



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

Intelligent data-driven model using machine learning to protect in-vehicle communications

R ADINARAYANA , K SANTHOSHI

Abstract—

Electric vehicles rely heavily on communications, both inside and between vehicles, which might be problematic. The topic of cyber security is explored in detail here. To avoid hacker access, this paper provides a secure and trustworthy intelligent architecture to shield electric vehicles from attack. The proposed model is based on an improved support vector machine model for anomaly detection using the CAN bus protocol. To improve the model's ability to swiftly recognize and avoid assaults during offline training, we develop a unique optimization approach using the social spider (SSO) algorithm. A two-stage modification strategy is proposed to significantly enhance the search capabilities of the algorithm and avoid premature convergence. Finally, the simulation results on the real data sets show that the suggested methodology is very effective, reliable, and secure against Dos hacking in electric vehicles.

I. INTRODUCTION

Modern automobiles rely on hardware modules known as electronic control units (Ecus). Many different programs control these Ecus. The electronic control unit (ECU) of a vehicle receives information from all of the vehicle's sensors, analyses that information, and then sends orders to the relevant actuators depending on the findings [1]. For such a sophisticated hardware-software procedure, it may be necessary to use network protocols such as CAN, LIN, Flex Ray, and MOST [2]. CAN bus has become the most popular of these protocols because to its adaptability and dependability, making it extensively utilized not just in vehicles but also in medical equipment, agricultural machinery, etc. Data transfer speeds of up to 1Mbps, simpler wiring that cuts down on installation time and costs, automated retransmission of missed messages, and the ability to discover and repair transmission faults [3] are just a few of the advantages of the CAN bus standard. Due to its development at an age when automobiles were mostly unmanned, the CAN bus protocol has several vulnerabilities in the context of today's dynamic smart grids. The result will be an increase in the number of assaults launched on EVs by people who aren't criminals by introducing malicious code into

their Ecus. Many cyber intrusion scenarios involving plug-in electric cars are modeled and used in [4] to assess the security of these vehicles and the possible effects of their incorporation into the electrical grid. In order to identify [5] develops a unique classification scheme for cyber-intrusions in automobiles. Increasing the rate at which CAN bus messages are sent or misusing CAN message IDs are two indicators that might point to a cyber attack, and both are taken into account while designing the data intrusion detection system described in [6]. This will allow the driver to immediately stop the vehicle upon realizing an attack has taken place.

The authors of [7] argue that all CAN communications should be routed via a data management system to shield them from hacking attempts. For instance, [8] uses an algorithmic strategy to prevent attacks. signs of service disruption or vehicle errors. According to [9], the attestation process should be handled by a master ECU that is selected during vehicle construction. As shown in [10], cyber attack commands to the may bus may be thwarted by inserting a firewall between the May bus and the communication system.

The authors of [11] suggest using entropy analysis on CAN bus data to identify network intrusion in automobiles. An anomaly detection technique is developed in [12] that may discover known and new sorts of faults without requiring expert parameterization.

II Cybersecurity flaws in electric vehicle technology

Electric vehicle makers depend heavily on the CAN standard because it allows for low-cost communications in units with a high number of components (up to 500 million). chips. Due to its construction, CAN provides sufficient robustness and noise-resistance in the automobile industry.

Since CAN bus specifications don't guarantee confidentiality and authentication to CAN data frames, hackers may get access to the automobile system through wired or wireless methods. If you prefer a hardwired setup, you may connect to the car's may bus using the on-board diagnostics (OBD) II connector, which is often located in the center console.

This port is designed for engine and vehicle diagnostics, however hackers may use a low-cost scanner to collect may signals.

ECOM application programming interfaces (APIs) like CANReceiveMessage and Neurotransmitter [10] simplify CAN bus communication reading and writing from this point forward. The cyber interference is the same whether a wireless attack is made on an ECU, but the point of entry is not the OBD-II port. In most circumstances, the car has to be within range of a compromised Wi-Fi network in order to be hacked wirelessly. Reverse engineering might potentially be used to crack the security of transponders used in keyless vehicles. Analysis has shown a number of problems with the cipher's architecture, particularly with its user authentication mechanisms and the way they are actually implemented.

