UDC 004.415.53:519.711
DOI: 10.15587/1729-4061.2022.269128

DEVISING A PROCEDURE FOR DEFINING THE GENERAL CRITERIA OF ABNORMAL BEHAVIOR OF A COMPUTER SYSTEM BASED ON THE IMPROVED CRITERION OF UNIFORMITY OF INPUT DATA SAMPLES

Serhii Semenov
Doctor of Technical Sciences, Professor
Department of «Cyber Security and Information Technologies»*

Oleksandr Mozhaiev
Corresponding author
Doctor of Technical Sciences, Professor**
E-mail: mozhaev1957@gmail.com

Nina Kuchuk
Doctor of Technical Sciences, Professor***

Mykhailo Mozhaiev
Doctor of Technical Sciences, Head of Laboratory
Laboratory of Copyright and Information Technologies****

Serhii Tiulieniev
PhD, Director****

Yurii Gnusov
PhD, Associate Professor**

Dmytro Yevstrat
PhD, Associate Professor
Department of Information Systems*

Yuliia Chyrva
PhD, Associate Professor
Department of Information Systems*

Heorhii Kuchuk
Doctor of Technical Sciences, Professor***

*Simon Kuznets Kharkiv National University of Economics
Nauki ave., 9-A, Kharkiv, Ukraine, 61166
**Department of «Cyber Security and DATA Technologies»
Kharkiv National University of Internal Affairs
L. Landau ave., 27, Kharkiv, Ukraine, 61080
***Department of «Computer engineering and programming»
National Technical University «Kharkiv Polytechnic Institute»
Kyrpychova str., 2, Kharkiv, Ukraine, 61002
****Scientific Research Centre for Forensic on Intellectual Property
of the Ministry of Justice of Ukraine
L. Ukrainky blvd., 26, Kyiv, Ukraine, 01133

The object of this study was the process of detecting anomalies in computer systems. The task to timely detect anomalies in computer systems was solved, based on a mathematical model underlying which is the criteria for uniformity of samples of input data. The necessity and possibility to devise a universal and at the same time scientifically based approach to tracking the states of the system were determined. Therefore, the purpose of this work was to develop a methodology for determining the general criterion of anomaly in the behavior of a computer system depending on the input data. This will increase the reliability of identifying the anomaly in the behavior of the system, which, in turn, should increase its safety. To solve the problem, a mathematical model for detecting anomalies in the behavior of a computer system has been built. The mathematical model differs from the well-known ones in the possibility of isolating a series of observations, the results of which show the anomaly in the behavior of the computer system. This made it possible to ensure the necessary level of reliability of the results of monitoring and research. In the process of modeling, the criteria for uniformity of samples of input data have been investigated and improved. The expediency of using the improved criterion of uniformity of samples of input data in the case of a significantly unequal distribution of values from the sensors of computer systems has been proved. An algorithm for the functioning of the software test tool has been developed. The results of the study showed that the confidence probability that the value of the statistical values of the shift in a certain criterion does not deviate from the mathematical expectation by more than 0.05 is approximately equal to 0.94. The scope of the obtained results is systems for detecting anomalies of computer systems. A necessary condition for the use of the proposed results is the presence of a series of observations of the state of the computer system

Keywords: computer system, network activity, anomaly criterion, vector statistics, homogeneous sample

Received date 28.09.2022
Accepted date 11.12.2022
Published date 30.12.2022

***How to Cite:** Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Y., Yevstrat, D., Chyrva, Y., Kuchuk, H. (2022). Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples. Eastern-European Journal of Enterprise Technologies, 6 (4 (120)), 40–49. doi: https://doi.org/10.15587/1729-4061.2022.269128*

1. Introduction

The issue related to the constant updating, improvement, and spread of malicious cyber influence is considered in many scientific works, in particular in [1, 2]. It should be noted that this occurs against the background of a significant technological and methodological lag in methods and means of detecting anomalous situations in computer systems (CS) [3].

This trend has so far continued despite the fact that recently separate means of detecting abnormal behavior have been developed and used (Awake Security Adversarial Modeling, Cisco Stealthwatch, Flowmon NBAD, etc.). This confirms that under the conditions of rapid development of information technologies and, as a result, constant modernization of CS software and hardware, solving only partial problems of detecting anomalies cannot ensure the safety of CS [4]. A universal and at the same time scientifically based approach to tracking the states of the system is needed.

This approach may consist of several components. One of the elements of this approach should be a procedure for determining the general criterion of anomaly in CS behavior, depending on the input data. Therefore, studies aiming at constructing a mathematical model for detecting anomalies in the behavior of a computer system based on the improved criterion of uniformity of samples of input data are relevant.

2. Literature review and problem statement

Most studies that consider the timely detection of anomalies in computer systems are based on the development of a general criterion for the anomaly in CS behavior. Thus, in [2], there are a number of fundamental facts that make it possible to assert the expediency and validity of the use of criteria for the power of scattering samples in the construction of the criterion of anomaly. Research into this area is complemented and confirmed by the results in [3], which considers a multi-parametric mathematical model of observations on the behavior of the system. In [3], from a mathematical point of view, the result of observations of the selected characteristics is considered and the asymptotic behavior of many statistics of various criteria, including the criterion of homogeneity, is considered. It should be noted at the same time that works [2, 3] have a more general theoretical character and do not directly relate to the rules of CS conduct.

In [4], a mathematical model of data formation has been built, taking into account statistical patterns that, according to the author's assumption, accompany the functioning of CS software that serves the workplace of the user or server. In this case, a series of sequences of observations is modeled by implementations of a series of independent random variables.

In particular, this is reported in studies on the abnormal behavior of users of social networks, when the sequence of their expressions in the semantic text is modeled by a sequence of independent random variables. These random variables have the same distribution within the series, on the set of possible values of the results, that is, the probability of distribution may be the same or different, depending on the assumptions in this model.

In the simulation of CS, the interpretation of the set of possible results can also be used. In particular, one can consider all possible options for the states of the sensors.

Such series of observations in the mathematical literature [5] are called schemes of independent polynomial tests or polynomial schemes. The central question formulated in [5] for polynomial schemes is the determination by observations of a series of tests of coincidences of probabilistic distributions. The main hypothesis corresponds to the model of no impact on the system or models of regular work on the segments of observations. To identify the main hypothesis from observations of random variables, in [5] the criterion of uniformity chi-square, based on quadratic statistics, is used.

The criterion based on the application of the selected statistics in the problem of determining the homogeneity of samples is the best in the sense of application for constructing estimates of maximum plausibility. The effectiveness of the chi-squared criterion is estimated asymptotically. When the main hypothesis is confirmed, the distribution of statistics asymptotically tends to the distribution of chi-square, which is tabulated, for example, in [6].

In [6], it is emphasized and practically proved that the convergence of statistics to the distribution of chi-squared random variables takes place only for a distribution vector with independent coordinates, otherwise the number of degrees of freedom in the boundary distribution increases. However, when conducting an experiment, the number of nonzero coordinates of the probability vector is initially unknown.

In [7], it is shown that usually for the theoretical analysis of the situation the conditions of infinite time are modeled. In this case, the asymptotic behavior of the distribution of chi-squared statistics is similar to the asymptotic for chi-squares with an increasing number of degrees of freedom. Also in [7], for the study of homogeneity and the identification of components that differ, vector statistics are considered according to the criterion of homogeneity, which is useful precisely for assessing the uniformity of distributions, but less convenient.

