Method of Arrival Process Description for Packet Switch

Konstantin E. Samouylov[a], Andrey K. Levakov[b], Nikolai A. Sokolov[c] and Vitaliy S. Zaitcev[d]

[a]*Peoples' Friendship University of Russia Mikluho-Maklaya St., 6, Moscow, 117198, Russia*

[b]*Center Macro-regional Branch (Center MRF) of PJSC Rostelecom, "Comcity", household 6, build. 1, Kievskoe shosse 22km, 108811, Moscow*

[c]*Saint-Petersburg Branch of Central Science Research Telecommunication Institute, 11, Warshavskaya, St. Petersburg, 196128, Russia*

[d]*The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, 22/1 Prospect Bolshevikov, St. Petersburg, 193232, Russia*

Abstract

The arrival process description is important problem for analysis of the packet switches that can be considered as queueing systems. Such queueing systems are the appropriate models for the study of the stochastic characteristics. The accuracy of the results of the model study is largely determined by the correctness of the description of the arrival process. This paper proposes a method to solve the problem of choosing a distribution type to describe arrival process at the teletraffic system input. The authors introduce a criterion for choosing the distribution type based on the error minimization in the estimates of the mean value and coefficient of variation in the delay times in the teletraffic system. Sometimes measurement results characterizing the arrival process are available. In this case, the correctness of the proposed distribution is also checked using the goodness-of-fit test. The paper provides case studies on how the proposed method can be applied to the process of choosing the distribution type in question.

Keywords

queueing system, packet switch, distribution function, average delay time, coefficient of variation, goodness-of-fit test, relative error

1. Problem Statement

The stated problem can be considered a special case of estimating the distance between functions [1]. In the teletraffic theory, this problem has some peculiarities. In this article, an entity that should be serviced by the queueing system is called a request. Typical example of the request is IP packet that is processed by a packet switch. For systems with queues [2, 3], the reliable information about the requests arrival interval distribution at the input of the object

Workshop on information technology and scientific computing in the framework of the X International Conference Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems (ITTMM-2020), Moscow, Russian, April 13-17, 2020

ksam@sci.pfu.edu.ru (K. E. Samouylov); levakov1966@list.ru (A. K. Levakov); nicksokolov@hotmail.com (N. A. Sokolov); zaitsev@protei.ru (V. S. Zaitcev)

0000-0002-6368-9680 (K. E. Samouylov)

under study is normally obtained based on measurements of packet traffic. If measuring is not possible, the proposed hypothesis should be thoroughly reasoned.

Based on measurements, step function $F(t)$ is formed, for which an interval of constant time τ is usually selected on the abscissa axis. In some cases, function $A(t)$ can be convenient to use if the initial distribution replacement simplifies the further model analysis. Function $A(t)$ is usually chosen from a set of known random value distributions [4].

For the analysis of most teletraffic models, it is sufficient to know the first and second moments of the request delay time in the system — $S^{(1)}$ and $S^{(2)}$. Interestingly, most of the known relations [2, 3] rely on delay time coefficient of variation k_S instead of the second moment. That's why the proximity of values $S^{(1)}$ and k_S to the values obtained using measurements of function $F(t)$ determines whether distribution $F(t)$ has been chosen appropriately. Therefore, relative errors in evaluating values $S^{(1)}$ and k_S denoted below as δ_1 and δ_2, respectively, should not exceed the predefined thresholds.

We can choose value τ for function $F(t)$ based on the considerations given in [5]. Values of function $A(t)$ at points that are multiples of τ are known. Figure 1 shows distribution functions $F(t)$ and $A(t)$ up to value 8τ on the abscissa axis. For example, parameter d_3 defines the difference between two distributions at point 3τ. In this example, we observe the maximum difference between functions $F(t)$ and $A(t)$ at point 8τ.

