Access System Modeling by Queues with Compound Customers

Dr. Olga I. Bogoiavlenskaia

Department of Computer Science, University of Petrozavodsk

Lenin St., 33, Petrozavodsk, Republic of Karelia, 185640, Russia

E-mail: olbgvl@mainpgu.karelia.ru

Abstract

An approach to access system modeling by queues with a compound customer population is presented. The proposed queuing system is assumed to contain an infinite number of heterogeneous servers. The corresponding stochastic Markovian process is constructed, theorems on its properties are proved, and its stationary distribution is derived. The results are used to model the modem pool as an access server system. The estimation of parameters is based on empirical measurements of real modem pool traffic. The fitness tests allow to conclude that the model does not contradict measured data.

1 Introduction

The intensive growth of the data communication networks has greatly increased the importance of the analysis and planning problems related with the structural elements of these networks. A customer–access server (ACS) typically consists of communication channels and a device connecting the ACS to a server node of the network. Today the ACS's are widely used to provide a mass access to the information resources. The ACS's determine to a great extent the quality of service the providers are able to deliver to remote users.

Modern ACS's have some special features which must be taken into account in the course of their mathematical modeling. One of the features is the compound structure of the ACS's traffic. The traffic constitutes of transfers of separate independent blocks of data, the lengths and/or transmission times of which may have either random or determined values. These structural units are transferred trough the ACS channels one by one.

Another important feature of the ACS's is heterogeneity. In particular, there exist a diversity of parameters for the various system components. In addition, the same ACS can be used to serve various kinds of traffic. Sometimes separate parts of an ACS are predestined for some special types of traffic, or remote users may prefer certain channels of it.

Hence, the features important for the analysis of the ACS steady–state characteristics are the compound structure of the traffic and the heterogeneous nature of the ACS parameters. Besides, the ACS's are multichannel, and they are able to serve a number of remote calls simultaneously.

The composite structure of the ACS traffic can be treated embedding the heterogeneity of customers into the process model. The compound customers consist of a sequence of independent units. Hence, one can transform the customer behavior to the service behavior and consider the service process as consisting of a sequence of corresponding consecutive steps. The problem is to construct and analyze distributions which reflect this structure of the service process. Furthermore, there are also other characteristics of the ACS's, which are important; for example, the distribution of the number of busy channels or the loss probability of remote calls are of vital interest.

Thus, the analyst has to use mathematical schemes which allow obtaining information about the system's behavior at different levels of detail, and if necessary, make it possible to consider general and detailed characteristics all together.

There exists a big amount of results on multichannel queues with loss (for instance, results considering the Erlang scheme and its extensions [1, 2, 3, 4, 5]). Some general methods have also been worked out [6, 7, 8], and these allow to make qualitative and quantitative analysis of a wide class of queues. However, most of these methods treat the service process as a continuous–time process, and, due to the natural properties of the ACS traffic, the considered case requires a discrete approach in the analysis of the service process.

In the present paper we propose and investigate a queue which service scheme is constructed to satisfy the ACS features described above. We have constructed a corresponding Markovian discontinuous stochastic process, and we have proved theorems on its properties and on its steady–state characteristics. We have also considered the applicability of the results of the analysis in modeling the modem pool of the Federal Petrozavodsk RUNNet Node (FPRN).

The rest of the paper is organized as follows. The queuing system is defined in Section 2; this section also contains the proof of the process regularity. Section 3 contains a description of the balance equations system of the stochastic process. Theorems on solutions of the system are presented in Section 4. In Section 5 the results have been applied in a real ACS environment, the FPRN modem pool has been used as the experimental basis. Results of this modeling effort are presented. These results also include a fitness test for the distributions obtained from modeling and the measured data of the real traffic.

2 Description of the queuing system

The queuing system to be considered consists of several groups of servers. All servers in a group are of equal capacity. The number of servers in a group can be either finite or infinite. Let us define the system state vector $N = (n_1, \ldots, n_s)$, where s is the total number of groups and n_i, $(i = 1, \ldots, s)$ is the number of busy servers in the group i. If, at a time t, the system is in the state $(n_1, \ldots, n_s)$ then the probability of a new arrival into the group i by the time $t + \Delta t$ is $\lambda_i(n_1, \ldots, n_s)\Delta t + o(\Delta t)$. The probability of two or more arrivals during the period Δt is of the order $o(\Delta t)$. Let m_i be the number of servers in a group of finite size. An arriving customer chooses randomly the server of the group it will occupy, there are no priorities associated with the servers inside the groups.

The customers have a compound structure: each customer consists of a finite number of independent units. The number of units in a customer is denoted by the random variable ξ, with the distribution

$$P\{\xi = k\} = \phi_k \qquad k = 1, 2, \ldots . \tag{1}$$

Obviously,

$$\sum_{k=1}^{\infty} \phi_k = 1$$

and $\phi_0 = 0$. Let us assume that the expectation of ξ is finite, $\mathsf{E}\xi < \infty$. All units of a customer are served without delay one by one, in the order of arrival.

The single unit service time in the group i is denoted by the random variable η_i. Variables η_i are distributed exponentially

$$B_i(x) = 1 - e^{-\mu_i x} \quad i = 1, \dots, s. \tag{2}$$

A customer which arrives in a service group all servers busy is lost. Obviously, a loss of customer can happen only if the group consists of a finite number of servers. Let us denote a queuing system specified above with Σ^{∞}. Its scheme is presented in Figure 1. The special case of one service group with processor sharing discipline is investigated in [9].

