma et al. (2012) the partitioning algorithms (like k-means and expectation-maximization) are recommended for a huge dataset while hierarchical algorithms show good results when using with small data. Verma et al. (2012) have also found that density-based clustering methods (i.e., DBSCAN and OPTICS) do not perform well on small datasets.

According to the above-raised discussion, the hierarchical clustering seems best to fit the small size data of this study. The hierarchical cluster is a popular clustering method. The method is generally based on constructing a hierarchy cluster by recursively partitioning the data according to some given similarity measures (Everitt et al. 2011). Since clustering is the grouping of similar objects, similarity measure methods should be applied to determine whether two objects are similar or dissimilar. Following the objective of this paper, it seems to be reasonable to classify in-vehicle network data based on the similarity in their vulnerability level. With this, it becomes possible to identify and compare the most susceptible in-vehicle network topologies from a security point of view.

The hierarchical process can be performed based on either a bottom-up or a top-down approach. In the case of the bottom-up clustering method, each object initially represents a cluster on its own. Then, fewer clusters are created by merging similar objects. This method is called the agglomerative hierarchical clustering. The inverse process is the top-down or division hierarchical clustering. The result of the hierarchical clustering is a dendrogram, which presents the nested groups of clusters at different similarity levels. The desired cluster structure of the data is obtained by cutting the dendrogram at a suitable level.

The result and process for classifying vehicles by their vulnerability values using the hierarchical agglomeration clustering method can be summarized, as in the dendrogram (Fig. 1), as follows: Every vehicle has a single vulnerability value (e.g., v_1, v_2, v_3, v_4, v_5). The network vulnerability value for each vehicle (object) initially represents a cluster by itself (Fig. 1a). Then converged clusters are successively merged in smaller groups but higher members, until the desired cluster structure is obtained (Fig. 1b–d respectively). The height in the dendrogram (Fig. 1) represents the distance at which each fusion is made. The desired cluster structure of the data is obtained by cutting the dendrogram at a suitable level. For example, cutting the hierarchy at the second-level, as shown in Fig. 1, results in three groups; (v_1, v_2), (v_3), and (v_4, v_5).

However, choosing the right clustering level or the appropriate number of clusters are crucial aspects in the case of this technique. The optimal number of clusters is selected based on previous experiences depending on the type of study. The selection process is usually supported by some statistical techniques. To determine the optimal number of clusters statistically, the Elbow method can be applied. According to this approach, it is recommended to merge clusters in the cluster space until the marginal increase in the variance of cluster elements would exceed the acceptable level (Ketchen and Shook 2002).

The application of the hierarchical method, in this study, is very attractive since it can be more efficient in case of a relatively small sample size (Baker and Hubert 1975)

compared to other clustering methods (e.g., partitioning methods).

3.3 Ordinal logistic regression

We developed a generalized ordered logistic regression model to understand the variability of in-vehicle network vulnerability patterns and investigate their main contributing variables. Like all regression, the ordered logistic regression is a predictive analysis used to describe data and to explain the relationship between one dependent variable and one or more explanatory variables (Quinn et al. 2001). What distinguishes the ordered logistic regression from the linear regression is the response variable. The ordered logistic regression allows using an ordered set of categories to define the dependent variable. Developing such a model can exactly fit the objective of the current analysis since it allows us to reveal in-vehicle network security characteristics for each different vulnerability level separately.

The ordinal model provides a cumulative probability representation; the sum of all possible outcomes is equal to 1. The number of categories is C , and c indicates the certain index of the investigated category. The probability of being in the c -th category is π_c (for every $c = 1, 2, \dots, C$). Accordingly, using the expected cumulative log link function (η_c) the log odd ratio compares the probability of being included by category c (π_c) and its complementary probability ($1 - \pi_c$). The log odd for this ratio is as follows.

$$\eta_c = \log \left(\frac{\pi_c}{1 - \pi_c} \right) \quad (1)$$

The odd ratio can be derived by the exponentiations of the log-odd coefficients [$\exp(\eta_c)$]. When the odd ratio is less than one, the probability of being included by the given category is less than its complementary probability, and if the probability of being included by the investigated category c is more than the probability of the complementary set, it results in a higher odd ratio, above one.