It should be noted that works [2–7] are based on the time-tested provisions of probability theory and mathematical statistics and have a great degree of reliability. But, as the previous analysis showed, at the same time, part of the problem related to the correctness of replacing the «to the limit» distributions with «limit» ones has not yet been finally resolved. The problem of taking into account the correctness of the replacement of «to the boundary» distributions is not considered in [8–11]. Consequently, the task of devising a general criterion for the anomaly in CS behavior, which takes into account the correctness of replacing «up to the boundary» distributions with «limit» ones, depending on the input data, becomes relevant.

3. The aim and objectives of the study

The aim of this work is to devise a methodology for determining the general criterion of anomaly in CS behavior, depending on the input data. This will increase the reliability of detecting the anomaly in CS behavior, which, in turn, should increase its safety.

To accomplish the aim, the following tasks have been set:

– to investigate the criteria for uniformity of samples of input data;

– to build a mathematical model for detecting anomalies in the behavior of a computer system based on an improved criterion of uniformity of samples of input data;

– to investigate the conditions of use of the mathematical model constructed to detect anomalies in the behavior of a computer system.

4. The study materials and methods

The study object: the process of detecting anomalies in computer systems.

The main hypothesis of the study assumes that the integral incoming traffic of a computer system acquires significant fluctuations at different time intervals.

This study has the following limitations:

– availability of a series of observations on the state of the computer system;

– significantly unequal distribution of incoming traffic.

Considering the need for operational treatment of the resulting statistical materials from a large number of input sources, vector *r*-dimensional statistics were chosen for their analysis [8]:

$$\bar{\chi}_T^2 = \left(\chi_T^2(1),\ldots,\chi_T^2(r)\right) = = \sum_{j=1}^{N} \left(\frac{\left(\nu_{dj} - \frac{(\nu+_j)T_1}{T}\right)^2}{\frac{(\nu+_j)T_1}{T}}, \ldots, \frac{\left(\nu_{dj} - \frac{(\nu+_j)T_r}{T}\right)^2}{\frac{(\nu+_j)T_r}{T}} \right), \qquad (1)$$

where r – the number of series of observations, $\nu+_j = \sum_{d=1}^{r} \nu_{dj}$, ν_{dj} – the frequency of the j-th result during the implementation of the scheme, $\sum_{j=1}^{N} \nu_{dj} = T_d$, T – the number of tests, $T=T_1+\ldots+T_r$, $d=1, \ldots, r$, N – the number of experiments.

Criteria for uniformity of samples were investigated in [9]. For this purpose, the behavior of values $\max_{1\le j\le N}(\nu_j)$ and $\min_{1\le d\le r}(\nu_j)$ was investigated and their useful properties and advantages of their application in solving the membership problem were defined. In particular, such an advantage appeared in the case of a significantly unequal-probable distribution. But these values were formed on the basis of the composition of one-dimensional statistics. Therefore, the main task of applying these criteria is the need to justify their use when using vector *r*-dimensional statistics.

It is fundamentally possible with the further development of the anomaly detection system to supplement it with criteria based on statistics such as spacing [10, 11]. The article proposes to dwell on the use of statistics from Cressy and Reed since it would be more fundamental to obtain experimental confirmation of the correspondence of the concepts of «homogeneity of theoretical-probabilistic schemes» and «regular operation of the system».

In order to study the laws of changing the behavior of software operating in CS, as well as to assess the practical applicability of the analyzed method, a set of specialized software tools was developed. The main orientation of these tools is to ensure timely and complete accumulation of statistics on changes in the specified characteristics of the software. It is also important to interpret the data obtained to ensure the possibility of processing the information obtained by means of the studied mathematical apparatus. In addition, one needs to pay attention to the implementation of the algorithms necessary for the study.

The general flowchart of the algorithm for the functioning of the software for monitoring the state of CS is shown in Fig. 1.

The implementation of steps related to the registration of software processes and the processing of system signals differs depending on the platform on which the software was used. Individual aspects of the algorithms of the sensors are also implemented differently for different platforms. This is achieved by breaking the system into modules, each of which contains a functional hierarchy implemented using C++ language classes, which also facilitates the solution of problems of the extensibility of the software package and the logical organization of the program.

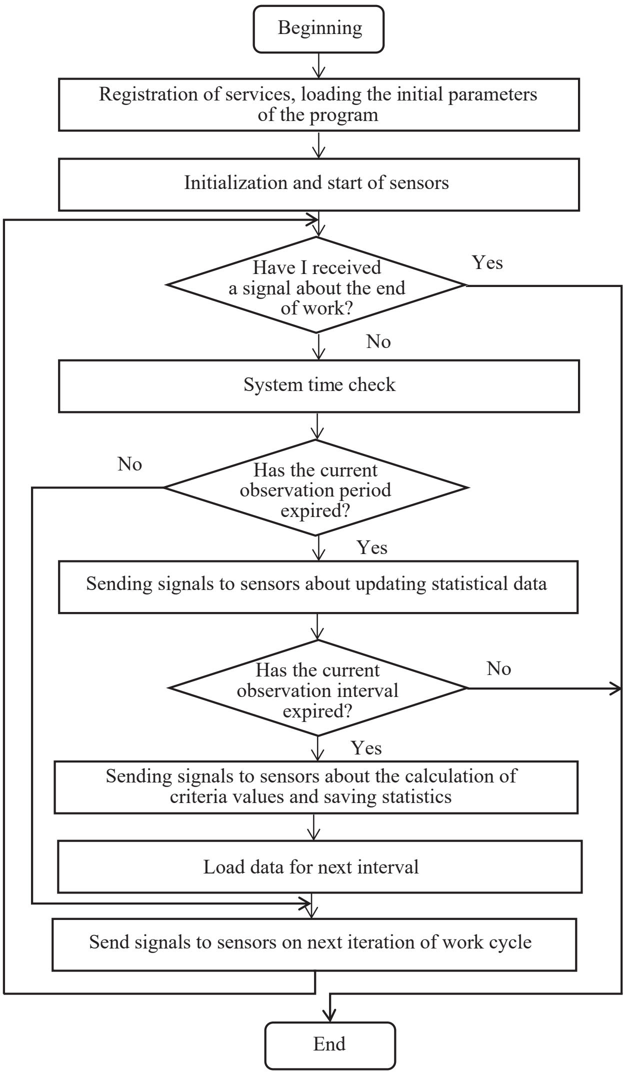
Beginning
Registration of services, loading the initial parameters of the program
Initialization and start of sensors
Have I received a signal about the end of work?
Yes
No
System time check
No
Has the current observation period expired?
Yes
Sending signals to sensors about updating statistical data
Has the current observation interval expired?
No
Yes
Sending signals to sensors about the calculation of criteria values and saving statistics
Load data for next interval
Send signals to sensors on next iteration of work cycle
End

Fig. 1. Algorithms for the functioning of the software tool for testing the application of the proposed approach

The software tool designed to test the proposed approach includes the following modules: network sensor, local activity sensors, implementation of a mathematical apparatus, service part. Each module of the software package solves a separate class of implementation problems.

The network sensor tests the preparation of a network adapter for the procedure of monitoring traffic, capturing, and filtering packets transmitted through the observed network interface. Also, with its help, the accumulation of statistical information takes place within one interval of observations and its subsequent transformation into a single type of representation of a statistical sample. This information is transmitted for processing to the module of the end of the observation interval [12–15].