We consider only the case where the distribution of the vector N does not depend on the distribution of ξ $(\{\phi_i\}_{i=1}^{\infty})$, so called invariance property. For the case where the number of servers is finite I. N. Kovalenko has formulated a necessary and sufficient condition of the invariance on the values $\lambda_i(N)$ [7]. The condition is formulated in the following way. For any set $\{n_1, \dots, n_s\}$ and for any two sets $\{i_1, \dots, i_{n_1+\ldots+n_s}\}$; $\{i'_1, \dots, i'_{n_1+\ldots+n_s}\}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$ the equality

$$\begin{aligned} &\prod_{l=1}^{n_1+\ldots+n_s} \lambda_{i_l}\Big(\sum_{m=1}^{l-1} \delta_{1,i_m}, \dots, \sum_{m=1}^{l-1} \delta_{s,i_m}\Big) = \\ &= \prod_{l=1}^{n_1+\ldots+n_s} \lambda_{j'_l}\Big(\sum_{m=1}^{l-1} \delta_{1,i'_m}, \dots, \sum_{m=1}^{l-1} \delta_{s,i'_m}\Big). \end{aligned} \tag{3}$$

must be true.

It is easy to see that the stochastic process generating $N(t) = (n_1(t), \dots, n_s(t))$ does not posses, in the general case, the Markovian property. Let us consider the following vector of a variable dimension

$$(N(t), R(t)) = (n_1(t), \dots, n_s(t), r_1^1(t), \dots, r_{n_1}^1(t), \dots, r_1^s(t), \dots, r_{n_s}^s(t)).$$

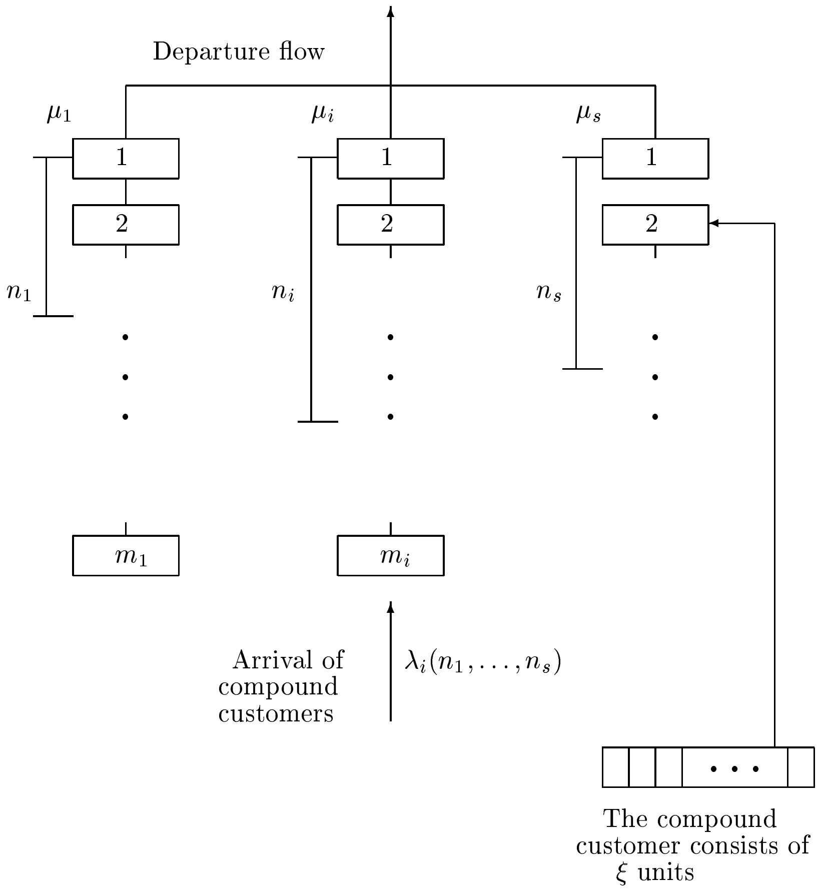

Figure 1. Queue Σ^∞

The dimension of this vector is the sum of two components: the number of groups and the total number of busy servers. Its coordinates $r^i_j(t)$ are the number of units which rest to the customer service completion and must be served (transmitted) by the busy servers at time t.

The enumeration of the busy servers in a group is arbitrary. The state space Q of the thus constructed random process $\theta(t) = \{(N(t), R(t))\}_{t>0}$ consists of variable dimension vectors and can be presented as follows

$$Q = \mathbb{N} \cup (\mathbb{N} \times \mathbb{N}) \cup (\mathbb{N} \times \mathbb{N} \times \mathbb{N}) \cup \ldots, \tag{4}$$

where $\mathbb{N}$ is the set of natural numbers. The set Q is numerable.

Let

$$p(N(t), R(t)) = p(n_1(t), \ldots, n_s(t), r^1_1(t), \ldots, r^1_{n_1}(t), \ldots, r^s_1(t), \ldots, r^s_{n_s}(t)) \tag{5}$$

be a function whose value is the probability of the system being at a time t in the state

$$(n_1(t), \ldots, n_s(t), r^1_1(t), \ldots, r^1_{n_1}(t), \ldots, r^s_1(t), \ldots, r^s_{n_s}(t)).$$

Generally, there exist discontinuous Markovian processes (in particular processes with a numerable set of states) which cannot be correctly described by the direct (second) equations of Kolmogorov [10, 11]. In cases where the minimal solution of the corresponding back equations is not purely stochastic, the existence of other, different solutions becomes possible. This includes the cases where a stochastic process makes an infinite number of transitions during a finite period of time (a so called finite accumulation point), or the process can leave the state space (for instance through reaching the infinity) with a positive probability.

According to [10], a Markovian process with a discrete state space is regular, if the moments of the transitions do not have a finite accumulation point with probability 1. Notice that for a regular Markovian process the solution of back Kolmogorov equations is purely stochastic and, hence, unique and satisfies the direct equations.