4 Results and discussion

4.1 Vehicle network vulnerability results

In-vehicle network security can be strongly affected by the degree of access protection considering the three investigated aspects; remote endpoints, segmented network domains, and the location of critical ECUs. Any unprotected access points without a properly designed and installed gateway may leave a serious security gap for perpetrators to attack the in-vehicle network and thus threaten passenger safety.

Fig. 1 Hierarchical dendrogram. The height represents the smallest distance of point-point, point-group or group-group

The methodology applied for measuring the vulnerability level of the in-vehicle network is based on the percentage of unprotected accesses in case of the three considered aspects (i.e., remote endpoints, segmented network domain, and critical ECUs) for every in-vehicle network respectively. All of the three considered components represent the most sensitive parts of the network, especially considering the field of cyberattacks and pose a high level of threat to the vehicle and passenger's safety. In another words, they all have the same high level of importance and consequently assumed to have the same weight and influence on the vulnerability of the in-vehicle network for malicious attacks, and can be added in a single value. For example, referring to Appendix 1, total network vulnerability for Audi A3 (i.e., ID=29 in Appendix 1) is equal to 1.83. This resulted from the summation of the ratios of unprotected accesses in case of the segmented network domains (i.e. $3 \div 6 = 0.5$), remote endpoints (i.e. $1 \div 1 = 1$), and critical ECUs (i.e. $3 \div 9 = 0.3$).

Figure 2 presents security vulnerability for the top 50 in-vehicle network topology out of the 114 data sample, ordered by their ID (see Appendix 1) and vulnerability level.

As seen from Appendix 1, different network architectures are provided by each car manufacturer and each car. This means that each car has different security vulnerability level, as shown in Fig. 2.

The presented vulnerability by Fig. 2 does not mean that low vulnerable vehicles are more secured than high vulnerable vehicles since this is also influenced by the technical characteristic of the installed gateways and other network elements and vice versa. On the other hand it provides a reliable method, how structural vulnerability can be described and compared.

According to the obtained results (Fig. 2 and Appendix 1), the five most vulnerable vehicles are presented below (in descending order):

1. 2018 Toyota IQ (vulnerability = 3.67)

2. 2019 Toyota Aygo (vulnerability = 3.00)
3. 2010 Infinity G37 (vulnerability = 3.00)
4. 2014 Jeep Cherokee (vulnerability = 3.00)
5. 2007 Volkswagen Crafter (vulnerability = 3.00)

Based on their topology characteristics, the least vulnerable vehicles (vulnerability = 0.00 for all) are as follows (see also Appendix 1):

1. 2005 Volkswagen Transporter
2. 2002 Volkswagen Touareg
3. 2005 Audi A6
4. 2018 Toyota Corolla
5. 2017 Toyota CHR
6. 2018 Toyota Avensis

The most noticeable result of Fig. 2 is the different Toyota models appear both among the most and also among the least vulnerable models. All of these Toyota models are newly introduced in the markets (2017–2019). However, cost-effectiveness seems to play an important role in the designed topologies, where higher costs show lower network vulnerability. The more expensive "Toyota" models are also provided by a higher number of segmented network domains (i.e. 5), and ECUs with an extra type of remote endpoints (i.e. navigation or head unit) with fully protected accesses, as described in Appendix 1.

Generally, several factors that might affect the weakness of the topology of a car network, such as; the number of critical ECUs, segmented networks, remote endpoints, as well as the cost and age of the vehicle model, as the most notable factors. But it seems to be difficult to determine any specific pattern of these factors by the naked eye for the given sample size. Thus, it is required to introduce some auxiliary statistical techniques to help in classifying and understanding the most relevant factors influencing in-vehicle network vulnerability and their patterns.

Fig. 2 Vulnerability values for the most vulnerable 50 in-vehicle network topology

4.2 Hierarchical clustering results

As the first step, the heterogeneity of the data has to be revealed to facilitate understanding the vulnerability patterns of the investigated in-vehicle networks. The hierarchical clustering has been applied to classify and organize the vehicles (Appendix 1) by their internal network vulnerability values.

The process begins by applying the agglomerative hierarchical clustering with a suitable distance metric technique to measure the affinity between vulnerability values. The application of the IBM SPSS software for the given data has resulted in a dendrogram (i.e. the hierarchy of the partitioned data). Figure 3 represents the process of defining the suitable partition level (i.e., number of clusters) by the Elbow method.

