

# Approximate Method for Estimating Characteristics of Joint Service of Real-time Traffic and Elastic Data Traffic in Multiservice Access Nodes

S. N. Stepanov<sup>\*,a</sup> and M. S. Stepanov<sup>\*,b</sup>

*\*Moscow Technical University of Communications and Informatics, Moscow, Russia*

*e-mail: <sup>a</sup>s.n.stepanov@mtuci.ru, <sup>b</sup>m.s.stepanov@mtuci.ru*

Received July 19, 2023

Revised August 29, 2023

Accepted September 4, 2023

**Abstract**—A mathematical model of joint servicing of priority real-time traffic and elastic data traffic in multiservice access nodes has been constructed and studied. Definitions of quality indicators for joint servicing of incoming requests for information services are provided. A system of statistical equilibrium equations is formed and its use for calculating the exact values of the introduced characteristics is considered. A method for approximate calculation of characteristics is proposed, based on the construction of a system of simplified equilibrium equations. It has been established that the obtained estimates of customer service indicators are asymptotically accurate in the region of large and small losses. The use of the developed method is shown to solve the problem of estimating the volume of traffic offloaded in an overload situation to other access nodes or to other frequency ranges in order to achieve values of specified QoS indicators and to solve the problem of planning the volume of required transmission resource of a multiservice access node.

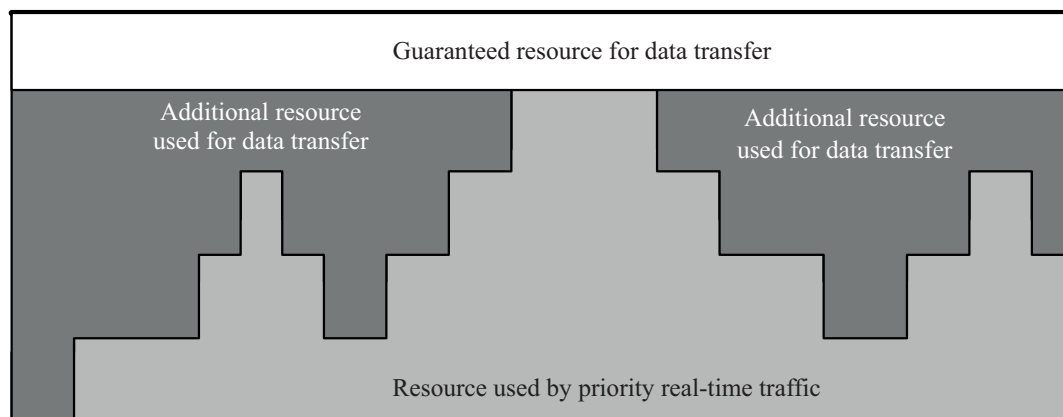
*Keywords:* multiservice traffic, real-time traffic, elastic traffic, recursive algorithms, asymptotically exact characteristics, offloading of excess traffic, transmission resource planning

**DOI:** 10.25728/arcRAS.2024.39.17.001

## 1. INTRODUCTION

The development of communication networks is moving towards expanding the number of services. The range of services received by the subscriber and their quality should not depend on where the subscriber is, how and at what speed he moves, and what access and information transfer technologies are used. This provision follows from the development trends of the telecommunications market and is enshrined in the recommendations of the International Telecommunication Union [1–6]. It is clear that each type of service, be it voice, video or data, has its own characteristics that should be taken into account when studying the conditions for their joint provision. In the simplest case, services differ only in the size of the required information transmission resource. In more complex situations, it is necessary to take into account the details of the formation of input streams of requests, the conditions of access to the network, the presence of priority and the possibility of resource redistribution during the service process, etc. The considered features of the functioning of current and future communication systems are analyzed in the family of so-called multiservice models. They are of great importance for practical applications and are the subject of intensive research by specialists in the field of communications [4–10].

The most important segment of a multiservice network of both fixed and mobile communications is the access node, which performs the function of concentrating subscriber traffic. Information flows entering access nodes can be divided into two categories: traffic of real-time services and traffic of



**Fig. 1.** Allocation of access node capacity during joint transmission of priority traffic of real-time services and elastic data.

elastic data transfer services. The transmission of real-time service traffic occurs with a preliminary reservation of a resource, which is retained for the entire duration of servicing the request [1–3, 6]. The quality of service is characterized by the proportion of lost requests and the average volume of occupied resources. To forward elastic traffic, a minimum guaranteed resource is allocated, as well as all or part of the resource that remains free from passing traffic from real-time services [8–11]. The quality of service of elastic data is measured by the average file transfer time and the average received throughput of the access node.

It is generally assumed that real-time traffic has resource occupancy advantage over elastic data. It is expressed in reducing the elastic data transmission rate to a predetermined minimum value if performing this procedure facilitates the acceptance of a request for a real-time service. When a free resource becomes available, the data transfer rate increases. Resource redistribution occurs at times determined by the mechanism used to adapt the speed of information transfer to the conditions of servicing requests. Usually these are the moments of receipt of requests and the end of their servicing. Let us call such a distribution procedure dynamic [7, 8, 10]. An example of its implementation is shown in Fig. 1.

The analyzed method of resource division allows to significantly increase its occupancy. This effect is especially important for cellular mobile communication networks due to the limited range of radio frequencies allocated for the formation of radio channels. According to experts, the gain can amount to several tens of percent of the total volume of the resource used. The noted effect is obtained as a result of the use of traffic control procedures that speed up the transmission of elastic data in situations where the number of telecommunications service users being served is decreasing. In mobile communication networks, dynamic resource allocation is performed by a packet scheduler and occurs under the control of the RRM (Radio Resource Management) software and hardware complex. Control decisions of the complex are made based on information about the state of the channel, the number and types of requests being serviced, the level of interference, etc. [12–15].

The relevance of the problem has attracted the attention of specialists of various professional backgrounds, which range from engineers involved in the design and operation of communication systems to mathematicians. Giving a general description of the published works, it should be noted that most of the studies are engineering developments based on the results of simulation modeling and full-scale experiments [1–3]. Theoretical analysis of the features of joint servicing of real-time traffic and elastic data usually ends with the construction of a mathematic model, determination of the characteristics of the quality of service of coming requests and the implementation of numerical

algorithms for their assessment, based on solving a system of equilibrium equations using any standard linear algebra method [4, 5, 7–10, 13, 15, 16].

For practical applications, in particular for setting up procedures for dynamic resource allocation, it is necessary to construct approximate procedures that are easy to implement and have acceptable accuracy. Unfortunately, few such methods are presented in published works. We can only mention the studies [11, 14]. Of particular importance are approximate algorithms that

- are based on general principles and can be easily generalized to other models for the formation of input flows of requests and resource allocation procedures;
- have good accuracy for values of input parameters that correspond to practical applications, in particular for values of small losses, where the problem of planning the volume of information transmission resource required by the load is solved, and for values of large losses, where the problem is solved estimating the volume of traffic offloaded to other access nodes or to other frequency bands in order to achieve specified QoS indicators;
- are constructed using easy-to-implement analytical expressions and recursive procedures.

This work is devoted to solving these problems. Section 2 discusses the mathematical description of the analyzed access node model, using the example of which the principles of approximate assessment of the characteristics of the quality of joint service of real-time service traffic and elastic data traffic will be formulated. In Section 3, a system of equilibrium equations is constructed and an algorithm for its numerical solution is proposed. The results obtained are further used to estimate the error of approximate methods. In Sections 4 and 5, the principles of constructing approximate methods for estimating the characteristics of the considered access node model are formulated and their error is numerically studied, in particular, it is shown that the estimates are asymptotically accurate in the region of large and small losses. Section 6 discusses the possibility of using the model to solve the problem of determining the volume of offloaded traffic and the problem of estimating the required amount of resource. The last section formulates conclusions based on the results of the study.