The sufficient condition of regularity for the process $\theta(t)$ is formulated in the following

Theorem 1 *Let us assume that the following series diverges*

$$\sum_{i=1}^{\infty} \frac{1}{L(i)} = \infty, \tag{6}$$

where

$$L(i) = \max_{|N|=i} \sum_{j=1}^{s} \lambda_j(N), \tag{7}$$

$|N| = \sum_{i=1}^{s} n_i$. *Then the random process* $\theta(t)$ *is regular.*

Proof. Let us define the random process $\theta'(t)$. The states of $\theta'(t)$ are set by the vector $Q(t)$ the dimension of which is equal to the dimension of the vector $R(t)$. The first s coordinates of Q(t) are equal to the coordinates of $N(t)$. The rest of the coordinates are defined as the vector $X(t)$; the values of its components correspond to the remaining service times of the customers.

The random process $\theta'(t)$ thus constructed in this way is a piecewise linear Markovian process [7, 12], and the rank of its states can be defined as

$$|N(t)| = \sum_{i=1}^{s} n_i(t).$$

One can show, using criteria based on the testing function methods [12], that if the condition (6) of the theorem is true the $\theta'(t)$ process is regular. Notice that the first s coordinates of the processes $Q(t)$ and $R(t)$ are equal for each fixed realization. Moreover, for each realization the transition moments of the process $\theta'(t)$ coincide with the transition moments of the process $\theta(t)$. In comparison with the process $\theta'(t)$ the process $\theta(t)$ has also 'additional' transitions; these appear because of the discrete interpretation of the service process. Let us consider the conditional probability of the following event. The random process $\theta(t)$ makes i transitions during a customer service time $t < T < \infty$, starting at the moment of its arrival. Then

$$p(i,T) = P(\xi = i/t \le T) = \frac{P(\xi = i \quad t \le T)}{B_j(T)} = \frac{\phi_i B_j^{i*}(T)}{B_j(T)}, \tag{8}$$

where $B_j^{i*}(t)$ is the i–fold convolution of exponential distributions with parameters μ_j $j = 1 \ldots s$, and $B_j(t)$ is the distribution function of the service time in the group j.

Now the formula (8) can be directed to the limit under $i \to \infty$. In accordance with our assumptions $0 < \mu_j < \infty$ and, hence

$$\lim_{i \to \infty} B_j^{i*}(T) = 0 \qquad \forall T < \infty. \tag{9}$$

Therefore

$$\lim_{i \to \infty} p(i, T) = 0 \tag{10}$$

is true. Thus, the regularity of the process $\theta'(t)$, which is ensured by the condition (6) of the theorem, leads to the regularity of the source process $\theta(t)$.
□

3 The balance equation system

Let us consider the limit

$$p(N, R) = \lim_{t \to \infty} p(N(t), R(t)). \tag{11}$$

To simplify notation we denote the state $(N, R_1, \ldots, R_s)$ with (N, R), where $N = (n_1, \ldots, n_s)$ is a vector the coordinates of which are equal to the number of busy servers in the groups, $R_i = (r_1^i, \ldots, r_{n_1}^i)\ i = 1, \ldots, s$ and $R = (R_1, \ldots, R_s)$. Let us also use below the unit vectors e_{ij}, with the dimensions n_i. The j–th coordinate of e_{ij} is equal to 1, and the rest $n_i - 1$ coordinates are equal to zero.

Other unit vectors e^i have the dimension s. Their i–th coordinate is equal to 1 and the rest are equal to zero.

Let us now consider possible ways of transitions for the random process $\theta(t)$. After a new customer arrival the system moves from the state $(N - e^i, R_1, \ldots, R_i - r_j^i e_{ij}, \ldots, R_s)$ into the state (N, R); this transition has the intensity $\lambda_i(N - e^i)\phi_{r_j^i}$. If a customer arrives into a group that has several servers each of which having to serve an equal number of units, the transitions corresponding to each of the servers do not differ. The system can also enter the state (N, R) from the states $(N + e^i, R_1, \ldots, 1, R_i, \ldots, R_s)$ and $(N, R_1, \ldots, R_i + e_{ij}, \ldots, R_s)$. Denote the number of least equal coordinates of the vector R_i as s_1^i, the number of coordinates bigger than the least but less than the rest as s_2^i, etc. It is obvious that $0 < s_j^i \leq n_i$, and the total number of s_j^i is not bigger than n_i. Let us denote with k_i the total number of the groups with equal coordinates r_j^i. If $k_i < n_i$, let us set $s_j^i = 0$ $j = k_i, \ldots, n_i$. Denote with a_j^i the values of the coordinates which belong to coordinate group numbered j. Under the taken notation the intensities of transitions $(N, R_1, \ldots, R_i + e_{ij}, \ldots, R_s) \to (N, R)$ and $(N + e^i, R_1, \ldots, 1, R_i, \ldots, R_s) \to (N, R)$ can be written as $(s_{j+1}^i + 1)\mu_i$ and $(s_1^i + 1)\mu_i$, correspondingly.

Figure 2 shows the part of the transition diagram in the case where the queuing system consists of two service groups. The first group has an infinite number of servers, and the second has two of them. The part of the transition diagram which is shown in the figure consists of all the ways through which the system can reach, in one transition, the state [1 (1) 2 (33)], that is $N = (12)$, $R_1 = (1), R_2 = (33)$.