Based on Fig. 3, it can be observed that the selected value of partition—located on the x-axis—is 111. Following the Elbow method, this value represents the optimal case from where the further reduction of the number of clusters would result in an unfavorable level of variance and cluster heterogeneity. Therefore, the optimal number of clusters is 3 (i.e., the difference between the total sample size -114- and the optimal value -111-). In other words, in-vehicle networks have been divided into three groups (or clusters) by vulnerability level (see Appendix 1). Cluster number 1 (CL1), includes 9 vehicles represent the least secured networks ($2.50 \leq \text{vulnerability} \leq 3.67$). Cluster number 2 (CL2) has a moderate vulnerability ($1.25 \leq \text{vulnerability} \leq 2.25$) compared with the given dataset and includes 31 vehicles. At the same time, the last 74 vehicles are the least susceptible to cyberattacks ($0.00 \leq \text{vulnerability} \leq 1.00$) and are classified within cluster number 3 (CL3).

4.3 Modelling results

We applied the IBM SPSS software to complete the model development process, and we used the ordinal logistic regression method to estimate the model parameters.

Fig. 3 Identifying the optimal number of clusters

Several models have been tested, including numerous combinations of different explanatory variables to find the most efficient predictor variable set. Eventually, the vehicle average cost, number of critical ECUs, network domains, and vehicle age (i.e., measured by subtraction the year of production of each vehicle from the newest one in the dataset) have been selected as the best predictors of the in-vehicle network vulnerability change. The model calibration results are shown in Table 1.

Although most of the selected variables (Table 1) do not reach the required significance level still, they are the most relevant variables related to the investigated response variable (i.e., vulnerability). The low significance can be interpreted by the relatively small sample size (i.e., 114 data objects) used in developing the model. The factor of small sample size can strongly contribute to the difficulty of statistical model development (Wisz 2008). Nevertheless, the whole model is statistically significant for all of its parameters as presented in the resulted goodness-of-fit tests shown in Table 2.

The overall model validation process has been completed through using three tests: (1) model fitting, (2) goodness-of-fit, (3) and, pseudo-R-squared test, as presented in Table 2. The model-fitting approach tests the fit of the resulted model with a full complement of predictors relative to a null model with no predictors (the baseline or ‘Intercept only’ model). The statistical significance of this test ($p < 0.05$) indicates that the final model gives a significant improvement over the baseline intercept-only model. In the case of the developed model, the resulted values of the fitting test (Table 2) indicate that the model gives better predictions than if it

Table 1 Model parameter estimates

Variables	Estimate	Sig
Threshold		
Vulnerability = 1.00	− 2.581	0.000
Vulnerability = 2.00	− 0.618	0.003
Vehicle MSRP	− 0.474	0.043
Number of critical ECUs	0.348	0.187
Vehicle age	0.333	0.117
Number of network domains	0.341	0.183

Table 2 Model goodness-of-fit tests results

Statistical test	−2 Log likelihood	Chi-square	Sig
(1) Model fitting	181.1	9.32	0.045
(2) Goodness-of-fit			
Pearson	—	190.4	0.926
Deviance	—	181.1	0.974
(3) R-squared (Nagelkerke)	0.29		

is just based on the marginal probabilities for the outcome categories. The $-2 \log$ -likelihood can be used in comparisons of nested models. Table 2 also shows the resulted goodness-of-fit. The goodness-of-fit test provides additional information concerning the overall fit of the model. The statistical insignificance of the model is also proved by the chi-square statistic ($p > 0.05$), as the achieved results show (Table 2), which is also an important indicator for proper model fitting.

The third goodness-of-fit test refers to measuring pseudo-R-squared. The “goodness” or “acceptability” of pseudo-R-squared value depends upon the nature of the investigated factors and the explanatory variables. Here, pseudo-R-squared values (e.g. Nagelkerke = 0.29) indicate that the given explanatory variables explain 29% of the variation between the three vulnerability levels. This is in accordance with our expectations since, for surly, there are other variables that can have an impact on in-vehicle network vulnerability. However, the “pseudo” R-squared values do not have the same interpretation as standard R-squared values from OLS regression (Long and Freese 2005). Though, this does not dispute the fact that explanatory variables have a statistically significant and relatively large influence on the investigated vulnerability level. Generally, the results of the three goodness-of-fit statistics indicate good fitting of the developed model for the given data.