"Other wireless entry points for vehicles include the Tire Pressure Monitoring System (TPMS), "Add-on technologies, entertainment system (gaming), smart key," and "Internet, smart infrastructures."

The full complement of security measures organized by category, together with their respective solutions, are shown in Table I below. During the authentication process, it is determined whether or not the transmitting and receiving devices really are who they claim to be. Data mining and machine learning-based approaches are used by intrusion

prevention and detection systems to detect and react to irregularities in network traffic.

TABLE I
SECURITY MEASURES IN CAN BUS
PROTOCOL IN VEHICLES [15]

Category	Solutions
Authentication	Membership, MAC (Message Authentication Code)
Intrusion Detection and Prevention	Anomaly Detection, Signature-based and Anomaly-Based
Encryption	AES (Advanced Encryption Standard), ALE (AES-Based Lightweight Authenticated Encryption), Central Gateway Encryption, Hybrid Cryptosystem
Restricted Physical Access	Central Gateway Isolation
Software Security	HSM (hardware security modules)

Denial-of-service In a denial-of-service (Dos) assault on a CAN bus, the attacker sends a flood of messages with the lowest possible identifier (ID), clogging up the network in the process. frequency to create an active environment inside the vehicle. This might be disastrous since it takes up the car's only communication route. In Fig. 1, we see a full representation of a CAN-bus data frame. The frames consist of a 12-bit SOF, a 7-bit Arbitration Field, a 6-bit Control Field, a 0-to-64-byte Data Field, a 16-bit CRC Field, a 2-bit Acknowledgement Field, and a 1-bit End of Frame.

The ECU's decision on whether or not to process a given message frame is based on its priority and a unique identity. Priority is lower for IDs with lower numbers. If two messages are published and sent concurrently, the ECU will give preference to the one with the lower ID. To prevent other messages or instructions from taking effect, attackers may use a technique called "message arbitration," in which they transmit a malicious message with a low ID but high frequency. In order to avoid such an issue, this research presents a method for identifying and evading abnormalities by using the identification and frequency of the linked message frames.

In the following paragraphs, we'll go through the suggested strategy in further depth.

There are four metrics that may be used to assess the quality of the results produced by the classifier model. The proportion of correct rejections (CR), the hit rate (HR), the false alarm rate (FR), and the miss rate (MR). To further comprehend these indices, we may look at Figure 3, which provides a conceptual representation that matches the actual expert observation from an experiment and the proposed anomaly detection model output. The HR, FA, MR, and CR indices may be used to evaluate any anomaly detection algorithm, where CA and CN stand for the authentic malicious data and normal data sets, respectively.

$$HR = |H_i| |C_A|^{-1} ; H_i = \{X \in D | X \in C_A \& X \in C_O\} \quad (10)$$

$$FR = |F_A| |C_N|^{-1} ; F_A = \{X \in D | X \in C_N \& X \in C_O\} \quad (11)$$

$$MR = |M_i| |C_A|^{-1} ; M_i = \{X \in D | X \in C_A \& X \in C_I\} \quad (12)$$

$$DR = |C_R| |C_N|^{-1} ; C_R = \{X \in D | X \in C_N \& X \in C_I\} \quad (13)$$

the whole datasets are denoted by D, outlier datasets by CA, normal datasets by CN, outlier datasets by CI, and outlier datasets by CO.

		Actual Value From Experiment	
		positives (C _A)	negatives (C _N)
Anomaly Detection Model Response	positives (C _O)	<u>Hit Rate</u> True Positive (TP)	<u>False Alarm Rate</u> False Positive (FP)
	negatives (C _I)	<u>Miss Rate</u> False Negative (FN)	<u>Correct Rejection Rate</u> True Negative (TN)

Fig. 3: The suggested anomaly detection model's confusion matrix

IV. AN OPTIMIZATION METHOD DEPENDING ON A DEVIATED SSO

Each spider in our anomaly detection model consists of the four variables [v, C, p], which are the optimal parameters for the SVM model.