Local activity sensors collect data on the degree of use of system resources. In particular, these are the following data:

– current processor usage;

– amount of free/occupied virtual memory;

– adjustment of frequency characteristics of the intensity of use of each type of resource;

– converting them into a single type of representation of a statistical sample;

– transmission of statistical data for processing by the module of the mathematical apparatus after the end of the observation interval.

5. Mathematical model for detecting anomalies in the behavior of a computer system based on an improved criterion of uniformity of samples of input data

5. 1. Research and improvement of criteria for uniformity of samples of input data

The distribution of statistics (1) at infinite time and fixed $T_1, \ldots, T_r \to \infty$, N=const will correspond to the distribution of the r-dimensional random vector, each coordinate of which is distributed over a chi-square with (N–1) degrees of freedom and which are in some way dependent on each other.

The use of this r-dimensional distribution for error calculations is problematic [10]. We use the results of work [8] where the case of the limiting behavior of the distribution of statistics (1) at $T_1, \ldots, T_r \to \infty$ and $N \to \infty$ is considered. In [8], it is shown that the limiting distribution for a vector random variable of the number of tests,

$$T\left(\frac{\chi_T^2(1)-N(1-a_1)}{\sqrt{2N(1-a_1)}},\ldots,\frac{\chi_T^2(1)-N(1-a_r)}{\sqrt{2N(1-a_r)}}\right), \qquad (2)$$

is an r-dimensional normal law with an average of 0 and a covariant matrix Q of size r*r, which is denoted by $N(0, Q)$, where

$$Q=\begin{pmatrix} 1 & & & & \frac{a_d\cdot a_s}{(1-a_a)\cdot(1-a_s)} \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ \frac{a_d\cdot a_s}{(1-a_a)\cdot(1-a_s)} & & & & 1 \end{pmatrix}, \qquad (3)$$

$$a_s=\frac{T_s}{T},\ d,s=1\ldots r.$$

Thus, the centered and normalized statistics χ_T^2 are concentrated accordingly in the case of the validity of the hypothesis H_0 inside the r-dimensional ellipse centered in O, and the size of the ellipse is determined mainly by the probability of errors.

Recall that $\chi_T^2=\left(\chi_T^2(1),\ldots,\chi_T^2(r)\right)$ and the convergence of statistics in (4) does not contradict the description of the convergence of statistics (5). Indeed, according to Taylor's formula, we get:

$$\frac{\chi_T^2-(r-1)\cdot(N-1)}{\sqrt{2(r-1)\cdot(N-1)}}=\frac{\chi_T^2(1)-(r-1)N}{\sqrt{2(r-1)N}}\left(1+O\left(\frac{1}{\sqrt{N}}\right)\right). \qquad (4)$$

Then we shall carry out the transformation:

$$\frac{\chi_T^2(1)-(r-1)N}{\sqrt{2(r-1)N}}=$$

$$=\sum_{s=1}^{r}\left(\frac{\chi_T^2(s)-N(1-a_s)}{(1-a_s)\sqrt{2(r-1)N\cdot(1-a_s)}}(1-a_s)\right), \qquad (5)$$

from which it follows that a random variable is given as a finite (r-terms) sum of random variables, in the aggregate converging according to representation (5) to the r-dimensional normal law. Then the weighted sum of the coordinates of a random vector converges to a normal already one-dimensional law with an average O and variance, therefore:

$$\left((1-a_1),\ldots,(1-a_r)\right)\cdot Q\begin{pmatrix}(1-a_1)\\ \vdots \\ (1-a_r)\end{pmatrix}=\sum_{s=1}^{r}(1-a_s)^2+2\sum_{s<d}a_d a_s;$$

$$\sum_{s=1}^{r}(1-a_s)^2+2\sum_{s<\mathrm{d}}a_d a_s=\sum_{s=1}^{r}\left(1-2a_s+a_s^2\right)+$$

$$+\left(\sum_{s-1}^{r}a_s\right)^2-\sum_{s-1}^{r}a_s^2=(r-2+1)=(r-1).$$

Thus, from the convergence of distribution (5) the convergence of the distribution (4) follows, which is more convenient due to one-dimensionality. At the same time, a random vector variable (5) is useful for its centering (zero value for all coordinates), and in the centering parameters. Normalization is absent, as in the formula for calculating statistics.

A certain development of the criteria was carried out in work [9], where, for various purposes, it is proposed to use criteria based on the statistics of scattering power:

$$I_T(\lambda)=\sum_{d=1}^{r}I_T(\lambda,d),$$

$$I_T(\lambda,d)=\begin{pmatrix}\lambda+1\\ 2\end{pmatrix}^{-1}\sum_{j=1}^{N}\left(\nu_{dj}\left(\left(\frac{\nu_{dj}T}{\nu+jT_d}\right)^{\lambda}-1\right)\right).$$

Statistics $I_T(\lambda)$ is the sum of the coordinates of vector statistics $\bar{I}_T(\lambda)=\left(I_T(\lambda,1),\ldots,I_T(\lambda,r)\right)$. Adding to the parameter λ some specific values leads, as indicated in [8], to reducing statistics $I_T(\lambda)$ to the corresponding known classical criteria. For example, $\bar{I}_T(1)=\chi_T^2$, and $I_T(1/2)$ meets Bernstein's statistical criterion. Also, in [9], it is noted that at $\lambda \to -1,0$, similar statistics for the criterion of consent:

$$I_\lambda(c)=\begin{pmatrix}\lambda+1\\ 2\end{pmatrix}^{-1}\sum_{j=1}^{N}\left(\nu_j\left(\left(\frac{\nu_j}{Tp_j}\right)^{\lambda}-1\right)\right)$$

converge at fixed N, T and $\nu_1, \ldots, \nu_N$ – the set of frequencies of the results in the polynomial scheme $M(T, P)$ to the statistics of the criteria for the plausibility ratio:

$$\lim_{\lambda\to 0}J_\lambda(c)=2\sum_{j=1}^{N}\nu_j\log\left(\frac{\nu_j}{Tp_j}\right), \qquad (6)$$

$$\lim_{\lambda\to 0}J_\lambda(c)=2\sum_{j=1}^{N}Tp_j\nu_j\log\left(\frac{Tp_j}{\nu_j}\right). \qquad (7)$$

This property is determined by the validity of the existence of the limit,

$$\log(t)=\lim_{h\to 0}\left(\frac{t^h-1}{h}\right), \qquad (8)$$

at valid h and t>0.

Let's consider what the statistics $I_T(\lambda)$ of the homogeneity criterion at λ→–1.0 λ→–1.0 will look like. From (8), the following ratios follow:

$$\lim_{\lambda\to 0} I_T(\lambda)=2\sum_{j=1}^{N}\sum_{d=1}^{r}\left(\nu_j\log\left(\frac{T\nu_{dj}}{T_d\nu+j}\right)\right), \qquad (9)$$

$$\lim_{\lambda\to 1} I_T(\lambda)=2\sum_{j=1}^{N}\sum_{d=1}^{r}\left(\frac{\nu+jT_d}{T}\ln\left(\frac{T\nu_{dj}}{T_d\nu+j}\right)\right). \qquad (10)$$

In the right parts of expressions (9) and (10), polynomial schemes $M(T_1,\ P_1)$,..., $M(T_r,\ P_r)$ are absolutely used, which are independent of the parameters P_1,..., P_r, as is the case with rounding statistics $I_T(\lambda)$.

