

Modeling the change detection process state of objects in monitoring data

Alexey Bryukhovetskiy¹ and Dmitriy Moiseev¹

¹ Sevastopol state university, 33 Universitetskaya str., Sevastopol, 299053, Russia

Abstract

A model for detecting anomalous data based on the research of sample uniformity is proposed. The method is designed to solve the detecting changes problem in the flow' state of controlled data using normal and gamma distributions models based on the Spearman nonparametric statistics criterion. The research results about intensity values that influence on requests generating, the intensity of servicing applications, the system load, the volume of samples, the time points of measuring characteristics and significance levels on the change the control object state are presented. The method can be used to control the process of detecting changes in UMV resources states. 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 are the most effective counter-measure and the most reliable approach to ensure the protection of automotive networks or traditional computer networks.

Keywords

Simulation modeling, queuing system, uniformity of samples, heteroskedasticity effect, complex systems modeling

1. Introduction

Research 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.

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

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

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



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. Problem statement

It is necessary to develop a software simulation model of a critical object in the form of a queuing system and conduct a series of experiments on it in order to test the effectiveness of the proposed intelligent methods for detecting changes in the monitoring objects state, as well as to investigate the critical parameters influence on changes in the object state:

- request generation rate – λ and $\sigma(\lambda)$;
- application service intensity – μ and $\sigma(\mu)$;
- system load – ρ ;
- sample sizes – n ;
- significance levels – p .
- time moments of characteristics measurements – τ ;

The influence of these parameters is estimated using normal and gamma distribution models based on Spearman's nonparametric statistics criterion. To simulate the changing object state process, a discrete-event model of the general-purpose simulation environment Anylogic was used [3], the information metrics calculations was carried out in MATLAB. In terms of the theory of queuing, the model under study is proposed to be considered as a queuing system (QS) of type G/G1/M/N [4]. The changes detection in the samples uniformity is based on the effect of heteroscedasticity (G-effect), as an observations heterogeneity, expressed in the unequal random error variance of the regression model [5]. The scheme and procedure of modeling are presented in [6 - 10].

3. Conducting model research experiments

Let's conduct an experiments series in which the receipt intensity and applications processing in the system are distributed according to the normal law [10 - 11].

Consider a single-channel QS with the following characteristics:

- request generation rate $M(\lambda)=0.7$, $\sigma(\lambda)=0.09$;
- application service intensity $M(\mu)=0.8$ and $\sigma(\mu)=[0.17; 0.2; \dots ; 0.4]$;
- storage capacity $N=5$;
- method of organizing and manipulating applications – *FCFS*.

The critical values of the Student's t-test for the given sample sizes and confidence probabilities are presented in Table 1.

Table 1

xr attribute discretization scheme

P	n=20	n=40	n=60
0.01	2.87844	2.711558	2.663287
0.005	3.196574	2.980293	2.918394
0.001	3.921646	3.565678	3.466329

Table 2 presents the experimental results in the form of the Student's t-test observed values. We distinguish the observed values that exceed the Student's critical values and correspond to the detection of the G-effect at given confidence probabilities p , as follows: for $p=0.001$ (blue), $p=0.005$ (green), $p=0.01$ (red). On the graph, the lines of confidence probabilities p are colored in the same colors.

Table 2

Experimental results- $t(n)$ for intensity different deviations $\sigma(\mu)$ and sample volumes – n

$\sigma(\mu)$	$t(20-2)$	$t(40-2)$	$t(60-2)$
0.17	2.272589	2.834956	3.129703
0.2	2.244760	3.098965	3.325995
0.24	2.584886	3.267130	3.503136

0.26	2.742413	4.003362	3.982638
0.32	2.908371	3.879590	3.952908
0.33	3.240325	3.929465	4.049291
0.35	4.149874	3.995526	4.152842
0.37	3.956550	3.956550	4.159725
0.38	3.921756	3.921756	4.165989
0.4	4.149874	3.937187	4.230893

The experimental results from Figure 1: are shown at the diagram on Figure 3:. Critical values are shown for the sample $n=40$. As we can see from the graphs with the sample size $n=40$, and significance level $p=0.01$, the model is constantly located in the hetero region, which calls into question the objectivity of the results obtained. In this example, we recommend using the significance level $p=0.001$.

