mathematics

MDPI

Article

Analysis of a Semi-Open Queuing Network with a State Dependent Marked Markovian Arrival Process, Customers Retrials and Impatience

Chesoong Kim [1,*], Sergey Dudin [2,3], Alexander Dudin [2,3] and Konstantin Samouylov [3]

[1] Department of Industrial Engineering, Sangji University, Wonju, Kangwon 26339, Korea
[2] Department of Applied Mathematics and Computer Science, Belarusian State University, 4, Nezavisimosti Ave., 220030 Minsk, Belarus
[3] Department of Applied Informatics and Probability Theory, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, 117198 Moscow, Russia
* Correspondence: dowoo@sangji.ac.kr

Received: 10 July 2019; Accepted: 5 August 2019; Published: 7 August 2019

Abstract: We consider a queuing network with single-server nodes and heterogeneous customers. The number of customers, which can obtain service simultaneously, is restricted. Customers that cannot be admitted to the network upon arrival make repeated attempts to obtain service. The service time at the nodes is exponentially distributed. After service completion at a node, the serviced customer can transit to another node or leave the network forever. The main features of the model are the mutual dependence of processes of customer arrivals and retrials and the impatience and non-persistence of customers. Dynamics of the network are described by a multidimensional Markov chain with infinite state space, state inhomogeneous behavior and special structure of the infinitesimal generator. The explicit form of the generator is derived. An effective algorithm for computing the stationary distribution of this chain is recommended. The expressions for computation of the key performance measures of the network are given. Numerical results illustrating the importance of the account of the mentioned features of the model are presented. The model can be useful for capacity planning, performance evaluation and optimization of various wireless telecommunication networks, transportation and manufacturing systems.

Keywords: queuing network; retrials; state-dependent marked Markovian arrival process; wireless telecommunication networks

1. Introduction

The theory of queuing networks has a wide range of applications for modeling various real-world systems including telecommunication and logistic networks, health care, public transportation, production and manufacturing systems, see, for example, References [1–4] and so forth.

The queuing networks with homogeneous customers are usually classified (see, e.g., Reference [5]) into three categories: open networks where customers arrive from the outside and depart from the network after receiving service; closed networks where the number of customers circulating in the network is constant; and semi-open networks where customers arrive from the outside and depart from the network but only the finite number of customers can stay inside the network at any time.

In our paper, we deal with a queuing network which belongs to a relatively new category of semi-open queuing networks that recently were applied for the analysis of various real-world systems. For the review of the state of the art and the references see, for example, References [5–11].

Salient features of the considered in our paper model are the following:

- **Account of retrial phenomenon**. We assume that at most N customers can receive service in the network simultaneously. If a primary customer (customer arriving from the outside) arrives when N customers receive service, the customer joins the so-called orbit having an infinite capacity from which he/she retries to obtain access to the network after a random amount of time. A customer from the orbit can enter the network if the number of customers receiving service at the retrial moment is less than N. The theory of retrial *queues* is essentially less developed than the theory of queues with losses or buffers due to the higher complexity of the processes defining the behavior of the system, for references see, for example, References [12,13]. To the best of our knowledge, the results devoted to the exact analysis of the retrial *queuing networks*, which are more general than the tandem queues, are absent in the existing literature except the recent paper, Reference [14], which is briefly cited below. Here and thereafter, we only occasionally cite the papers where the corresponding networks are analyzed by means of approximations rather than exact solutions, see, for example, Reference [15].
- **More complex customers arrival process**. We consider a more general and realistic arrival process than those known in the literature. The overwhelming majority of the existing queuing networks consider the input as a stationary Poisson process. However, in most real-world systems, the input rate is time dependent. It is already well-recognized in the literature that the flows in modern telecommunication systems and networks are bursty and inadequately modelled by a stationary Poisson process. Instead, the so-called Markovian arrival process (MAP), see, for example, References [16,17], is a much better choice for description of real-world arrival processes which exhibit variation of the instantaneous arrival rate and correlation of inter-arrival times. Surveys on queuing *systems* with the MAP can be found in References [16,18]. Concerning the queuing *networks*, there are only a few papers on this topic, see, for example, Reference [19]. This paper deals with an approximation of the queuing networks with the MAP and phase-type distribution of service times. Exact results are known only for a specific kind of queuing networks, namely, the tandem queues, with the MAP, see, for example, in References [20–23]. Note that tandem queues considered in References [22,23] take into account retrials of customers. In this paper, we consider more general than the MAP-marked Markovian arrival process ($MMAP$), see, for example, Reference [24]. This flow is heterogeneous and has several types of customers. Type defines the node of the network at which the customer arrives.
- **More complex process of customers retrials and dependence of arrivals of primary customers and retrials**. Traditionally, it is assumed in the literature that the processes of customers arrivals and retrials are independent. More, it is usually assumed (except the special case when it is suggested that only one customer from the orbit can make retrials) that, under a fixed number of customers in orbit, the inter-retrial times have the exponential distribution with a fixed parameter. We can refer only to Reference [25] where the $BMAP/SM/1$ retrial queue is studied under the assumption that the intensities of the individual retrials are modulated by a continuous-time Markov chain. This Markov chain is independent of the underlying Markov chain of the $BMAP$ arrival process of primary customers. Such independence is not very realistic in real-world systems because when the arrival rate of primary customers fluctuates depending on the time of a day or a night or due to some external factors, it is very likely that the rate of retrials also depends on the time and the same external factors. In our paper, we consider the model with dependent processes of the arrival of primary customers and customers from the orbit.
- **Account of possible impatience of customers during staying in the orbit and waiting times in the nodes as well as non-persistence of customers staying in the orbit.** Impatience of customers, that is, a possibility of abandonment during the waiting time after some period of waiting and non-persistence of customers staying in the orbit, that is, a possibility to renege from the orbit after any unsuccessful retrial, are typical for many real-world systems and networks. Therefore, they should be taken into account during performance evaluation and capacity planning.

Semi-open queuing networks with MAP arrival process are analysed in References [7,10]. However, the retrial phenomenon is not taken into account in those papers. The model considered in Reference [7] is simpler than the one studied in Reference [10] because it is assumed in Reference [7] that the arrival flow is described by the $MMPP$ (Markov Modulated Poisson Process), which is a particular case of the MAP considered in Reference [10] and the network has a linear topology, that is, a type of network topology in which each node is connected one after the other in a sequential chain. An arbitrary topology is supposed in Reference [10]. Exact algorithmic results are obtained in Reference [7] only for the case of a tandem consisting of two stations. In the case of a larger number of stations, approximate results are obtained. The analysis presented in Reference [10] is exact algorithmic for an arbitrary finite number of nodes, network topology and customer routing. In the recent paper, Reference [14], the model from Reference [10] is generalized to the case when there is no input buffer in the network and a customer arriving to the network when N customers receive service moves to the orbit and makes the retrials to obtain service. A classical retrial strategy is applied. This strategy assumes that the total retrial rate is proportional to the number of customers staying in the orbit.

In the model considered in this paper, we significantly extend the results of References [10,14] to the networks with state dependent processes of arrival of primary customers and retrials, account of impatience and non-persistence of customers in the orbit and customers impatience in the buffers of the nodes of the network. Considered mutual dependence of the flows of primary customers and retrials is typical for many real-world systems but is not studied in the existing literature even for simple queuing systems, not to mention queuing networks.

Examples of potential applications of the obtained results to the analysis of real-world systems can be found, for example, in References [7,10].

The rest of the paper has the following structure. The mathematical model of the queuing network is completely described in Section 2. The multidimensional stochastic process describing the dynamics of the orbit and nodes of the network is presented in Section 3. This process is a Markov chain with one denumerable and several finite space components. It belongs to the class of level-dependent Quasi-Birth-and-Death processes. The generator of this Markov chain is presented. The problem of computing the stationary distribution of the chain is discussed in this section. In Section 4, expressions for the main performance indicators of the network are presented. Illustrative numerical examples giving insights into behavior of the network are provided in Section 5. Section 6 contains some concluding remarks.