Statistics (9) and (10) resemble statistics of the most powerful criteria for the ratio of plausibility (8).

Statistics $I_T(\lambda)$ have a property by analogy with statistics $J_\lambda(c)$:

$$\lim_{\lambda\to+\infty}\left[1+\frac{\lambda(\lambda+1)J_\lambda(c)}{2T}\right]^{\frac{1}{\lambda}}=\max_{1\le j\le N}\left(\frac{\nu_j}{Tp_j}\right),$$

$$\lim_{\lambda\to-\infty}\left[1+\frac{\lambda(\lambda+1)J_\lambda(c)}{2T}\right]^{\frac{1}{\lambda}}=\min_{1\le j\le N}\left(\frac{\nu_j}{Tp_j}\right).$$

Consider the limits:

$$\lim_{\lambda\to+\infty}\left[1+\frac{\lambda(\lambda+1)I_\lambda(\lambda)}{2T}\right]^{\frac{1}{\lambda}}=\lim_{\lambda\to+\infty}\left[\sum_{j=1}^{N}\sum_{d=1}^{r}\left(\frac{\nu_{dj}}{T}\left(\frac{T\nu_{dj}}{T_d\nu+j}\right)^{\lambda}\right)\right]^{\frac{1}{\lambda}},$$

hence

$$\lim_{\lambda\to+\infty}\left[1+\frac{\lambda(\lambda+1)I_\lambda(\lambda)}{2T}\right]^{\frac{1}{\lambda}}=\max_{1\le d\le r}\left(\max_{1\le j\le N}\left(\frac{\nu_{dj}T}{\nu+jT_d}\right)\right),$$

$$\lim_{\lambda\to-\infty}\left[1+\frac{\lambda(\lambda+1)I_T(c)}{2T}\right]^{\frac{1}{\lambda}}=\min_{1\le d\le r}\left(\min_{1\le j\le N}\left(\frac{\nu_{dj}T}{\nu+jT_d}\right)\right).$$

So, we come to the criteria close to those studied in work [9], where the behavior of $\max\limits_{1\le j\le N}(\nu_j)$ and $\min\limits_{1\le d\le r}(\nu_j)$ was investigated.

It can be expected that similar properties of statistics $I_T(\lambda)$ at large values of λ will also manifest themselves in the criteria of homogeneity. At the same time, to assess errors in the hypothesis H_1, it is necessary to consider the distribution of statistics $I_T(\lambda)$ for heterogeneous schemes.

5. 2. Mathematical model for detecting anomalies in the behavior of a computer system based on an improved criterion

The presence of heterogeneity can be considered as the absence of homogeneity and in this sense, to conclude that there is an invasion or change in the behavior of CS in case of inconsistency with the hypothesis H_0 of the observed r samples of observations of volumes T_1, ..., T_r. This approach will make it possible to obtain an estimate of the probability of the so-called error of the first kind, that is, the probability of the event: the criterion mistakenly rejected homogeneous samples because statistics $I_T(\lambda)$ exceeded the set limit. This event corresponds to the declaration of a false alarm. If such events happen often, then this harms the operation of the software system, which conducts observation in the sense of implementing the functionality embedded in it.

An error of the second kind, that is, the probability of skipping an invasion or extraneous actions of the operator, cannot be calculated without specifying the hypothesis H_1 alternative to H_0. At the same time, when a non-compliance with the requirement of homogeneity is detected, a logical question arises about the place of discrepancy. This is due to the definition of the period in which heterogeneity takes place, that is, with the identification of sample numbers in which the probability distribution of the results differs from other.

Such a statement of the problem mathematically is still quite undefined. Therefore, we consider two variants of the alternative hypothesis H_1, which relate to hypotheses close to the main hypothesis H_0. With such hypotheses, it is possible to use the results regarding the limiting behavior of statistics $I_T(\lambda)$ and its vector r-dimensional analog $\overline{I}_T(\lambda)=\left(I_T(\lambda,1),\ldots,I_T(\lambda,r)\right)$.

Based on (6), it is possible to propose an initial version of the algorithm for isolating samples that differ, that is, that do not meet the requirements of uniformity of distribution of probabilities of results and corresponding to hypothesis H_1. This algorithm implies referring to samples with numbers that differ, for example, for which

$$\frac{\left|I_T(\lambda,s)-N(1-a_s)\right|}{(1-a_s)\sqrt{2N}}>C\cdot\left(\frac{\beta}{r}\right), \qquad (11)$$

where $C(\beta)$ is the level of significance, depending on the magnitude of the error of the second kind of β and satisfying the equality:

$$\beta=\frac{1}{\sqrt{2\pi}}\int_{C(\beta)}^{\infty}e^{-\frac{x^2}{2}}\,\mathrm{d}x. \qquad (12)$$

Then, to attribute as a whole a set of observations to the hypothesis H_1, we can propose an algorithm similar to (11), that is, if:

$$\frac{\left|I_T(\lambda)-N(r-1)\right|}{\sqrt{2N(r-1)}}>C\cdot(\beta), \qquad (13)$$

then we consider the hypothesis H_1 to be true.

The proposed algorithms are based on inequalities (11) and (13), which are synthesized on the basis of intuitive ideas about the behavior of the distribution of statistics $I_T(\lambda)$ with an increase in the volume of observations T_1, ..., T_r with hypotheses other than H_0. Questions about the construction of an algorithm for distinguishing a complex hypothesis H_0 versus a complex alternative to H_0, has not yet been solved in studies on mathematical statistics. The problem of finding algorithms in one sense or another can be solved only with a significant specification (narrowing) of the hypotheses H_0 and H_1.

Paper [2] considers the marginal behavior of the distribution of statistics $I_T(\lambda)$ at $T_1 \ldots T_r\to\infty$ and $N\to\infty$. The conditions obtained in [2] for the convergence to the normal law of centered and normalized accordingly statistics $I_T(\lambda)$ are very complex. To obtain a qualitative picture that gives an idea of the parameters of the boundary distributions, consider a special case that can significantly simplify the type of convergence conditions.

An alternative hypothesis H_1 is defined in the form:

$$H_1\colon p_{dj}=p_j\left(1+\varepsilon_{dj}\right),\ -1<\varepsilon_{dj},\ d=1,\ldots,r,\ j=1,\ldots,N. \qquad (14)$$

This representation clearly shows the deviation of alternative hypotheses from the main H_0, in which $\varepsilon_{dj}=0$ for all possible values of d and j.

In [3], an alternative hypothesis is represented in a different way:

$$H_1:\ p_{dj}=\left(1+\delta(j)\right)\sum_{s=1}^{r}a_s p_{sj}, \tag{15}$$

where $a_s=T_r/T$, $-1\le\delta_d(j)$, $d=1,\ldots,r$, $j=1,\ldots,N$ and the conditions of asymptotic normality of the distribution of statistics $I_T(\lambda)$ are formed through the values of $\delta_d(j)$.

Representations (14) and (15) make it possible to determine the deviation of $\delta_d(j)$ through ε_{dj}.

Indeed, from equality:

$$p_{dj}=p_j\left(1+\varepsilon_{dj}\right)=\left(1+\delta_d(j)\right)\sum_{s=1}^{r}a_s p_j\left(1+\varepsilon_{dj}\right),$$

the following ratio follows:

$$\delta_d(j)=\frac{\varepsilon_{sj}-\sum_{s=1}^{r}a_s\varepsilon_{sj}}{1+\sum_{s=1}^{r}a_s\varepsilon_{sj}}. \tag{16}$$

We shall formulate one of the results that directly follows from work [3], and is necessary to describe the possibility of using statistics $I_T(\lambda)$ when distinguishing between the hypotheses H_0 and H_1.