The Original SSO Algorithm (Type A) The SSO approach generates a random population of spider SKs, each of which may be the best possible answer to the optimization problem at hand. Since both male and female spiders will be present in the population, the algorithm predicts that there will be a total of NS spiders. After the objective function is calculated for each spider Sk, the best spider Sb and the worst spider SW are recorded. Now, we may compare each spider by its "weight."

$$W_k = \frac{f_N - f_k}{f_N - f_b} \quad (14)$$

$$S_{k,F}^{iter-1} = S_{k,F}^{iter} \pm \theta_1 W_k e^{-d_k} (S_b - S_{k,F}^{iter}) \pm \theta_2 W_k e^{-d_k} (S_b - S_{k,F}^{iter}) - (\theta_3 - 0.5) \quad (15)$$

The same may be said about men generally: it's time for an improvement. To find the most influential males, it is necessary to sort the male population, according to what the criteria function yields. The median is meant to be an average or middle number. The spider with the highest objective function value relative to the others in the group is the strongest. The present male social order looks like this:

$$S_{k,DM}^{iter-1} = S_{k,DM}^{iter} + \theta_5 W_F e^{-d_k} (S_c^F - S_{k,DM}^{iter}) + (\theta_6 - 0.5) \quad (16)$$

Conforming to the weighted mean of the male population (Mw) is used to update spider populations outside of the dominating males:

$$S_{k,NM}^{iter-1} = S_{k,NM}^{iter} + \theta_7 (M_w - S_{k,NM}^{iter}) \quad (17)$$

The spiders' mating method in (16) is based on a roulette wheel where each spin is assigned a mating probability value (rk). The following formula is used to estimate the probability that a given pair will fall inside the lower (lz) or upper (uz) boundaries.

$$r_k = \frac{1}{2n_v} \sum_{i=1}^{n_v} (u_i - l_i) \quad (18)$$

$$S_i^{next} = S_{i1} + \theta_2 (S_{i2} - S_{i3}) \quad (19)$$

$$S_i^{next} = \begin{cases} S_{i2,z} & ; \theta_3 < \theta_9 \\ S_{i2,z} & ; \theta_9 \geq \theta_9 \end{cases} \quad (20)$$

$$S_i^{next} = \begin{cases} S_{i2,z} & ; \theta_9 < \theta_{10} \\ S_{i2,z} & ; \theta_9 \geq \theta_{10} \end{cases} \quad (21)$$

The second strategy makes advantage of a small-step walking equation to continually update the spider's position.

$$S_i^{iter+1} = S_i^{iter} + \varepsilon \Delta_i \quad (22)$$

V. Discussions and Results from Computer Models

The previous paragraphs have mostly covered the proposed model, theories, and settings. In this case, data from a trial of electric vehicles is used to assess the efficiency of the proposed approach. Due to the importance of DoS attacks on vehicle communication, this study assesses them. Denial-of-service attacks aim to prevent legitimate users (the driver) from gaining access. Even though cars are mobile, they are still vulnerable to DDoS assaults since they might cause significant vehicle accidents or losses. As an example, a denial-of-service (DoS) assault on a car may cause the driver to suddenly apply the brakes, pull the steering wheel to the left or right, lose power to the engine, unlock the doors, etc. The proposed anomaly detection algorithm might potentially learn these frequencies by analyzing CAN communication data collected during the course of a typical 10-minute drive.