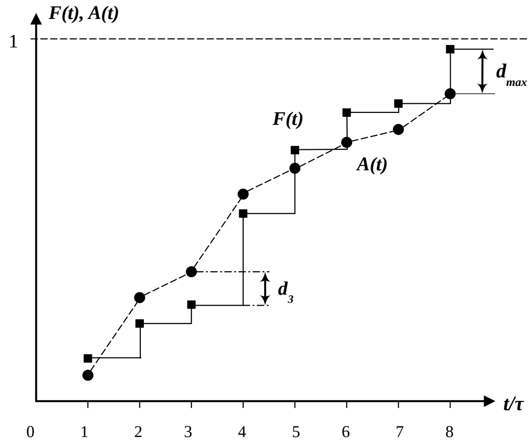

Figure 1: Example of distribution functions $F(t)$ and $A(t)$

The proximity of functions $F(t)$ and $A(t)$ is normally measured by checking if they belong to the same distribution class using an appropriate goodness-of-fit test [6]. This approach does not allow us to make assertions about the values of errors in evaluating the characteristics at the teletraffic system output.

Figure 2 shows the simplest model of a teletraffic system as a so-called "black box". Function $B(t)$ represents the distribution of the request processing times in the teletraffic system. In other words, it defines the set of operations on functions $F(t)$ or $A(t)$. Two distributions at the model output are of practical interest: request delay time $S(t)$ and time interval $D(t)$ at which the processed requests leave the teletraffic system.

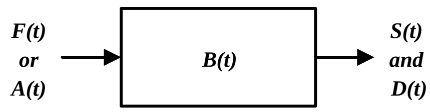

Figure 2: Teletraffic system modeled as a black box

This paper only covers distribution $S(t)$. Moreover, the analysis of function $S(t)$ is limited to estimating values $S^{(1)}$ and k_S.

2. Choosing Distribution Based on Measurements

Let us assume that the measurements were made correctly and we obtained step function $F(t)$ with increment p_k at point $k\tau$. Some increments can evaluate to zero. Generally, index k varies from zero to n, that is, for $t \geq n\tau$, condition $F(t) \equiv 1$ is true. It is convenient to represent function $F(t)$ as the Laplace-Stieltjes transform [7] denoted as $\varphi(s)$.

The first and second moments of the distribution, $F^{(1)}$ and $F^{(2)}$, are determined according to the corresponding rules by differentiation of function $\varphi(s)$. Standard deviation σ_F and coefficient of variation k_F are calculated based on values $F^{(1)}$ and $F^{(2)}$ [4].

Approximating distribution $A(t)$ is normally chosen using the least squares method [8]. In some cases, it is reasonable to use the weighted least squares method [9]. In this case, normally, the following inequalities are true: $A^{(1)} \neq F^{(1)}, A^{(2)} \neq F^{(2)}, \sigma_A \neq \sigma_F$ and $k_A \neq k_F$. These inequalities introduce additional errors in the evaluation of characteristics $S^{(1)}$ and k_S.

A methodological approach based on the following operations can help us minimize these errors:

- First, the most appropriate type of two-parameter distribution $A(t)$ is selected using a suitable goodness-of-fit test [6].
- Then, we determine such parameters of distribution $A(t)$, for which equations $A^{(1)} = F^{(1)}$ and $k_A = k_F$ (or $\sigma_A = \sigma_F$ if it simplifies the calculations) are true.

So, the distribution parameters are calculated by solving a system of two equations.

3. Choosing Distribution Using Goodness-of-Fit Test

Pearson's chi-squared test [6], also known as χ^2, is often used to test the hypothesis that the sample belongs to the theoretical distribution $A(t)$. Some researchers prefer the Kolmogorov-Smirnov test for this purpose [10]. Some other tests may also be chosen.

A goodness-of-fit test is an important step in solving the stated problem, which should be described in terms of "necessity and sufficiency" [11]. Replacing function $F(t)$ with distribution $A(t)$ should be interpreted as a necessity but cannot be considered sufficient. Indeed, goodness-of-fit tests cannot provide numerical estimates of errors that occur when further operations are performed on distribution $A(t)$. Alternatively, if function $F(t)$ and distribution $A(t)$ are not close to each other [1], an acceptable difference in characteristics $S^{(1)}$ and k_S can be unstable within the load range under study. Besides, this narrows the application scope of the proposed method for choosing distribution $A(t)$.

However, we cannot insist that the condition, which was hereinafter treated as necessary, is indispensable. When solving some specific tasks, the mentioned tests may indicate that the chosen hypothesis is false while the evaluation accuracy of characteristics $S^{(1)}$ and k_S can be acceptable. A simple example, in which two moments are the same while distribution functions differ significantly, will be given below in this paper. Therefore, it appears that the obtained results will be still more valuable if we use proven laws of mathematical statistics.