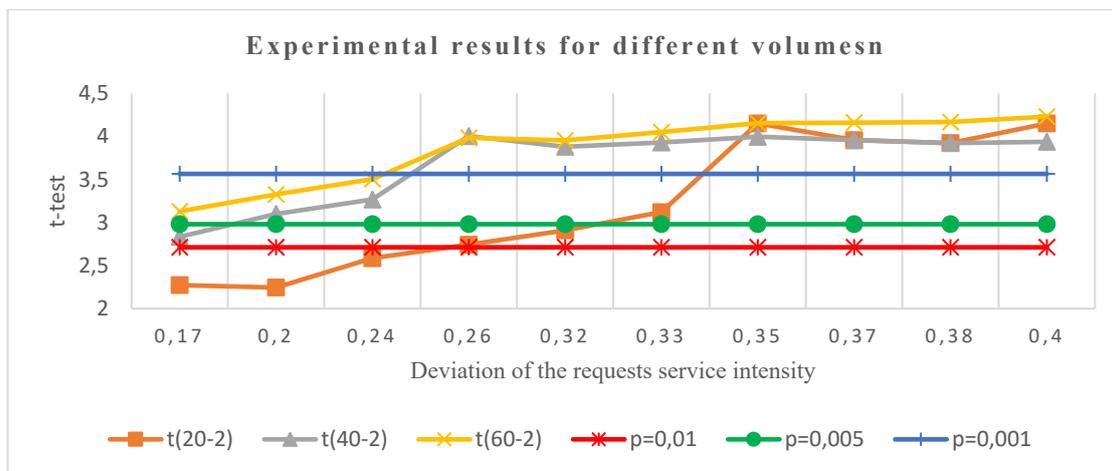


Figure 1: Plot of the dependence of the observed values of the Student's t-test at $\lambda=0.7$ $\mu=0.8$ for different sample sizes n

We will conduct a series of experiments on the simulation model, varying the time points $\tau = 21, 31, 41$ of characteristics, for samples with volume $n = 20, 40, 60$. The generating requests intensity $M(\lambda)$ and servicing requests intensity $M(\mu)$ don't change. Critical values are shown for the sample $n=20$. The experimental results - $t(n)$ for different time points τ , at $n=20$ are shown on Figure 2:. For the time points $\tau=21, 41$ at $n=20$ G- the effect does not appear.