The identities and frequencies of CAN buses are shown in Table II. To ensure that our model was learning every possible ID number, we had to include a lot of safeguards. checking for any and all signs After analyzing the traffic log and tracing the messages' paths, we learned that a 10-minute driving scenario captures the vast majority of the frequently recurring messages since most CAN connections are periodic. Thus, the produced model can be seen as proof-of-concept that shows how the proposed anomaly detection model can learn the existing pattern in the CAN signals to discern between normal and abnormal behavior during testing. To provide a realistic driving test, the CAN traffic file mimics the following scenarios: After turning the key to "on," the driver waited a few seconds for the engine to warm up, and then shifted into "D."

Thereafter, you drive the vehicle around for around 8 minutes on a public road. The brake pedal will be pressed many times during the trip. To park, the car is placed into park and the gear selection is moved to the reverse position ("R"). Once the vehicle has stopped entirely, the transmission is placed in "P," and the engine is turned off.

TABLE II A RANGE OF CAN BUS IDs AND CHANNELS

CAN Identifier	6FF	308	340	2A0
Frequency	101.010101	85.74311927	50	48.7804878
CAN Identifier	670	3F0	D21	210
Frequency	99.00990099	100.1666667	38.7804878	51.02040816
CAN Identifier	238	410	200	A7F
Frequency	108.6956522	93.45794393	61.02040816	49.01960784
CAN Identifier	B61	212	240	4EB
Frequency	10	68.54368932	78.01010101	113.6363636
CAN Identifier	2C1	312	5AE	1A3
Frequency	110.3595506	50	80.01960784	43.2449244

Fig. 4 displays the outcomes of using the suggested anomaly detection model based on support vector MSSO and the machine to identify outliers in the message frame training set, including ID and frequency. After performing the requisite feature selection procedure, it was determined that just frequency and frame Id were needed for a strong and reliable anomaly detection model. All of the vehicle's CAN bus message frames are captured by the support vectors as sit is monitored. Here, the boundary between the outliers and the rest of the BUS data is shown by the zero-valued contour. The fraction of unfavorable evaluations in the cross-validated data is therefore revealed to be relatively low, hovering around 0.13 percent. Therefore, it can be shown that the proposed model has excellent discriminative power. Many mathematically based benchmarks with diverse qualities including regularity, separability, and multi modality are used here to evaluate the search capabilities of the proposed MSSO method. Here, we evaluate it against five different criteria

-Sphere Function:

$$\min F = \sum_{n=1}^K x_n^2 \quad (23)$$

- Rosenbrock Function:

$$\min F = \sum_{n=1}^{K-1} [100(x_n^2 - x_{n+1})^2 + (1 - x_n)^2] \quad (24)$$

- Rastrigin Function:

$$\min F = \sum_{n=1}^K [x_n^2 - 10 \cos(2\pi x_n) + 10] \quad (25)$$

- Griewank Function:

$$\min F = \frac{1}{4000} \sum_{n=1}^K x_n^2 - \prod_{n=1}^K \cos\left(\frac{x_n}{\sqrt{m}}\right) + 1 \quad (26)$$

[16]:

where K describes the relative magnitude of the evaluation. Table III displays the optimal values for the functions and the ranges of the variables. Table IV shows the optimization results for a number of popular approaches, providing the mean and standard deviation (SD) value for each. It is clear from these results that the proposed MSSO approach successfully maximizes all objective functions across all dimensions.

Table III contains optimization problem benchmarks and their distinguishing characteristics.

Bench. no	Function	Range	Global optimum
1	Sphere	[-100, 100]	$F_{min} = 0, X = (0, 0, \dots)$
2	Rosenbrock	[-30, 30]	$F_{min} = 0, X = (1.1, \dots)$
3	Rastrigin	[-5.12, 5.12]	$F_{min} = 0, X = (0, 0, \dots)$
4	Griewank	[-600, 600]	$F_{min} = 0, X = (0.0, \dots)$

