Розглядаються способи підвищення ефективності моніторингу процесів хмарної інфраструктури, які полягають в зниженні витрат обчислювальних ресурсів при збереженні необхідного рівня точності вимірювань. У даній роботі отримав подальший розвиток спосіб організації моніторингу процесів хмарної інфраструктури, заснований на апроксимації вимірювань, що накопичуються. Сформована необхідна і достатня множина апроксимуючих функцій, що відповідають ключовим властивостями спостережуваних процесів. Розроблено метод вибору апроксимуючих функцій для спостережуваних процесів хмарної інфраструктури. Метод складається з оцінювання властивостей спостережуваного процесу та вибору його апроксимуючої функції.

Практична цінність роботи полягає в можливості зниження витрат обчислювальних ресурсів за рахунок зменшення кількості планових вимірювань при допустимому рівні зниження їх точності. Оригінальність підходу полягає у використанні апріорних даних про спостережувані процеси з метою отримання більш точних оцінок їх властивостей. Практична реалізація запропонованого методу показує 20–40 % зниження кількості планових вимірювань при збережені точності моніторингу на рівні не нижче 95 %. Таким чином, запропонований метод дозволяє знизити навантаження на компоненти хмарної інфраструктури, зменшити використання процесорного часу і заощадити дисковий та оперативний простір фізичних і віртуальних вузлів. Результати дослідження можуть бути використані для програмної реалізації систем моніторингу хмарної інфраструктури

Ключові слова: моніторинг хмарної інфраструктури, комп'ютерна мережа, апроксимація функцією, витрати обчислювальних ресурсів

UDC 004.738.52:004.031
DOI: 10.15587/1729-4061.2020.200372

DEVELOPMENT OF A METHOD FOR SELECTING THE APPROXIMATIMG FUNCTIONS FOR THE OBSERVABLE PROCESSES OF CLOUD INFRASTRUCTURE

O. Grytsenko
Postgraduate Student*
E-mail: alegrys@ukr.net
V. Sayenko
PhD*
E-mail: visan@ukr.net
*Department of Information Control System
Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Received date 05.03.2020
Accepted date 03.04.2020
Published date 30.04.2020

1. Introduction

The main purpose of monitoring any information system (IS) is to obtain up-to-date and complete information about its state. The effectiveness of decision making, directed to ensure the required quality of the IS operation, depends on the success of achieving this goal. The applied monitoring techniques must correspond to the technical conditions of the information environment, meet the requirements for the quality of collected information and the level of the computational burden.

At present, many software solutions [1] have been developed to monitor a cloud-based IS. Many of them implement the function of continuous monitoring with the help of standardized technologies such as SNMP [2] or WMI [3]. These technologies are deeply integrated with existing operating systems and are characterized by high scalability, adaptability, and predictability of overheads.

The distributed multi-level architecture of the cloud based IS implies using the client-server technologies to collect the values of observable variables. Monitoring a large number of distributed objects of the physical and virtual levels and transmission of large arrays of collected data to centralized warehouses cause significant computational burden and increase the costs of operation of a cloud-based IS. It is required to find solutions to reduce the excessive computational burden generated by monitoring the cloud infrastructure processes. Computational resources such as the bandwidth of network communication channels, the use of processor time by the nodes of monitoring sites, and data storage space, are expected to be saved.

2. Literature review and problem statement

One of the challenges in the organization of monitoring cloud infrastructure processes is the issue of the generated excessive computational burden. In paper [4], it is argued that continuous monitoring of various types of services of a hybrid cloud-based IS can lead to the reduction of its total bandwidth and increase the cost of its operation. To reduce the operating burden, the optimal topology of monitoring agents should be developed and the level of data redundancy should be minimized. In [5], it is said that the cloud based IS is the future of cloud computing. The functions of cloud resources management are classified. The importance of ensuring the required level of monitoring cost-effectiveness is indicated. In [6], the task of the organization of cost-effective monitoring in the context of minimizing delays and analyzing large data arrays in real time is considered. Paper [7] proposes a new monitoring architecture, which shows a high level of cost-effectiveness under limited conditions: a 0.05 % degradation of the response of the offered service while monitoring a small number of indicators of the performance of virtual machines.

Various methods of technical and organizational nature were developed to solve this problem: load redistribution by the communication channels [8, 9], compression of transmitted data [10], statistical processing of measurements [11] and others [12]. However, there is a not solved issue of ensuring the required accuracy of measurements when achieving the goal of reducing the total computational burden. The reason for this is the objective difficulty of finding effective ways to implement monitoring that is independent of the changing conditions of the cloud environment.

Typically, the process of reducing the computational burden remains outside the scope of standardized technologies of cloud infrastructure monitoring. The task of improving the cost-effectiveness of monitoring is solved by developing additional methods and technologies. The issue of organizing the platform-independent cost-effective monitoring of large hybrid cloud-based IS remains unresolved.

A variant of overcoming these difficulties may be using monitoring methods based on approximation of measurements. However, the effectiveness of such methods depends significantly on the adequacy of the chosen approximating function. Research into the problems of approximation of values of observable processes of cloud infrastructure is appropriate. The formal approach can be based on the information models and the methods for their estimation proposed in [13].

3. The aim and objectives of the study

The aim of this study is to develop an approach to reducing the computational burden generated by monitoring the processes of the cloud infrastructure based on the approximation of measurements. The research is aimed at improving the accuracy of approximation of the values of observable processes by selecting an adequate approximating function.