The interpretation of the estimated coefficients (Table 1) is measured with respect to CL3 (i.e., the lowest vulnerable vehicle network group) as a reference variable. The vehicle MSRP is the only significant variable for the 95% confidence level (p value = 0.043). The negative coefficient of the number of remote endpoints (-0.471) indicates that expensive vehicles are more likely to have a considerably vulnerable internal network from a security point of view. In other words, as the vehicle cost increases, the security of the vehicle network for malicious intervention decreases by 0.47. This is an interesting result. The explanation is that the increase in cost is always linked to the multiple functionalities of the vehicle. According to Henniger et al. (2009), as vehicles adopt more and more technological applications and implemented connectivity functions to the external world, threats on electronic functionalities are highly rising. The positive coefficients related to the number of critical ECUs and network domains (significance is discussed above) indicate a proportional relationship between their numbers and the in-vehicle network safety. At the same time it, can also be concluded that newer (less age) and more expensive vehicles are getting more susceptible to external attacks, in general.

5 Conclusion

This study has investigated the vulnerability of the in-vehicle network topologies to cyberattacks. The methodology proposes a new statistical approach for measuring, classifying, and modeling in-vehicle networks with respect to their vulnerability to cyberattacks.

We analyzed the dataset has to facilitate understanding vulnerability patterns of in-vehicle networks through three stages: vulnerability identification, classification, and modeling. Vehicle vulnerability to cyberattacks may be strongly influenced by the characteristics of in-vehicle networks. According to literature, the remote endpoints, network domains, and critical ECUs represent the most sensitive part of the network, especially considering the field of cyberattacks. The remote endpoints can provide easier access for intruders. While network segmentation and the protection of critical ECUs can directly improve the security of the vehicle. The insufficient attention paid to these three aspects could lead to an increased security threat. Therefore, the vulnerability level of in-vehicle networks has been defined based on the total percentages of unprotected accesses for those three components (i.e., remote endpoints, segmented network domain, and critical ECUs) in case of every investigated vehicle network.

In the second stage, the hierarchical clustering has been applied to classify and organize the vehicles by their internal network vulnerability values. This stage is applied to reveal the heterogeneity and facilitate understanding the vulnerability patterns for the given in-vehicle networks. With hierarchical clustering, the data has been classified into three groups (CL1, CL2, and CL3 from high to low vulnerability) that have primarily been differentiated based on the level of vulnerability. The ordinal logistic regression has finally been applied to describe the impact of the influencing factors affecting vulnerability.

The result reveals that the most susceptible vehicle is Toyota IQ (2018), followed by Toyota Aygo (2019), Infinity G37 (2010), Jeep Cherokee (2014), Volkswagen Crafter (2007). While the least sensitive vehicles are: Volkswagen Transporter (2005), Volkswagen Touareg (2002), Audi A6 (2005), Toyota Corolla (2018), 2017 Toyota CHR (2017), and 2018 Toyota Avensis (2018). This finding has revealed the presence of such influencing factors that are mainly affected by the price category and the target consumer group of the vehicles. For instance, Toyota models appear to have the most and least secured internal networks. Accordingly, the topology of in-vehicle networks in case of the more expensive Toyota models shows higher network security compared to cheaper Toyota models. Applying the introduced modeling technique has helped in more understanding of vulnerability related factors. The results have shown

that the in-vehicle security gets more vulnerable to newer vehicles. On the other hand, more expensive vehicles, due to the increased number of unprotected access points have generally been proved to be more vulnerable to cyberattacks. Besides this, the increasing number of segmented network domains affects network security in a positive way.

Generally, the application of the developed combined classification and estimation model has helped in analyzing the processed complex dataset.

Acknowledgements The research reported in this paper was supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Artificial Intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI/FM).

Funding Open access funding provided by Budapest University of Technology and Economics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Table 3.