2. Mathematical Model

A semi-open queuing network consists of L nodes which are single-server queuing systems with finite buffers. The structure of the network is presented on Figure 1. We assume that the capacity of the buffer at the lth node is equal to $N_l - 1$, $1 \le N_l \le \infty$. The capacity of the network, that is, the maximum number of customers, which can be processed in the network at the same time, is N, $1 \le N < \infty$. We assume that $N \le \min\limits_{l \in \{1,2,\ldots,L\}} N_l$. This guarantees that customers admitted to the network are not lost during processing in the network due to a buffer overflow.

The admission of arriving (primary) customers to the network is implemented as follows. If the capacity of the network is not exhausted on customer's arrival, that is, the number of customers in the network is less than N, the customer enters the service. Otherwise, the customer moves to the orbit and retries for service later. The capacity of the orbit is assumed to be infinite. The process of generation of primary customers and retrials of customers from the orbit is defined by the underlying process ψ_t, $t \ge 0$, with a finite state space $\{0, 1, \ldots, W\}$. The intensities of transitions of the process ψ_t depend on the current number i_t of customers in the orbit, $i_t \ge 0$.

When $i_t = 0$, that is, the orbit is empty, the intensities of transitions are given by the set of square matrices D_l, $l = \overline{1, L}$, of size $\bar{W} = W + 1$ and by the non-diagonal entries of the matrix D_0. Here and thereafter, the notation $l = \overline{1, L}$ means that the parameter l admits the values from the set $\{1, 2, \ldots, L\}$.

The diagonal entries of the matrix D_0 are negative such as the relation $\sum_{l=0}^{L} D_l \mathbf{e} = \mathbf{0}^T$ holds true where $\mathbf{e} = (1, 1, \ldots, 1)^T$, $\mathbf{0} = (0, 0, \ldots, 0)$ and ^{T} is the vector transpose symbol. Transitions, the intensities of which are given by the non-diagonal entries of the matrix D_0, do not cause the generation of primary customers. Transitions, the intensities of which are given by the entries of the matrix D_l, cause the generation of a type-l customer, $l = \overline{1, L}$. If the capacity of the network is not exhausted at the moment of type-l customer generation, this customer enters the lth node of the network. If the server of this node is idle, the customer starts service. Otherwise, the customer enters the buffer of this node. By comparing the presented description of the arrival process when the orbit is idle with the definition given in Reference [24] we can conclude that this process coincides with the $MMAP$.