To achieve the set goal, the following tasks are to be solved:

– to analyze the methods for approximation of observable processes and describe the necessary and sufficient conditions for choosing the optimal approximating function;

– to develop the method for selecting approximating functions for observable processes of cloud infrastructure;

– to test the developed method.

4. Studying the techniques for selecting the approximating functions for the observable processes of cloud infrastructure

4. 1. Provisions of the approximation of observable processes of cloud infrastructure

Any observable process of cloud infrastructure can be considered a random process. Therefore, some approximating function may be chosen for it. We will formulate the statements that are relevant to solve the problem of selecting the approximation functions for the observable processes of the cloud infrastructure.

Statement 1. The true function $f(t)$ of any observable process of the cloud infrastructure can be approximated by a pre-selected function $\hat{f}(t)$ with adequacy not lower than A.

The feature of cloud infrastructure monitoring is that many of the measured indicators describe well-studied and understandable processes. This allows pre-selection (before measurements) of the structure of an approximating function based on the most significant properties of an observable process. Thus, during measurements, it remains only to choose the parameters of the function by ensuring the required degree of approximation adequacy.

If the set of functions $\hat{\Phi}$ with different properties is formed, according to [14, 15], there can be found at least one function $\hat{f}(t)\in\hat{\Phi}$ that allows approximating the values of an observable process with the adequacy that is not worse than A, provided that an observable process has a functional dependence.

Statement 2. To approximate the observable processes, it is enough to choose the finite set of approximating functions $\hat{\Phi}$ in accordance with the established requirements for approximation adequacy not lower than A.

The structural adequacy of the approximating function $\hat{f}(t)$ is determined by the degree of its proximity to the true function of the process $f(t)$ by their properties. For all possible combinations of a priori selected values of the observable process, it is possible to form the finite set of approximating functions $\hat{\Phi}$. In this case, according to statement 1, there can be found at least one function $\hat{f}(t)\in\hat{\Phi}$, that allows approximating the values of an observable variable with adequacy not worse than A.

Statement 3. To approximate any observable cloud infrastructure process, a limited set of approximating functions $\hat{F}\subset\hat{\Phi}$ are necessary and sufficient.

It is enough to select the functions non-recurrent by properties out of the set of approximating functions $\hat{\Phi}$, that correspond to pre-selected properties of an observable process. The resulting set of approximating functions $\hat{F}\subset\hat{\Phi}$ is sufficient to approximate any observable process of the cloud infrastructure.

4. 2. Assessing the properties of the observable processes of cloud infrastructure

Paper [16] described the key properties of the observable processes of the cloud infrastructure for the selection of approximating functions. Formally, these properties are represented in the form of a priori information model $C_{-1}=\{d_\varphi,l_\varphi,s_\varphi,t_\varphi\}$ and a posteriori information model $C_1=\{d_\varphi,l_\varphi,s_\varphi,t_\varphi,l_f,c_f,m_f,u_f\}$. Their parameters are presented in Table 1.

The represented models were revised in the current paper. Thus, to increase informativeness and brevity, the «lifetime» parameter of the a priori model was renamed «variability» and the «bulge/concave» parameter of the posteriori model was renamed «bulge». The «continuity» parameter was excluded from a posteriori model as not relevant. The values of the «monotony» and «bulge/concave» were reduced to set {0, 1}, where 0 is no monotony or bulge, respectively, 1 – the existence of monotony or bulge, respectively. The values of other parameters remained the same. The designation of a priority model was changed from C_1 to I^-, of a posteriori information from C_2 to I^+, the index was removed for all parameters, and parameter l_f was renamed for p.

Table 1

Values of parameters of models $C_{-1}=\{d_\varphi,l_\varphi,s_\varphi,t_\varphi\}$ and $C_1=\{d_\varphi,l_\varphi,s_\varphi,t_\varphi,l_f,c_f,m_f,u_f\}$

Parameter	Value
Dynamicity d_φ	0 – low dynamicity, 1 – high dynamicity
Linearity l_φ	0 – non-linearity, 1 – linearity
Stationarity s_φ	0 – stationary, 1 – non-stationary
Lifetime t_φ	0 – no changes, 1 – existence of changes
Non-linearity degree l_f	0 – no non-linearity, 1 – first degree non-linearity (existence and permanence of a derivative of first order), 2 – second degree non-linearity (existence and permanence of second order and higher)
Continuity c_f	0 – piece-continuous function with a break of first kind, 1 – piece-continuous function with a break of second kind, 2 – continuous function that has no breaks
Monotony m_f	–1 – monotonous decrease, 1 – monotonous increase, 2 – no monotony
Bulge/concave u_f	–1 – bulge, 1 – concave, 2 – no bulge/concave

The new records of a priori and a posteriori information models $I^-=\{d,l,s,t\}$ and $I^+=\{d,l,s,t,p,m,u\}$ respectively, where d is dynamicity, l is linearity, s is stationarity, t is the variability, p is linearity degree, m is monotony, u is the bulge.

Estimation of models $I^-=\{d,l,s,t\}$ and $I^+=\{d,l,s,t,p,m,u\}$ is based on the method [13]. The source data are formed in the form of the monitoring specification $M=\{G,F,S,T,V\}$, where G is the aims of monitoring, F is the functional tasks of monitoring, S is the characteristics of an observable system, T is the monitoring tools, V is the set of observable variables (processes).