The novelty of the results obtained is as follows:

- Methods are formulated for constructing approximate procedures for estimating the characteristics of joint servicing of priority traffic of real-time services and elastic data traffic. The proposed calculation procedures are based on the use of a system of simplified equilibrium equations.
- It is established that the obtained characteristics estimates are asymptotically accurate in the region of large and small losses.
- It is shown that the ideas underlying the approximate methods are of a general nature and can easily be generalized to other models for the formation of input flows of requests and resource allocation procedures.

## 2. MATHEMATICAL DESCRIPTION OF THE MODEL

Let us denote by  $C$  the capacity of a multiservice access node created by the fixed or wireless communication standard used and expressed in bits/s. The access node services  $n$  Poisson flows of requests for the transmission of real-time service traffic and one Poisson flow of requests for the transmission of elastic data. Requests of the  $k$ th flow for the transmission of real-time traffic arrive with an intensity of  $\lambda_k$  and require a capacity reservation of  $c_k$  bit/s for the entire service time, which has an exponential distribution with the parameter  $\mu_k$ ,  $k = 1, \dots, n$ . Requests for the transfer of elastic data (files) are received with an intensity of  $\lambda_d$ . The file size has an exponential distribution with mean  $F$  expressed in bits.

Let us introduce the concept of a virtual channel, which is applicable to numerically estimate the size of the information transmission resource provided to users. Let us denote by  $c$  the information

transmission rate of one channel, expressed in bits/s. The transition to virtual channels simplifies the modeling of the process of resource occupation by requests. The smaller the value of  $c$ , the more accurate the approximation of transmission rates. However, in this situation the number of model states increases. The choice of the value  $c$  depends on the formulation of the problem, for example, you can use the following expressions:  $c = \min(c_1, \dots, c_n)$  or  $c = \text{GCD}(c_1, \dots, c_n)$ , where abbreviation GCD means greatest common divisor. In the first case, the approximation of the bit rate requirements is rougher than in the second. The total number  $v$  of available virtual channels and the number  $b_k$  of virtual channels required to service the request of the  $k$ th flow are found from the relations:

$$v = \left\lfloor \frac{C}{c} \right\rfloor, \quad b_k = \left\lceil \frac{c_k}{c} \right\rceil.$$

Elastic data is maintained in accordance with the provisions of the Processor Sharing discipline. Let's consider the implementation of this procedure. For simplicity, let's assume that the minimum amount of bandwidth that can be used to transfer a file is one virtual channel. Let us denote by  $\mu_d$  the parameter of the exponential distribution of file transmission time by one channel. Let  $i$  be the number of virtual channels used to service real-time traffic, and let  $d > 0$ —be the number of files being transmitted. According to the model description  $(v - i)$  channels are used to serve them. Let  $s = \left\lfloor \frac{v-i}{d} \right\rfloor$ . Free channels are divided between  $d$  files according to the following rule. To serve each of the  $(v - i - sd)$  files,  $(s + 1)$  channels are used, and to serve each of the remaining  $((s + 1)d - (v - i))$  files,  $s$  channels. As a result of this procedure, all  $(v - i)$  channels are occupied. It is easy to show that in the situation under consideration, the time until the end of the transfer of one of the  $d$  files has an exponential distribution with the parameter  $(v - i)\mu_d$ .

An incoming request for real-time traffic transmission has priority in resource occupation, reducing, if necessary, the bandwidth used by one file to one channel. Let  $i_k(t)$ ,  $k = 1, \dots, n$ —the number of requests of the  $k$ th flow for the transmission of traffic of real-time services at time  $t$ , and  $d(t)$ —the number of files being serviced at time  $t$ . The dynamics of changes in the number of serviced requests is described by the Markov process  $r(t) = (i_1(t), \dots, i_n(t), d(t))$ , defined on finite state space  $S$ , which includes states  $(i_1, \dots, i_n, d)$ , with components

$$\begin{aligned} i_1 = 0, 1, \dots, \left\lfloor \frac{v}{b_1} \right\rfloor; \quad \dots \quad i_n = 0, 1, \dots, \left\lfloor \frac{v - i_1 b_1 - \dots - i_{n-1} b_{n-1}}{b_n} \right\rfloor; \\ d = 0, 1, \dots, v - i_1 b_1 - \dots - i_n b_n. \end{aligned} \quad (1)$$

The quality of service for requests from the  $k$ th flow for the transmission of real-time traffic is determined by the values of the share of lost requests  $\pi_k$  and the average number of occupied virtual channels  $m_k$ . The value of the last characteristic makes it possible to calculate the average number of requests of the  $k$ th flow being serviced,  $y_k = m_k/b_k$ , and the average capacity of the access node they occupy  $z_k = m_k c$ . Let us denote by  $z_r$  the average amount of node capacity occupied by real-time traffic  $z_r = \sum_{k=1}^n z_k$ . The quality of service for requests for the transfer of elastic files is determined by the values of the share of lost files  $\pi_d$ , the average number of occupied virtual channels  $m_d$ , the average number of files being serviced  $y_d$ , the average value of used node bandwidth  $z_d = m_d c$ , average file transfer time  $h_d$ , average bitrate used for file transfer  $c_d$ , average number of virtual channels  $b$  used for file transfer.

The introduced indicators can be determined and calculated using the values of stationary probabilities  $p(i_1, \dots, i_n, d)$  of states  $(i_1, \dots, i_n, d) \in S$ . For the state  $(i_1, \dots, i_n, d)$ , let  $i$  denote the number of virtual channels used to service real-time traffic  $i = i_1 b_1 + \dots + i_n b_n$ . Let us present the

calculated expressions

$$\begin{aligned}
 \pi_k &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid i+d+b_k > v\}} p(i_1, \dots, i_n, d); \\
 m_k &= \sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d) i_k b_k; \quad k = 1, \dots, n; \\
 \pi_d &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid i+d+1 > v\}} p(i_1, \dots, i_n, d); \\
 m_d &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid d > 0\}} p(i_1, \dots, i_n, d) (v - i); \\
 y_d &= \sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d) d; \\
 h_d &= \frac{y_d}{\lambda_d (1 - \pi_d)}; \quad c_d = \frac{F}{h_d}; \quad b = \frac{m_d}{y_d}.
 \end{aligned} \tag{2}$$

### 3. SYSTEM OF STATISTICAL EQUILIBRIUM EQUATIONS

To evaluate the characteristics specified by expressions (2), it is necessary find the values of  $p(i_1, \dots, i_n, d) \in S^1$ . To do this, it is enough to construct and solve a system of statistical equilibrium equations, connecting the unnormalized probabilities of the model. It looks like this:

$$\begin{aligned}
 P(i_1, \dots, i_n, d) &\left\{ \sum_{k=1}^n (\lambda_k I(i + d + b_k \leq v) + i_k \mu_k) + \lambda_d I(i + d + 1 \leq v) + (v - i) \mu_d I(d > 0) \right\} \\
 &= \sum_{k=1}^n P(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) + P(i_1, \dots, i_n, d - 1) \lambda_d I(d > 0) \\
 &\quad + \sum_{k=1}^n P(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + d + b_k \leq v) \\
 &\quad + P(i_1, \dots, i_n, d + 1) (v - i) \mu_d I(i + d + 1 \leq v), \quad (i_1, \dots, i_n, d) \in S.
 \end{aligned} \tag{3}$$

Here and below  $I(\cdot)$ —indicator function defined by relation

$$I(\cdot) = \begin{cases} 1, & \text{if the condition is met, formulated in brackets,} \\ 0, & \text{if this condition is not met.} \end{cases} \tag{4}$$

Resulting solutions of (3)  $P(i_1, \dots, i_n, d)$  needs to be normalized.