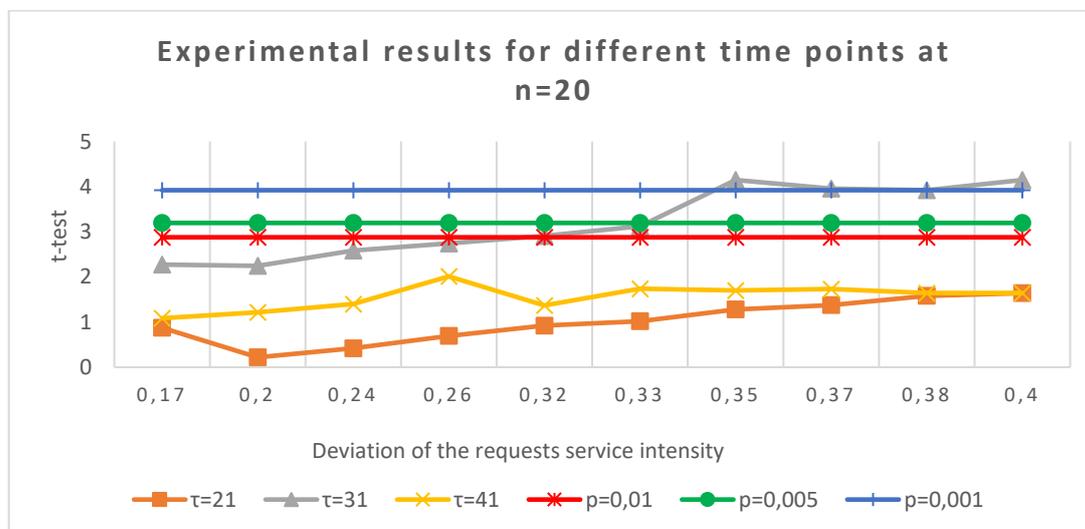


Figure 2: Experimental results - $t(n)$ for different time points τ at $n=20$

In other experiments, for the time moments $\tau=21, 31, 41$ at $n=40$ and at $n=60$ the G-effect manifests itself, and is stable for $\tau=31, 41$. At $\tau=21$ the G-effect is unstable for the specified sample volumes.

Let's conduct a series of experiments with the changed values of the mathematical expectation and the standard deviation:

- the intensity of requests received $M(\lambda)=0.8, \sigma(\lambda)=0.02$;
- request service intensity $M(\mu)=0.9$ and $\sigma(\mu)=[0.17; 0.2; \dots; 0.43]$.

Experimental results - $t(n)$ for $n=[20, 40, 60]$ are shown at Figure 3:. Critical values are shown for the sample $n=40$.

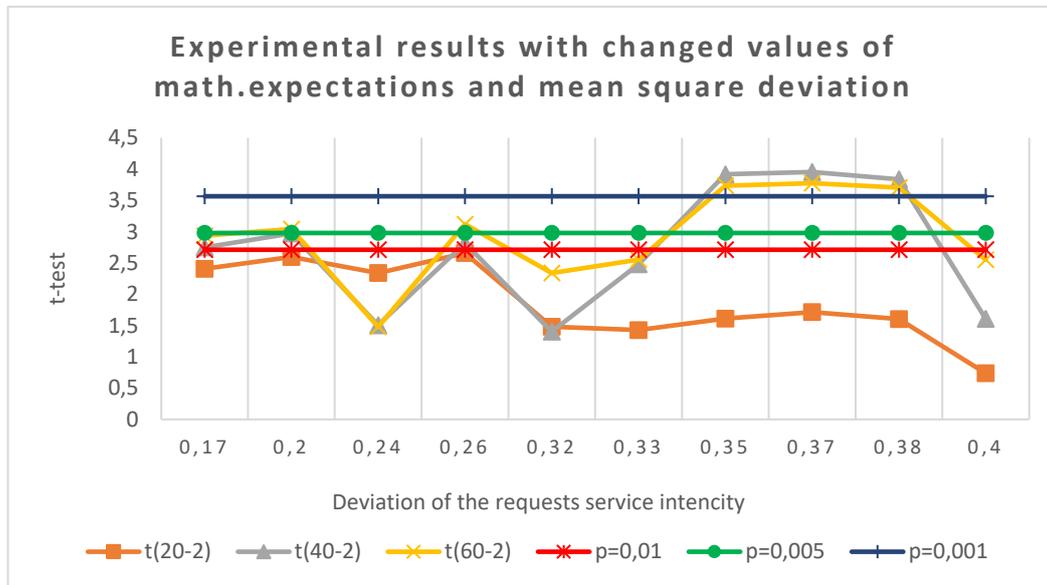


Figure 3: Experimental results - $t(n)$ for different sample volume n

Based on the results obtained, we can conclude that in the experiment, for a sample of 40 values and a significance level of $p=0.001$ we observe only one region of heteroscedasticity. When the significance level of p increases, for example, to 0.005 and $p=0.01$, two more heterogeneous regions are observed. When changing $\sigma(\lambda)$ o a value of 0.02 compared to $\sigma(\lambda)=0.09$ (Figure 1:) the G- effect s observed less frequently and the ranges of changes in the values of $\sigma(\mu)$ have decreased. So, the range $\sigma(\mu)$, in which the G-effect is observed for $n=40$ and $n=60$, is: $\sigma(\mu)=0.35-0.38$.