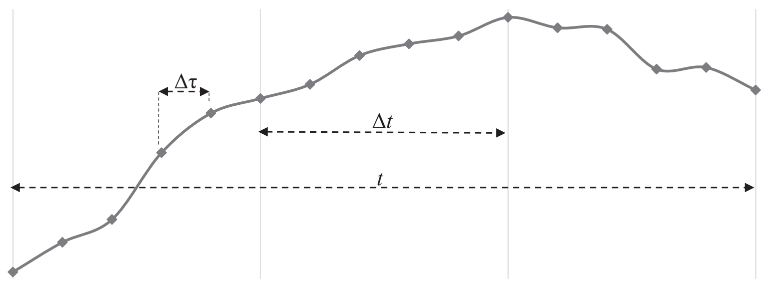

Fig. 1. Relations between measurement duration t, measurement interval $\Delta\tau$ and assessment interval Δt

To assess a priori model $I^-=\{d,l,s,t\}$, information about the character of an observable process is collected. A short-term measurement session for obtaining the sample of values of observable process X^- is carried out. Based on X^-, a priori analytical data $A(X^-)$ – basic statistical indicators and diagrams for graphic-visual analysis are formed. The parameters of model $I^-=\{d,l,s,t\}$ are calculated based on $A(X^-)$ and rules R. The formed a priori model $I^-=\{d,l,s,t\}$ can be used both for the choice of the approximating function for an observable process, and more detailed analysis of its properties and the formation of a posteriori model $I^+=\{d,l,s,t,p,m,u\}$.

A posteriori model $I^+=\{d,l,s,t,p,m,u\}$ is assessed based on the data obtained after longer measurements of an observable process. The plan for measurements P, containing a set of measurement sessions non-intersecting in time $p_i\in P$ is chosen. The specification $p_i=\{t,t_0,\Delta\tau\}$ is chosen for each session, where t is the measurement duration, t_0 is the time of the beginning of measurements, $\Delta\tau$ is the measurement interval. P is formed expertly based on the pre-selected a priori model $I^-=\{d,l,s,t\}$ and specification of monitoring $M=\{G,F,S,T,V\}$. Specification of data post-processing $N=\{\mathfrak{R},\Delta t\}$ is formed, where $\mathfrak{R}$ is the rules of converting the measured values, Δt is the assessment interval. The elements of the analyzed series of values are shown in Fig. 1.

As a result of the implementation of measurement plan P, a series of measured values X^+ is formed for an observable process. In accordance with Δt, series X^+ is split into segments, for each of which an instance of model $I^+=\{d,l,s,t,p,m,u\}$ is selected. The assessment of a posteriori model is based on rules R^+. All the resulting instances of model $I^+=\{d,l,s,t,p,m,u\}$ are recorded in the form of a frequency distribution table Δ^+.

4. 3. Selection of the set of approximating functions $\hat{F}\subset\hat{\Phi}$

In accordance with section 4.1, we will select the initial set of approximating functions $\hat{\Phi}$.

We will consider one of the basic sets of functions for the approximation of a time series [17]. These include such functions as linear, quadratic, cubic, power, exponential, fractional-linear, logarithmic, hyperbolic, fractional-rational and moving average. According to statement 3, it is possible to choose from the described set of functions $\hat{\Phi}$ such set $\hat{F}\subset\hat{\Phi}$, that will correspond to all possible variants of instances of a posteriori model. As a result, out of all mathematically possible combinations of the values of parameters of a posteriori models, we obtain the following admissible combinations: 0101010, 0110000, 0001111, 0001101, 0001201, 0000201, 0000200, 0000111, 0000210, 0010200, 0001110, 0001210, 0001200, 0011200, 1001200, 1001100, 0001100, 0101111, 1001111. Following the principle of the choice of the simplest approximating function and excluding redundancy, we will choose the final set of functions $\hat{F}$:

1) linear: $\hat{f}_1(t)=a_1t+a_0$;
2) quadratic: $\hat{f}_2(t)=a_2t^2+a_1t+a_0$;
3) cubic: $\hat{f}_3(t)=a_3t^3+a_2t^2+a_1t+a_0$;
4) exponential: $\hat{f}_4(t)=a_1e^{a_0t}$;
5) moving average: $\hat{f}_5(t)=(f(t-n)+...+f(t-1))/n$.

For each $\hat{f}_i(t)\in\hat{F}$, write down the corresponding instances of a posteriori model in the form of set Θ_i (Table 2).

The formed sets Θ_i consist of instances $I^+=\{d,l,s,t,p,m,u\}$, obtained after verification and validation of all the values of parameters of the model separately and in combination with each other.

Table 2

Set Θ_i of instances $I^+=\{d,l,s,t,p,m,u\}$ for $\hat{f}_i(t)\in\hat{F}$

Θ_1	Θ_2	Θ_3	Θ_4	Θ_5
0101010	0001111	0000201	0001110	1001111
0110000	0001101	0000200	0001210	
	0001201	0000111	0001200	
		0000210	0011200	
		0010200	1001200	
			1001100	
			0001100	
			0101111	

4. 4. The rule of choosing an approximating function for an observable process of cloud infrastructure

