Research of the model for detecting UMV interfaces vulnerabilities based on information criterion

Alexey Bryukhovetskiy [1], Vera Miryanova [1] and Dmitriy Moiseev [1]

[1] *Sevastopol state university, 33 Universitetskaya str., Sevastopol, 299053, Russia*

Abstract

An approach related to the development of methods for ensuring unmanned vehicles computer security is considered. The approach is based on the statistical distance estimation between the probability distributions of a random variable. The Jensen-Shannon information criterion, which provides a symmetric version of the Kullback-Leibler divergence, is proposed as an evaluation criterion. Vulnerability detection is performed on the basis of processing the UMV resources state values. The features of UMV state monitoring give rise to new problems characterized by data flows with variable intensity, heterogeneous information flows in conditions of a lack of a priori information and noisy data. When solving this problem, there are problems of big data processing: high computational complexity due to the processing of huge data amounts; high dynamics of controlled objects; non-stationary information situation of the objects state and the environment; ensuring high speed of query processing; providing metrics of information resources in real time.

Keywords

resources unmanned vehicles, detection of vulnerabilities, information criterion, statistical distance

1. Introduction

Car interfaces (CAN, LIN, FlexRay, and MOST) are vulnerable to various cybersecurity attacks. Through the on-board diagnostic (OBD) port or USB port, attackers can stop the engine or affect the vehicle's braking system and cause an accident [1]. The "replay" attack and the "simulation" attack on the CAN bus are described in [2]. The authors [3] simulated a "Sybil" spoofing attack on the FlexRay bus.

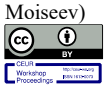
Each incident that is recorded in the UMV is characterized by an entry point that the attacker uses to perform the attack. Open ports are the main vulnerability point through which viruses penetrate and spread. A port in network technologies refers to a virtual "door" that can be accessed. To fix the vulnerability, identify suspicious processes that use ports [4]. Therefore, controlling them is a top priority for security.

Since different technologies are used in vehicles, therefore, there are relatively many interfaces for both internal and external communication with the UMV. Due to the fact that the complexity of interface technologies varies widely, therefore, knowledge of these technologies is necessary to evaluate the methods of implementing attacks and, consequently, to determine the feasibility of attacks. This is especially important also for assessing the possible damage caused by the implementation of the attack [5 - 13].

III International Workshop on Modeling, Information Processing and Computing (MIP: Computing-2021), May 28, 2021, Krasnoyarsk, Russia

EMAIL: a.alexir@mail.ru (Alexey Bryukhovetskiy); VNMiryanova@sevsu.ru (Vera Miryanova); dmitriymoiseev@mail.com (Dmitriy Moiseev)

ORCID: 0000-0002-2612-2968 (Alexey Bryukhovetskiy); 0000-0002-2941-2765 (Vera Miryanova); 0000-0002-3141-1529 (Dmitriy Moiseev)



2. Problem statement

The proposed method is intended for detecting vulnerabilities of UMV interfaces. The approach is based on the estimation of the statistical distance between the probability distributions of a random variable over different time intervals. The Jensen-Shannon information criterion [6], which provides a symmetric and normalized version of the Kullback-Leibler divergence, is proposed as an evaluation criterion. Vulnerability detection is carried out on the basis of processing UMV resources state values, such as, communication channel, processor and memory. The monitored features include: resource loading and the rate of change in resource loading.

Let's introduce the notation D_{KL} (P, Q) to calculate the Kullback-Leibler divergence between two distributions Q(x) and P(x). Then the divergence is defined as

$$D_{KL}(P, Q) = \sum_{i=0}^{n} Pi(x) * (\log (P_i(x) / Q_i(x)))$$

It should be noted that the KL divergence value is not symmetric:

$$D_{KL}(P, Q) \neq D_{KL}(Q, P)$$

Therefore, it is proposed to use the Jensen-Shannon – JS divergence, which allows us to estimate the discrepancies between the two probability distributions. In this case, the divergence is used to calculate the normalized estimate, which is symmetric. This means that the divergence of P from Q is the same as Q from P, i.e.