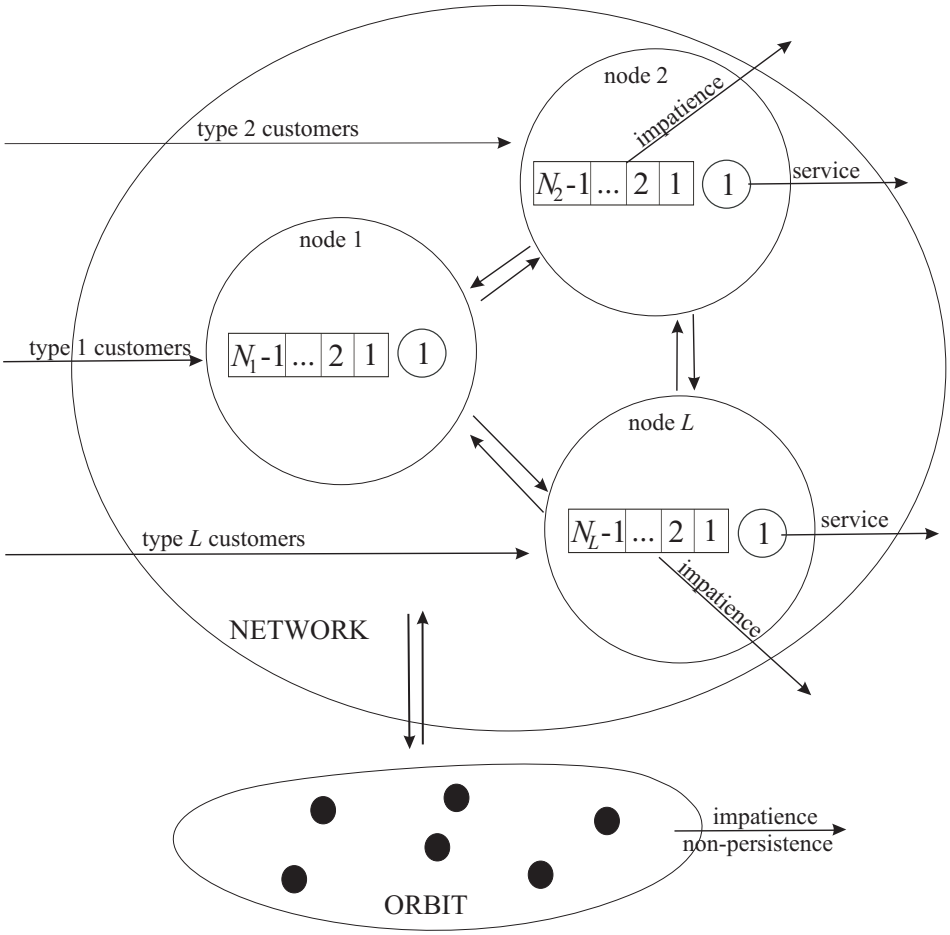

Figure 1. Retrial queuing network under study.

Under any fixed number i, $i > 0$, of customers in the orbit, the transitions of the process ψ_t can cause either the arrival of a primary customer or a retrial from the orbit. These transitions are given by a set of the same matrices D_l, $l = \overline{0, L}$, and the matrix $D^{(i)}$. As above, the intensities of transitions, which are given by the entries of the matrix D_l, cause the generation of a primary type-l, $l = \overline{1, L}$, customer. Transitions, the intensities of which are given by the entries of the matrix $D^{(i)}$, cause the generation of a repeated attempt from the orbit. The non-diagonal entries of the matrix $D_0^{(i)} = D_0 - \mathrm{diag}\{D^{(i)}\mathbf{e}\}$ define the intensities of transitions which do not cause the generation of primary customers or customers from the orbit. Here, $\mathrm{diag}\{\mathbf{b}\}$ denotes the diagonal matrix with the diagonal entries given by the entries of the vector $\mathbf{b}$. The matrix $\sum_{l=1}^{L} D_l + D^{(i)} + D_0^{(i)}$ is the generator. Let us denote by $\boldsymbol{\theta}^{(i)}$ the left stochastic eigenvector of this matrix. Conditional on the fact that i customers are staying in the orbit, the arrival rate of type l primary customers

is equal to $\lambda_l^{(i)} = \boldsymbol{\theta}^{(i)} D_l \mathbf{e}$, $l = \overline{1, L}$. The conditional arrival rate of customers from the orbit is equal to $\lambda^{(i)} = \boldsymbol{\theta}^{(i)} D^{(i)} \mathbf{e}$, $i > 0$.

We assume that customers staying in the orbit are impatient and non-persistent. Impatience means that after a time interval, whose duration is exponentially distributed with parameter γ, $\gamma > 0$, the customer from the orbit leaves the network without service and is lost. The non-persistence means that after unsuccessful retrial attempts, the customer leaves the network forever with probability h, $0 \leq h \leq 1$, and returns to the orbit with the complementary probability. Analogously, the customers waiting in the buffers of the network are impatient. After the time interval, whose duration is exponentially distributed with the parameter β_l, a customer waiting in the buffer of the l-th node leaves the node and departs from the network without service.

A retrial customer, which is admitted for service, moves to the l-th node with probability p_l, $l = \overline{1, L}$, $\sum\limits_{l=1}^{L} p_l = 1$.

We assume that the service time at the node l has an exponential distribution with parameter μ_l, $0 < \mu_l < \infty$, $l = \overline{1, L}$. When this time expires, the serviced customer can move to another node or leave the network forever. The probability of the transition of the serviced customer from the l-th node to the l'-th node is defined as $q_{l,l'}$, $l' = \overline{1, L}$, $l' \neq l$. The probability of leaving the network after service in the l-th node is $q_{l,0}$, $q_{l,0} = 1 - \sum\limits_{l'=1, l' \neq l}^{L} q_{l,l'}$.