For the rule of choosing $\hat{f}_i(t)\in\hat{F}$, designated as $\rho(\Theta,\Delta^+)$, the original data are: Θ –the set of admissible instances $I^+=\{d,l,s,t,p,m,u\}$ for all $\hat{f}_i(t)\in\hat{F}$ (Table 2), Δ^+ – frequency table of instances distribution $I^+=\{d,l,s,t,p,m,u\}$ for $v_i\in V$. There is a search for the closest by properties $\hat{f}_i(t)\in\hat{F}$ for $v_i\in V$ based on the normalized indicator of Euclidean distance between instances $I^+=\{d,l,s,t,p,m,u\}$ from sets Θ and Δ^+:

$$D=N\left(\Sigma\left(\min\left(d_{1j},d_{2j},\ldots,d_{kj}\right)\cdot\delta\left(I^+_{\Delta^+j}\right)\right)\right), \quad (1)$$

where

$$d_{kj}=\sqrt{\sum_{i=1}^{7}\left(I^+_{\theta k}[i]-I^+_{\Delta^+j}[i]\right)^2},$$

$I^+_{1k}[i]$ is the i-th parameter of the k-th instance of the model from Θ, $I^+_{\Delta^+j}[i]$ is the i-th parameter of the j-th instance of the model from the table Δ^+, N is the normalizing function.

The criterion of choosing $\hat{f}_i(t)\in\hat{F}$ is the minimization of indicator (1).

5. Formation of the necessary conditions for the adequate approximation of the observable processes of cloud infrastructure

5. 1. Description of a method for the selection of approximating functions for the observable processes of cloud infrastructure

The method for selecting approximating functions for observable processes of cloud infrastructure with partial a priori and a posteriori certainty is proposed.

The original data of the method are the specification of monitoring of cloud infrastructure $M=\{G,F,S,T,V\}$, where G is the monitoring aims, F is the functional tasks of monitoring, S is the characteristic of an observable system, T is the monitoring tools, V is the set of observable processes.

The result of the method application is the choice of approximating function $\hat{f}_i(t)\in F$ for each observable process $v_i\in V$.

The method consists of twelve steps, carried out for each $v_i\in V$:

1) a short-term measurement session for the observable process v_i is carried out to form the sample of values X^-. The recommended session duration is 30 minutes with the measurement interval from 1 to 10 seconds;

2) based on X^-, a priori analytical data $A(X^-)$ – basic statistical indicators and diagrams – are formed;

3) graphic and visual analysis of $A(X^-)$ is performed, within which model $I^-=\{d,l,s,t\}$ is estimated with the help of rules R^-;

4) the hypothesis about the necessity of a detailed analysis of properties of the observable process v_i is verified. If the hypothesis is rejected, proceed to step 12, otherwise, to step 5;

5) the plan of measurements P for the observable process v_i is formed; for each session $p_i\in P$, specification $p_i=\{t,t_0,\Delta\tau\}$, is described, where t is the session duration, t_0 is the time of the beginning of the session, $\Delta\tau$ is the measurement interval. Plan P is formed in an expert way based on the monitoring specification $M=\{G,F,S,T,V\}$ and model $I^-=\{d,l,s,t\}$;

6) within the implementation of the plan of measurement P of observable process v_i, series of measured values X^+ is formed;

7) the rules of converting the measured values $\Re$ are selected;

8) the series of values X^+ is processed according to rules $\Re$ for a whole series of values;

9) estimation interval Δt is selected in the range from 10 to 15 values;

10) the whole set X^+ is split into sections $x_i\in X^+$ according to Δt;

11) the cycle of analysis of all $x_i\in X^+$ is implemented. The values are converted according to $\Re$ and instances of a posteriori model $I^+=\{d,l,s,t,p,m,u\}$ are selected with the help of rules R^+. The frequency table of distribution Δ^+ of all computed instances of the model is formed;

12) $\hat{f}_i(t)\in\hat{F}$ is selected using rule $\rho(\Theta,\Delta^+)$.

5. 2. Verifying the adequacy of approximating functions of the observable processes

The adequacy criteria $\hat{f}_i(t)\in\hat{F}$ are based on indicators from [18]:

1. The existence of a trend in the training sample. The hypothesis about the absence of a trend is checked using the method for checking the average levels' differences. Within this method, using the Fischer criterion $F=\hat{\delta}^2_X/\hat{\delta}^2_Y$, where $\hat{\delta}^2$ is the selective variance, the hypothesis about the homogeneity of variances is verified. If this hypothesis is accepted, we proceed to the next stage of verification of the trend existence, otherwise, the method does not give an answer to the question of whether there is a trend or not. The final verification of the hypothesis about the absence of a trend is made using a two-selective t-criterion by Student:

$$t=\frac{\left|\bar{Y}_1-\bar{Y}_2\right|}{\delta\sqrt{\frac{1}{n_2}+\frac{1}{n_2}}}, \quad (2)$$

where δ is the root mean square deviation of the difference of the mean.

If the hypothesis is accepted, the trend is absent, otherwise, the trend is proved.

2. The randomness, normal distribution, zero mathematical expectation and lack of autocorrelation of residuals. A series criterion is used to verify the residual's randomness:

$$\begin{cases} k_{\max}<\left[3.3\left(\lg n+1\right)\right] \\ v>\left[0.5\left(n+1-1.96\sqrt{n-1}\right)\right], \end{cases} \quad (3)$$

where $k_{\max}$ is the duration of the longest series, v is the total number of series, n is the length of the sample of values, square brackets designate the whole part of a number.

If at least one inequality of (3) is broken, the approximating function is considered inadequate.

The peak criterion is also used to verify the randomness:

$$p > \left[\bar{p} - 1.96\sqrt{\delta_p^2}\right], \qquad (4)$$

where square brackets designate the integer part of the number, p is the number of peak points, $\bar{p}$ is the mathematic expectation of the number of peak points, δ is the variance of the number of peak points.