The system of equations (3) does not have any special properties that provide a recursive estimate of stationary probabilities. For this reason, standard linear algebra methods are used to solve (3). Based on the experience of solving similar systems, it is recommended to use the iterative Gauss–Seidel algorithm to estimate the probabilities of states. The use of this approach makes it possible to calculate stationary probabilities for models of communication systems with a number of states of up to several million. The standard implementation of Gauss–Seidel recursion for the solution (3) is not guaranteed to converge, but in most cases it does. Convergence is studied by indirect methods based on the analysis of the proximity of successive approximations and the fulfillment of known theoretical relationships connecting the values of the characteristics of the studied model of the

<sup>1</sup> We will use lowercase letters to denote the normalized values of the probabilities of states and characteristics and uppercase letters to denote their non-normalized values.

communication system. Such relations include Little's formula. Details of the implementation of this approach can be found in [10, 17–19]. To ensure convergence, it is enough to set one of the unknowns equal to unity, remove the corresponding equation (3) and proceed to solving an inhomogeneous system of linear equations. After the transformations performed, the Gauss–Seidel recursion always converges (due to the presence of weak diagonal dominance), but requires a significantly larger number of iterations for its implementation.

The Gauss–Seidel algorithm will be further used to estimate the error of the approximate method constructed in Section 4 for calculating the quality characteristics of joint servicing of incoming requests for the node model under study. Sufficiently substantiated approximate methods play a major role in the development of engineering techniques aimed at solving the problems of estimating the maximum sufficient load and the minimum required volume of information transmission resource of a multiservice access node. Subsequent sections of the work will be devoted to the formulation, analysis and examples of the use of approximate methods for calculating the characteristics of the constructed access node model.

#### 4. ESTIMATION OF CHARACTERISTICS USING SIMPLIFIED EQUILIBRIUM EQUATIONS

The principle of simplified equilibrium equations was first formulated and used in constructing the boundaries of truncated state spaces, providing a given error in calculating the characteristics of models, taking into account the effect of repeated calls [20, 21]. The idea of the method is to form a system of relationships for the approximate calculation of stationary probabilities of model states based on the requirement that they comply with local conservation laws, which for the model under study are satisfied in individual macrostates, selected in accordance with the structure of the state space used and the physical meaning of individual state components.

For the constructed model, we refer to such macrostates as  $(i)$ , where all states of the model are combined with the number of occupied virtual channels equal to  $i$ ,  $i = 0, 1, \dots, v$ , and  $(d)$ , where all states of the model are combined with the number of transferred files equal to  $d$ ,  $d = 0, 1, \dots, v$ . Proceeding in a similar way, we also introduce the macrostate  $(i, d)$  and determine the stationary probability of this state from the expression

$$p(i, d) = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} p(i_1, \dots, i_n, d),$$

$$i = 0, 1, \dots, v; \quad d = 0, 1, \dots, v - i.$$

If the values of  $p(i, d)$  are known, then the values of the characteristics specified by the expressions (2) can be calculated. To do this, it is enough to use the relations

$$\pi_k = \sum_{i=0}^v \sum_{d=0}^{v-i} p(i, d) I(i + d + b_k > v); \quad m_k = \frac{\lambda_k}{\mu_k} (1 - \pi_k) b_k; \quad (5)$$

$$\pi_d = \sum_{i=0}^v \sum_{d=0}^{v-i} p(i, d) I(i + d + 1 > v); \quad m_d = \sum_{i=0}^v \sum_{d=1}^{v-i} p(i, d) (v - i);$$

$$y_d = \sum_{i=0}^v \sum_{d=1}^{v-i} p(i, d) d.$$

The remaining elastic traffic service characteristics are determined from expressions (2).

Let us obtain the form of local conservation laws for the macrostates  $(i)$ ,  $(d)$ , which will be further used for an approximate estimate of  $p(i, d)$ . Let's start with the macrostate  $(d)$ . Equating



the intensity of the exit  $r(t)$  from  $(d)$  to the intensity of the transition to the macrostate  $(d)$ , we obtain a system of relations of the following form:

$$\begin{aligned} & \left( P(0, d) + P(1, d) + \dots + P(v - d - 1, d) \right) \lambda_d \tag{6} \\ & + \left( P(0, d)v + P(1, d)(v - 1) + \dots + P(v - d, d)d \right) \mu_d I(d > 0) \\ & = \left( P(0, d - 1) + P(1, d - 1) + \dots + P(v - d, d - 1) \right) \lambda_d I(d > 0) \\ & + \left( P(0, d + 1)v + P(1, d + 1)(v - 1) + \dots + P(v - d - 1, d + 1)(d + 1) \right) \mu_d, \\ & \quad d = 0, 1, \dots, v - 1. \end{aligned}$$

Having considered (6) sequentially for  $d = 0, 1, \dots, v - 1$  and performing simple algebraic transformations, we reduce (6) to the following form:

$$\begin{aligned} & \left( P(0, d) + P(1, d) + \dots + P(v - d - 1, d) \right) \lambda_d \tag{7} \\ & = \left( P(0, d + 1)v + P(1, d + 1)(v - 1) + \dots + P(v - d - 1, d + 1)(d + 1) \right) \mu_d, \\ & \quad d = 0, 1, \dots, v - 1. \end{aligned}$$

Now we obtain the form of local conservation laws for the macrostate  $(i)$ . Equating the intensity of the exit  $r(t)$  from  $(i)$  to the intensity of the transition to the macrostate  $(i)$ , we obtain a system of relations of the following form:

$$\begin{aligned} & \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=0}^n \left( P(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) \right. \tag{8} \\ & \quad \left. + P(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + b_k + d \leq v) \right) \\ & = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=0}^n \left( P(i_1, \dots, i_n, d) \lambda_k I(i + b_k + d \leq v) \right. \\ & \quad \left. + P(i_1, \dots, i_n, d) i_k \mu_k I(i_k > 0) \right), \quad i = 0, 1, \dots, v. \end{aligned}$$

Let's move on to constructing a system of simplified equilibrium equations, which we will further use to approximately calculate the characteristics of the access node model under study. Let us denote by  $\hat{P}(i_1, \dots, i_n, d)$ ,  $(i_1, \dots, i_n, d) \in S$  the resulting estimates of the probabilities of stationary states of the considered model. Let us save for estimates of the introduced characteristics (2), probabilities  $P(i, d)$ , etc., obtained using  $\hat{P}(i_1, \dots, i_n, d)$ , previously introduced notations, just add the symbol  $\hat{\phantom{x}}$  to their notation. The quantities  $\hat{P}(i_1, \dots, i_n, d)$  are found from the fulfillment requirement for  $\hat{P}(i_1, \dots, i_n, d)$  local conservation law (7)

$$\begin{aligned} & \left( \hat{P}(0, d) + \hat{P}(1, d) + \dots + \hat{P}(v - d - 1, d) \right) \lambda_d \tag{9} \\ & = \hat{P}(0, d + 1)v \mu_d + \hat{P}(1, d + 1)(v - 1) \mu_d + \dots + \hat{P}(v - d - 1, d + 1)(d + 1) \mu_d, \\ & \quad d = 0, 1, \dots, v - 1, \end{aligned}$$

and relations

$$\begin{aligned} & \hat{P}(i_1, \dots, i_n, d) i_k = \hat{P}(i_1, \dots, i_k - 1, \dots, i_n, d) I(i_k > 1) a_k, \tag{10} \\ & \quad (i_1, \dots, i_n, d) \in S, \quad a_k = \frac{\lambda_k}{\mu_k}, \quad k = 1, \dots, n. \end{aligned}$$

The expressions (10) coincide in form with the detailed balance relations in the formation and servicing of requests for the transmission of real-time traffic in the Erlang multiservice model [7, 10, 22, 23]. They determine the approximate nature of the resulting estimates, since they are not satisfied for the original model.

Using equalities (10), we can show that for  $\hat{P}(i_1, \dots, i_n, d)$  the local conservation law (8) is satisfied, i.e. the following relations are valid

$$\begin{aligned} & \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=1}^n \hat{P}(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) \\ & + \hat{P}(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + b_k + d \leq v) \\ = & \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=1}^n \hat{P}(i_1, \dots, i_n, d) \lambda_k I(i + b_k + d \leq v) \\ & + \hat{P}(i_1, \dots, i_n, d) i_k \mu_k I(i_k \geq v), \quad i = 0, 1, \dots, v. \end{aligned} \tag{11}$$

Next, we will show that the relations (9), (10) allow us to uniquely determine the values of  $\hat{P}(i_1, \dots, i_n, d)$ ,  $(i_1, \dots, i_n, d) \in S$ , and with them the introduced estimates of the characteristics of joint servicing of requests in the studied model of the access node. Thus, estimates of stationary probabilities  $\hat{P}(i_1, \dots, i_n, d)$ ,  $(i_1, \dots, i_n, d) \in S$ , as well as their exact values  $P(i_1, \dots, i_n, d)$ ,  $(i_1, \dots, i_n, d) \in S$  satisfy local conservation laws of the same form (7), (8) and (9), (11), which allows us to expect good estimation accuracy. Based on the probabilistic interpretation (9), (10), let's call these relations as system of simplified equilibrium equations.