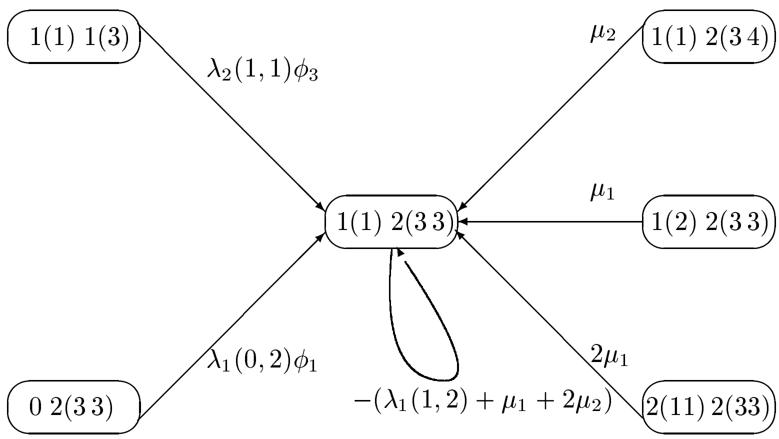

Figure 2. The part of transition diagram (Queue Σ^∞)

Let us formulate the balance equation system (BES) for the $p(N, R_1, \ldots, R_s)$. The zero state equation is

$$-p_0 \sum_{i=1}^{s} \lambda_i(\mathbb{O}) + \sum_{i=1}^{s} \mu_i p(\mathbb{O} + e^i, 1) = 0, \tag{12}$$

where the vector $(\mathbb{O})$ corresponds to the idle state, and in the state $(\mathbb{O} + e^i, 1)$ one server of the i–th group is serving the last unit of the customer.

For the states of $N \neq \mathbb{O}$ the balance equations are as follows:

$$\sum_{i=1}^{s}\sum_{j=1}^{k_i} \lambda_j(N-e^i)\phi_{r_j^i} p(N, R_1, \ldots, r_1^i, \ldots, r_{j-1}^i, r_{j+1}^i, \ldots, r_{n_i}^i, \ldots, R_s) -$$

$$-\sum_{i=1}^{s}(\lambda_i(N) + n_i\mu_i)p(N,R) +$$

$$+\sum_{i=1}^{s}\sum_{j=1}^{k_i}(s_{j+1}^i + 1)\mu_i p(N, R_1, \ldots, R_i + e_{ij}, \ldots, R_s) +$$

$$+\sum_{|U_i|>0}(s_1^i + 1)\mu_i p(N + e^i, R_1, \ldots, 1, R_i, \ldots, R_s) = 0. \quad (13)$$

The norming condition is considered together with the equations (12)—(13)

$$\sum_{Q} P(N, R_1, \ldots, R_s) = 1. \quad (14)$$

4 The solution of the balance equations and its analysis.

Let us start with the following theorem.

Theorem 2 *If the condition (3) is true then the system of equations (12)—(13) has a nontrivial solution*

$$p(N,R) = p_0 \frac{\Lambda(N)}{\mu_1^{n_1}\ldots\mu_s^{n_s}} \prod_{i=1}^{s} \frac{1}{s_1^i!\cdots s_{n_i}^i!} F_{a_1}^{s_1^i}\cdots F_{a_{n_i}}^{s_{n_i}^i}, \quad (15)$$

where $F_i = \sum\limits_{j=i}^{\infty}\phi_j$ *and*

$$\Lambda(N) = \prod_{l=1}^{n_1+\ldots+n_s} \lambda_{i_l}\left(\sum_{m=1}^{l-1}\delta_{1,i_m}, \ldots, \sum_{m=1}^{l-1}\delta_{s,i_m}\right).$$

Proof. We show that the expressions (15) transform the system (12)—(13) to identity. Denote

$$D(R_i) = \frac{1}{s_1^i! \cdots s_{n_i}^i!} F_{a_1^i}^{s_1^i} \cdots F_{a_{n_i}^i}^{s_{n_i}^i}. \tag{16}$$

After substitution of (15) in the equations (13) we obtain for all states $(N, R) \in Q \setminus \mathbb{O}$:

$$\sum_{i=1}^{s} \sum_{j=1}^{k_i} \lambda_i(N - e^i)\phi_{a_j^i} \rho(N - e^i) \frac{D(R_i - r_j^i e_{ij})}{D(R_i)} -$$

$$- \sum_{i=1}^{s} (\lambda_i(N) + n_i \mu_i)\rho(N) + \sum_{i=1}^{s} \sum_{j=1}^{k_i} \mu_i(s_{j+1}^i + 1)\rho(N) \frac{D(R_i + e_{ij})}{D(R_i)} +$$

$$+ \sum_{|U_i|>0}^{s} \mu_i(s_1^i + 1)\rho(N + e^i) \frac{D[(1, R_i)]}{D(R_i)} = 0, \tag{17}$$

where

$$\rho(N) = \frac{\Lambda(N)}{\mu_1^{n_1} \ldots \mu_s^{n_s}}. \tag{18}$$

After collecting the terms the expression (17) can be divided into two parts:

$$A = \sum_{i=1}^{s} \left[\sum_{j=1}^{k_i} \lambda_i(N - e^i)\phi_{a_j^i} \rho(N - e^i) \frac{D(R_i - r_j^i e_{ij})}{D(R_i)} + \right.$$

$$\left. + \sum_{j=1}^{k_i} \mu_i(s_{j+1}^i + 1)\rho(N) \frac{D(R_i + e_{ij})}{D(R_i)} - n_i \mu_i \rho(N) \right] \tag{19}$$

and

$$B = \sum_{i=1}^{s} \left[\mu_i(s_1^i + 1)\rho(N + e^i) \frac{D[(1, R_i)]}{D(R_i)} - \lambda_i(N)\rho(N) \right]. \tag{20}$$