If inequality (4) is met, the approximating function is considered adequate.

To verify the normal distribution of the residual value, indicators of asymmetry and excess indicators are used:

$$\begin{cases} \hat{\gamma}_1 < 1.5\sigma_{\hat{\gamma}_1}, \\ \left|\hat{\gamma}_2 + \dfrac{6}{n+1}\right| < 1.5\sigma_{\hat{\gamma}_2}, \end{cases} \qquad (5)$$

$$\begin{cases} \hat{\gamma}_1 > 2\sigma_{\hat{\gamma}_1}, \\ \left|\hat{\gamma}_2 + \dfrac{6}{n+1}\right| \geq 2\sigma_{\hat{\gamma}_2}, \end{cases} \qquad (6)$$

where $\hat{\gamma}_1$, $\hat{\gamma}_2$ are selective characteristics of asymmetry and excess, $\sigma_{\hat{\gamma}_1}$, $\sigma_{\hat{\gamma}_2}$ are corresponding root mean square errors.

If both inequalities in (5) are true, the hypothesis about the normal character of the distribution of residual component is accepted. If at least one of the system's inequalities (6) is true, the hypothesis of the normal character of distribution is rejected and the approximating function is considered inadequate.

The verification whether a mathematical expectation of residual component is equal to zero is based on the single-selective t-criterion by Student:

$$t = \frac{|\bar{\varepsilon}|}{S}\sqrt{n}, \qquad (7)$$

where $\bar{\varepsilon}$ is the mathematic expectation, S is the standard deviation. If (7) is smaller than the tabular value with significance α and the number of freedom of powers n–1, the hypothesis about the equality of mathematical expectation to zero is accepted, otherwise this hypothesis is rejected and the approximating function is considered inadequate.

Verification of independence of values of the residual component is performed using the d-criterion by Darbin-Watson.

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}, \qquad (8)$$

The value (8) is compared with tabular values of d_1 and d_2. If $d > d_2$, the hypothesis about the independence of the levels of the residual sequence is accepted – the approximating function is adequate. If $d < d_1$, the hypothesis is rejected and the approximating function is not recognized as adequate. At values $d_1 < d > d_2$, it is impossible to draw any conclusion.

6. Example of implementing the method for the selection of approximating functions for the observable processes of cloud infrastructure

6. 1. Description of specification of cloud infrastructure monitoring

Table 3 gives the specification of the planned monitoring of cloud infrastructure $M = \{G, F, S, T, V\}$, G is the aims of monitoring, F is the functional tasks of monitoring, S is the characteristics of an observable system, T is the monitoring tools, V is the observable processes.

Table 3

Specification of planned monitoring of cloud infrastructure

Parameter	Meaning
G	Ensuring the required level of performance of virtual cloud infrastructure nodes
F	Control of the loading level of processors of virtual nodes of cloud infrastructure
S	Cloud infrastructure based on VMware vSphere technology, consisting of a set of virtual servers based on Windows Server 2008 and Windows Server 2012
T	Monitoring tools within VMware ESXi
V	Variable «CPU utilization», showing an instant value of percentage of using processor time

Description of the specification of the planned monitoring allows proceeding to the beginning of selecting approximating functions for observable processes.

6. 2. Implementation of the steps of the developed method

Step 1. Form a set of values X^- by measuring the process v_1 every second within 5 minutes.

Step 2 Prepare $A(X^-)$. Based on sample X^-, calculate the basic statistical indicators (Table 4), plot the diagram of measured values X^- (Fig. 2) and the diagram of the corresponding histogram (Fig. 3).

Table 4

Statistical indicators of observable process v_1

Minimum	Maximum	Mean	Standard deviation
15.68357	86.82031	22.72008	7.647295

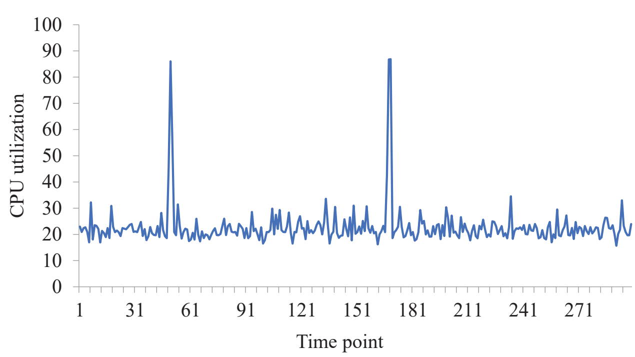

Fig. 2. Diagram of the set of values X^- of the CPU utilization variable

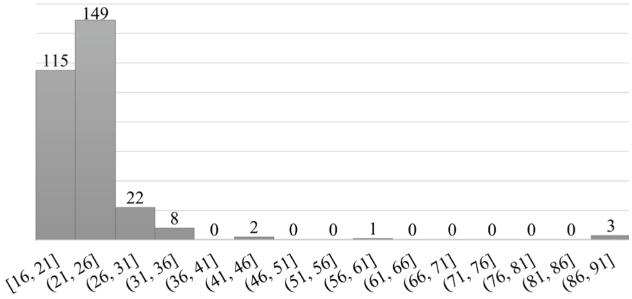

Fig. 3. Histogram of the set of values X^{-} of the CPU utilization variable

Step 3. Choose parameters $I^{-}=\{d,l,s,t\}$ for v_1 by means of graphic-visual analysis $A(X^{-})$ according to rules R^{-}. The result is given in Table 5.