Table 3 The columns of Appendix 1 can be defined as follows: (1) Year of production, (2) Original MSRP (.000 USD), (3) Number of segmented network domains, (4) Number of un-protected accesses to all network domains, (5) Number of remote endpoints (Wifi, Bluetooth, wireless), (6) Number of un-protected accesses to remote endpoints, (7) Number of critical ECU, (8) Un-protected critical ECU, (9), Total percent of un-protected access points (vulnerability level), and (10) Cluster_ID

ID	Manufacturer	Type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Toyota	IQ	2018	15	3	5	1	1	6	6	3.67	CL1
2	Toyota	Aygo	2019	15	1	1	1	1	5	5	3.00	
3	Infinity	G37	2010	57	2	2	3	3	7	7	3.00	
4	Jeep	Cherokee	2014	46	5	5	4	4	8	8	3.00	
5	Volkswagen	Crafter	2007	31	2	2	1	1	5	5	3.00	
6	Cadillac	Escalade	2015	80	6	4	4	4	7	7	2.67	
7	Dodge	RAM 3500	2014	42	4	3	4	3	7	7	2.50	
8	Dodge	RAM 1500	2009	35	3	3	2	1	6	6	2.50	
9	Skoda	Fabia	2014	21	2	1	2	2	5	5	2.50	
10	Mitsubishi	Outlander	2010	33	4	1	3	3	7	7	2.25	CL2
11	Honda	Accord	2014	30	6	1	4	4	8	8	2.17	
12	Mercedes	W463	2018	130	5	4	3	3	10	3	2.10	
13	Chrysler	300C	2012	48	2	2	1	0	8	8	2.00	
14	Ford	Fusion	2013	22	4	3	1	1	12	3	2.00	
15	VOLVO	XC90	2001	35	5	0	1	1	8	8	2.00	
16	Volkswagen	Phaeton	2002	75	6	3	2	2	8	3	1.88	
17	Volkswagen	Golf 7	2014	26	6	3	2	2	9	3	1.83	
18	Mercedes	W246	2011	50	6	2	3	3	7	3	1.76	
19	Mercedes	W212	2010	50	7	3	3	3	10	3	1.73	
20	Volkswagen	Scirocco	2009	35	4	2	2	2	9	2	1.72	
21	Mercedes	W176	2013	35	6	2	3	3	8	2	1.58	
22	Volkswagen	Beetle	2012	23	4	1	3	3	7	2	1.54	
23	Mercedes	W222	2013	150	12	2	3	3	14	5	1.52	
24	Ford	Fusion	2006	18	2	1	1	1	6	0	1.50	
25	Mercedes	C218	2018	100	8	2	3	3	10	2	1.45	
26	Ford	F150	2008	30	5	2	1	1	8	0	1.40	
27	MINI	Cooper	2005	21	5	2	1	0	6	6	1.40	
28	Rover	Sport	2006	60	5	2	2	2	6	0	1.40	
29	Audi	A3	2004	35	6	3	1	1	9	3	1.83	
30	Rover	Sport	2010	63	5	2	2	2	9	0	1.40	
31	Ford	Transport	2014	25	6	2	3	1	7	5	1.38	
32	Chevrolette	Aveo	2009	15	5	1	1	1	7	1	1.34	

Table 3 (continued)