Let us demonstrate that the expressions A and B are identical to zero. The following equalities are true because of the condition (3)

$$\lambda_i(N - e^i)\rho(N - e_i) = \mu_i\rho(N) \quad (21)$$

and

$$\lambda_i(N)\rho(N) = \mu_i\rho(N + e^i). \quad (22)$$

At the same time

$$F_{a^i_j}(s^i_{j+1} + 1)D(R_i + e_{ij}) = F_{a^i_{j+1}} s^i_j D(R_i), \quad (23)$$

and also

$$F_{a^i_j} D(R_i - r^i_j e_{ij}) = s^i_j D(R_i). \quad (24)$$

Therefore, the expression A can be transformed as follows

$$\sum_{i=1}^{s} \left[\mu_i\rho(N) \sum_{j=1}^{k_i} s^i_j \frac{\phi_{a^i_j} + F_{a^i_j+1}}{F_{a^i_j}} - n_i\mu_i\rho(N) \right] \equiv 0 \quad (25)$$

The identity (25) is evident, since $\sum\limits_{j=1}^{k} s^i_j = n_i$ and $F_{a^i_j+1} + \phi_{a^i_j} = F_{a^i_j}$. As for the expression B, it was noted above, that if $|U_i(N)| = 0$, then $\lambda_i(N) = 0$ and hence $B \equiv 0$.

□

Now consider the question of the existence of a steady–state distribution of the stochastic process $\theta(t)$. Let us define the set V. Its elements are the vectors $N = (n_1, \ldots, n_s)$, which correspond to the states of the queue Σ^∞, i.e. $0 \le n_i \le m_i$, if the number of servers in group i of queue Σ^∞ is finite, and $n_i \in \mathbb{N} \cup \mathbb{O}$ otherwise.

Theorem 3 *If the following series diverges*

$$\sum_{i=1}^{\infty} \frac{1}{L(i)} = \infty, \quad (26)$$

and the series

$$\sum_{N \in V} \rho(N) \frac{(\mathsf{E}\xi)^{|N|}}{n_1! \ldots n_s!} < \infty \quad (27)$$

converges, then the stochastic process $\theta(t)$ has a steady–state distribution.

Proof. Let us assume that the condition (26) of the theorem is true. Then, according to Theorem 1 the process $\theta(t)$ is regular, and the direct (second) Kolmogorov system of equations can be applied to the process. Let us demonstrate, that if the condition (27) is true, then unknown constant p_0 from (15) can be defined to transform the $\{p(N,R)\}_{(N,R\in Q)}$ into probabilities distribution. Let consider following sums

$$S(N) = p(\mathbb{O}) \sum_{R\in Q(|N|)} \frac{p(N,R)}{p_0}, \tag{28}$$

where $Q(k)$ is

$$Q(k) = \underbrace{\mathbb{N} \oplus \ldots \oplus \mathbb{N}}_{k}.$$

After having taken into account (15) the expression (28) can be transformed into the following form

$$S(N) = \sum_{R\in Q(|N|)} \rho(N) \prod_{i=1}^{s} D(R_i). \tag{29}$$

After collecting the terms of (29), $S(N)$ becomes

$$S(N) = \rho(N) \left(\sum_{R_1\in Q(n_1)} D(R_1) \right) \left(\sum_{R\in Q(|N|)} \prod_{i=2}^{s} D(R_i) \right). \tag{30}$$

Since

$$\sum_{i=1}^{s} s_i^1 = n_1, \tag{31}$$

the extracted factor yields

$$\sum_{R_1\in Q(n_1)} D(R_1) = \sum_{s_1+\ldots+s_k=n_1}^{s} \frac{1}{s_1!\ldots s_{k_1}!} F_{a_1}^{s_1} \ldots F_{a_{n_1}}^{s_{n_1}} =$$

$$= \frac{1}{n_1!} \left(\sum_{i=1}^{\infty} F_i \right) = \frac{1}{n_1!} (\mathsf{E}\xi)^{n_1} \tag{32}$$

using a polynomial formula. The upper indices s_j^i and a_j^i are missing in last transformation as those are equal to 1. One can apply the described transformations to $S(N)$ s times, and this yields the totally transformed expression

$$S(N) = \rho(N)\frac{(\mathsf{E}\xi)^{|N|}}{n_1!\dots n_s!}. \tag{33}$$

Hence, the series formed by the BES (12)–(13) solution is as follows

$$\sum_{(N,R)\in Q} p(N,R) = p_0[1 + \sum_{N\in V\setminus\mathbb{O}} S(N)] =$$
$$= p_0\left[1 + \sum_{N\in V\setminus\mathbb{O}} \rho(N)\frac{(\mathsf{E}\xi)^{|N|}}{n_1!\dots n_s!}\right] = 1. \tag{34}$$

The series

$$\sum_{N\in V} \rho(N)\frac{(\mathsf{E}\xi)^{|N|}}{n_1!\dots n_s!} \tag{35}$$

converges because of the condition (27). Therefore, based on the norming condition, the p_0 can be calculated as

$$p_0 = \left[1 + \sum_{N\in V\setminus\mathbb{O}} \rho(N)\frac{(\mathsf{E}\xi)^{|N|}}{n_1!\dots n_s!}\right]^{-1} \tag{36}$$

and

$$p(N) = p_0\rho(N)\frac{(\mathsf{E}\xi)^{|N|}}{n_1!\dots n_s!}. \tag{37}$$

Thus, the satisfaction of the theorem conditions implies the existence of a steady–state distribution for the random process $\theta(t)$; this distribution is given in the formulas(15) and (36).
□

5 The real Access Server modeling

This section is devoted to modeling of the modem pool on the base of results given above. The pool is Customer–Access server of the FPRN. The pool traffic measurement data was used for the estimation of the model parameters. The section also contains the figures of the fitness test, which was made using the estimation.