You may judge how effectively the proposed MSSO approach aids convergence of the anomaly detection model if you know the value of the Lagrangian objective function, recorded in each cycle using different techniques. All algorithms begin with a population size of 40, and all stop conditions utilize a value of 100 as their threshold. We used a mutation rate of 0.08% and a crossover rate of 0.80% in our GA. With an inertia weight factor of 0.8 and a maximum speed of 2 m/s, PSO is somewhat slow. MSSO and the previous SSO have similar requirements. These results indicate that the suggested MSSO approach may achieve convergence more quickly than the GA, PSO, and SSO methods. The MSSO was also successful in breaking out of local optima, whereas the other algorithms were unable to do so. When the original SSO was created, it was unable to optimize the Lagrangian function to its full potential because to difficulties with early convergence at about the number 50. Convergence curves like this demonstrate conclusively that the suggested MSSO is an effective tool for improving the support vector machine's ability to classify data.

Table IV displays the mean and standard deviation (SD) of optimization results across a variety of

methods.

Bench. No	K	IBC [19]		SSO		Proposed MSSO	
		Mean	SD	Mean	SD	Mean	SD
1	5	4.30 e-17	1.07 e-17	5.76 e-22	5.74 e-23	8.38 e-29	6.47 e-30
	30	4.69 e-16	1.07 e-16	7.38 e-19	2.12 e-21	7.44 e-28	1.64 e-29
	50	1.19 e-15	4.68 e-16	3.85 e-18	5.66 e-19	5.82 e-25	8.53 e-27
	100	1.99 e-06	2.26 e-06	5.36 e-16	5.85 e-17	5.34 e-23	3.84 e-23
2	5	2.33 e-01	2.24 e-01	7.06 e-02	6.94 e-03	3.65 e-05	5.09 e-06
	30	9.98 e-01	1.52 e+00	4.34 e-01	5.59 e-02	5.36 e-04	4.84 e-05
	50	4.33 e+00	5.48 e+00	7.96 e+00	4.37 e+00	7.47 e-03	2.36 e-04
	100	1.12 e+02	6.92 e+01	2.24 e+02	7.74 e+01	1.55 e+00	5.65 e-03
3	5	4.34 e-17	1.10 e-17	5.73 e-24	6.63 e-27	0.00 e+00	0.00 e+00
	30	4.80 e-05	2.43 e-04	5.59 e-10	6.73 e-14	0.00 e+00	0.00 e+00
	50	4.72 e-01	4.92 e-01	5.83 e-06	3.45 e-08	0.00 e+00	0.00 e+00
	100	1.46 e+01	4.18 e+00	3.33 e-04	4.24 e-05	0.00 e+00	0.00 e+00
4	5	4.04 e-17	1.12 e-17	5.73 e-10	5.46 e-11	7.62 e-19	7.83 e-25
	30	5.82 e-06	3.13 e-05	5.44 e-07	6.63 e-08	6.84 e-17	3.04 e-23
	50	5.72 e-01	9.22 e-01	6.78 e-04	3.28 e-05	4.55 e-12	4.58 e-22
	100	1.31 e+01	6.30 e+00	7.23 e-02	6.56 e-03	7.03 e-10	6.40 e-21