Let us construct a recursive algorithm for calculating the values of  $\hat{P}(i, d)$  for a fixed  $d$ . Let us introduce an auxiliary characteristic

$$\hat{Y}_k(i, d) = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \hat{P}(i_1, \dots, i_n, d) i_k, \quad i = 0, 1, \dots, v; \quad d = 0, 1, \dots, v - i.$$

Summing (10) over all  $(i_1, \dots, i_n, d) \in S$  satisfying the condition  $i_1 b_1 + \dots + i_n b_n = i$ , we obtain the equality

$$\hat{Y}_k(i, d) = \hat{P}(i - b_k, d) a_k I(i \geq b_k). \tag{12}$$

Let's multiply (12) by  $b_k$  and sum over  $k = 1, 2, \dots, n$ . Rearranging the order of summation on the left, we obtain a recursive relation connecting successive values  $\hat{P}(i, d)$ ,

$$\begin{aligned} & \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \hat{P}(i_1, \dots, i_n, d) \sum_{k=1}^n i_k b_k = \hat{P}(i, d) i \\ = & \sum_{k=1}^n a_k b_k \hat{P}(i - b_k, d) I(i \geq b_k), \quad d = 0, 1, \dots, v - 1; \quad i = 1, \dots, v - d. \end{aligned} \tag{13}$$

Relations (7) and (13) allow us to construct a recursive algorithm for estimating the values of  $\hat{P}(i, d)$ ,  $d = 0, 1, \dots, v$ ;  $i = 1, \dots, v - d$ . Let us list the sequence of actions for its implementation.

(1) Let's set  $d = 0$ , and the value  $\hat{P}(0, 0) = 1$ .

(2) Let's express values  $\hat{P}(i, 0)$ ,  $i = 1, \dots, v$ , through  $\hat{P}(0, 0)$ , using the relation (13) with  $d = 0$

$$\hat{P}(i, 0) = \frac{1}{i} \times \sum_{k=1}^n \hat{P}(i - b_k, 0) I(i \geq b_k)$$



and successively increasing  $i$  from 1 to  $v$ . For fixed  $i$  values of estimates  $\hat{P}(i - b_k, 0)$ ,  $k = 1, \dots, n$ , or already expressed through  $\hat{P}(0, 0)$  (for  $i - b_k \geq 0$ ), or equal to 0 (for  $i - b_k < 0$ ).

(3) Let's put  $d = 1$  and the value  $\hat{P}(0, 1) = x$ .

(4) Let us express values  $\hat{P}(i, 1)$ ,  $i = 1, \dots, v - 1$ , through  $x$ , using the relation (13) with  $d = 1$

$$\hat{P}(i, 1) = \frac{1}{i} \times \sum_{k=1}^n \hat{P}(i - b_k, 1) I(i \geq b_k)$$

and successively increasing  $i$  from 1 to  $v - 1$ .

(5) Set in (9)  $d = 0$ . We get the ratio

$$\begin{aligned} & (\hat{P}(0, 0) + \hat{P}(1, 0) + \dots + \hat{P}(v - 1, 0)) \lambda_d \\ &= \hat{P}(0, 1) v \mu_d + \hat{P}(1, 1) (v - 1) \mu_d + \dots + \hat{P}(v - 1, 1) \mu_d, \end{aligned} \tag{14}$$

which allows us to express the value of  $x$  in terms of  $\hat{P}(0, 0)$ . Using the obtained relation and the results of step 4, we obtain expressions for  $\hat{P}(i, 1)$ ,  $i = 0, 1, \dots, v - 1$ , through  $\hat{P}(0, 0)$ .

(6) Next, we set  $d = 2, 3, \dots, v$  and, having implemented the sequence of actions formulated in the above stages of the proposed algorithm, we find expressions for  $\hat{P}(i, d)$ ,  $d = 0, 1, \dots, v$ ;  $i = 1, \dots, v - d$  via  $\hat{P}(0, 0)$ .

(7) We find an expression in terms of  $\hat{P}(0, 0)$  for the normalization constant

$$N = \sum_{d=0}^v \sum_{i=0}^{v-d} \hat{P}(i, d).$$

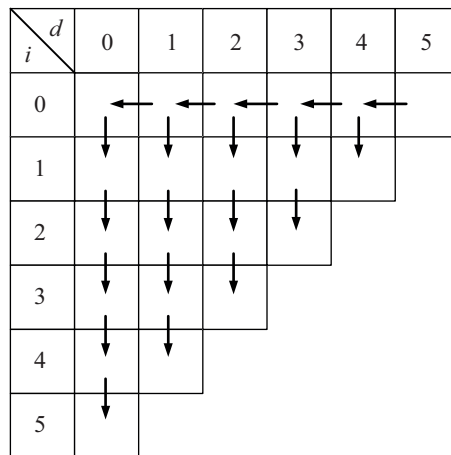
(8) We determine the normalized values of probability estimates  $\hat{p}(i, d)$ :

$$\hat{p}(i, d) = \frac{\hat{P}(i, d)}{N}, \quad d = 0, 1, \dots, v; \quad i = 1, \dots, v - d.$$

(9) Using  $\hat{p}(i, d)$  and expressions (5), we find the values of estimates of the characteristics of joint servicing of requests in the studied model of a multiservice access node.

In Fig. 2, arrows indicate the sequence of calculating  $\hat{p}(i, d)$  using the constructed algorithm for  $v = 5$ .

From a computational point of view, the implementation of the constructed algorithm does not cause any difficulties. The complexity of the procedure is comparable to the repeated use of a recursive algorithm for assessing the characteristics of a multiservice Erlang model [7, 10, 22, 23].



**Fig. 2.** Sequence of implementation of the algorithm for  $v = 5$ .

Let's compare the amount of calculations when estimating the characteristics of the model using the Gauss–Seidel method and the constructed approximate algorithm. At each step of the iterative solution method (3), the approximate values of the probabilities of all states of the model  $(i_1, \dots, i_n, d) \in S$  belonging to the space  $S$  are determined, given by the relations (1). As a result of applying the approximate method, probability estimates  $P(i, d)$  are found for states  $(i, d) \in \hat{S}$ , where  $i = 0, 1, \dots, v$ ;  $d = 0, 1, \dots, v - i$  (see Fig. 2) and calculations are performed only once. From this follows an approximate assessment of the effectiveness of the developed method. The volume of calculations in comparison with the iterative method is reduced by a number of times equal to the number of iterations in the implementation of the Gauss–Seidel method, multiplied by the ratio of the number of states in the space  $S$  to the number of states in the space  $\hat{S}$ . Depending on the model parameters and calculation conditions, this value can significantly exceed several thousand times.

## 5. ERROR IN ESTIMATING THE CHARACTERISTICS OF AN ACCESS NODE

Let us conduct a numerical study of the accuracy of the approximate method for estimating the characteristics of an access node, constructed in the previous section. Let's choose the following values of the input parameters:  $C = 100$  Mbit/s;  $n = 2$ ;  $c_1 = 2$  Mbit/s;  $c_2 = 5$  Mbit/s. Based on the accepted assumptions, we obtain the structural parameters of the model:  $c = 1$  Mbit/s;  $v = 100$  virtual channels (v.c.);  $b_1 = 2$  v.c.;  $b_2 = 5$  v.c. Let's assume that  $F = 80$  Mbit. The average file transfer time using one channel is 80 s. When performing calculations, this time will be taken as time unit. Hence  $\mu_d = 1$ . For real-time services, we select service time parameters from the expressions  $\mu_1 = 0.5$  and  $\mu_2 = 0.5$ .