5.1 The modem pool of FPRN and its traffic measurements

The FPRN modem pool is used to serve the remote users' calls. The pool consists of a multichannel telephone located at a city telephone exchange. The telephone is connected by wirelines to the modems located at FPRN. The modems are linked with a server or a router. Usually there are a number of messages accumulated at the connection establishment time; all of them are transmitted during one session, in both directions. The connections are initiated by remote users. According to many protocols (including UUCP) several files are transmitted during one session.

The pool traffic measurements were undertaken to estimate the parameters of the Σ^{∞} queue as the model of the FPRN modem pool. First we present a set of general characteristics of the pool traffic (see Table 1).

One can see that the amount of data transferred from the FPRN is essentially greater than the amount transferred to the FPRN. The data exchange is characterized by a relatively small amount of data transferred per a session. A rather great number of incorrectly broken sessions demonstrates the low quality of the communication lines used.

The remote call arrival rate is one of the most important characteristic of the pool workload. It is natural to suppose that the rate changes during the day. Figure 3 visualizes the daily changes. It presents the average number of calls arrived in the pool during each day, avearaged over the period from 17.02.97 to 28.03.97.

This type of characterization is of big interest for modeling. Naturally the arrival and service rates are the main parameteres for this class of models. As the steady–state analysis assumes that the model parameters are constant, a characterization of the arrival rate allows to fix apropriate time windows for further analysis.

Table 1. The general characteristics of the modem pool traffic during 17.02.97–28.03.97

Number of calls	37259
Number of transmitted files	139222
Total amount of data	632.2 Mbytes
Amount of data transmitted to the FPRN	91.9 Mbytes
Amount of data transmitted from the FPRN	540.2 Mbytes
Number of incorrectly completed sessions	1542
Total number of zero groups	17461
Average size of a group	4.7 files
Average file length	4.761 Kbytes
Average file processing time	13.5 sec.

Since all calls are served by the same resource of the server, the service rate is likely to depend on the number of calls processed simultaneously.

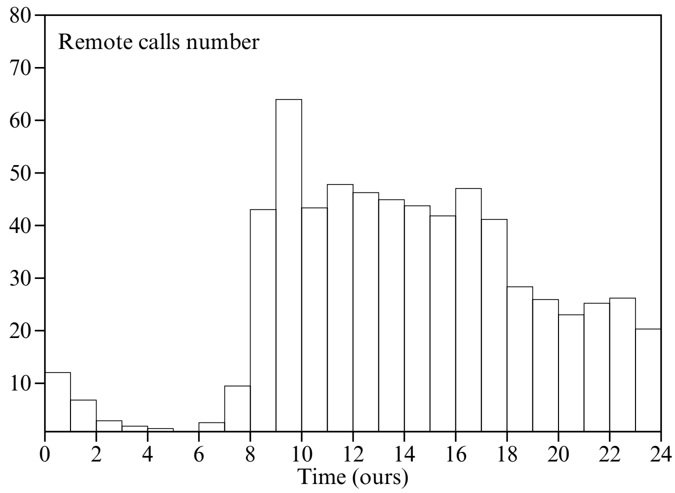

Figure 3. The average calls number came to the the modem pool in the each of 24 ours.

An additional analysis of the data has shown that in practice the service rate in the pool does not change (Figure 4).

Let us treat the whole transmission session as a customer, and each file transmitted, in any direction, as a unit of customer. The measured traffic data allows to calculate the relative frequencies characterizing the random variable ξ (see (1)), the number of files (i.e. units) transmitted during the session (Figure 5).

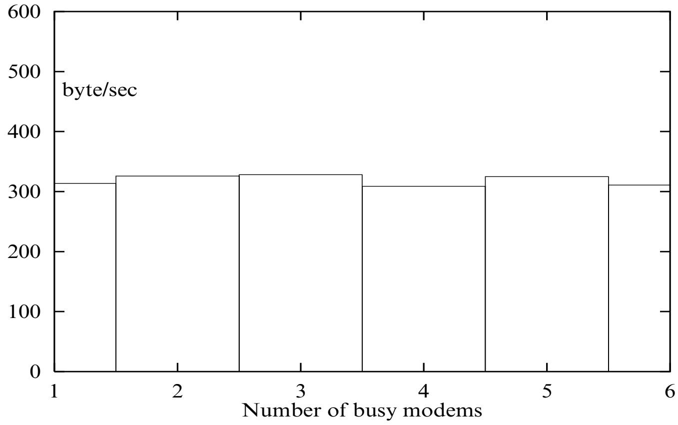

Figure 4. Service rate

According to the UUCP–protocol each message is transmitted by two files. Therefore groups containing an odd number of files correspond to incorrectly broken sessions. Another specific feature is that remote users often establish a connection just to check whether there is any incoming mail. In the case of mail absence an empty group is formed: it does not contain any files. Let us call it a zero group. However, the transmission time of the zero groups is positive, as the FPRN server and the remote node always execute the start and completion procedures. Let us consider the procedures as one additional unit. Thus, the total number of units is equal to the number of files transmitted in the session plus one.