ID	Manufacturer	Type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
33	Opel	Vectra-C	2012	26	3	1	0	0	6	6	1.33	
34	Volkswagen	Golf 5	2004	18	6	2	2	2	9	0	1.33	
35	Mercedes	R231	2013	92	8	1	3	3	10	2	1.33	
36	Audi	Q2	2017	28	7	3	3	2	10	2	1.30	
37	Ford	Focus	2011	17	2	1	2	1	7	2	1.29	
38	Mercedes	C216	2014	130	8	2	3	3	7	0	1.25	
39	Mercedes	R172	2019	60	8	1	3	3	8	1	1.25	
40	Opel	Astra-H	2004	18	4	1	0	0	6	6	1.25	
41	BMW	X3	2014	44	6	3	4	2	8	0	1.00	CL3
42	Infinity	G35	2010	45	2	2	0	0	5	0	1.00	
43	Volkswagen	Passat	2006	26	4	2	0	0	10	5	1.00	
44	Volkswagen	Polo	2002	18	3	2	1	0	8	2	0.92	
45	Volkswagen	Polo	2011	21	3	2	1	0	8	2	0.92	
46	Volkswagen	Vento	2011	15	3	2	1	0	8	2	0.92	
47	Volkswagen	UP	2011	21	3	2	1	0	8	1	0.79	
48	Rover	Evoque	2014	49	7	2	4	1	8	2	0.79	
49	BMW	I12	2014	136	8	2	4	2	11	0	0.75	
50	Mitsubishi	Grandis	2003	22	4	1	2	1	7	0	0.75	
51	Mitsubishi	Lancer X	2008	35	4	1	2	1	7	0	0.75	
52	BMW	X5	2011	53	6	2	3	1	9	0	0.67	CL3
53	BMW	3 series F30	2014	37	5	2	4	1	8	0	0.65	
54	Volkswagen	Passat	2012	25	4	1	2	0	10	4	0.65	
55	Suzuki	Kizashi	2012	25	5	3	1	0	5	0	0.60	
56	Volkswagen	Tiguan	2008	32	6	2	1	0	9	2	0.56	
57	Audi	A6	2012	45	6	3	4	0	10	0	0.50	
58	Toyota	Auris	2018	30	4	2	2	0	7	0	0.50	
59	Chevrolette	Spark	2009	19	4	1	1	0	6	1	0.42	
60	Chevrolette	Lacetti	2009	22	7	2	1	0	8	1	0.41	
61	Audi	A1	2010	32	5	2	2	0	7	0	0.40	
62	Audi	TT	2007	47	5	2	4	0	10	0	0.40	
63	Audi	A7 Sportback	2010	64	8	3	4	0	10	0	0.38	
64	Audi	A8	2010	75	8	3	4	0	10	0	0.38	
65	Audi	A8	2013	77	8	3	4	0	10	0	0.38	
66	Audi	A8	2014	79	8	3	4	0	10	0	0.38	
67	Mitsubishi	Triton	2016	25	8	3	3	0	8	0	0.38	
68	Audi	A3	2013	38	6	2	2	0	6	0	0.33	
69	Chrysler	300	2014	40	3	1	4	0	8	0	0.33	
70	Dodge	Viper	2014	102	3	1	4	0	8	0	0.33	
71	Mitsubishi	Outlander	2007	25	3	1	3	0	7	0	0.33	
72	Volkswagen	Jetta	2004	22	3	1	0	0	6	0	0.33	
73	Volkswagen	Touareg	2002	57	3	1	3	0	8	0	0.33	
74	Chevrolette	Captiva	2010	25	6	1	1	0	8	1	0.29	
75	Chevrolette	Tacoma	2003	15	4	1	0	0	7	0	0.25	
76	Citroen	Berlingo	2008	28	4	1	0	0	7	0	0.25	
77	Citroen	C5	2000	29	4	1	0	0	9	0	0.25	
78	Seat	Alhambra	2011	59	4	1	2	0	7	0	0.25	
79	Seat	Altea	2004	30	4	1	2	0	8	0	0.25	
80	Seat	AlteaXL	2006	33	4	1	2	0	8	0	0.25	
81	Seat	Leon	2012	24	4	1	2	0	8	0	0.25	
82	Skoda	Octavia	2004	27	4	1	2	0	6	0	0.25	

Table 3 (continued)