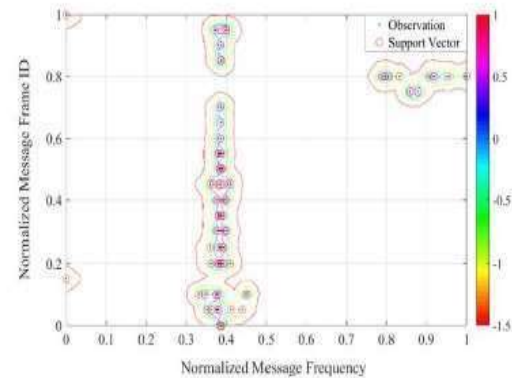


Fig. 4: Identifying frames in a message and finding frequency outliers

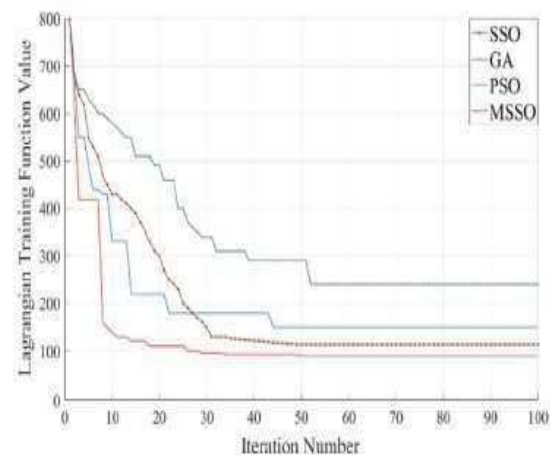
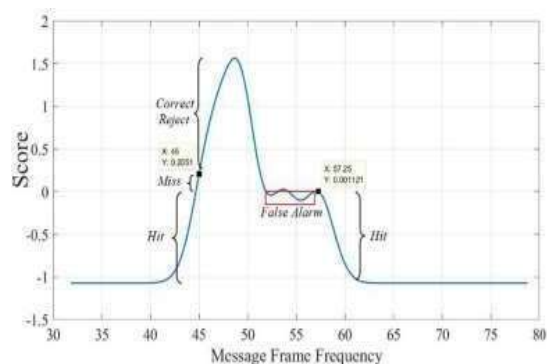


Figure 5: Different algorithms have different convergence behaviors.

The proposed anomaly detection approach is subjected to a series of simulated attacks, both good and bad.

In this situation, a new message frame with ID 2A0 is sent to the ECU. This ID is picked out because of its wide frequency range and sophisticated model response. The impact these signals have on the detection model is seen in Fig. 6. The model's ability in drawing trustworthy positive and negative inferences is shown by the high hit rate and accurate reject rate shown in this image. Real-world hackers, hoping to win arbitration and expose the information on the bus, often launch assaults at a pace almost twice as fast as the messages. As can be seen in Fig. 6, such behavior will be immediately recognized as harmful and blocked.



For a certain CAN ID in the car, Fig. 6 depicts the hit, miss, correct reject, and false alarm zones generated by the device.

The overall effectiveness of the proposed anomaly detection methodology is shown in Tab V for various IDs of regular and attack message frames. Multiple classifiers, including the nearest algorithm [17], a decision tree-based model [18], a radial basis function (RBF) neural network [19], a classic support vector machine, a modified support vector machine based on support vector singularity elimination (SSO), and a modified support vector machine based on multi-source support singularity elimination (MSSO), all have their simulated confusion matrices provided. Based on these results, it seems that the proposed anomaly detection model has better HR% and CR% than the other techniques. Since the data set is both non linear and complicated, the proposed TTS-BASED model stands a better chance of making reliable forecasts.

Different Anomaly Detection Models' Confusion Matrix Values

Outlier Algorithm	HR(%)	MR(%)	FR (%)	CR (%)
k-Nearest Neighbor [17]	81.63%	18.37%	19.55%	80.45%
Decision Tree-Based detection [18]	80.09%	19.91%	20.29%	79.71%
RBF Neural Network [19]	82.18%	17.82%	18.73%	81.27%
Conventional support vector model	83.12%	16.88%	18.07%	81.93%
Support vector machine based on SSO	89.47%	10.53%	12.24%	87.76%
Proposed Model	96.1%	3.9%	6.45%	93.55%

False positive decision values are harmless unless they are derived from a reliable source, which is why the proposed model has such low false positive decision values. uncommon CAN message structure. Understanding this argument requires reference to the paper's simulated counterattacks. The simulations include a wide range of frequencies, from a massive increase/decrease of 200 percent to a negligible one-hundredth of a percent.

The developed detection model may be tested against a wide range of message frequencies, as shown.