During the experiments, it was found that under load $\rho=0.875, \rho=0.889$ the G-effect is observed. For $\rho=0.5$ and $\rho=0.375$ the model is permanently located in the hetero region.

4. Gamma distribution of requests intensities

Let's conduct a series of experiments in which the intensity of receipt and processing of applications in the system is distributed according to the gamma law. The random value generator using the gamma distribution –gamma (α, β) has two parameters: α – shape parameter > 0 , β – scale parameter > 0 . The scale parameter β and shape parameter α of this distribution are related to the mathematical expectation and variance by two equations:

$$\begin{cases} \alpha * \beta = M \\ \alpha * \beta^2 = \sigma^2 \end{cases}$$

The solution of this system of equations allows us to determine the maximum plausible values of the parameters of the gamma distribution. We convert the parameters values of the normal distribution to the gamma distribution. We get the following request generation rate: for $M(\lambda)=0.7, \sigma(\lambda)=0.09 \Rightarrow \alpha=5.5, \beta=0.128$.

Values of the requests service intensity are shown at the Table 3.

Table 3

Converting the parameters of the normal distribution to the gamma distribution

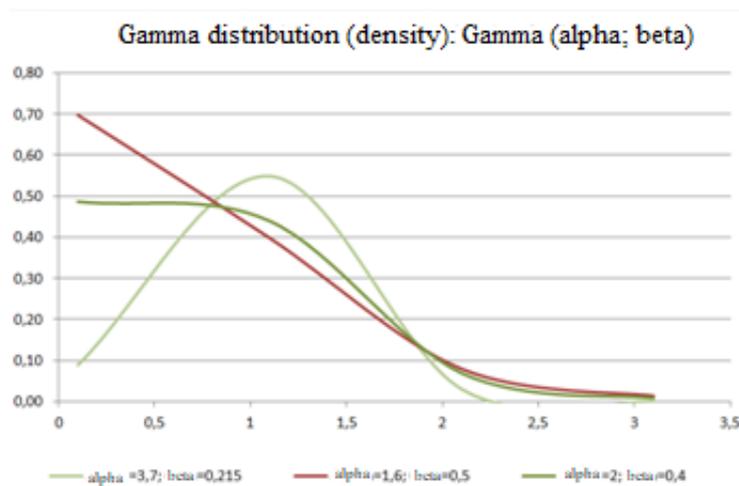
μ	$\sigma(\mu)$	α	β
0.8	0.17	3.7	0.215
0.8	0.2	3.2	0.25
0.8	0.24	2.7	0.297
0.8	0.26	2.45	0.326
0.8	0.32	2	0.4
0.8	0.33	1.92	0.416
0.8	0.35	1.8	0.445
0.8	0.37	1.73	0.463
0.8	0.38	1.7	0.473
0.8	0.4	1.6	0.5

We present gamma distribution graphs for some examples of parameter values presented in Table 4

Table 4

Examples of parameter values for a gamma distribution

μ	$\sigma(\mu)$	α	β
Parameter	Example 1	Example 2	Example 3
alpha	3.7	2	1.6
beta	0.215	0.4	0.5
mathematical expectation	0.80	0.80	0.80
dispersion	0.17	0.32	0.40
mode	0.58	0.40	0.30
x	alpha =3.7; beta =0.215	alpha =2; beta =0.4	alpha =1.6; beta =0.5
0.1	0.08867	0.48675	0.69773
1.1	0.54899	0.43950	0.39805
2.1	0.03005	0.06887	0.07940
3.1	0.00082	0.00835	0.01357

**Figure 4:** Examples of distribution function' graphs with different parameters

After converting the parameters, we will perform a system simulation for the parameters from Figure 3:. The simulation results for different samples are presented on Table 5, where we note the observed values for $p=0.001$ – blue, $p=0.005$ – green, $p=0.01$ – red, which correspond to the G-effect.