4. Proposed Method for Choosing Distribution

Measuring traffic multiple times at different packet switches shows that function $F(t)$ belongs to a class of distributions that are defined on a limited interval. They are denoted with the lower index "l" (the first letter in word "limited"). The lower index "u" (the first letter in word "unlimited") is used to denote distributions that take possible values along the entire positive semiaxis. The applied approximations $A_u(t)$ introduce an error, which is usually very difficult to evaluate. Therefore, it is appropriate to look for an approximation to function $F(t)$ in the $A_l(t)$ class.

A particular interest in the functions of the $A_l(t)$ class is related to a beta distribution [12, 13]. It can be useful in studying functions $A_l(t)$ with a high value of the coefficient of variation, which is typical of packet multiservice networks. Other distributions [4], such as parabolic, uniform, and others, are also relevant. When the derivative of function $F(t)$ has several extrema, we can use a combination of two or more distributions that belong to the $A_l(t)$ class.

To get a conclusive estimate, it is sufficient to consider an example of using a beta distribution defined on the interval [0;1]. In this case, its density $a(x)$ is determined by the following relation [4]:

$$a(x) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} x^{u-1}(1-x)^{v-1}. \tag{1}$$

Variable x is a dimensionless value. It can be defined as time t divided by value $n\tau$. The following conditions are true for the distribution parameters in formula (1): $u > 0, v > 0$. The relations for calculating the mathematical expectation of the request arrival interval value $A^{(1)}$ and its coefficient of variation k_A are given, for example, in [4]:

$$A^{(1)} = \frac{u}{u+v}, \quad k_A = \sqrt{\frac{v}{u(u+v+1)}}. \tag{2}$$

It is evident that $A^{(1)} < 1$. Having fixed value $A^{(1)}$, we change the parameters of the chosen approximating distribution to obtain the necessary values for coefficient of variation k_A. To calculate parameters u and v, it is necessary to solve a system of two equations, which provides the following result:

$$v = \frac{\left[1 - A^{(1)}\left(1 + k_A^2\right)\right]\left[1 - A^{(1)}\right]}{A^{(1)} k_A^2}, \quad u = \frac{1 - A^{(1)}\left(1 + k_A^2\right)}{k_A^2}. \tag{3}$$

Figure 3 shows an example of distribution $F(x)$ obtained by measuring the packet traffic characteristics. The corresponding step function has five increments. The monotonically increasing curve corresponds to an approximation of the obtained dependence by function $A(x)$, which is a beta distribution of the first type with the following parameters: $u \approx 0.052, v \approx 0.212$.

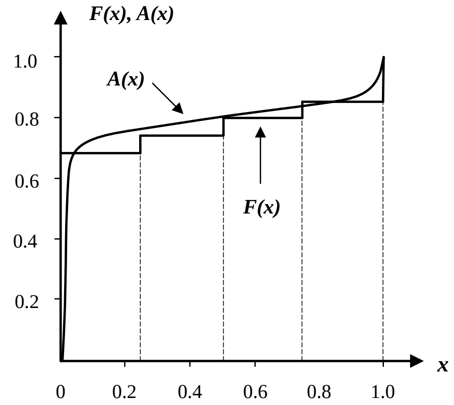

Figure 3: Function $F(x)$ and its approximation by distribution $A(x)$

Comparing two distributions with a 5% significance level according to Pearson's chi-squared test showed that the beta distribution can be used in further research. Then, we should choose a teletraffic system model, which would allow us to evaluate errors in calculating values $S^{(1)}$ and k_S. In this paper, a packet switch is considered a teletraffic system. The request (IP packet) processing time can be safely considered a constant value [14, 15]. Therefore, in the Kendall's notation [3], the model under study can be represented as $Beta/D/1$. The designation "*Beta*" in the first position specifies the nature of the arrivals process, as determined by the beta distribution. If the arrivals process is defined based on measurements, the letter "*G*" [3] should be put in the first position.

Later, we will describe the impact of request processing time distribution $B(t)$ on the evaluation accuracy of values $S^{(1)}$ and k_S. For this reason, for the sake of generality, the processing time for the $Beta/D/1$ model is denoted below as moment $B^{(1)}$. Values $B^{(1)}$ should be chosen in such a way as to investigate the dependence of errors δ_1 and δ_2 on model load ρ. According to the above-mentioned designations, load ρ is defined by the relation of $B^{(1)}$ to $A^{(1)}$. In this paper, the load range of $0.1 \leqslant \rho \leqslant 0.9$ is selected based on two considerations. The load of less than 0.1 is of no practical interest in terms of compliance with the Quality of Service targets. The load greater than 0.9 is not typical of the teletraffic system's operating processes and requires additional research.