Let for the vectors of results $P_1, \ldots, P_r$ in the corresponding polynomial schemes $M(T_1, P_1)$, ..., $M(T_r, P_r)$ there are constants c_1, c_2, c_3 in which $0<c_1<Np_{dj}<c_2<\infty$, $P=(p_{d1}, \ldots, p_{dN})$ and test volumes $T_1, \ldots, T_r$ and the numbers of possible results are related by the ratio $T_d/T=a_d$, $0<c_1<a_d<c_3<1$ and $T^{2/3}N^{-1}\to\infty$.

Alternative to hypothesis H_0 and H_1, we shall call close in the sense of fulfilling the condition:

$$\max_{d,j}\delta_d^2(j)=0\left(T^{-1/2}\right). \tag{17}$$

Then, in the hypothesis $H_0:\delta_d(j)=0$, $\delta_d(j)=0$, $d=1, \ldots, r$, $j=1, \ldots, N$ the distribution of the random variable:

$$\left(I_T(\lambda)-N(r-1)\right)\left(2N(r-1)\right)^{-1/2}$$

converges to the normal law $N(0,1)$.

With a close alternative to H_1, the distribution of a random variable is already:

$$\left(I_T(\lambda)-N(r-1)-A(T)\right)\left(2N(r-1)+B(T)\right)^{-1/2},$$

$B(T)=2A(T)-2\lambda\sum_{j=1}^{N}\sum_{d=1}^{r}\delta_d(j)$ converges to the same normal law, where

$$A(T)=T\cdot\sum_{d=1}^{r}\left(a_d\sum_{j=1}^{N}\left(\delta_d^2(j)\sum_{s=1}^{r}a_s p_{sj}\right)\right)+\lambda\sum_{j=1}^{N}\sum_{d=1}^{r}\delta_d(j).$$

It can be seen from this that the differences in the parameters of the boundary laws here are determined by the value of $A(T)$, depending on λ, and the value of $B(T)$, which does not depend on λ.

Thus, replacing the domezhny representations of histograms of the distribution of statistics $I_T(\lambda)$ with the limiting densities of normal laws, we obtain a visual illustration of asymptotic densities, as shown in Fig. 2.

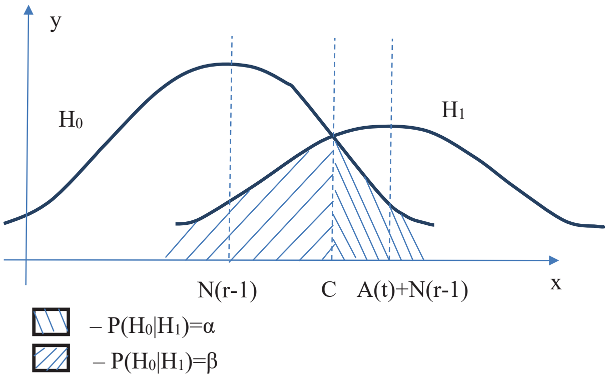

Fig. 2. Asymptotic densities and error values of the first and second kind

Here C is the limit of decision making, which can be shifted to the left or right, from the point of intersection of the density, with the corresponding correction of errors α and β.

The values of α and β are determined by relations $A(T)(\sqrt{2N(r-1)})^{-1}$ and $A(T)(\sqrt{2N(r-1)+B(T)})^{-1}$.

The larger these values, the less errors α and β.

At the same time, the type of parameters $A(T)$ and $B(T)$ is quite complicated in presenting the hypothesis H_1 using equality (15). In this regard, it is proposed to use the form (16) or (14). Then,

$$A(T)=T\sum_{s,d=1}^{r}\left(a_d a_s\sum_{j=1}^{N}p_j\frac{\left(\varepsilon_{sj}-\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}\right)^2}{1+\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}}\right)+$$
$$+\lambda\cdot\sum_{j=1}^{N}\sum_{d=1}^{r}\frac{\varepsilon_{dj}-\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}}{1+\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}}+\bar{0}\left(T^{-\frac{1}{4}}\right), \tag{18}$$

$$B(T)=2T\sum_{s,d=1}^{r}\left(a_d a_s\sum_{j=1}^{N}p_j\frac{\left(\varepsilon_{dj}-\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}\right)^2}{1+\sum_{\nu=1}^{r}a_\nu\varepsilon_{\nu j}}\right)+0\left(T^{-\frac{1}{4}}\right), \tag{19}$$

if required to fulfill (17) equality $\max_{d,j}\varepsilon_{dj}^2=o\left(T^{-1/2}\right)$.

Equalities (18), (19) make it possible to formulate conditions under which errors α and β can be infinitesimal and therefore a criterion based on statistics $I_T(\lambda)$ can be considered acceptable for the separation of hypotheses H_0 and H_1 with small errors. Indeed, if

$$\frac{T\cdot\min_{d,j}\left(\varepsilon_{dj}-\sum_{s=1}^{r}a_s\varepsilon_s\right)^2}{\sqrt{N}}\to\infty,\ \text{at } N\to\propto, \tag{20}$$

then the error $a=P\left(H_1/H_0\right)$ will be as small as one likes. The coefficient α can be provided with a set of a sufficiently large number of observations simulating the condition $T/\sqrt{N}\to\infty$ and (20).

So, equations (11) to (20) define a mathematical model for detecting anomalies in the behavior of a computer system based on the improved criterion given in the previous subsection.

The possibility of increasing the effectiveness of the criterion by changing the parameter λ can be tested during the implementation of experiments to identify the heterogeneity of samples $\bar{\nu}_1,...,\bar{\nu}_r$.

We also note that if the basic hypothesis H_0 is also equally probabilistic, then there will be no gain by increasing the value of the parameter λ because $\sum_{o=1}^{T}\varepsilon_{dj}=0$, where $d=1, ..., r$.

Thus, the possibility of effective application of the criterion for identifying the heterogeneity of samples based on statistics $I_T(\lambda)$ was justified if the heterogeneity is, so to speak, close, i.e., $p_{dj}=p(1+o(1)$. It is intuitively clear that the criterion will distinguish between the hypotheses H_1 and H_0 even better if p_{dj} is more different from p_j for some d, that is, the hypotheses H_0 and H_1 will not be close.

During the experiments, the performance of the criterion was demonstrated when «observing» the hypotheses H_0 and H_1, for those that differ significantly in p_{dj} and p_j.

The mathematical model for detecting anomalies in the behavior of a computer system based on the improved criterion of uniformity of samples of input data is the basis of the proposed methodology. A module that implements the mathematical component of the method, which is a mathematical model, performs the following functions:

– separate storage of statistical data on observation intervals corresponding to the previous observation stages for each type of observation;

– loading of statistical data on previous similar intervals when the observation interval changes;

– application of the criterion of the degree of uniformity of samples to statistical data obtained in the process of observations at the current and previous similar observation intervals.

The module of the mathematical model, designed to solve service problems, performs the following functions:

– start and stop the processes of tracking individual sets of system characteristics;

– loading of initial monitoring parameters;

– tracking the current system time and, in accordance with it, sending signals to other modules;

– setting up the environment for the program to work, ensuring formal fixation, etc.

The collection of statistical information on changes in the selected characteristics and the calculation of criteria values is performed by activity sensors and a mathematical module.