$$JS(P, Q) = JS(Q, P)$$

The JS divergence can be defined as follows [6]:

$$JS(P, Q) = 1/2 * D_{KL}(P, M) + 1/2 * D_{KL}(Q, M)$$

where the distribution M is calculated as

$$M = 1/2 * (P + Q).$$

The Jensen — Shannon divergence is limited to the value of one for two probability distributions if the Kullback-Leibler divergence uses the base 2 logarithm:

$$0 <= JS(P, Q) <= 1$$

The main advantage of the proposed method is that it provides a smoothed and normalized version of the Kullback divergence with estimates from 0 (there is no discrepancy between the distributions) up to 1 (maximum different distributions).

The Jensen — Shannon divergence of the distribution *P* with respect to *Q* can be estimated as:

$$JS(P, Q) <= Z – \text{no divergence},$$

$$JS(P, Q) > Z – \text{observing sample divergence},$$

where Z is the limit value of the distance that depends on the criticality of the object parameter value and is set by the expert. Then the null hypothesis H_0 holds for JS (P, Q) <= Z – no divergence. Otherwise, the hypothesis H_1 is accepted there is a qualitative change in the information state of the controlled object parameter.

In order to compare the estimates of divergences between the two probability distributions, a model for detecting changes in resource state was studied. The Jensen — Shannon distance is used as an information measure. We introduce the concept of the estimation zone of the divergence value. For the sake of certainty, we will consider the following boundaries of the recognition zones: {0, Z_1, Z_2, 1}. Then the zones are defined by the following intervals:

$$[0; Z_1), [Z_1; Z_2), [Z_2; 1].$$

Depending on the belonging of the current distance value JS (P, Q) $\in Z_i$ $(i=1, k)$ we will classify the following object information states:

$$0 <= JS(P, Q) < Z_1 – \text{no divergence (normal state of the object)},$$

$$Z_1 <= JS(P, Q) < Z_2 – \text{unstable region (precritical state of the object)},$$

$$Z_2 <= JS(P, Q) <= 1 – \text{observation of divergence (critical condition of the object)}.$$

In general, the number of Z_i recognition zones is determined by the expert and depends on the object criticality, the dynamics of its state, the requirements for the quality of its characteristics control, possible losses during control, etc.

The proposed model includes the following main modules:

- generating random samples according to a given distribution law with a priori specified parameters,
- setting input data,

- setting the information criterion,
- training and configuring model parameters,
- processing of statistical data,
- plotting histograms,
- acceptance of the hypothesis H_0 on the samples homogeneity or an alternative hypothesis H_1, confirming the samples heterogeneity.

The problem of estimating the samples divergence is solved according to the following algorithmic scheme:

- Input data is set: V – sample size, k – number of histogram intervals, $[Z_i; Z_{i+1}]$ – width of recognition zones, cr – information criterion.
- Sets the critical areas of the null hypothesis test criteria-areas of no divergences / unstable areas of divergences / areas of divergences.
- Samples X of a given volume V are generated from a general population having a given distribution law.
- Histograms of the distribution of the values of the controlled parameter are formed for two given samples P and Q.
- The distance value is calculated as the square root of the information measure JS (P, Q) and its belonging to the specified recognition zones is determined.

The above steps are repeated for other values of the input data and a conclusion is made about the evaluation of the homogeneity of other pairs of samples P and Q.

In accordance with the tasks set, experiments were conducted, during which the influence of a number of parameter values on changes in the state of resources was determined: the volume of samples - V, the number of histogram intervals– k, the width of the zones $[Z_i; Z_{i+1}]$ of the resource state assessment. The simulation results are presented below.

Research of the impact of sample size – V. Set: resource state classification zones $[Z_i;Z_{i+1}]$ for three states that differ in the width of the intervals. The intervals number k=3. We compared the estimates of discrepancies for the distributions P and Q for cases where the boundaries of the recognition zones differed slightly (homogeneous) and significantly (heterogeneous). The following recognition zone boundaries were set:

- intervals-1 {0; 0.50; 0.80; 1},
- intervals -2 {0; 0.40; 0.60; 1},
- intervals -3 {0; 0.30; 0.70; 1},
- intervals -4 {0; 0.10; 0.50; 1}.