The proposed range of load change ρ is sufficient to analyze the operation modes of a model being a teletraffic system, when the object under study functions in standard conditions, which allow us to assume that the number of waiting places in a queue is unlimited. This hypothesis is true if the real capacity of the buffer memory is rated for the loss probability at the approximate level of 0.001, as stated by the International Telecommunication Union Standardization Sector in Recommendation Y. 1541 [16]. The validity of this assumption was established in [17].

For the model under study, the values of errors δ_1 and δ_2 in the given range did not exceed 1.5%. This is quite acceptable for the tasks in the telecommunication network design. Several similar models with other step function types have shown acceptable estimates for errors δ_1 and δ_2 in the given range of the load change.

The change in distribution $B(t)$, which allows us to analyze the change in values δ_1 and δ_2 when the coefficient of variation of the request processing time is increased to 2.0, showed that the corresponding errors are within the same range. It means that the obtained estimates of errors δ_1 and δ_2 are almost invariant with the request processing time distribution.

The type of function $F(x)$, for which Pearson's chi-squared test discards the hypothesis that this function is similar to beta distribution $A(x)$, was chosen artificially using variations in values of increments p_k. The analysis of such functions $F(x)$ and $A(x)$ showed that errors δ_1 and δ_2 begin to grow significantly and often exceed 20%. Generally, this value is not considered acceptable for the analysis of the teletraffic system characteristics.

The results confirm an intuitive conclusion that the positive result of the goodness-of-fit test should be considered a "necessary" condition for applying the proposed method for choosing function $A(t)$. This statement is based on the understanding that the goodness-of-fit test is indicative of a relatively small distance between the functions [1].

However, as the analysis was limited to using only one distribution, it does not allow us to apply this statement to all types of functions $A(t)$. If we limit the types of function $A(t)$ to the distributions that passed the goodness-of-fit test, it will meet the "beauty in science" criterion [18].

5. Errors Related to Different Types of Distributions

The condition that the two moments of functions $S(t)$ and $A(t)$ should be equal can be met for several types of the approximating distribution. This brings up the question about the preferred type of distribution $A(t)$. It may happen that some distributions will show very close values of errors δ_1 and δ_2.

This assumption is based on the results given in [19]. This monograph presents a graph of Hurst exponent H dependence [20] on coefficient of variation k_A for two types of distributions, a gamma distribution and Weibull distribution [4]. The mentioned graph is reproduced in Figure 4. According to the graphs, the maximum deviation of the corresponding dependencies does not exceed 5%.

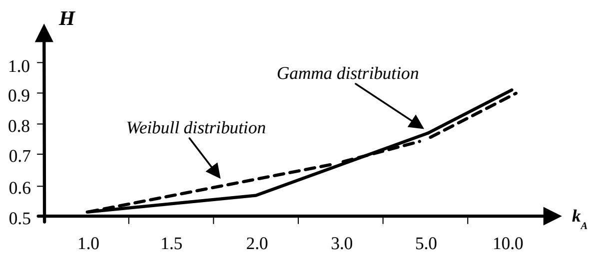

Figure 4: Relationship between coefficient of variation k_A and Hurst exponent H

Errors of type δ_1 and δ_2 for different distributions $A(t)$ but with identical first and second moments, respectively, are of practical interest. The variance or the coefficient of variation can be used instead of the second moment if this simplifies the required calculations.

Let us consider three types of distribution $A(t)$. The first and second types are the same as the distributions shown in Figure 4. The third type is a hyperexponential distribution [4]. All the described functions belong to the $A_u(t)$ family. For all three distributions, $A^{(1)} = 1$ and $k_A = 2$. Table 1 shows the values of the skewness and kurtosis [4], which significantly differ for the distributions under study.