Table 5

Values of parameters of model $I^{-}=\{d,l,s,t\}$ for an observable process

Parameter	Value	Comment
D	0	Low indicator of standard deviation rate, as well as the absence of a large number of strong curve fluctuations, indicates the lack of high dynamism
L	0	Existence of the points of the graph inflection and curvature characterizes the process as non-linear
S	0	Inconstant character of variance and the mean indicates the lack of stationarity
T	1	Observable process is long-term because its values differ at any moment of time

Thus, $I^{-}=\{0,0,0,1\}$.

Step 4. Consider that the observable process v_1 requires more detailed analysis since the existence of a weakly predictable noise component does not allow determining quite exactly the character of the process based on a priori data. Proceed to step 5.

Step 5. Choose the measurement plan P for process v_1. Based on the description of the system, the basic period of activity of the database servers under consideration is from 9 to 18 hours. Carry out 6 measurement sessions $p_i \in P$, widely embracing the specified period of activity. Select the specification of measurement sessions $p_i=\{t,t_0,\Delta\tau\}$, where t is the session duration, t_0 is the time of the beginning of the session, $\Delta\tau$ is the measurement interval: s_1={30 min, 11:30, 1 s.}, s_2={30 min, 12:30, 1 s}, s_3={30 min, 14:30, 1 s}, s_4={30 min, 15:30, 1 s}, s_5={30 min, 16:30, 1 s}, s_6={30 min, 17:30, 1 s}.

Step 6. Perform the measurements of the observable process v_1 according to plan P and form the sample of values X^{+} with the capacity of 10800 values.

Step 7. Select the rules of converting the measured values $\mathfrak{R}$: sharply outstanding values (exhausts) should be eliminated using the Irvine criterion [19] and exponential smoothing with alpha constant 0.8 should be performed in each estimation interval.

Step 8. As rules $\mathfrak{R}$, applied to the entire X^{+} were not selected, proceed to the next step.

Step 9. Select the estimation interval Δt equal to 10 values.

Step 10. Split set X^{+} into sections $x_i \in X^{+}$ according to Δt, where each subsequent section x_i is taken by means of shifting the value window having length Δt by one position to the right. As a result, we form 10791 sections $x_i \in X^{+}$.

Step 11. For each $x_i \in X^{+}$, convert the values in accordance with rules $\mathfrak{R}$ and estimate model $I^{+}=\{d,l,s,t,p,m,u\}$ using rules R^{+}. As a result, form the frequency table of distribution Δ^{+} of estimated instances of model $I^{+}=$ $=\{d,l,s,t,p,m,u\}$ (Table 6).

Table 6

Frequency table Δ^{+} – frequency f of model instances $I^{+}=\{d,l,s,t,p,m,u\}$

I^{+}	0001200	0001100	0001110	0011200	0001101	0101000	0101010
f	91.16 %	8.42 %	0.24 %	0.08 %	0.07 %	0.01 %	0.01 %

Step 12. Select the approximating function $\hat{f}_i(t) \in F$ using rule $\rho(\Theta,\Delta^{+})$. Calculate normalized indicator D from formula (1) for each tested function (Table 7).

Table 7

Value of indicator D for tested approximating functions

$\hat{f}_i(t)$	Cubic	Quadratic	Moving average	Exponential	Linear
D	0	0.42	0.43	0.83	1

In accordance with the presented indicator, it is recommended to select a cubic approximating function.

6. 3. Assessing the adequacy of the selected approximating function

Verify the adequacy of the approximating function empirically. Form test samples of the values of observable process v_1 and using the rules described earlier, compare the chosen approximating function with the other functions of set $\hat{F}$.

Test data were collected for a week on weekdays. Measurements of the observable process v_1 were performed in accordance with the measurement plan P. 39 samples of 1.800 each, united into a single set of values of the capacity of 70.200, were formed.

The adequacy of approximating function $\hat{f}_i(t) \in \hat{F}$ was estimated within estimation interval Δt, selected at step 9 of the method testing. The estimation interval is shifted by one position to the right after each calculation of adequacy indicators. For the value set of the capacity of 70200, the approximating function will be tested 70199 times. Using rules from p. 5. 2, calculate indicator A – the percentage of intervals, in which corresponding $\hat{f}_i(t)$ is adequate (Table 8).

Table 8

Adequacy indicator A of the tested approximating functions

$\hat{f}_i(t)$	Cubic	Quadratic	Exponential	Linear	Moving average
A, %	41.32	37.25	34.24	33.90	29.42

The resulting adequacy indicator is the best for the cubic approximation function, which proves the adequacy of the developed method. Moreover, the order of adequacy value of other approximating functions is also almost in line with the previous theoretical calculations. Only the moving average function got a relatively worse result than it was predicted.

The presented evidence of the adequacy of the developed method, calculated based on theoretical indicators, requires additional practical verification. Testing will be carried out, which will involve the implementation of the monitoring process in accordance with the previously described specification (Table 3). The results of monitoring without approximation and with the approximation of each function $\hat{f}_i(t) \in \hat{F}$ will be estimated. We expect the best indicators for the previously selected cubic function.

Table 9 gives the results of calculations of computational burden C and the approximation error MAPE after conducted test measurements. The C indicator is calculated as an average bitrate of the data transmitted within the monitoring process. The MAPE is calculated as the average absolute approximation error percent.