ID	Manufacturer	Type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
83	Toyota	Rav4	2015	25	4	1	0	0	5	0	0.25	
84	Toyota	Prius	2015	22	8	2	1	0	7	0	0.25	
85	Toyota	Yaris	2018	18	4	1	1	0	7	0	0.25	
86	Volkswagen	Amarok	2010	35	4	1	3	0	6	0	0.25	
87	Volkswagen	Golf 6	2009	32	4	1	3	0	8	0	0.25	
88	Volkswagen	Passat CC	2009	28	4	1	3	0	8	0	0.25	
89	Audi	Q3	2010	37	5	1	3	0	12	0	0.20	
90	Audi	Q7	2005	53	5	1	3	0	11	0	0.20	
91	Audi	Q7	2017	53	10	2	4	0	14	0	0.20	
92	Seat	Ibiza	2008	22	5	1	3	0	7	0	0.20	
93	Skoda	Octavia	2012	34	5	1	2	0	7	0	0.20	
94	BMW	5 series E60	2008	50	6	1	1	0	8	0	0.17	
95	Chevrolette	Rezzo	2009	25	6	1	1	0	6	0	0.17	
96	Mitsubishi	Mirage	2012	15	6	1	3	0	5	0	0.17	
97	Subaru	WRX Sti	2007	25	6	1	3	0	7	0	0.17	
98	Volkswagen	Sharan	2011	30	6	1	3	0	9	0	0.17	
99	Audi	R8	2015	110	7	1	4	0	12	0	0.14	
100	Audi	TT	2014	41	7	1	2	0	9	0	0.14	
101	Chevrolette	Epica	2009	18	7	1	1	0	7	0	0.14	
102	Seat	Ateca	2016	40	7	1	4	0	9	0	0.14	
103	Ford	Crown Vic	2003	28	8	1	0	0	9	0	0.13	
104	Audi	A6	2005	45	7	0	4	0	7	0	0.00	
105	Infinity	Q50	2014	42	3	0	4	0	9	0	0.00	
106	Mazda	Series 3	2009	18	6	0	2	0	8	0	0.00	
107	Mitsubishi	MIEV	2009	23	2	0	0	0	5	0	0.00	
108	Seat	Ibiza	2002	18	4	0	0	0	6	0	0.00	
109	Seat	Leon	2001	22	3	0	0	0	6	0	0.00	
110	Toyota	Avensis	2018	35	5	0	2	0	8	0	0.00	
111	Toyota	CHR	2017	30	5	0	2	0	9	0	0.00	
112	Toyota	Corolla	2018	25	5	0	2	0	9	0	0.00	
113	Volkswagen	Touareg	2011	58	3	0	0	0	6	0	0.00	
114	Volkswagen	Transporter	2005	30	4	0	0	0	7	0	0.00	

References

- Abbas OA (2008) Comparisons between data clustering algorithms. *Int Arab J Inf Technol* 5(3):320–325
- Ahmed-Zaid F et al (2011) Vehicle safety communications—applications (VSC-A) final report. DOT HS 811 492, no. September, pp 1–102
- András S, Levente B, Zsolt S (2015) Hacking cars in the style of Stuxnet. Retrieved from <http://www.hit.bme.hu/~buttyan/publications/carhacking-Hacktivity-2015.pdf>. Accessed 2019
- Baker FB, Hubert LJ (1975) Measuring the power of hierarchical cluster analysis. *J Am Stat Assoc* 70(349):31–38
- Boehner M (2019) Security for connected vehicles throughout the entire life cycle. *ATZelectronics Worldw* 14(1–2):16–21
- Burns A (1996) Broadening real-time systems research. *ACM Comput Surv* 28(4):178
- Cheah M, Shaikh SA, Haas O, Ruddle A (2017) Towards a systematic security evaluation of the automotive Bluetooth interface. *Veh Commun* 9:8–18
- Checkoway S, McCoy D (2011) Comprehensive experimental analyses of automotive attack surfaces. In: *Proceedings of the 20th USENIX conference on Security*, p 6
- Chevrolet (2009) Workshop manuals. Detroit, Michigan, US. <https://www.automotive-manuals.net/chevrolet/>. Accessed 2019
- Everitt BS, Landau S, Leese M, Stahl D, Shewhart W, Wilks SS (2011) *Cluster analysis*, 5th Edn
- Fraga-Lamas P, Fernández-Caramés TM (2019) A review on blockchain technologies for an advanced and cyber-resilient automotive industry. *IEEE Access* 7:17578–17598
- Hegde R, Kumar S, Gurumurthy KS (2013) the impact of network topologies on the performance of the in-vehicle network. *Int J Comput Theory Eng* 5:405–409
- Hennig C, Liao TF (2013) Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. *J R Stat Soc* 62C:309–369
- Henniger O, Aprville L, Fuchs A, Roudier Y, Ruddle A, Weyl B (2009) Security requirements for automotive on-board networks. In: *2009 9th International conference on intelligent transport systems telecommunications, ITST 2009*, pp 641–646