The width of the intervals-1, 2, 3 differs slightly from each other, and the width of the intervals-4 differs significantly from the rest. Figure 1: shows the values of the distance JS (P, Q) when comparing the sample distributions for different values of the intervals: exp 1-2, exp 1-4 at V=30.

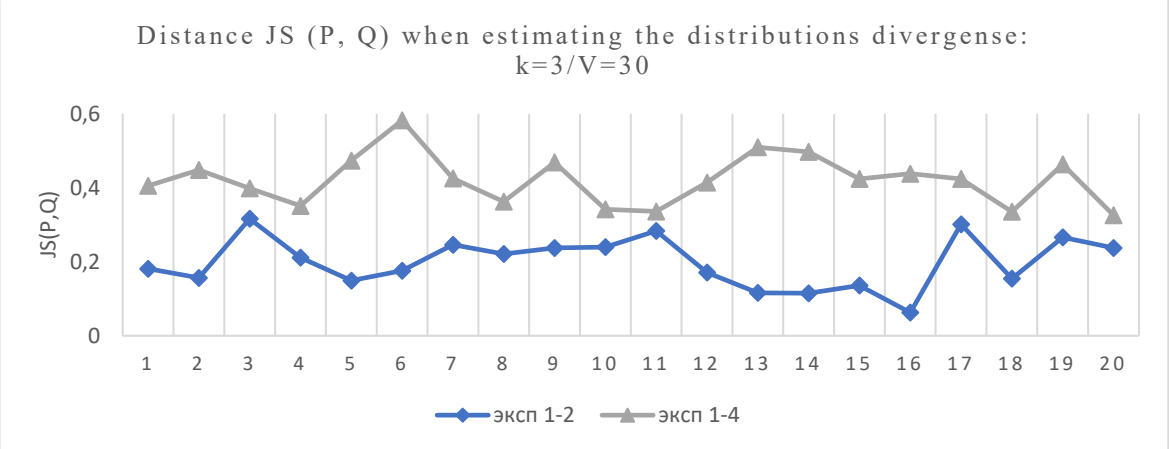


Figure 1: The distance JS (P, Q) when estimating the divergence of the distributions exp 1-2, kspp 1-4: k=3 / V=30

In 20 experiments, when comparing the intervals 1-2-1-1-2, the maximum distance was 0.32, and the minimum was 0.06, while in experiments 1-4 it was 0.58 and 0.33, respectively. Similar experiments were performed when estimating discrepancies for samples V=40, 60, and 100.

Figure 2: shows the values of the distance JS (P, Q) when comparing the sample distributions for different intervals: exp 1-2, exp 2-3, exp 1-4 at V=100.