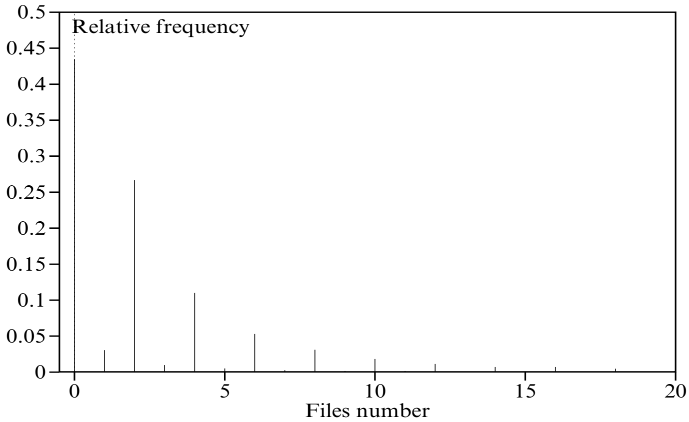

Figure 5. The relative frequency of units number inside the customer

5.2 The model of the modem pool and figures of the fitness test.

Let us now consider the issues of the modem pool model development and its parameter estimation. The modem pool is interpreted as a Σ^∞ queue. Let $\lambda_i(N) = \lambda$, $\mu_i = \mu$. According to the results obtained above the steady–state characteristics of the model are determined by the following parameters: λ—the Poisson arrival rate of the customer (call) stream, μ—the unit service rate, and $\{\phi_i\}_{i=1}^{\infty}$—the distribution of the unit quantity. The estimates of the parameters can be obtained from the traffic measurement data. The system workload can be defined as

$$\rho = \frac{\lambda}{\mu}\mathsf{E}\xi.$$

Using the formulas (37) and (36) one can find out the distribution of the number of busy modems.

At the same time one can obtain the relative frequencies of the number of busy modems $\{\tilde{p}_n\}_{n=0}^{m}$, where m is the number of modems in the pool

$$\tilde{p}_n = \frac{t_n}{t}. \tag{38}$$

Here t is the total length of the observation period, and t_n is the sum of intervals (inside the observation period) during which n modems were simultaneously busy. It is obvious, that during the time window, chosen as the period t, the system's workload should not change significantly.

During the day one can extract two such windows (see Figure 3). The first lasts from 9.00 till 10.00, the second lasts from 10.00 till 18.00. On the base of measured data we have calculated, for each of the two time windows, the relative frequencies of the number of busy modems. As a model we have used the distribution calculated by the formula

$$p_n = \frac{\rho^i}{i!} e^{-\rho}, \tag{39}$$

which is special case of (37). We have also used means of relative frequencies (by total period of measurements) to compare the theoretical and empirical values. The values received by formula (39) and from measurement data (38) are presented in the tables 2 and 3. We have set $t = 60$ to calculate the relative frequencies for the first time window and $t = 480$ for the second one. The parameters of the modem pool model were estimated from the sample and have the following values $\lambda = 0.018$, $\mu = 0.068$ $\mathsf{E}\xi = 4.67$, $\rho = 1.245$, for the period of 9.00–10.00, and $\lambda = 0.0125$, $\mu = 0.067$ $\mathsf{E}\xi = 3.98$, $\rho = 0.726$ for the period of 10.00–18.00.

Table 2. Comparative data of the number of busy channels for the first period of 9.00–10.00

number of channels	0	1	2	3	4	5
theoretical values	0.288	0.358	0.223	0.093	0.029	0.007
measurements	0.292	0.351	0.226	0.094	0.028	0.007

Table 3. Comparative data of the number of busy channels for the second period of 10.00–18.00

number of channels	0	1	2	3	4	5
theoretical values	0.483	0.352	0.128	0.031	0.006	0.0008
measurements	0.480	0.358	0.127	0.032	0.007	0.0005

As the tables 2 and 3 show, the obtained values start to differ at the third digit after the decimal comma. This accuracy is satisfactory for all practical needs. The Figures 6 and 7 present diagrams of relative frequencies $\tilde{p}_n$ (solid line) and probabilities p_n (dot line) obtained by the modeling.

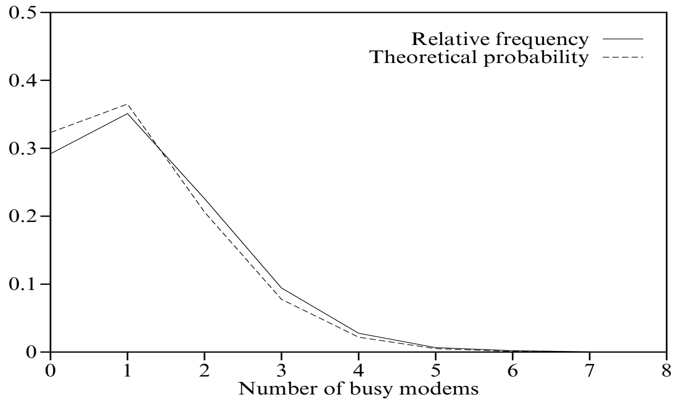

Figure 6. Relative frequencies and theoretical probabilities of the number of busy channels. Period of 9.00–10.00.

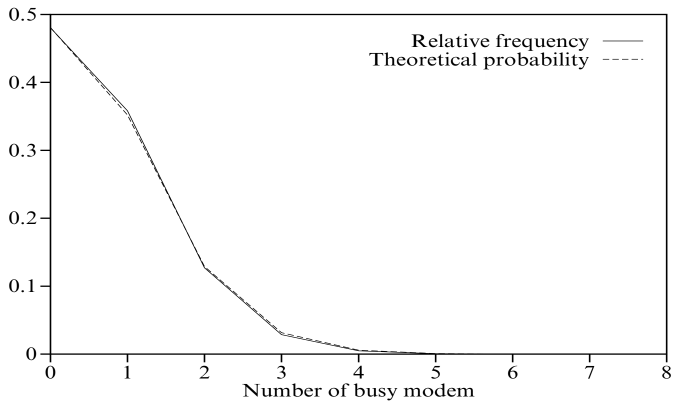

Figure 7. Relative frequencies and theoretical probabilities of the number of busy channels. Period of 10.00–18.00.

We have calculated the statistics of χ^2 and the Kolmogorov–Smirnov criterions to test the fitness of the theoretical and empirical distributions. The obtained values of the statistics are given in Table 4.