Table 5

Experimental results for different volumes of n

α	β	$\sigma(\mu)$	$t(20-2)$	$t(40-2)$	$t(60-2)$
3.7	0.215	0.17	2.71026823	3.4551096	5.84330347
3,2	0.25	0.2	3.82052841	3.31260798	4.77561313
2.7	0.297	0.24	1.79406021	3.53239149	4.67969141
2.45	0.326	0.26	4.79421496	6.1844856	7.35062798
2	0.4	0.32	3.83611859	2.84421792	4.97467211
1.92	0.416	0.33	6.74909246	6.70973064	6.68590119
1.8	0.445	0.35	3.85179474	6.50216333	6.65408968
1.73	0.463	0.37	3.78960209	5.31104172	4.34127791
1.7	0.473	0.38	4.04699038	6.27757451	9.63782085
1.6	0.5	0.4	1.92696007	4.29294489	4.06779261

We present the results on Table 5 in the form of graphs for $n=20, 40, 60$ (Figure 5):

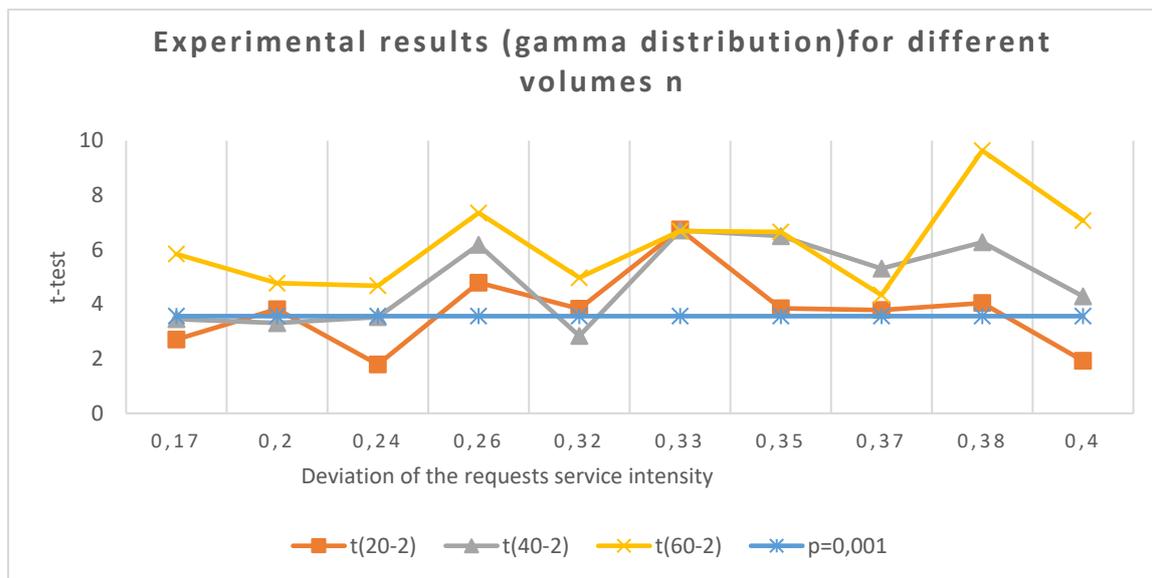


Figure 5: Experimental results - $t(n)$ for different sample sizes n

When the sample size is $n=60$ the model is constantly located in the hetero region. When comparing the results obtained for the gamma distribution with the normal distribution, the most stable areas of the G-effect were observed, as before, with a sample size of $n=40, p=0.001$.