Let's introduce the  $\rho$  parameter, which we will use to estimate the potential load of one virtual channel. Let us determine  $\rho$  from the expression

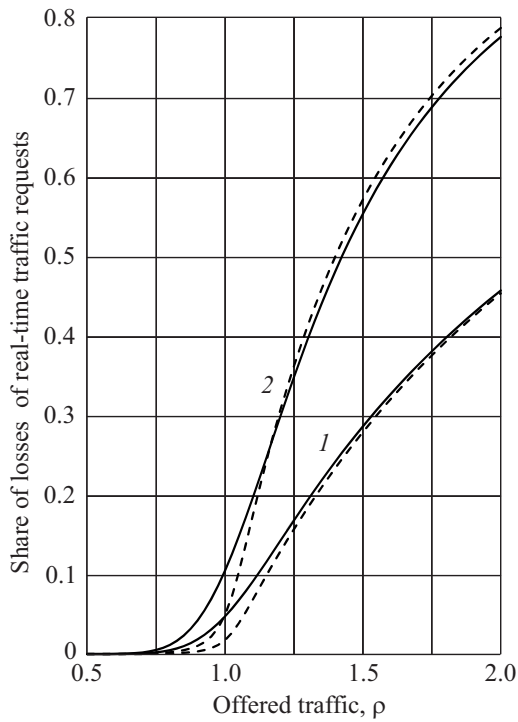
$$\rho = \frac{a_1 b_1 + a_2 b_2 + \frac{\lambda_d}{\mu_d}}{v}. \quad (15)$$

For elastic data, we calculate the potential resource load from the condition of using one channel for file transfer<sup>2</sup>. We will assume that all three flows create the same potential resource load. From here follow the expressions for estimating the intensities of incoming requests:

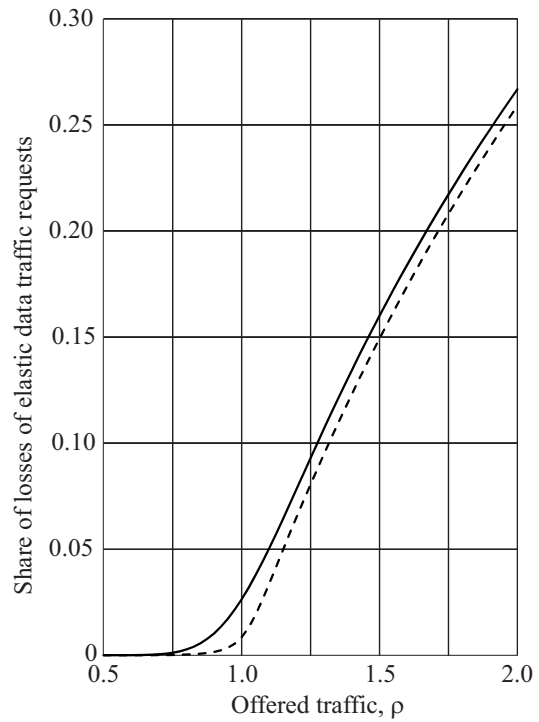
$$\lambda_1 = \frac{v\rho\mu_1}{3b_1}; \quad \lambda_2 = \frac{v\rho\mu_2}{3b_2}; \quad \lambda_d = \frac{v\rho\mu_d}{3}. \quad (16)$$

For selected values of input parameters, Figs. 3–8 shows the dependence of the exact and approximate calculation of the main characteristics of the model from changes in  $\rho$  potential load of the virtual channel. The exact values of the characteristics are obtained from solving the system of equilibrium equations (3) using the iterative Gauss–Seidel method and using definitions (2). Approximate values were found from the relations (5) after substituting into them instead of  $p(i, d)$  the estimates  $\hat{p}(i, d)$  obtained as a result of implementing a recursive algorithm based on the use of simplified equilibrium equations (9), (10) (see Section 4). The corresponding curves are indicated by dotted lines. The number next to the curve indicates the number of the flow for real-time services. In Figs. 3–4 the error in calculating the share of lost requests for servicing real-time service traffic (Fig. 3) and the share of lost requests for servicing elastic data (Fig. 4) was estimated. In Figs. 5–6, the error in calculating the average value of the bitrate used for servicing traffic of real-time services (Fig. 5) and the average value of the bitrate used for servicing elastic data (Fig. 6) was estimated. In Figs. 7–8 the error in calculating the average file transfer time (Fig. 7) and the average number of channels used to transfer one file (Fig. 8) was estimated.

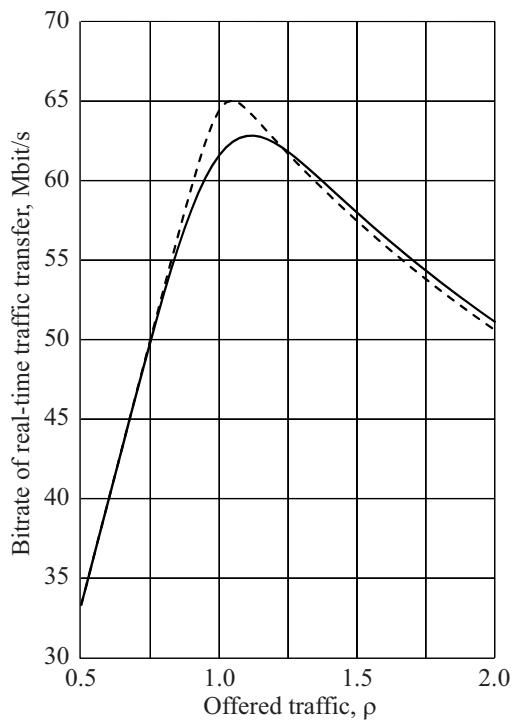
<sup>2</sup> Worst case scenario for servicing elastic data.



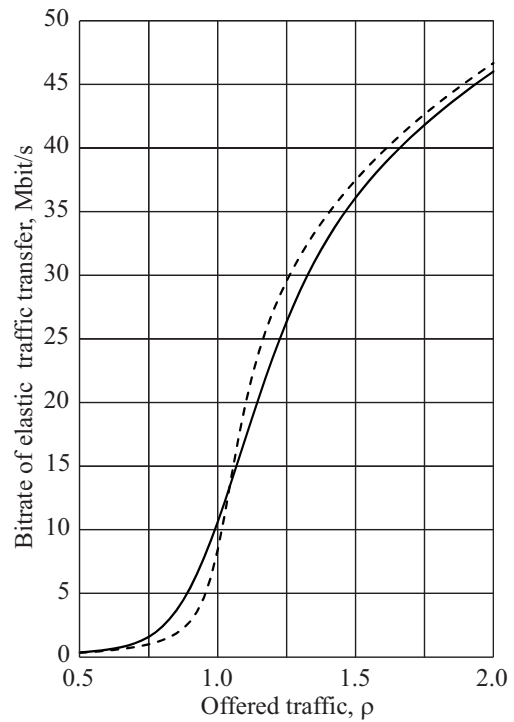
**Fig. 3.** Error in estimating the share of lost requests for servicing real-time traffic.



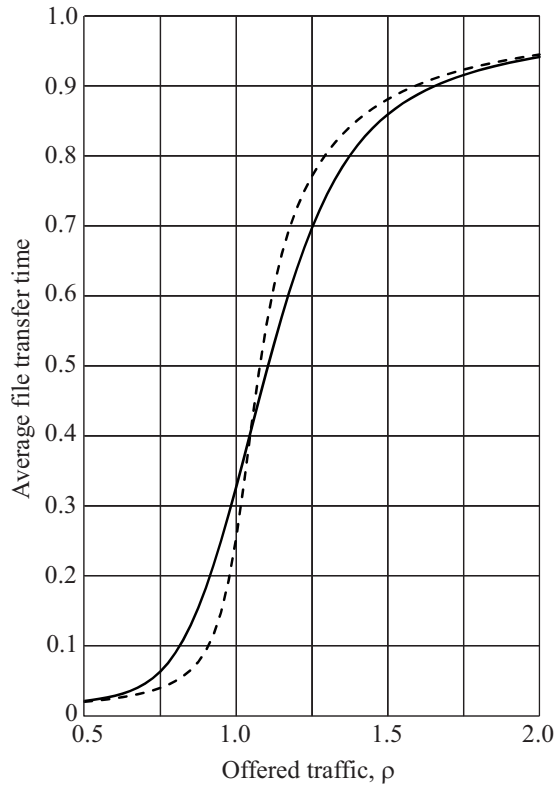
**Fig. 4.** Error in estimating the share of lost requests for servicing elastic data.