3. Process of System States

It is easy to see that the dynamics of the considered network is described by the continuous-time Markov chain

$$\xi_t = \{i_t, n_t, \psi_t, n_t^{(1)}, \ldots, n_t^{(L)}\}, \; t \geq 0,$$

where, at the moment t, $t \geq 0$,

- i_t is the number of customers in the orbit, $i_t \geq 0$;
- n_t is the total number of customers in the network, $n_t = \overline{0, N}$;
- ψ_t is the state of the underlying process of the arrival process of primary and retrial customers, $\psi_t = \overline{0, W}$;
- $n_t^{(l)}$ is the number of customers in the l-th node, $l = \overline{1, L}$, $n_t^{(l)} = \overline{0, n_t}$, $\sum\limits_{l=1}^{L} n_t^{(l)} = n_t$.

To compute the stationary distribution of the states of the Markov chain ξ_t, $t \geq 0$, we need to derive the generator of this chain. To this end, we use the matrix-analytic methods. Therefore, to simplify derivation of the generator it is useful to deal not with separate states of the chain but with whole groups of the states having the same value of the two first components of the Markov chain. We call such a group having the value (i, n) of these components a macro-state (i, n). There are many possibilities to enumerate the states of the Markov chain that belong to a fixed macro-state. Here, we assume that the states are numerated in the reverse lexicographic order of the components $n_t^{(1)}, \ldots, n_t^{(L)}$ and the direct lexicographic order of the component ψ_t. We call the set of macro-states $\{(i, 0), (i, 1), \ldots, (i, N)\}$ as the level i, $i \geq 0$.

To simplify derivations, in what follows, we use the following notation:

- I is the identity matrix and O is a zero matrix of an appropriate dimension;
- $\otimes$ and $\oplus$ are the symbols of Kronecker product and sum of matrices, respectively, see Reference [26];
- $\mathrm{diag}\{\ldots\}$ is the diagonal matrix with the diagonal entries listed or defined by the entries of the vector in the brackets;
- $\mathrm{diag}_+\{\ldots\}$ is the updiagonal matrix with the updiagonal entries listed in the brackets;
- $\mathrm{diag}_-\{\ldots\}$ is the subdiagonal matrix with the subdiagonal entries listed in the brackets;

- $\boldsymbol{\beta}$ is the column vector defined as $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)^T$;
- $\mathbf{p}$ is the row vector defined as $\mathbf{p} = (p_1, \ldots, p_L)$;
- $\mathbf{a}_l$ is the row vector of size L that has all zero components except the lth component which is equal to 1;
- $\mathbf{q}$ is the column vector defined as $\mathbf{q} = (q_1, q_2, \ldots, q_L)^T = (q_{1,0}\mu_1, q_{2,0}\mu_2, \ldots, q_{L,0}\mu_L)^T$;
- $T_n = \binom{n+L-1}{L-1} = \frac{(n+L-1)!}{n!(L-1)!}$ is the number of states of the process $(n_t^{(1)}, \ldots, n_t^{(L)})$ when $\sum\limits_{l=1}^{L} n_t^{(l)} = n$, $n = \overline{0, N}$.

Because we will operate with macro-states, we need to analyse the vector process $\mathbf{n}_t = \{n_t^{(1)}, \ldots, n_t^{(L)}\}$, $t \geq 0$, that defines the dynamics of the number of customers in the nodes of the network. The following events can change this number:

(1) An admission to the network of a customer from the orbit. Let us denote as $P_n(\mathbf{p})$, $n = \overline{0, N-1}$, the matrix consisting of probabilities of the transitions of the process $\mathbf{n}_t$ during the epoch when n, $n < N$, customers are obtaining service in the network and a customer from the orbit makes a retrial attempt;