5. Conclusions

Experiments were carried out to determine the influence of a number parameter values on changes in objects state using the gamma distribution. The results obtained allow us to state that the established parameters of the gamma distribution allow us to ensure the adequacy of the simulation model. The simulation results for the gamma distribution confirm the stable G-effect detection regions obtained for the normal law of intensity distribution. Moreover, in a model based on the gamma distribution, the hetero regions have a larger time span compared to the model using the normal distribution law.

Experiments have shown that the most optimal choice for the gamma distribution, which provides a high confidence in the detection of the G-effect, is to use a sample size of $n=40$. When using a sample

of $n=20$, the heteroscedasticity is unstable, and when using a sample of $n=60$, the model was constantly in the hetero region.

Various points of time τ for measuring the characteristics were investigated. Thus, it was found that for $\tau = 21$ and $\tau = 41$ the G-effect was rarely observed, and for $\tau=51$ – it was practically not observed. Based on the above, it is recommended to use $\tau=31$ in order to increase the reliability of monitoring objects.

The general recommendation for evaluating the detection of the G-effect is the following. Depending on the monitored system purpose, whether it is a critical purpose or not, the expert has the right to set values of the significance level – p , for which, on the one hand, high reliability of the characteristics values of the monitored objects will be ensured, on the other hand, the minimum errors number of the first and second kind is achieved, which means that the risks of making erroneous decisions will be reduced, which in turn will ensure minimal losses.

The results obtained in the conducted experiments with models based on various distribution laws of the requests flow intensity and the criteria used are used as input information for intelligent decision support systems for choosing a monitoring strategy, depending on the such parameters values as the required volume random variables samples; frequency, depth, monitoring frequency; moments of measurement of the characteristics of the control objects; the environment state, and others.

6. 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.

7. References

- [1] 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.
- [2] Mohammed Ali Hezam Al Junaid, Syed A. A. et al. Classification of Security Attacks in VANET: A Review of Requirements and Perspectives, *MATEC Web of Conferences* 150, (2018) 06038, <https://doi.org/10.1051/mateconf/201815006038>.
- [3] V. D. Boev, *Conceptual system design in Anylogic 7 and GPSS World*, M., NOI, p. 556, 2016. ISBN: 978-5-9556-0161-8.2015.
- [4] F. Sakiz, S. Sen, A survey of attacks and detection mechanisms on intelligent transportation systems: Vanets and IoV Ad Hoc Networks 61 (2017) 33–50.
- [5] S. A. Aivazian, V. S. Mhitaryan, *Applied statistics and fundamentals of econometrics*, Moscow, High School, Publ «Yunity», 1998, p. 1000.
- [6] Skatkov A., Bryukhovetskiy A., Shevchenko V. Monitoring of qualitative changes of network traffic states based on the heteroscedasticity effect, *Application of Information and Communication Technologies, AICT 2016, Conference Proceedings*, Baku, 12-14 Oct. 2016. 7991765.
- [7] P. D. Zegzhda, M. A. Poltavtseva, D. S. Lavrov, Systematization of cyberphysical systems and assessment of their security? *Problems of information security Computer system 2* (2017) 127–138.
- [8] H. Hasrouny, A. Samhat, C. Bassil, A. Laouiti, Vanet security challenges and solutions: A survey *Vehicular Communications* 7 (2017) 7–20.
- [9] B. Mokhtar, M. Azab, Survey on security issues in vehicular adhoc networks *Alexandria Engineering Journal* 54(4) (2015) 1115–1126.
- [10] A. V. Skatkov, A. A. Bryukhovetskiy, D. V. Moiseev, Intelligent monitoring system for solving large-scale scientific tasks in cloud computing environments, *Information and control systems 2* (2017) 19–25.
- [11] A. V. Skatkov, A. A. Bryukhovetskiy, D. V. Moiseev, Kullback measure in dynamic clustering problems of environment state observations, *Environmental monitoring systems*, 3(37) (2019) 35–38.