Table 1
Characteristics of three types of distributions $A(t)$

Parameter	Gamma distribution	Weibull distribution	Hyperexponential distribution
skewness	4.00	5.58	5.87
kurtosis	24.00	57.75	49.17

However, the pattern of change in the three curves that represent the density graphs of the distributions under study, has a common nature, which is illustrated in Figure 5 and confirmed by the χ^2 test. In other words, functions $f_{u1}(x)$, $f_{u2}(x)$ and $f_{u3}(x)$ are close to each other [1]. For this reason, any distribution can be chosen as an approximating dependence if values $A^{(1)}$ and k_A are the same.

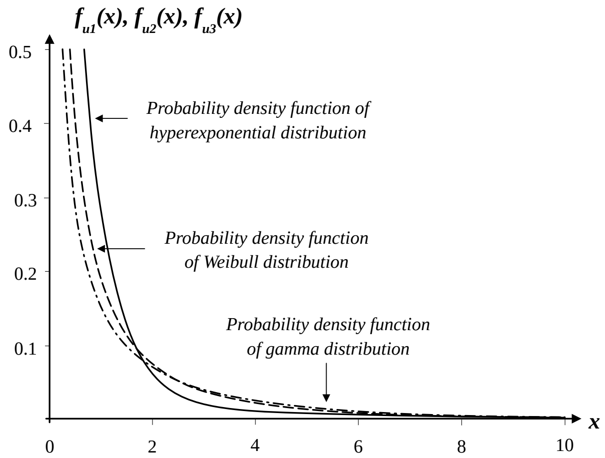

Figure 5: Probability densities for three distributions $A_u(t)$

The last statement was verified by modeling a teletraffic system of type $G/D/1$ [3]. As in the previous experiment, the load change was selected in the range of $0.1 \leqslant \rho \leqslant 0.9$. The values of errors δ_1 and δ_2 do not exceed 11%, which is quite acceptable for solving most of the practical problems. It seems appropriate that, among the alternative functions of type $f_{uj}(x)$, we select a function with skewness closest to a similar value obtained by measuring the parameters of the approximated distribution. This statement relies on the use of the skewness in the relation in order to evaluate the quantile of the IP packet delay time, as recommended in [16].

If the measurements do not match the approximating distribution but they have identical values $A^{(1)}$ and k_A, there may be significant errors in further analysis of the teletraffic models. Figure 6 shows an example of two such functions for distributions $A_l(t)$. It should be emphasized that, while two densities are clearly different, they have the same values $A^{(1)}$ and k_A. Moreover, both distributions have the same values of the skewness coefficient, which is equal to zero due to the symmetry of functions $f_{l1}(x)$ and $f_{l2}(x)$.

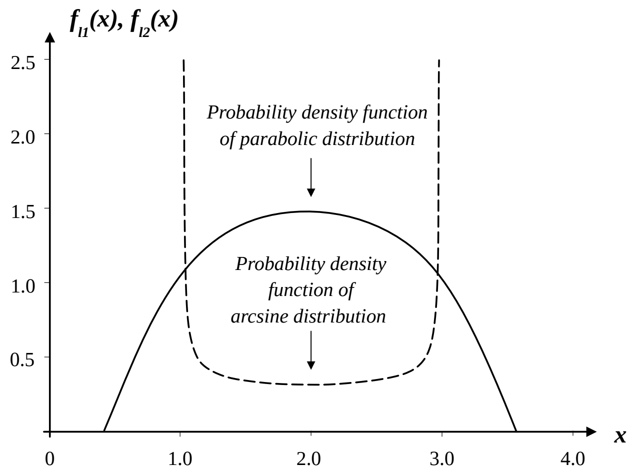

Figure 6: Probability densities for two distributions $A_l(t)$

This example with functions $f_{l1}(x)$ and $f_{l2}(x)$ demonstrates a radical divergence of distributions where values $A^{(1)}$ and k_A are the same. For these two functions, error δ_1 is small. It amounts to a few percent for the given range of the model load. The situation with error δ_2 is different: the error almost reaches 100% when the model is under high load. This indicates that comparing only the mean values of random variables can yield false results.

6. Discussion of Results and Further Research Directions

The proposed method of choosing distribution $A(t)$ is characterized by very high accuracy in evaluating the indicators of the quality of service for the multiservice traffic service presented as a set of IP packets. This method is similar to the procedure proposed in [21] for the analysis of the stochastic characteristics of models with queues. This fact also indicates that the proposed method for calculating characteristics $A^{(1)}$ and k_A is acceptable.