(2) An admission to the network of an arriving type-l customer. Let us denote as $P_n(\mathbf{a}_l)$, $n = \overline{0, N-1}$, the matrix consisting of the transitions probabilities of the process $\mathbf{n}_t$ during the epoch when n, $n < N$, customers are obtaining service in the network and a type-l customer arrives to the system;

(3) A customer finishes the service in one of the nodes and transits to some another one. Let us denote as B_n, $n = \overline{1, N}$, the matrix which components define the intensities of the process $\mathbf{n}_t$ transitions in this case, conditioned on the fact that n customers are in the network;

(4) A customer finishes the service in some node and leaves the network. Let us denote as $S_n(\mathbf{q})$, $n = \overline{0, N-1}$, the matrix which components define the intensities of the process $\mathbf{n}_t$ transitions in this case, conditioned on the fact that there are $N - n$ customers in the network;

(5) A customer leaves some node due to impatience. The matrix $Z_n(\boldsymbol{\beta})$, $n = \overline{1, N}$, defines the intensities of the process $\mathbf{n}_t$ transitions in the considered case when there are n customers in the network.

It is worth noting that the matrices P_n, $S_n(\mathbf{q})$, $n = \overline{0, N-1}$, and B_n, $n = \overline{1, N}$, can be computed based on the use of the corresponding results from Reference [10]. Derivation of matrices $Z_n(\boldsymbol{\beta})$ is novel because the impatience of the customers staying in the buffers of the network nodes was not considered in the related literature previously. This derivation can be performed as follows.

Step (1) We compute the matrices $Z_n^{(k)}(\boldsymbol{\beta})$ using the recursive formulas:

$$Z_n^{(0)}(\boldsymbol{\beta}) = (n-1)\beta_L,\ n = \overline{1, N},$$

$$Z_n^{(k)}(\boldsymbol{\beta}) = \begin{pmatrix} (n-1)\beta_{L-k}I & O & \cdots & O \\ Z_1^{(k-1)}(\boldsymbol{\beta}) & (n-2)\beta_{L-k}I & \cdots & O \\ O & Z_2^{(k-1)}(\boldsymbol{\beta}) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & 0\beta_{L-k}I \\ O & O & \cdots & Z_n^{(k-1)}(\boldsymbol{\beta}) \end{pmatrix}, k = \overline{1, L-1},\ n = \overline{1, N}.$$

Step (2) Compute the matrices $Z_n(\boldsymbol{\beta})$ as $Z_n(\boldsymbol{\beta}) = Z_n^{(L-1)}(\boldsymbol{\beta})$, $n = \overline{0, N-1}$.

Also, let us introduce the matrix Γ_n. It is the diagonal matrix the diagonal entries of which define the intensities of exit of the process $\mathbf{n}_t$ from its states when n customers obtain service in the network. The matrices Γ_n are defined as follows:

$$\Gamma_0 = 0,\ \Gamma_n = -\mathrm{diag}\{B_n\mathbf{e} + S_{N-n}(\mathbf{q})\mathbf{e} + Z_n(\boldsymbol{\beta})\mathbf{e}\},\ n = \overline{1, N}.$$

We denote as $\mathbf{G}$ the infinitesimal generator of the Markov chain $\xi_t,\ t \geq 0$. The matrix $\mathbf{G}$ has an infinite size. It consists of the blocks $\mathbf{G}_{i,j},\ i,j \geq 0$, defining transition intensities from the level i to the level j. In turn, each matrix $\mathbf{G}_{i,j}$ consists of sub-blocks $\mathbf{G}_{i,j}^{(n,n')}$ of transition intensities from the macro-state (i,n) to the macro-state (j,n'). The diagonal entries of the blocks $\mathbf{G}_{i,i}^{(n,n)}$ are negative and are equal, up to the sign, to the rates of the exit of the Markov chain ξ_t from the corresponding states.