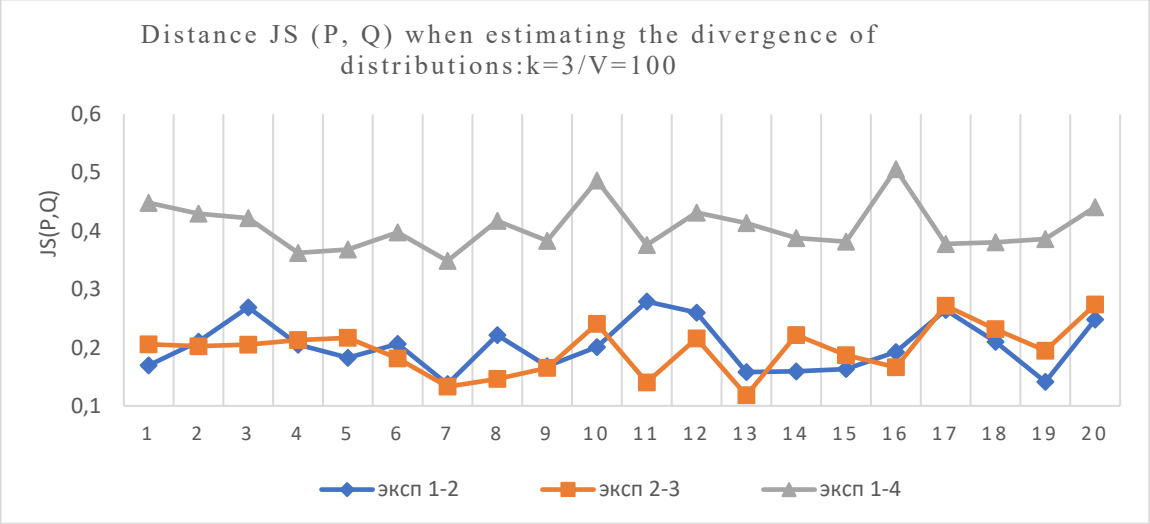


Figure 2: Distance JS (P, Q) when estimating the divergence of distributions exp 1-2, exp 2-3, exp 1-4: k=3 / V=100

In 20 experiments exp1-2, exp2-3, the maximum distance was 0.28, and the minimum distance was 0.12, while in experiments 1-4 it was 0.51 and 0.35, respectively. Thus, as the sample size increases, we observe an increase in the distance between homogeneous and inhomogeneous distributions. In this case, max (JS (P, Q)) =0.28 for homogeneous distributions 1-2, 2-3 is less than the minimum distance for inhomogeneous distributions 1-4: min (JS (P, Q)) =0.35. This fact indicates an increase in the reliability of the classification of object states and a decrease in the number of errors of the first and second kind.

Figure 3: shows the dependence of the mean square deviation of the co-ordinate distance JS (P, Q) between homogeneous 1-2, 2-3 and inhomogeneous 1-4 distributions, depending on the sample size at k=3. The figure shows that there is a tendency to decrease the value of the standard deviation from the sample size in all experiments.

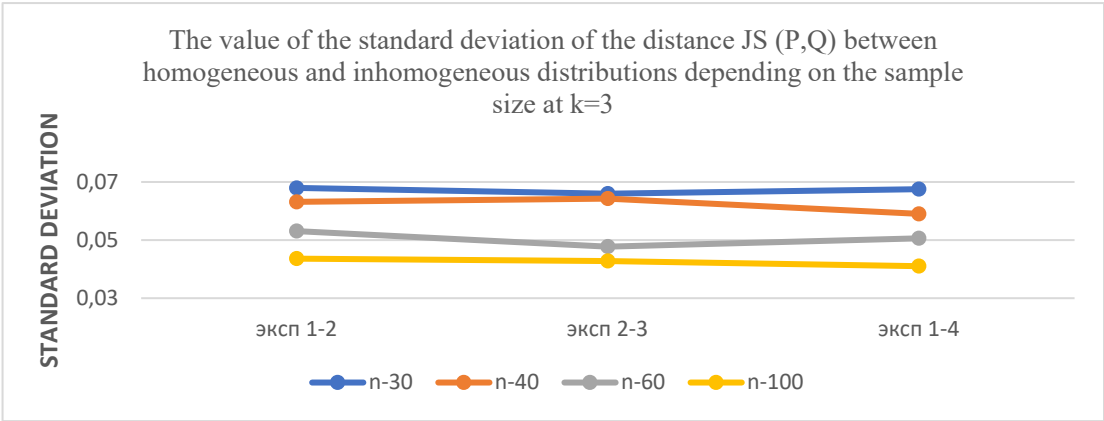


Figure 3: The value of the standard deviation of the distance JS (P, Q) between homogeneous and inhomogeneous distributions, depending on the sample size at k=3

Research of the influence of the sample volume-V at the changed number of intervals k=4. Set: resource state classification zones $[Z_i; Z_{i+1}]$ for four states that differ in the width of the intervals. The intervals number k=4. The following recognition zone boundaries were set:

- intervals-1 {0; 0.30; 0.60; 0.90; 1},
- intervals-2 {0; 0.20; 0.40; 0.80; 1},
- intervals -3 {0;0.10; 0.30; 0.50; 1}.