5. 3. Studying the conditions of using the mathematical model constructed to detect anomalies in the behavior of a computer system

The study was performed of the possibility of isolating a series of observations that differ in anomaly, which is associated with determining the period of observation when there is an attack on the system or an extraordinary type of user actions. For this purpose, vector r-dimensional statistics $\bar{I}_T(\lambda)$ and a rule based on inequality (10) were used. As before, the expression of vector statistics $\bar{I}_T(\lambda)$ is replaced by the marginal normal law in accordance with the results of work [8]. then with accuracy to this boundary junction $P(H_1/H_0)\leq a$, since there are a coordinates in the vector $\bar{I}_T(\lambda)$ and the probability of going abroad $C(a/r)$ in (10) at least one of the values among $I_T(\lambda,1)$, ..., $I_T(\lambda,r)$ with the hypothesis H_0 no more than $\lambda/r+r$. It should be noted that the use of a boundary distribution instead of a pre-limit one is associated with the risk of not taking into account the following circumstances:

1. The application of rule (10) is a heuristic guideline. The final conclusion about the presence or absence of heterogeneity is given by a detailed analysis of the event log conducted by the researcher in a heuristic way.

2. The amount of data to work with, is a variable. Since there are no estimates of the convergence rate of distributions of vector statistics, for example, the $T/\sqrt{N}\to\infty$, condition $d=1, ..., r$ will be considered fulfilled if $T/\sqrt{N}>100$ at $d=1, ..., r$.

For the purpose of heuristic analysis, it is proposed to consider the criterion of homogeneity r of independent polynomial schemes with N consequences and $T_1, ..., T_r$ – test volumes based on $I_T(\lambda)$ scattering power statistics, where $T=T_1+...+T_r$, λ – a valid parameter. To apply the criterion in the practice of detecting abnormal software operation, the parameters of the mean $LI_T(\lambda)$ and the dispersion of $DI_T(\lambda)$ are important [16–18]. In assumptions $N=\text{const}$, $T_1, ..., T_r\to\propto$ expressions are found for the mean and variance with estimates of residual terms, allowing the calculation of error at specific values of $T_1, ..., T_r, N$.

Let $M(T_1, P_1), ..., M(T_r, P_r)$ be r independent polynomial schemes with the same number of N results in each, probability distributions $P=(p_{d1}, ..., p_{dN})$ of probability of the appearances of results in d-scheme, $d=1, ..., r$ and $T=T_1+...+T_r$ total volume of observations. Denote through i $(\nu_{d1}, ..., \nu_{dN})$ the frequency vector of the results observed in the d-th polynomial scheme $M(T_d, P_d)$, $\nu_{+j}=\sum_{d=1}^{r}\nu_{dj}$, $\sum_{j=1}^{r}\nu_{dj}=T_d$.

A generalization of the standard criterion of uniformity chi-square is a criterion based on the statistics of scattering power:

$$I_T(\lambda)=\sum_{d=1}^{r}I_T(\lambda,d),$$

$$I_T(\lambda,d)=\binom{\lambda+1}{2}^{-1}\sum_{j=1}^{N}\nu_{dj}\left(\left(\frac{\nu_{dj}T}{\nu_{+j}T_d}\right)-1\right). \tag{21}$$

The distribution of scattering power statistics in the criteria for belonging to the sample of a particular law converges to the central distribution of chi-squared at $N=\text{const}$, $T\to\propto$ [19, 20].

At the same time, the exact expressions for $LI_T(\lambda)$ and $DI_T(\lambda)$ have a rather complex form, and to select the required volumes of observations, when instead of pre-boundary expressions, one can use limit expressions, one needs to estimate the magnitude of the error when replacing $LI_T(\lambda)$, $DI_T(\lambda)$ with boundary expressions.

Denote by:

$$\omega_{dj}=\left(\nu_{dj}-T_d p_{dj}\right)\cdot T^{-1/2}. \tag{22}$$

Let $T_d/T=a_d$ and there are constants c_1, c_2, for which inequalities are satisfied:

$$0<c_1<a_d<c_2<1;\ 0<c_1<p_{dj}<c_2<1, \tag{23}$$

at $T\to\propto$, $N=\text{const}$.

Hereafter, we agree to skip the result and assume that the number N is reduced by one if $\nu_{dj}=0$ for at least one pair of numbers d, j. In fact, under conditions (23) and $T\to\propto$, the probability of such an event is infinitesimal.

In this case, the distribution of the random variable ω_{dj} is asymptotically normal with the mean 0 and variance $a_d p_{dj}(1-p_{dj})$. This fact will be denoted as $\omega_{dj}\sim N(0, a_d p_{dj}(1-p_{dj})$. Then the joint distribution of random vectors $(\omega_{d1}, ..., \omega_{dN})=W_d$ is asymptotically normal with mean 0 and covariant matrix

$a_d\left[diag\left(p_{d1},...,p_{dN}\right)-P_d^T p_d\right]$, where diag($p_{d1}$, ..., p_{dN}) – diagonal matrix with a diagonal (p_1, ..., p_N) and P^T is a vector-column transposed with respect to a vector-string P.

Having received the necessary justifications and theoretical assessments of the possibilities of the proposed methodology, we proceed to its testing in practice. To verify the practical value, there are all the necessary results for this, namely: the choice of research material is carried out, there is a theoretical justification for the applicability of the criteria, the expected sensitivity of the method, and the probability of errors are indicated.

The practical application of the proposed methodology on the basis of the mathematical model constructed was investigated on two types of machines: client PCs and servers. The article reports the results of using the described software for client machines.

Studies on computer systems such as «workstation» were carried out on machines with sets of three types. The first type of workstations involves active work on the Internet, the use of file servers of the local network, work with office programs.

The second type of workstations is aimed at software development and related processes, network activity is average, system resource utilization is high, but z peak. The third type of workstations implements the «home computer» profile, that is, the systems of this perform the functions of an Internet directory, a client of file-sharing networks, a game console, etc.

The experiments were carried out under minimal changes in the topology and settings of local networks, equipment for routing *and* local network servers. Software sets were recorded at each beginning of a new series of experiments and changed throughout the series. For each of the workstations of the first two types, 1–2 users worked, for the workstations of the third type 3–5 users.

Criteria corresponding to different values of the parameter λ (21) were investigated. The ranges of values 0.01–0.1, 0.1–0.5, 0.5–1 and the value λ>1 were investigated.

The average number of active states of network activity sensors for the first type of system was in the range of 30–40, the second type – in the range of 25–45, for the third type – 60–80. The limits of theoretical homogeneity are equal to $\pm 3\sqrt{2r(N-1)}$, $r=5$. The approximate limits of the theoretical homogeneity of these sensor sets are given in Table 1.

Table 1

Approximate limits of theoretical homogeneity of the studied sensor sets

CS type	Theoretical boundaries
CS for general (office) use	Approximately ±53–±59
CS – software developer workstation	Approximately ±50–±59
CS for home use	Approximately ±70–±80

For system resource flow sensors, the average number of active states was 150, the corresponding theoretical limits are in the range of 110–130.

During the experiment, such anomalies were modeled that can affect the observed characteristics. Thus, for CS of general (office) use, an imitation of an attack of the «network worm» type, distributed via e-mail (VBS LoveLetter, Cassandra, RedZONE 7.1, Melissa and others) was simulated [21–23]. For software developer workstations, SynFlood and Teardrop attacks were simulated. For CS for domestic use, user anomalies and attacks were used that used known software vulnerabilities (MS03-39 Kaht2 (135), MS03-043 Remote SYSTEM, nOD-ms04011-lsasrv-expl LSASS (MS04-011)) [24, 25].