Lemma 1. *The infinitesimal generator* $\mathbf{G}$ *of the Markov chain* $\xi_t,\ t \geq 0$, *has a block-tridiagonal structure:*

$$\mathbf{G} = \mathrm{diag}\{\mathbf{G}_{i,i},\ i \geq 0\} + \mathrm{diag}_+\{\mathbf{G}_{i,i+1},\ i \geq 0\} + \mathrm{diag}_-\{\mathbf{G}_{i,i-1},\ i \geq 1\}. \tag{1}$$

The non-zero blocks $\mathbf{G}_{i,j},\ i,j \geq 0$, *have the following form:*

$$\mathbf{G}_{i,i} = \mathrm{diag}\{\mathbf{G}_{i,i}^{(n,n)},\ n = \overline{0,N}\} + \mathrm{diag}_+\{\mathbf{G}_{i,i}^{(n,n+1)},\ n = \overline{0,N-1}\} + \mathrm{diag}_-\{\mathbf{G}_{i,i}^{(n,n-1)},\ n = \overline{1,N}\},\ i \geq 0, \tag{2}$$

$$\mathbf{G}_{i,i-1} = \mathrm{diag}\{\mathbf{G}_{i,i-1}^{(n,n)},\ n = \overline{0,N}\} + \mathrm{diag}_+\{\mathbf{G}_{i,i-1}^{(n,n+1)},\ n = \overline{0,N-1}\},\ i \geq 1, \tag{3}$$

$$\mathbf{G}_{i,i+1} = \mathrm{diag}\{O,\ldots,O,\sum_{k=1}^{L} D_k \otimes I_{T^N}\},\ i \geq 0, \tag{4}$$

where

$$\mathbf{G}_{i,i}^{(0,0)} = D_0^{(i)} - i\gamma I_{\bar{W}}, \tag{5}$$

$$\mathbf{G}_{i,i}^{(n,n)} = (D_0^{(i)} - i\gamma I_{\bar{W}}) \oplus (B_n + \Gamma_n),\ n = \overline{1,N-1}, \tag{6}$$

$$\mathbf{G}_{i,i}^{(N,N)} = (D_0^{(i)} + (1-h)D^{(i)} - i\gamma I_{\bar{W}}) \oplus (B_N + \Gamma_N), \tag{7}$$

$$\mathbf{G}_{i,i}^{(n,n+1)} = \sum_{k=1}^{L} D_k \otimes P_n(\mathbf{a}_k),\ n = \overline{0,N-1}, \tag{8}$$

$$\mathbf{G}_{i,i}^{(n,n-1)} = I_{\bar{W}} \otimes (S_{N-n}(\mathbf{q}) + Z_n(\boldsymbol{\beta})),\ n = \overline{1,N}, \tag{9}$$

$$\mathbf{G}_{i,i-1}^{(n,n+1)} = D^{(i)} \otimes P_n(\mathbf{p}),\ n = \overline{0,N-1}, \tag{10}$$

$$\mathbf{G}_{i,i-1}^{(n,n)} = i\gamma I_{\bar{W}T_n},\ n = \overline{0,N-1}, \tag{11}$$

$$\mathbf{G}_{i,i-1}^{(N,N)} = i\gamma I_{\bar{W}T_N} + hD^{(i)} \otimes I_{T_N}. \tag{12}$$

A brief proof of the Lemma is as follows. Block tridiagonal structure (1) of the generator $\mathbf{G}$ is explained by the fact that the probability of two or more customers arrival or departure from the orbit during the interval of the infinitesimal length is negligible. The matrices $\mathbf{G}_{i,i}, \mathbf{G}_{i,i-1}, \mathbf{G}_{i,i+1}$ have block structures (2)–(4) were the blocks define the intensities of transitions that lead to the corresponding change of the number of customers presenting in the network.

The most simple form (4) has the matrix $\mathbf{G}_{i,i+1}$ defining the intensities of customers arriving at the orbit. Because a customer joins the orbit only when the capacity of the server is exhausted (the number of customers in the network is equal to N), only one block of the matrix $\mathbf{G}_{i,i+1}$, namely, $\mathbf{G}_{i,i+1}^{(N,N)}$, is not equal to zero. This matrix block corresponds to the arrival of a primary customer of any type when the number of customers in the network is equal to N. This customer moves to the orbit and the number of customers in the network does not change.