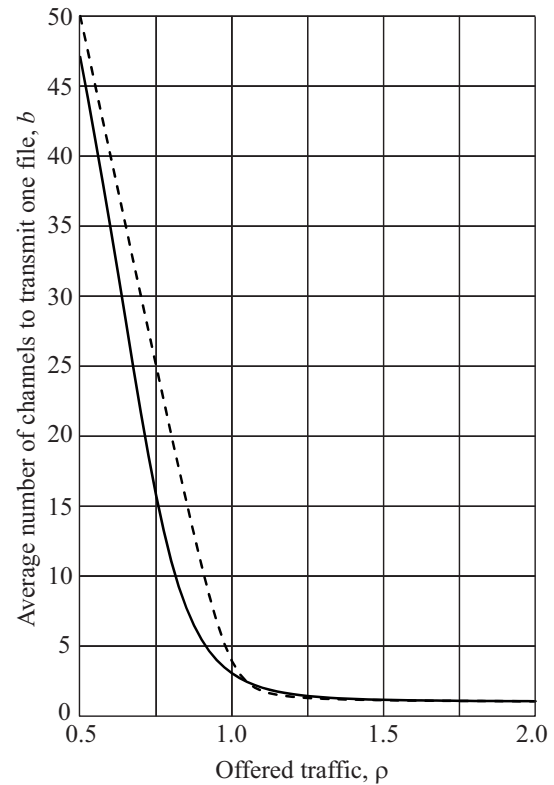
**Fig. 5.** Error in estimating the average bitrate used to service real-time traffic.



**Fig. 6.** Error in estimating the average bitrate used to service elastic data.



**Fig. 7.** Error in estimating the average file transfer time.



**Fig. 8.** Error in estimating the average number of channels used to transmit one file.

From the analysis of numerical data it follows that the obtained estimates have the following properties.

- (1) They have good accuracy, which increases in the area with low and high loads. These results will be examined in more detail below.
- (2) As the load on the channel increases, the values of the analyzed characteristics increase with the exception of the average value of the bitrate used to service the traffic of real-time services (see Fig. 5). This property is explained by the fact that as  $\rho$  increases, elastic data increasingly uses one channel to transfer a file, thereby displacing real-time traffic, which requires more channels for servicing.
- (3) As a rule, the obtained approximate values of the characteristics at low loads give a lower estimate for the characteristics of the access node model under study, and at a heavy load they provide an upper estimate.

As noted earlier, the accuracy of estimates increases in the region of large and small losses. Let us present numerical data that confirm this conclusion. Tables 1 and 2 show the values of the quality of service characteristics of real-time traffic and elastic data for the access node model with the values of the input parameters used in calculating the curves shown in Figs. 3–8 and listed at the beginning of this Section. The presented numerical data confirm the asymptotic properties of the estimates. The error in estimates quickly decreases with an increase in  $\rho$  (Table 1) and with its decrease (Table 2). In the region of large losses, asymptotic properties manifest themselves for all characteristics listed in (2). In the low loss region—only for elastic data transmission characteristics.

The presented numerical data showed that the developed method has good accuracy for values of input parameters that correspond to practical applications, in particular for values of small losses,

**Table 1.** Accurate and approximate estimation of the characteristics of the access node model under heavy load conditions

| $\rho$ | $\pi_2$ |         | $\pi_d$ |         | $h_d$  |         | $z_r$ |         | $z_d$ |         |
|--------|---------|---------|---------|---------|--------|---------|-------|---------|-------|---------|
|        | Exact   | Approx. | Exact   | Approx. | Exact  | Approx. | Exact | Approx. | Exact | Approx. |
| 1.00   | 0.1051  | 0.0496  | 0.0264  | 0.0085  | 0.3270 | 0.2546  | 61.58 | 64.42   | 10.61 | 8.41    |
| 1.25   | 0.3484  | 0.3624  | 0.0930  | 0.0806  | 0.6972 | 0.7713  | 61.81 | 61.68   | 26.34 | 29.54   |
| 1.50   | 0.5538  | 0.5716  | 0.1600  | 0.1493  | 0.8594 | 0.8811  | 57.99 | 57.46   | 36.09 | 37.47   |
| 1.75   | 0.6877  | 0.7021  | 0.2173  | 0.2081  | 0.9157 | 0.9233  | 54.33 | 53.80   | 41.81 | 42.65   |
| 2.00   | 0.7760  | 0.7873  | 0.2668  | 0.2591  | 0.9414 | 0.9449  | 51.12 | 50.60   | 46.01 | 46.67   |
| 2.50   | 0.8778  | 0.8848  | 0.3493  | 0.3437  | 0.9649 | 0.9661  | 45.77 | 45.30   | 52.32 | 52.84   |
| 3.00   | 0.9290  | 0.9334  | 0.4154  | 0.4112  | 0.9756 | 0.9762  | 41.53 | 41.12   | 57.03 | 57.47   |
| 4.00   | 0.9723  | 0.9742  | 0.5147  | 0.5122  | 0.9853 | 0.9855  | 35.28 | 34.96   | 63.76 | 64.09   |
| 5.00   | 0.9875  | 0.9884  | 0.5853  | 0.5838  | 0.9897 | 0.9898  | 30.88 | 30.63   | 68.39 | 68.66   |
| 7.50   | 0.9974  | 0.9975  | 0.6956  | 0.6951  | 0.9942 | 0.9943  | 23.90 | 23.76   | 75.65 | 75.80   |

**Table 2.** Accurate and approximate estimation of service characteristics of elastic traffic under light load conditions

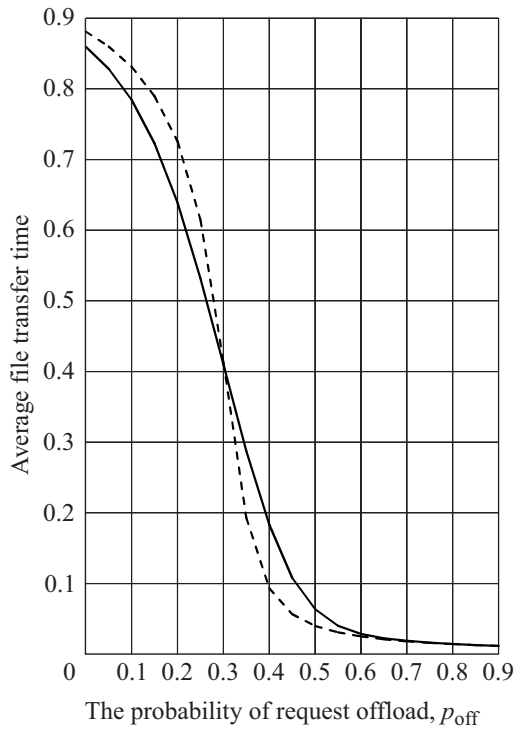
| $\rho$ | $h_d$   |         | $z_r$ |         | $z_d$   |         | $b$    |         |
|--------|---------|---------|-------|---------|---------|---------|--------|---------|
|        | Exact   | Approx. | Exact | Approx. | Exact   | Approx. | Exact  | Approx. |
| 0.500  | 0.02124 | 0.02000 | 33.33 | 33.33   | 0.35401 | 0.33333 | 47.078 | 50.000  |
| 0.400  | 0.01717 | 0.01666 | 26.66 | 26.66   | 0.22899 | 0.22222 | 58.225 | 60.000  |
| 0.300  | 0.01451 | 0.01428 | 20.00 | 20.00   | 0.14512 | 0.14285 | 68.908 | 70.000  |
| 0.200  | 0.01259 | 0.01250 | 13.33 | 13.33   | 0.08398 | 0.08333 | 79.378 | 80.000  |
| 0.150  | 0.01182 | 0.01176 | 10.00 | 10.00   | 0.05912 | 0.05882 | 84.565 | 85.000  |
| 0.100  | 0.01114 | 0.01111 | 6.666 | 6.666   | 0.03714 | 0.03703 | 89.728 | 90.000  |
| 0.075  | 0.01083 | 0.01081 | 5.000 | 5.000   | 0.02708 | 0.02702 | 92.302 | 92.500  |
| 0.050  | 0.01054 | 0.01052 | 3.333 | 3.333   | 0.01756 | 0.01754 | 94.871 | 95.000  |
| 0.025  | 0.01026 | 0.01025 | 1.666 | 1.666   | 0.00855 | 0.00854 | 97.437 | 97.500  |