The settings and startup features of the surveillance and analysis software remained unchanged for normal and abnormal system behavior. The results of observations are shown in Fig. 3–5.

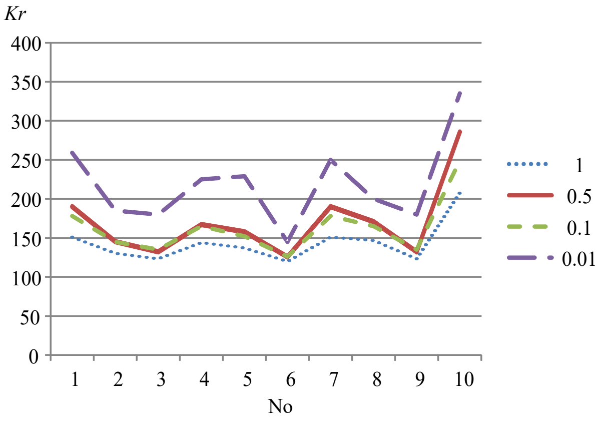

Fig. 3. Results of observations of abnormal behavior of a computer system for general (office) use from network activity sensors (network worm)

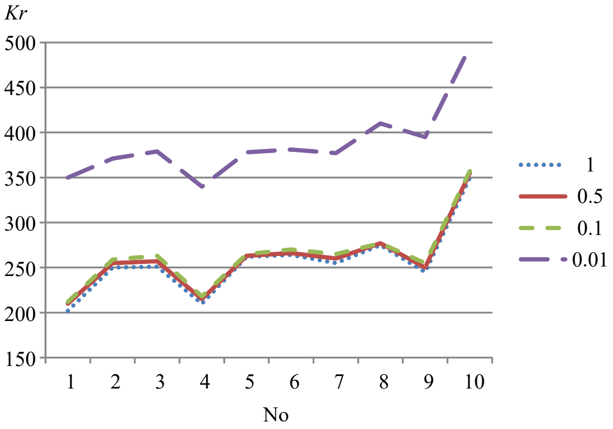

Fig. 4. Results of observations of abnormal behavior of the software developer's workstation from network activity sensors (DoS type attack)

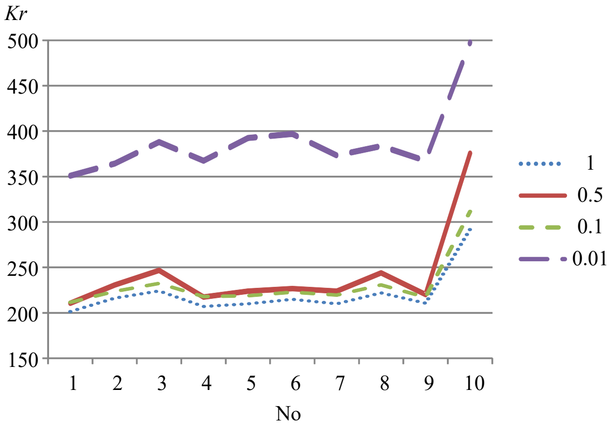

Fig. 5. Results of observations of abnormal behavior of a computer system of household use from network activity sensors (attack using software vulnerability)

In Fig. 3–5, the following notation is applied: Kr – the value of the criterion; No – the number of the experiment step. The obtained results of the use of the described software for client machines showed the possibility of detecting anomalies in the behavior of the computer system and conducting a preliminary forecast. Note that *the* use of an improved criterion provides the best results in identifying abnormal behavior of a computer system in the cases of significantly unequal distributions of incoming traffic.

6. Discussion of results of investigating anomalies in the behavior of a computer system

A feature of this study is the use of vector *r*-dimensional statistics to isolate a series of observations with a difference from other distributions.

The results of the experiments demonstrate visually visible detection of anomalies in all practical cases. Plots in Fig. 3–5 demonstrate a dynamic jump in criteria values in all the above cases when interacting with a computer network. The values of the criteria in the presence of anomalies go beyond the theoretical limits of homogeneity (Table 2), which confirms the hypothesis of heterogeneity (the anomalous behavior of the system generates the resulting samples with a distribution different from the sample distributions observed during regular work). This fact makes it possible to assert that to identify anomalies in CS, the proposed method may be applicable.

According to the results of a series of experiments, some of which are shown in Fig. 3–5, the results of the frequencies of the average shift of the determined criterion with the specified anomalies in CS. It is proved that for all the studied types of data, the confidence probability that the value of the statistical values of the shift of the determined criterion does not deviate from the mathematical expectation by more than 0.05 is approximately equal to 0.94. This confirms the reliability of the results of the detection of anomalies in CS and the results of scientific research.

The peculiarity of the proposed method is the use of an improved criterion of uniformity of samples of input data for a significantly unequal-probability distribution.

In contrast to the results obtained by the methods proposed in works [3, 4], the developed method revealed abnormal traffic behavior faster, by 3–4 times. The use of multidimensional statistics allowed us to exceed by 5–10 % the results of similar experiments reported in work [5]. The processing time of statistical data in the method given in [6] and in the proposed method did not differ significantly but the proposed method was more accurate in identifying anomalies in significantly unequal-probable distribution of traffic. Also, the best results for a significantly unequal-probable distribution of traffic were obtained by comparison with the methods proposed in [7].

It should be noted that the experiment revealed the following. For CS of general (office) use, it is advisable to use the criterion obtained using expression (20) by substituting the parameter λ from the interval [0.5–1]. For a software developer's workstation, approximately equal parameter variability indicators from the range 0.01–1 are visually observed, so it does not matter which value of the parameter λ should be selected. For computer systems of domestic use, the criteria, as well as for the first example, it is advisable to use the parameter λ from the interval [0.5–1]. At the same time, for all types of CS, each of the λ ranges demonstrates the practical applicability of the method.

This study has the following limitations:

– availability of a series of observations on the state of the computer system;

– significantly unequal distribution of incoming traffic.

The disadvantage of this study is the great computational complexity of the mathematical model. This disadvantage is planned to be eliminated through the use of approximate calculations.

The advancement of this study is to make it possible to process series from small samples.

7. Conclusions

1. The criteria for homogeneity of samples of input data have been investigated and improved. A feature of the study is the use of vector *r*-dimensional statistics to isolate a series of observations that differ from previous series. The results of the study of the behavior of criteria for maximizing and minimizing the frequencies of results showed the possibility of using the improved criterion of uniformity of samples of input data. The positive side of the proposed solution is the possibility of practical application of the improved criterion in the case of a significantly unequal-probability distribution *P*. This is due to the expansion of the capabilities of the proposed criterion, especially at high values of the indicator λ.

2. A mathematical model for detecting anomalies in the behavior of a computer system based on an improved criterion of homogeneity of samples of input data has been built. The model differs from those known by the possibility of isolating a series of observations, the results of which show the anomaly in CS behavior. This made it possible to provide the necessary level of reliability of the results obtained. The confidence probability that the value of the statistical quantities of the shift of a certain criterion does not deviate from the mathematical expectation by more than 0.05 is approximately equal to 0.94.