Table 9

Results of measurement of the observable variable CPU utilization

Indicator	$\hat{f}_i(t)$					
	Without approximation	Cubic	Quadratic	Exponential	Linear	Moving average
C, bite/s	360	193	267	319	337	359
$MAPE$, %	0	3.79	4.00	0.28	0.21	0.21

Fig. 4 shows a fragment of the curves of values of an observable variable without and with the approximation of the cubic function.

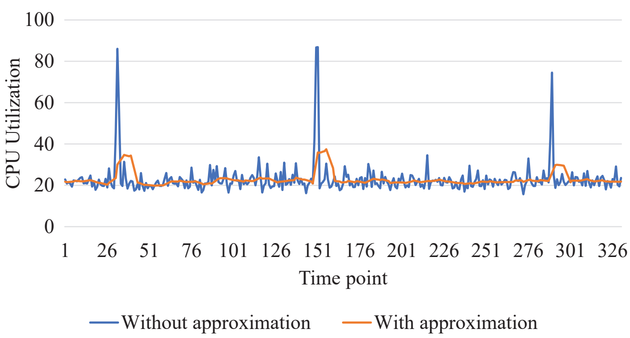

Fig. 4. True and approximated values of the observable process

The obtained results prove the adequacy of the developed method because the cubic approximating function received the best indicator of computational burden at an acceptable level of reduction in measurement accuracy (no more than 5 %). During monitoring a large number of processes, significant savings in computational burden can be achieved.

7. Discussion of results of studying the approximation of values for the observable processes in the cloud infrastructure

One of the criteria of effectiveness of monitoring the cloud infrastructure processes is to achieve the required indicators of the level of computational burden and degree of measurement accuracy. The monitoring effectiveness, based on the approximation of accumulated measurements, depends, in particular, on the adequacy of the chosen approximating function.

The studies of the ways of approximation of values of processes of the cloud infrastructure revealed that it is possible to choose an adequate approximating function based on seven properties of the approximating curve (information models $I^- = \{d,l,s,t\}$ and $I^+ = \{d,l,s,t,p,m,u\}$), in this case, five approximating functions (set $\hat{F}$) are enough for the approximation of any process. The selection criterion $\hat{f}_i(t) \in \hat{F}$ is the minimization of the indicator (1) – normalized Euclidean distance between the instances of model $I^+ = \{d,l,s,t,p,m,u\}$ of functions from set $\hat{F}$ (Table 2) and instances of model $I^+ = \{d,l,s,t,p,m,u\}$ calculated for an observable process.

The formal approach to the selection of $\hat{f}_i(t) \in \hat{F}$ is represented by twelve steps of the proposed method. Analysis of accumulated measurements according to the structure shown in Fig. 1, allows estimating the properties of an observable process in the form of a set of instances of model $I^+ = \{d,l,s,t,p,m,u\}$ and select $\hat{f}_i(t) \in \hat{F}$ according to rule $\rho(\Theta,\Delta^+)$. The results of the method testing (Table 7) are verified in the study by the criteria for verification of the adequacy of the approximating function (Table 8). Empirical indicators of monitoring efficiency (Table 9) prove the effectiveness of solving the problem of the generated excess computational burden. The positive result is reached due to missing the planned measurements and their replacement with approximated values.

A feature of the proposed method of selecting approximating functions for cloud infrastructure processes is the use of a priori information about observable processes. This made it possible to improve approximation results and ultimately reduce computational burden within the required measurement accuracy. The designed method is the development of the approaches proposed in [20, 21]. The improvement lies in enhancing the approximation accuracy by increasing the number of approximating functions.

The constraints of the method are:

1) the probability of obtaining 100 % accuracy of approximation of values of observable processes tends to zero. Using the method when solving the problem of reducing the computational burden is bound to lead to some decrease in measurement accuracy;

2) for an observable process with a poorly predictable behavior or often changing properties, the use of the method is not effective, as in this case, it is difficult to choose an adequate approximating function;

3) The method is focused on the quantitative one-dimensional observable processes of cloud infrastructure. Qualitative or multidimensional processes require additional transformations to enable the use of the proposed method.

The disadvantages of the developed method include possible omissions of short-term bursts of measured values, a low probability of achieving 100 % accuracy of measurements, the dependence of the method effectiveness on expert evaluation.

The proposed method can find its practical application in the implementation of cloud infrastructure monitoring systems. The described method for approximation of the values of observable processes can reduce the level of computational burden, provide the required level of measurement accuracy and promptness.

8. Conclusions

1. The set of key properties of an observable process of cloud infrastructure for the selection of an approximating function (information models $I^- = \{d,l,s,t\}$ and $I^+ = \{d,l,s,t,p,m,u\}$) was formed. The set of approximating functions $\hat{F}$ was selected. Criterion D and rule $\rho(\Theta,\Delta^+)$ of selecting the approximating function $\hat{f}_i(t) \in \hat{F}$ based on the properties of an observable process were developed.

2. The method for the selection of approximating functions for cloud infrastructure processes was developed. The method involves evaluation of the properties of an observable process, comparing these properties with the properties of pre-selected approximating functions, and selection of an approximating function using the proposed criterion. The specific feature of the method is the use of a priori information about observable processes to improve the accuracy of evaluation of their properties and the choice of an appropriate approximating function. Compared to similar solutions [20, 21], the difference of the proposed method is to enhance approximation accuracy through an expanded set of approximating functions and the developed method of selecting them for observable processes. The method solves the problem of an excessive computational burden by increasing the number of missing planned measurements.