where the problem of planning the volume of information transmission resource required by the load is solved, and for values of large losses, where the problem is solved estimating the volume of traffic offloaded to other access nodes or to other frequency bands in order to achieve specified QoS indicators. Let us give examples of solving the formulated problems.

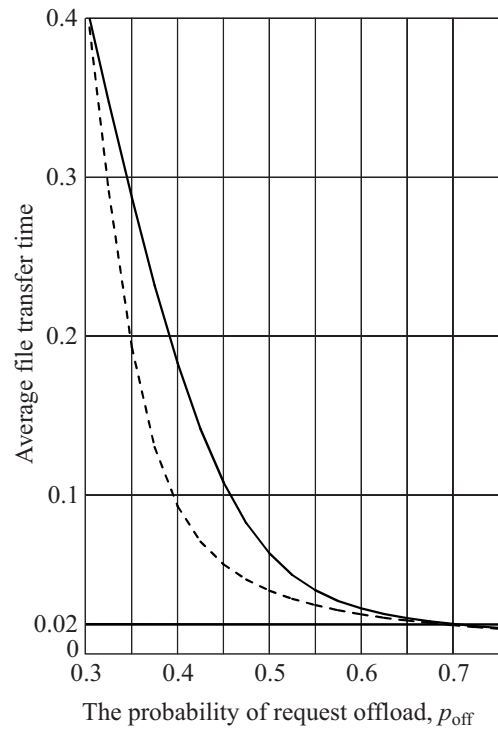
## 6. THE USAGE OF THE OBTAINED RESULTS IN PRACTICAL APPLICATIONS

Estimating the required resource of an access node for a given load and the maximum permissible load for a given resource value are similar problems and are solved by brute force. In the first case, starting from a certain initial value, the throughput of the access node increases until the required result in terms of quality of service is achieved; in the second case, in a situation of overload, the input flow of requests is reduced for the same purpose. Let's start by solving the second problem.

Let us change the description of the procedure for generating an input stream in the model of a multiservice access node under study. Let us assume that the access node services represent a Poisson flow of requests of intensity  $\lambda$ , divided into  $n + 1$  service categories. The first  $n$  categories represent requests for the transmission of real-time traffic. The last  $(n + 1)$ th category is requests for the transmission of elastic data traffic. Let us assume that  $\lambda = \lambda_1 + \dots + \lambda_n + \lambda_d$ . With probability  $p_k = \frac{\lambda_k}{\lambda}$ ,  $k = 1, \dots, n$  the request belongs to the  $k$ th category, requires  $c_k$  bit/s,  $k = 1, \dots, n$  and the resource occupies a random time, which has an exponential distribution with the parameter  $\mu_k$ . With probability  $p_d = \frac{\lambda_d}{\lambda}$  the request belongs to the category of elastic traffic and is served according to the corresponding rules introduced when describing the original node model (see Section 2). It is



**Fig. 9.** Dependence of the average file transfer time estimate on  $p_{off}$ .



**Fig. 10.** Estimation of the share of offloaded traffic to ensure specified QoS indicators.

clear that the considered change in the procedure for generating the input stream did not change the mathematical description of the access node model under study.

Assume that existing or scheduled traffic is causing the access node’s transmission capacity to become overloaded. This situation can be tracked with good accuracy using a method for assessing the characteristics of the quality of service for requests, based on the use of simplified equilibrium equations (see Section 4 and Table 1). Let’s assume that in the current situation it is impossible to simply increase the throughput of the access node. The difficulties that arise can be dealt with by redirecting (they also say offloading) some part of the input flow to other access nodes so that the quality of service characteristics of the remaining part of the traffic on the available resource do not exceed standard values [24]. Let us denote by  $p_{off}$  the probability of offloading an incoming request.

The formal formulation of the problem is as follows. It is necessary to determine the value  $p_{off}$ , which provides a given level of quality of service for the remaining requests in the form of fulfilling the inequalities

$$\pi = \max(\pi_1, \dots, \pi_n, \pi_d) < \pi^{norm}, \quad h_d < h_d^{norm}. \tag{17}$$

Here  $\pi^{norm}$  and  $h_d^{norm}$ —are the values required under the terms of the service agreement, respectively, of the maximum loss of requests and the average time of elastic data transmission.

Let us consider the access node model for the values of the input parameters used in calculating the contents of Table 1. The values of the characteristics of servicing requests for  $\rho = 1.5$  indicate that the node is in a state of overload and it is necessary to reduce the incoming flow of requests. Let’s take the following values of normative indicators:  $\pi^{norm} = 0.01$ ,  $h_d^{norm} = 0.2$ .

We solve the problem of estimating the probability of offloading in two stages. First, let’s find the value  $p_{off}$ , which ensures the fulfillment of the inequality (17) for the average file transfer time.

Note that the choice of  $h_d^{\text{norm}} = 0.2$  means that the average file transfer time is chosen 50 times less than the average file transfer time by one channel, and is 1.6 s. Figure 9 shows the change in  $h_d$  with increasing  $p_{\text{off}}$  in the range from 0 to 0.9. The exact value of the characteristic found as a result of solving the system of equilibrium equations (3) using the Gauss–Seidel iterative method, and its estimate found using simplified equilibrium equations (the curve is marked with a dotted line) are given. Note that the initial and final values of the range of changes  $p_{\text{off}}$  belong to the region where the characteristics estimates used are asymptotically accurate. This determines the high accuracy of calculating the characteristics of the original model, as evidenced by the curves shown in Fig. 9. A more detailed solution to the problem of determining  $p_{\text{off}}$  is shown in Fig. 10, where it is established that  $p_{\text{off}} \approx 0.7$ .

The second part of calculating the share of offloaded traffic is the need to satisfy the inequality  $\pi \leq 0.01$ . To establish this fact, we construct an upper bound for  $\pi$ , assuming that elastic files are transmitted using only one channel. This means that all requests arriving at the access node are serviced according to the traffic rules of real-time services. In this situation, the values of the characteristics can be calculated using recursion built for the multi-service Erlang model [7, 10, 22, 23]. The results obtained confirm the fulfillment of the inequality  $\pi \leq 0.01$ . Thus, the solution to the problem has been achieved for the offloading probability  $p_{\text{off}} \approx 0.7$ . The problem is solved in a similar way in a situation where only real-time service traffic or only elastic traffic is offloaded.

Estimation of the required resource of an access node for a given load is solved in the same sequence, so we will omit this part of the study.