3. The mathematical model constructed to detect anomalies in the behavior of a computer system is investigated. The results of the study showed the practical value of using the proposed model in the process of detecting anomalies of the computer system with significantly unequal and probable traffic distributions. The proposed model produces a special effect under the conditions of using the parameter λ approaching 1.

Conflicts of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was conducted without financial support.

Data availability

All data are available in the main text of the manuscript.

References

1. BasuMallick, C. (2022). 10 Network Behavior Anomaly Detection Tools in 2022. Available at: https://www.spiceworks.com/tech/networking/articles/network-behavior-anomaly-detection-tools/
2. Wang, C., Zhou, H., Hao, Z., Hu, S., Li, J., Zhang, X. et al. (2022). Network traffic analysis over clustering-based collective anomaly detection. Computer Networks, 205, 108760. doi: https://doi.org/10.1016/j.comnet.2022.108760
3. Ayensa-Jiménez, J., Pérez-Aliacar, M., Randelovic, T., Oliván, S., Fernández, L., Sanz-Herrera, J. A. et al. (2020). Mathematical formulation and parametric analysis of in vitro cell models in microfluidic devices: application to different stages of glioblastoma evolution. Scientific Reports, 10 (1). doi: https://doi.org/10.1038/s41598-020-78215-3
4. Meleshko, Y., Drieiev, O., Yakymenko, M., Lysytsia, D. (2020). Developing a model of the dynamics of states of a recommendation system under conditions of profile injection attacks. Eastern-European Journal of Enterprise Technologies, 4 (2 (106)), 14–24. doi: https://doi.org/10.15587/1729-4061.2020.209047
5. Bannai, E., Bannai, E., Ito, T., Rie, T. (2021). 6 P- and Q-polynomial schemes. Algebraic Combinatorics, Berlin, Boston: De Gruyter, 241–398. doi: https://doi.org/10.1515/9783110630251-006
6. Semenov, S., Davydov, V., Lipchanska, O., Lipchanskyi, M. (2020). Development of unified mathematical model of programming modules obfuscation process based on graphic evaluation and review method. Eastern-European Journal of Enterprise Technologies, 3 (2 (105)), 6–16. doi: https://doi.org/10.15587/1729-4061.2020.206232
7. Meleshko, Y., Raskin, L., Semenov, S., Sira, O. (2019). Methodology of probabilistic analysis of state dynamics of multidimensional semiMarkov dynamic systems. Eastern-European Journal of Enterprise Technologies, 6 (4 (102)), 6–13. doi: https://doi.org/10.15587/1729-4061.2019.184637
8. Kovalenko, A., Kuchuk, H. (2022). Methods to Manage Data in Self-healing Systems. Studies in Systems, Decision and Control. Cham: Springer, 113–171. doi: https://doi.org/10.1007/978-3-030-96546-4_3
9. Rempała, G., Wesołowski, J. (2023). Poisson limit theorems for the Cressie-Read statistics. Journal of Statistical Planning and Inference, 223, 15–32. doi: https://doi.org/10.1016/j.jspi.2022.07.004
10. Escalante, J. M., Skipetrov, S. E. (2018). Level spacing statistics for light in two-dimensional disordered photonic crystals. Scientific Reports, 8 (1). doi: https://doi.org/10.1038/s41598-018-29996-1
11. Tung, D. D., Rao Jammalamadaka, S. (2012). U-Statistics based on spacings. Journal of Statistical Planning and Inference, 142 (3), 673–684. doi: https://doi.org/10.1016/j.jspi.2011.09.007
12. Raskin, L., Sukhomlyn, L., Sagaidachny, D., Korsun, R. (2021). Analysis of multi-threaded markov systems. Advanced Information Systems, 5 (4), 70–78. doi: https://doi.org/10.20998/2522-9052.2021.4.11
13. Oleksenko, O., Khudov, H., Petrenko, K., Horobets, Y., Kolianda, V., Kuchuk, N. et al. (2021). The Development of the Method of Radar Observation System Construction of the Airspace on the Basis of Genetic Algorithm. International Journal of Emerging Technology and Advanced Engineering, 11 (8), 23–30. doi: https://doi.org/10.46338/ijetae0821_04
14. Semenov, S., Sira, O., Gavrylenko, S., Kuchuk, N. (2019). Identification of the state of an object under conditions of fuzzy input data. Eastern-European Journal of Enterprise Technologies, 1 (4 (97), 22–30. doi: https://doi.org/10.15587/1729-4061.2019.157085
15. Kovalenko, A., Kuchuk, H., Kuchuk, N., Kostolny, J. (2021). Horizontal scaling method for a hyperconverged network. 2021 International Conference on Information and Digital Technologies (IDT), 331–336. doi: https://doi.org/10.1109/idt52577.2021.9497534
16. Yaloveha, V., Hlavcheva, D., Podorozhniak, A., Kuchuk, H. (2019). Fire Hazard Research of Forest Areas based on the use of Convolutional and Capsule Neural Networks. 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON). doi: https://doi.org/10.1109/ukrcon.2019.8879867
17. Al-Anzi, F. S., Lababidi, H. M. S., Al-Sharrah, G., Al-Radwan, S. A., Seo, H. J. (2022). Plant health index as an anomaly detection tool for oil refinery processes. Scientific Reports, 12 (1). doi: https://doi.org/10.1038/s41598-022-18824-2
18. Sohn, H., Czarnecki, J. A.,Farrar, C. R. (2000). Structural Health Monitoring Using Statistical Process Control. Journal of Structural Engineering, 126 (11), 1356–1363. doi: https://doi.org/10.1061/(asce)0733-9445(2000)126:11(1356)
19. Lu, Y., Wang, T., Liu, T. (2020). Bayesian Network-Based Risk Analysis of Chemical Plant Explosion Accidents. International Journal of Environmental Research and Public Health, 17 (15). doi: https://doi.org/10.3390/ijerph17155364
20. An, S. H., Heo, G., Chang, S. H. (2011). Detection of process anomalies using an improved statistical learning framework. Expert Systems with Applications, 38 (3), 1356–1363. doi: https://doi.org/10.1016/j.eswa.2010.07.031
21. Pratama, A., Rafrastara, F. A. (2012). Computer Worm Classification. International Journal of Computer Science and Information Security, 10, 21–24. Available at: https://www.researchgate.net/publication/299580232_Computer_Worm_Classification
22. Raskin, L., Ivanchikhin, Y., Sukhomlyn, L., Sviatkin, I., Korsun, R. (2022). Evaluation model of the recovery processes of non-markovian systems, considering the elements unreliability under arbitrary distribution laws. Advanced Information Systems, 6 (3), 28–35. doi: https://doi.org/10.20998/2522-9052.2022.3.04
23. Joseph, U. M., Jacob, M. (2022). Real time detection of Phishing attacks in edge devices using LSTM networks. AIP Conference Proceedings. https://doi.org/10.1063/5.0103355
24. Mozhaev, O., Kuchuk, H., Kuchuk, N., Mozhaev, M., Lohvynenko, M. (2017). Multiservice network security metric. 2017 2nd International Conference on Advanced Information and Communication Technologies (AICT), 133–136. doi: https://doi.org/10.1109/aiact.2017.8020083
25. Guevara López, P., Delgado Reyes, G., Audelo González, J., Valdez Martínez, J., Perez Meana, H. (2015). Basic definitions for discrete modeling of computer worms epidemics. Ingeniería e Investigación, 35 (1), 79–85. doi: https://doi.org/10.15446/ing.investig.v35n1.44323