The experiments were carried out according to the scenario of the previous one. The first pair of samples 1-2 differs slightly in the width of the intervals, while the second pair 1-3 differs significantly. Figure 4: shows the values of the distance JS (P, Q) when comparing the sample distributions for different values of the intervals: exp1-2, exp1-3 with a sample size of V=100.

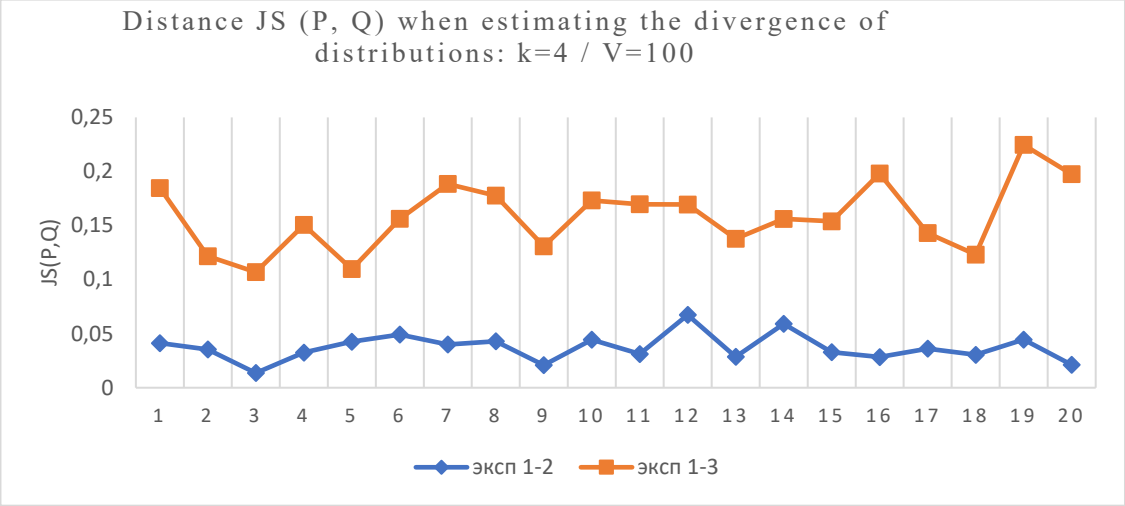


Figure 4: Distance JS (P, Q) when estimating the divergence of distributions exp 1-2, exp 1-3: k=4 / V=100

Figure 4: shows the values of the distance JS (P, Q) for k=4 and V=30, 100. It should be noted that the spread between the maximum and minimum values of the distance JS (P, Q) both with an increase in volume and with an increase in the value of k decreased: when comparing homogeneous samples, it was 0.05, and heterogeneous samples – 0.12 at V=100, and at V=30, respectively, it was 0.19 and 0.25

Table 1

Distance values JS (P, Q) for k=4 and V=30, 100

V	V=30		V=100V	
	exp 1-2	exp 1-3	exp 1-2	exp 1-3
max	0.312898	0.571586	0.067323	0.224536
min	0.124797	0.324762	0.013745	0.106922
Average	0.201451	0.429998	0.037132	0.158575
Δ	0.188101	0.246824	0.053578	0.117614

This trend is observed in other experiments when the number of intervals increases. Similar experiments were performed when estimating discrepancies for samples V=30, 40, 60, and 100 and the number of intervals k=5, 7, and others

3. Conclusions

The results obtained allow us to state that the application of the model based on the Jensen — Shannon information measure provides greater statistical stability in assessing changes in the UMV resources state with an increase in the samples volume and the intervals number. In the conducted experiments, the best estimates were obtained at V=100, K=5, and the worst at V=30, k=3. With a small

number of intervals and a small volume of samples, there were situations when individual intervals of the histogram contained zero values.