However, it should be noted that, in theory, there could be some specific models with lower accuracy in evaluating the indicators of the quality of service for the multiservice traffic. Therefore, from this perspective, additional research is necessary to solve the following three tasks.

The first task is to establish the relations between the values of errors δ_1 and δ_2 and values d_I, or, possibly, only d_{max}. The method for determining values d_i was shown in Figure 1.

The second task is to introduce the Mahalanobis distance as a measure of the distance between functions [22]. This approach appears to be very productive because it has shown good results in studying the systems similar to the model described in this paper.

The third task is to study the aspects of how the proposed method can be applied for step functions $F(t)$, with several extrema on the histogram. Changes in the packet multiservice traffic in emergencies [23] showed that sometimes several extrema are registered on the histograms that provide the basis for constructing step functions $F(t)$.

References

[1] A. O. Vatulyan, Measure distances between functions, Soros Educational Journal 6 (2000) 123–127.

[2] S. N. Stepanov, Teletraffic theory: concepts, models, applications, Hotline-Telecom, Moscow, 2015.

[3] L. Kleinrock, Queueing Systems: Volume I – Theory, Wiley, New York, NY, 1975.

[4] R. N. Vadzinsky, Handbook of probability distributions, Science, Saint-Petersburg, 2001.

[5] A. N. Sokolov, Methods for Analyzing the IP Packet Delays in a Next-Generation Network, Ph.D. thesis, The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, Saint-Petersburg, Russia, 2011.

[6] G. Kramer, Mathematical methods of statistics, Mir, Moscow, 1975.

[7] V. A. Ditkin, A. P. Prudnikov, Integral Transforms and Operational Calculus, Nauka, Moscow, 1974.

[8] G. A. Korn, T. M. Korn, Mathematical Handbook for Scientists and Engineers, Dover Publications, 2013.

[9] I. V. Bychkov, V. I. Zorkaltsev, A. V. Kazazaeva, Weight coefficients in the weighted least squares method, Siberia Journal of Computational Mathematics 18 (2015) 275–288.

[10] H. W. Lilliefors, On the kolmogorov-smirnov test for normality with mean and variance unknown, Journal of the American Statistical Association 62 (1967) 399–402.

[11] E. D. Bloch, Proofs and Fundamentals: A First Course in Abstract Mathematics, Springer, 2011.

[12] 3GPP, Technical Specification Group Radio Access Network; Study on RAN Improvements for Machine-type Communications, Technical Report TR 37.868 V11.0.0, 3GPP, Sophia Antipolis, 2011.

[13] A. K. Levakov, Character of incoming ip-packet flow under emergencies, Vestnik Sviazy (2017) 4–7.

[14] B. D. Cabrera, G. J., L. W., R. K. Mehra, On the statistical distribution of processing times in network intrusion detection, in: Proceedings of the 43rd IEEE Conference on Decision and Control, IEEE, Bahamas, 2004, pp. 75–80.

[15] M. Fukushima, T. Yakine, Performance Analysis of a Differentiated Service Router, Kyoto University, 2001.

[16] ITU-T, Network performance objectives for IP-based services, Recommendation Y.1541, ITU-T, Geneva, 2011.

[17] N. A. Sokolov, Telecommunication Network Planning Tasks, Communication technology, Saint-Petersburg, 2012.
[18] N. A. Sokolov, The aesthetic criterion application for analysis of the of telecommunication systems evolution scenarios, International Journal of Open Information Technologies 7 (2019) 48–52.
[19] N. F. Bakhareva, V. N. Tarasov, Approximate methods and models of queuing. The study of computer networks, Publishing House of Samara Scientific Center of the Russian Academy of Sciences, Samara, 2011.
[20] H. E. Hurst, Long-term storage of reservoirs: an experimental study, Transactions of the American Society of Civil Engineers 116 (1951) 770–799.
[21] T. I. Aliev, Approximation of probability distributions in queueing models, Scientific and Technical Journal of Information Technologies, Mechanics and Optics (2013) 88–93.
[22] G. J. McLachlan, Mahalanobis distance, Resonance (1999) 20–26.
[23] A. K. Levakov, N. A. Sokolov, Three algorithms for traffic limitation in emergencies, in: Selected Papers of the IX Conference "Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems", Moscow, 2019, pp. 84–91.