## 7. CONCLUSION

A mathematical model of joint servicing of real-time traffic and elastic data traffic in a multi-service access node has been developed and studied. The arriving of requests for all service categories obeys the Poisson law; the duration of servicing requests for the transmission of real-time traffic has an exponential distribution, as do the volumes of transmitted elastic data. Elastic data is maintained in accordance with the provisions of the Processor Sharing discipline. An incoming request for real-time traffic transmission has priority in resource occupation, reducing, if necessary, the value of resource used by one file to a predetermined minimum value. A Markov process is constructed that describes the change in model states. Definitions of quality indicators for joint servicing of incoming requests for information services are provided. A system of statistical equilibrium equations is formed and its use for estimating the introduced characteristics is considered. A method for approximate calculation of characteristics is proposed, based on the construction of a system of simplified equilibrium equations. The use of the developed methods is shown to solve the problem of estimating the volume of traffic offloaded in an overload situation to other access nodes or to other frequency ranges in order to achieve specified QoS indicators and to solve the problem of planning the volume of required transmission resource of a multiservice access node. The results obtained are also applicable to the estimation of the required volume of resource of access nodes in satellite communication systems [25], data centers [26], and information centers [21].

Let us note the positive characteristics of the developed approximate method.

- The formation of simplified equilibrium equations is based on the use of local conservation laws that are fulfilled in individual macrostates of the model, selected in accordance with the structure of the state space used and the physical meaning of individual state components. Thus, it can be argued that this approach is based on the general principles of the functioning of communication system models described by Markov processes, which makes it possible to generalize the results obtained to other models of the formation of input flows of requests and resource allocation procedures, in particular to models with group arrival of requests and waiting.



- The obtained estimates of the quality of service characteristics of incoming requests have good accuracy for values of input parameters that correspond to practical applications, in particular, for values of small losses, where the problem of planning the volume of information transmission resource required by the load is solved, and for values of large losses, where the problem of estimating the volume of traffic offloaded to other access nodes or to other frequency ranges is solved in order to achieve specified QoS indicators.
- The computational algorithms used in the process of implementing the method are based on the use of simple recursive procedures.

## REFERENCES

1. *Study on Scenarios and Requirements for Next Generation Access Technologies*, 3GPP Technical Report (TR) 138.913 version 15.0.0 Release 15. 2018.
2. *Network Slice Selection Services*, 3GPP Technical Specification (TS) 129.531 version 15.5.0. Release 15. 2019.
3. Evans, J. and Filstis, C., *Deploying IP and MPLS QoS for Multiservice networks. Theory and Practice*, Morgan Kaufmann Publishers, Elsevier, 2007.
4. Ross, K.W., *Multiservice Loss Models for Broadband Telecommunication Networks*. London: Springer, 1995.
5. Iversen, V.B., *Teletraffic Engineering and Network Planning*, Technical University of Denmark, May 2010.
6. *Broadband Network Traffic. Performance Evaluation and Design of Broadband Multiservice Networks. Final Report of Action COST 242*, James Roberts, Ed., Lecture notes in computer sciences, Springer, 1996.
7. Stepanov, S.N., *Osnovy Teletrafika Mul'tiservisnykh Setei* (The Foundation of Multiservice Network Teletraffic), Moscow: Ego-Trends, 2010.
8. Bonald, T. and Feuillet, M., *Network Performance Analysis*, United Kingdom, London: Wiley, 2011.
9. Basharin, G.P., Gaidamaka, Yu.V., and Samouylov, K.E., Mathematical Theory of Teletraffic and Its Application to the Analysis of Multiservice Communication of Next Generation Networks, *Autom. Control Comput. Sci.*, 2013, vol. 47, no. 2, pp. 62–69.
10. Stepanov, S.N., *Teoriya teletrafika: kontseptsii, modeli, prilozheniya* (Teletraffic Theory: Concepts, Models, Applications), Moscow: Goryachaya Liniya-Telekom, 2015.
11. Stepanov, S.N. and Stepanov, M.S., Planning Transmission Resource at Joint Servicing of the Multiservice Real Time and Elastic Data Traffics, *Autom. Remote Control*, 2017, vol. 78, no. 11, pp. 2004–2015.
12. Shorgin, S., Samouylov, K., Gaidamaka, Y., Chukarin, A., Buturlin, I., and Begishev, V., Modeling Radio Resource Allocation Scheme with Fixed Transmission Zones for Multiservice M2M Communications in Wireless IoT Infrastructure, *Lecture Notes Comput. Sci.*, Springer, Cham., 2015, vol. 9012, pp. 473–483.
13. Begishev, V., Petrov, V., Samouylov, A., Moltchanov, D., Andreev, S., Koucheryavy, Y., and Samouylov, K., Resource Allocation and Sharing for Heterogeneous Data Collection over Conventional 3GPP LTE and Emerging NB-IoT Technologies, —it Comput. Commun., 2018, vol. 120, no. 2, pp.93–101.
14. Naumov, V., Gaidamaka, Yu., Yarkina, N., and Samouylov, K., *Matrix and Analytical Methods for Performance Analysis of Telecommunication Systems*, Springer Nature, Switzerland, 2021.
15. Kallos, G.A., Vassilakis, V.G., Moscholios, I.D., and Logothetis, M.D., Performance Modelling of W-CDMA Networks Supporting Elastic and Adaptive Traffic, *Proc. of 4 Int. Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs '06)*, Ilkley, 2006.
16. Gudkova, I.A. and Samouylov, K.E., Modelling a Radio Admission Control Scheme for Video Telephony Service in Wireless Networks, *Lecture Notes Comput. Sci.*, 2012, vol. 7469, pp. 208–215.
17. Bonald, T. and Virtamo, J., A Recursive Formula for Multirate Systems with Elastic Traffic, *IEEE Communicat. Lett.*, 2005, vol. 9, no. 8, pp. 753–755.

18. Stepanov, S.N. and Stepanov, M.S., Planning the Resource of Information Transmission for Connection Lines of Multiservice Hierarchical Access Networks, *Autom. Remote Control*, 2018, vol., 79. no. 8, pp. 1422–1433.
19. Stepanov, S.N. and Stepanov, M.S., Efficient Algorithm for Evaluating the Required Volume of Resource in Wireless Communication Systems under Joint Servicing of Heterogeneous Traffic for the Internet of Things, *Autom. Remote Control*, 2019, vol., 80, no. 11, pp. 2017–2032.
20. Stepanov, S.N. and Stepanov, M.S., Methods for Estimating the Required Volume of Resource for Multiservice Access Nodes, *Autom. Remote Control*, 2020, vol., 81, no. 12, pp. 2244–2261.
21. Stepanov, S.N., Markov Models with Retrials: The Calculation of Stationary Performance Measures Based on the Concept of Truncation, *Math. Comput. Model*, 1999, no. 30, pp. 207–228.
22. Stepanov, S. and Stepanov, M., Estimation of the Performance Measures of a Group of Servers Taking into Account Blocking and Call Repetition Before and After Server Occupation, *Mathematics*, 2021, vol. 9, no. 21.2811, pp. 1–24. <https://doi.org/10.3390/math9212811>
23. Fortet, R., and Grandjean, Ch., Congestion in a Loss System When Some Calls Want Several Devices Simultaneously, *Electr. Communicat.*, 1964, vol. 39, no. 4, pp. 513–526.
24. Kaufman, J.S., Blocking in a Shared Resource Environment, *IEEE Transact. Communicat.*, 1981, vol. 29, no. 1, pp. 1474–1481.
25. Chen, J., Chang, Z., Guo, X., Li, R., Han, Z., and Hamalainen, T., Resource Allocation and Computation Offloading for Multi-Access Edge Computing With Fronthaul and Backhaul Constraints, *IEEE Transactions on Vehicular Technology*, 2021, vol. 70, no. 8, pp. 8037–8049.
26. Stepanov, S.N., Andrabi, U.M., Stepanov, M.S. and Ndayikunda, J., Reservation Based Joint Servicing of Real Time and Batched Traffic in Inter Satellite Link, *Proc. of 2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 2020, pp. 1–5. <https://doi.org/10.1109/IEEECONF48371.2020.9078542>
27. Volkov, A.O., Korobkina, A.V. and Stepanov, S.N., Development of Model and Algorithms for Servicing Real-Time and Data Traffic in a Cloud Computing System *Proc. of 2022 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, pp. 1–6. 2022. <https://doi.org/10.1109/IEEECONF53456.2022.9744289>

*This paper was recommended for publication by A.I. Lyakhov, a member of the Editorial Board*