Thus, the obtained results of the study of the model based on the Jensen-Shannon information measure confirm the facts of the presence of perturbations in the assessment of changes in the state of objects. The main advantages of the proposed method for assessing the UMV resources state are: sensitivity to changes in the resources state, low computational complexity, adaptability to external influences. It is the assessment of the resources state heterogeneity over different time periods that can be used to detect external influences on the UMV.

Conducted to date in the field of vehicle protection has solved a number of safety problems and offered a many solutions. However, there are still open problems that require further study. The need to solve problems that ensure the security of the critical information infrastructure in the "smart city" is due not only to the growth trends of traffic flows, but also to significant changes in the digital technologies field used on vehicles, when interaction with the environment is carried out through the network through interfaces: V2V, V2X, V2P ,V2G, V2D. The article considers a simulation model that allows you to simulate the changes dynamics in the objects states and can be considered as one of the possible approaches to improve the methods of protecting critical objects, in particular, intelligent vehicles in VANET networks.

Currently, the process of anomalies rapid detection in the monitoring data of critical infrastructure objects is a complex, time-consuming and difficult to formalize task. Intrusion detection systems (IDS) are the most effective counter-measure and the most reliable approach to ensure the protection of automotive networks or traditional computer networks [1, 2]. In complex information systems for monitoring critical objects, a decision-making support mechanism is implemented to identify the control object critical state. The combined use of operational monitoring tools, simulation modeling, and probabilistic models allows us to predict the dynamics of state changes and proactively perform corrective actions, thereby preventing the emergencies occurrence.

4. Acknowledgements

The research was carried out with the financial support of the RFBR in the framework of scientific projects № 19-29-06015 and № 19-29-06023.

5. References

[1] L. Pan, X. Zheng, H. Chen, et al., Cyber security attacks to modern vehicular systems, Journal of Information Security and Applications 36 (2017) 90–100.

[2] M. Markovitz, A. Wool, Field classification, modeling and anomaly detection in unknown can bus networks, Vehicular Communications 9 (2017) 43–52.

[3] D. K. Nilsson, U. E. Larson, F. Picasso, et al., A first simulation of attacks in the automotive network communications protocol flexray, Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08. Springer (2009) 84–91.

[4] Top 20 and 200 most scanned ports in the cybersecurity industry, SecurityTrails blog (2019) securitytrails team. https://securitytrails.com/blog/top-scanned-ports.

[5] S. Checkoway, D. McCoy, B. Kantor, et al., Comprehensive Experimental Analyses of Automotive Attack Surfaces. https://web.archive.org/web/20150221064614/http://www.autosec.org/pubs/cars-usenixsec2011.pdf.

[6] F. Nielsen, R. Nock, Total Jensen divergences: definition, properties and clustering, In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016–2020.

[7] I. Butun, S. D. Morgera, R. Sankar, A survey of intrusion detection systems in wireless sensor networks, IEEE communications surveys & tutorials 16(1) (2014) 266–282.

[8] Mohammed Ali Hezam Al Junaid, A. A. Syed, et al., Classification of Security Attacks in VANET: A Review of Requirements and Perspectives, MATEC Web of Conferences 150 06038 (2018), MUCET 201. https://doi.org/10.1051/matecconf/201815006038.

[9] V. D. Boev, Conceptual system design in Anylogic 7 and GPSS World, Moscow, NOI, p. 556, 2016. ISBN: 978-5-9556-0161-8.2015.

[10] L. Kleinrock, Queueing theory / from Engl I I Grushko / V I Neiman. M Mashinostroenie, p. 432, 1979.
[11] S. A. Aivazian, V. S. Mhitaryan, Applied statistics and fundamentals of econometrics. M: High School, Publ «Yunity», p. 1000, 1998.
[12] A. Skatkov, A. Bryukhovetskiy, V. Shevchenko, Monitoring of qualitative changes of network traffic states based on the heteroscedasticity effect, Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings, Baku (2016) 7991765.
[13] A. V. Skatkov, A. A. Bryukhovetskiy, D. V. Moiseev, Intelligent monitoring system for solving large-scale scientific problems in cloud computing environments, Information and control systems 2(87) (2017) 19–25.