Table 4. Statistics

criteria	9.00–10.00	10.00–18.00
χ^2	0.709	7.443
KS	0.0072	0.0040

The test statistics do not reach 5% level of significance. At the level of significance the critical value of the χ^2 criterion is 11.07, and the critical value of the Kolmogorov–Smirnov criterion is 0.032 for the first time window and 0.011 for the second one. Thus the conclusion is that the results obtained in the FPRN modem pool modeling do not contradict with measured data.

Our model allows to present the customer call loss probability (Figure 8) as a function of three parameters. These are the arrival rate of

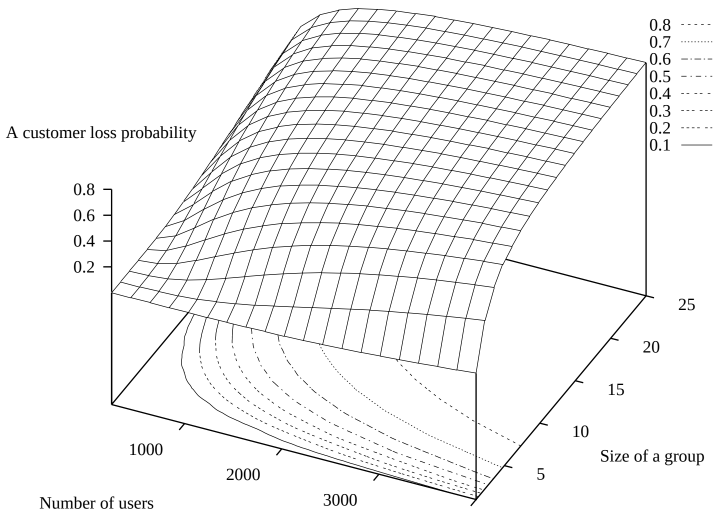

Figure 8. A customer loss probability

remote calls, the message group size, and the service rate, which depends on the hardware and software characteristics of the server and of the modem pool. The obtained dependence allows to plan when to update and how to update the modem pool to keep its performance at the required level.

6 Summary

We have constructed and analyzed a queuing model that reflects the key features of a modern access server. The model is a heterogeneous infinite–server queue with a compound structure for customer arrivals. To investigate the constructed queue we defined a discontinuous Markovian stochastic process. The sufficient conditions for the process regularity and the existence of the steady–state distribution are formulated and the corresponding theorems are proved. We have also obtained the steady–state distribution of the number of busy servers in groups of servers, and the steady–state joint distribution of customer units to be served and of the number of busy servers in the server groups. The distributions are given in explicit analytical product forms.

The results have been tried in a real access–server environment using the FPRN modem pool as an experimental basis. Results of the modeling are presented, including a fitness test for the distributions. The dependence of the call loss probability on the arrival rate of remote calls, on the message group size, and on the service rate has been given. The dependence model allows to plan when to update and how to update the modem pool to keep its performance at the required level.

Acknowledgments

This project was appreciably contributed during my research visit to the Department of Computer Science of the University of Helsinki. I express my deep gratitude to the former head of the department prof. M. Tienari for supporting the visit and to the members of the Performance Analysis group with prof. T. Alanko for fruitful consideration of the research and personally to T. Alanko for his editorial efforts.

I am grateful to prof. V.V. Kalashnikov for helpful discussions of the presented results.

I am also thankful to S. Matsko, the administrator of the FPRN, for detailed consultations, and to my graduate student O. Shtchoukin for the processing of measurement data.

References

[1] *van Doorn E. A., Panken F. J. M.* Blocking Probabilities In A Loss System with Arrivals in Geometrically Distributed Batches and Heterogeneous Service Requirements // IEEE / ACM Transactions on Networking.— 1993.— Vol.1, №6.— P. 664–667.

[2] *Gazdzicki P., Lambadaris I., Mazumdar R. R.* Blocking Probabilities for Large Multirate Erlang Loss Systems // Adv. Appl. Prob.— 1993.— Vol. 25.— P. 997–1009.

[3] *Kaufman J. S., Rege K. M.* Blocking in a Shared Resource Environment with Batched Poisson Arrival Processes // Performance Evaluation.— 1996.— Vol. 24.— P. 249–263.

[4] *Labourdette J.-F. P., Hart G. W.* Blocking Probabilities in Multitraffic Loss Systems:Insensitivity, Asymptotic Behavior, and Approximations // IEEE Transactions On Communications.— 1992.— Vol.40, №8.— P. 1355–1366.

[5] *Morrison J. A.* Blocking Probabilities for Multiple Class Batched Poisson Arrivals to a Shared Resource // Performance Evaluation.— 1996.— Vol.25.— P. 131–150.

[6] *Sevastyianov B. A.* An ergodic theorem for Markov process and its application to telephone systems with refusals // Theory Probab. Applic. — 1957.— 2.— P.104–112

[7] *Gnedenko B. V., Kovalenko I. N.* Queuing theory. — Jerusalem: Israel Program for Scientific Translation, 1968

[8] *Burman D. Y.* Insensitivity in queuing systems. // Adv. Appl. Prob. — 1981.— Vol 13. — P.846–859.

[9] *Bogoiavlenskaia O. I.* The Steady State Distribution of the Queue Length for a Queue with Bulk Arrival and Processor Sharing Discipline // Probabilistic Methods in Discrete Mathematics. Proceedings of the 4th International Conference, Petrozavodsk.— Utrecht: VSP, 1996.— P. 131–137.

[10] *Gihman I. L., Skorohod S. K.* Theory of Stochastic Process.— Berlin: Springer, 1975.

[11] *Feller W.* An introduction to Probability Theory, Vol.2. — New-York: Wiley, 1965.

[12] *Kalashnikov V. V.* Mathematical Methods in Queuing Theory. — Dordreecht: Kluwer Academic Publisher, 1994