VI. CONCLUSION

This study presents a novel intelligent and protected anomaly detection approach for detecting and preventing cyber assaults on electric vehicles. The proposed model is based on the MSSO method, which is used to strengthen a support vector machine model. The proposed paradigm for cybersquatting has the potential to detect harmful behaviors while letting legitimate message frames transmit over the CAN standard. The high HR% and FR% indices demonstrate that the proposed approach was used to make honest positive and negative evaluations. The model's dependability is shown by the low values of the MR% and CR% indices, which are often seen towards the upper and lower borders of the message frame frequency. In subsequent works, the authors will assess the impact of additional counterattacks on the performance of different anomaly detection strategies.

REFERENCES

In 2012, the journal IEEE Trans. Industrial Electronics published "Multi-source Software on Multi core Automotive Ecus—Combining Burnable Sequencing With Task Scheduling" by authors A.

Monet, N. Navel, B. Bayeux, and F. Simon-Lion. The article's page range was 3934–3942.

2007 International Conference on Control, Automation, and Systems, pp. 2844–2849, "Gateway system with diagnostic function for LIN, CAN, and Flex Ray" by T.Y. Moon, S.H. Sec, J.H. Kim, S.H. Hang, and J. Wook Jean. Third, see "Efficient Protocols for Secure Broadcast in Controller Area Networks" by B. Grozny and S. Murray in IEEE Transactions on Industrial Informatics, volume 9, issue 4, pages 2034–2042, 2013.

The article "Advancing cyber-physical sustainability through integrated analysis of smart power systems: A case study on electric vehicles" by B. Mohammedans, R. Al Muhammad, W. Sinus, T. Hemmer, and S. El Khat was published in the International Journal of Critical Infrastructure Protection, volume 23, pages 33–48, 2018.

A taxonomy and review of cyber-physical intrusion detection techniques for automobiles, Ad Hoc Networks, volume 84, pages 124–147, 2019. [5] G. Louisa, E. Epistolary, E. Panasonic, P. Sanitariums, T. Dugong.

Security concerns to automotive can networks. [6] Hoppe T, Kiltz S, Pittman J. real-world cases and carefully chosen interim fixes. 2011, volume 96, issue 1, pages 11–25, Trustworthy Eng Cyst Saf.

This is according to [7] Schultz, Pukall, Saake, Hoppe, and Dittmann. The importance of organized data in modern cars. printed in BTW, volume 144, issue 2, pages 217–226.

Ling C. and D. Feng. A method for identifying potentially harmful communications sent through can buses. Information technology and computer science annual meeting 2012. 2012; Atlantis Press.

Houma H, X. Higashiosaka, M. Nakanishi, R. Otsuka, and H. Imai. Shintoism. All-new secure in-car communication framework based on attestation. Conference on International Telecommunications, 2008. 1–6, 2008. IEEE GLOBECOM 2008. IEEE. IEEE.

For example, see "Cyber security attacks to modern vehicular systems" by L. Pan, X. Zheng, H. X. Chen, T. Luan, and L. Batten in the October 2017 issue of "Journal of Information Security and Applications" (vol. 36, pages 90–100).

"Intrusion detection system using deep neural network for in-vehicle network security," by M. J. Kang and J. W. Kang, was published in Plot one, volume 11, issue 6: e0155781 in 2016.

Theiler, A., "Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection", Knowledge-Based Systems, vol. 123, pp. 163–173, 2012.

For example, see [13] "A weighted one class support vector machine" by F. Zhu, J. Yang, C. Gao, S. Xu, and T. Yin in Computerizing, volume 189, pages 1–10, 12 May 2016.

"A simplex method-based social spider optimization algorithm for clustering analysis," by Y. Zhou, Y. Zhou, Q. Luo, and M. Abdel-Basset, was published in 2017 in Engineering Applications of Artificial Intelligence, volume 64, pages 67–82.

[15] Future Generation Computer Systems, G. De La Torre, P. Rad, K.K. Raymond Choo, "Driverless vehicle security: challenges and future research opportunities," Published (with corrections), online release date: 11 January 2018.