3. The computational research into the developed method was carried out by implementing the process of monitoring the CPU utilization variable of one of the cloud infrastructure servers. The result of the method implementation was verified by statistical criteria, and the obtained final assessment proved the adequacy of the chosen function. The indicators of computational burden and measurement accuracy calculated during monitoring the CPU utilization variable prove the effectiveness of solving the described problem of the research: the use of the bandwidth of communication channels was reduced by 40 %, the monitoring accuracy is ensured at the level of 95 %.

References

1. Network Monitoring Tools. Available at: http://www.slac.stanford.edu/xorg/nmtf/nmtf-tools.html
2. Savic, M., Ljubojevic, M., Gajin, S. (2020). A Novel Approach to Client-Side Monitoring of Shared Infrastructures. IEEE Access, 8, 44175–44189. doi: https://doi.org/10.1109/access.2020.2978172
3. Lavy, M., Meggitt, A. (2001). Windows Management Instrumentation (WMI). Indianapolis, Ind.: New Riders.
4. Natu, M., Ghosh, R. K., Shyamsundar, R. K., Ranjan, R. (2016). Holistic Performance Monitoring of Hybrid Clouds: Complexities and Future Directions. IEEE Cloud Computing, 3 (1), 72–81. doi: https://doi.org/10.1109/mcc.2016.13
5. Liaqat, M., Chang, V., Gani, A., Hamid, S. H. A., Toseef, M., Shoaib, U., Ali, R. L. (2017). Federated cloud resource management: Review and discussion. Journal of Network and Computer Applications, 77, 87–105. doi: https://doi.org/10.1016/j.jnca.2016.10.008
6. Ward, J. S., Barker, A. (2015). Cloud cover: monitoring large-scale clouds with Varanus. Journal of Cloud Computing, 4 (1). doi: https://doi.org/10.1186/s13677-015-0041-9
7. Alcaraz Calero, J. M., Aguado, J. G. (2015). MonPaaS: An Adaptive Monitoring Platformas a Service for Cloud Computing Infrastructures and Services. IEEE Transactions on Services Computing, 8 (1), 65–78. doi: https://doi.org/10.1109/tsc.2014.2302810
8. Kozat, U. C., Liang, G., Kokten, K., Tapolcai, J. (2016). On Optimal Topology Verification and Failure Localization for Software Defined Networks. IEEE/ACM Transactions on Networking, 24 (5), 2899–2912. doi: https://doi.org/10.1109/tnet.2015.2494850
9. Khalili, M., Zhang, M., Borbor, D., Wang, L., Scarabeo, N., Zamor, M.-A. (2019). Monitoring and Improving Managed Security Services inside a Security Operation Center. ICST Transactions on Security and Safety, 5 (18), 157413. doi: https://doi.org/10.4108/eai.8-4-2019.157413
10. Tudoran, R., Costan, A., Antoniu, G. (2016). OverFlow: Multi-Site Aware Big Data Management for Scientific Workflows on Clouds. IEEE Transactions on Cloud Computing, 4 (1), 76–89. doi: https://doi.org/10.1109/tcc.2015.2440254
11. Weingärtner, R., Bräscher, G. B., Westphall, C. B. (2015). Cloud resource management: A survey on forecasting and profiling models. Journal of Network and Computer Applications, 47, 99–106. doi: https://doi.org/10.1016/j.jnca.2014.09.018
12. Surianarayanan, C., Chelliah, P. R. (2019). Cloud Monitoring. Essentials of Cloud Computing, 241–254. doi: https://doi.org/10.1007/978-3-030-13134-0_8
13. Grytsenko, O. (2015). Evaluation method of information models of observed processes in computer network. Eastern-European Journal of Enterprise Technologies, 1 (2 (73)), 4–11. doi: https://doi.org/10.15587/1729-4061.2015.36277
14. Inozemtsev, K. O., Sharapov, M. P. (2014). O vybore approksimiruyushchey funktsii dlya opisaniya zavisimosti otnositel'nogo energeticheskogo razresheniya gamma-spektrometrov ot energii gamma-kvantov. Vestnik magistratury, 1 (7 (34)), 4–13.
15. Chatfield, C. (1995). Problem Solving: A statistician's guide. Chapman and Hall/CRC, 325. doi: https://doi.org/10.1201/b15238
16. Saenko, V. I., Gritsenko, A. I. (2014). Informatsionnye modeli nablyudaemyh protsessov dlya monitoringa komp'yuternyh setey. Visnyk Natsionalnoho tekhnichnoho universytetu «KhPI». Seriya: Novi rishennia v suchasnykh tekhnolohiyakh, 48, 55–66.
17. Lunt, M. (2013). Introduction to statistical modelling: linear regression. Rheumatology, 54 (7), 1137–1140. doi: https://doi.org/10.1093/rheumatology/ket146
18. Lukashin, Yu. P. (2003). Adaptivnye metody kratkosrochnogo prognozirovaniya vremennyh ryadov. Moscow: Finansy i statistika, 416.
19. Lemeshko, B. Yu. (1997). Robastnye metody otsenivaniya i otbrakovka anomal'nyh izmereniy. Zavodskaya laboratoriya, 63 (5), 43–49.
20. Saenko, V. I., Gritsenko, A. I. (2007). A sampling method for the processes of contiguous monitoring. Radioelektronika i informatika, 4, 113–118.
21. Hernandez, E. A., Chidester, M. C. George, A. D. (2001). Adaptive Sampling for Network Management. Journal of Network and Systems Management, 9, 409–434. doi: https://doi.org/10.1023/A:1012980307500