- Ketchen DJ Jr, Shook CL (2002) The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manag J* 17(6):441–458
- Köhl S, Plöger M, Otterbach R, Lemp D (2003) Steuergeräteverbundtest mittels Hardware-in-the-Loop-Simulation. *VDI Berichte* 1755:121–136
- Lee H, Kim J (2011) Attack resiliency of network topologies. pp 638–641
- Leen G, Heffernan D (2002) Expanding automotive electronic systems. *Computer (Long Beach)* 35(1):88–93
- Lin CW, Sangiovanni-Vincentelli A (2013) Cyber-security for the Controller Area Network (CAN) communication protocol. In: *Proceedings of the 2012 ASE International Conference on Cyber Security, CyberSecurity 2012*, pp 1–7
- Lo-Bello L, Mariani R, Mubeen S, Saponara S (2019) Recent advances and trends in on-board embedded and networked automotive systems. *IEEE Trans Ind Inform* 15(2):1038–1051
- Long JS, Freese J (2005) Regression models for categorical outcomes using Stata
- Maile M, Delgrossi L (2009) Cooperative intersection collision avoidance system for violations (CICAS-V) for avoidance of violation-based intersection crashes. In: *Proc. 21st Int. Tech. Conf. Enhanc. Saf. Veh.* 2009, Stuttgart, Ger., No. 09–0118, pp 1–14
- Mercedes (2019) Workshop manuals. Stuttgart, Germany, MA. <https://car-diagrams.jimdofree.com/mercedes-benz/>. Accessed 2019
- Miller C, Valasek C (2013) Adventures in automotive networks and control units. *IOActive Tech. White Pap.*, pp 1–99
- Miller C, Valasek C (2014) A survey of remote automotive attack surfaces
- Mirkovic J, Benzel TV, Faber T, Braden R, Wroclawski JT, Schwab S (2010) The DETER project: Advancing the science of cyber security experimentation and test. In: *2010 IEEE International Conference on Technologies for Homeland Security, HST 2010*, 2010, pp 1–7
- Mitsubishi (2019) Workshop manuals. Minato, Tokyo, Japan. <https://www.onlymanuals.com/mitsubishi>. Accessed 2019
- Quinn G, Keough M, Models EL, Regression L (2001) Generalized linear models and logistic regression. *Exp Des Anal Biol* 14:435–457
- Schoitsch E, Schmittner C, Ma Z, Gruber T (2015) The need for safety and cyber-security co-engineering and standardization for highly automated automotive vehicles. pp 251–261
- Sharma S, Kaul A (2018) A survey on Intrusion detection systems and honeypot based proactive security mechanisms in VANETs and VANET Cloud. *Veh Commun* 12:138–164
- SAE J (2016) 3061: Cybersecurity guidebook for cyber-physical vehicle systems. 2016. Society for automotive engineers
- Suzuki (2012) Suzuki training academy, Minami-ku, Hamamatsu, Japan
- Szalay Z, Tettamanti T, Esztergár-Kiss D, Varga I, Bartolini C (2017) Development of a test track for driverless cars: vehicle design, track configuration, and liability considerations. *Period Polytech Transp Eng* 46(1):29
- Ten CW, Manimaran G, Liu CC (2010) Cybersecurity for critical infrastructures: attack and defense modeling. *IEEE Trans Syst Man Cybern Part A Syst Hum* 40(4):853–865
- Toyota (2019) Workshop manuals, Toyota City, Japan
- Tsigkanos C, Pasquale L, Ghezzi C, Nuseibeh B (2015) Ariadne: topology aware adaptive security for cyber-physical systems. In: *Proceedings of the International Conference on Software Engineering*, 2015, vol 2, pp 729–732
- V. Group (2019) Self study program. Wolfsburg, Germany, MA. http://www.vaglinks.com/vaglinks_com/Docs/SSP/. Accessed 2019
- Valasek C, Miller C (2015) Remote exploitation of an unaltered passenger vehicle. *IO Active*, 2015. [Online]. Available: <https://illmatics.com/RemoteCarHacking.pdf>
- Verma M, Srivastava M, Chack N, Diswar AK, Gupta N (2012) A comparative study of various clustering algorithms in data mining. *Int J Eng Res Appl* 2(3):1379–1384
- von Luxburg U, Williamson R, Guyon I (2012) Clustering: science or Art? In: *JMLR: Workshop and Conference Proceedings* 27, 2012, pp 65–79
- Ward D, Ibarra I, Ruddle A (2013) Threat analysis and risk assessment in automotive cyber security. *SAE Int J Passeng Cars Electron Electr Syst* 6(2):507–513
- Wisn MS et al (2008) Effects of sample size on the performance of species distribution models. *Divers Distrib* 14(5):763–773
- Zöldy M, Zsombók I (2019) Influence of external environmental factors on range estimation of autonomous hybrid vehicles. *Syst Saf Hum Tech Facil Environ* 1(1):472–480

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.