The matrix $\mathbf{G}_{i,i-1}$ defining the intensities of customers departure from the orbit is the block two-diagonal matrix of form (3). It has the non-zero diagonal blocks corresponding to the case when the customer departing from the orbit does not enter the network but is lost. Such a departure occurs due to the impatience of the customers in the orbit when the number of customers in the network is equal to n, $n < N$, or the impatience and non-persistence of the customers staying in the orbit when

this number is equal to N. The corresponding intensities of the departure are given by the matrices of form (11) and (12), respectively. The matrix $\mathbf{G}_{i,i-1}$ has also the up-diagonal blocks corresponding to the case when a customer from the orbit makes a retrial when the number of customers in the network is equal to n, $n < N$, and enters the network. After entering the network, this customer joins some of the network nodes. The intensities of occurrence of these events are given by the matrices of form (10).

The matrix $\mathbf{G}_{i,i}$ has block tridiagonal structure (2). The diagonal entries of its diagonal blocks $\mathbf{G}_{i,i}^{(n,n)}$ are negative. Their values with the opposite sign define the rate of the exit of the Markov chain ξ_t from the corresponding states. These diagonal entries are the corresponding entries of the matrices of form (5), if $n = 0$, (6), if $n = \overline{1, N-1}$, and (7) if $n = N$. The non-diagonal entries of the diagonal blocks $\mathbf{G}_{i,i}^{(n,n)}$ are non-negative and define the intensities of the transitions of the Markov chain ξ_t that do not cause the change of the number of customers neither in the orbit and in the network. These entries are given by the corresponding entries of the matrices of form (5), if $n = 0$, (6), if $n = \overline{1, N-1}$, and (7) if $n = N$. The up-diagonal blocks $\mathbf{G}_{i,i}^{(n,n+1)}$ correspond to the arrival of a primary customer of any type when the number of the customers in the network is less than N and its entering the corresponding node of the network. These blocks are defined by formula (8). The sub-diagonal blocks $\mathbf{G}_{i,i}^{(n,n-1)}$ correspond to a customer departure from the network due to the service completion or due to impatience and have the form (9). Lemma is proven.

It can be shown that the Markov chain ξ_t, $t \geq 0$, belongs to the class of Asymptotically Quasi-Toeplitz Markov chains, see Reference [27]. Using the results from Reference [27], it is easily verified that because the customers in orbit are impatient, that is, $\gamma > 0$, the Markov chain ξ_t, $t \geq 0$, is ergodic. Then, the following limits (stationary probabilities) exist for any set of the network parameters:

$$\pi(i, n, \psi, n^{(1)}, \dots, n^{(L)}) = \lim_{t \to \infty} P\{i_t = i,\ n_t = n, \psi_t = \psi,\ n_t^{(1)} = n^{(1)}, \dots, n_t^{(L)} = n^{(L)}\},$$

$$i \geq 0,\ n = \overline{0, N},\ \psi = \overline{0, W},\ n^{(l)} = \overline{0, n},\ l = \overline{1, L},\ \sum_{l=1}^{L} n^{(l)} = n.$$

Let us denote by $\boldsymbol{\pi}(i, n)$ the row vectors of the stationary probabilities that belong to the macro-state (i, n) and by $\boldsymbol{\pi}_i$ the row vectors of the stationary probabilities that belong to the level i, $i \geq 0$.

It is well known that the probability vectors $\boldsymbol{\pi}_i$, $i \geq 0$, satisfy the following system of linear algebraic equations:

$$(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots)\mathbf{G} = \mathbf{0},\ \ (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots)\mathbf{e} = 1. \tag{13}$$

Comparing this system with the corresponding system for the probability vectors $\boldsymbol{\pi}_i$, $i \geq 0$, for the analogous queuing network with the buffer considered in Reference [10], we see the following. The size of the vectors $\boldsymbol{\pi}_i$ (and the size of the corresponding square blocks of the generator) in Reference [10] is equal to $\bar{W}T_i$ if $i = \overline{0, N}$, and $\bar{W}T_N$ if $i > N$. The size of all vectors $\boldsymbol{\pi}_i$, $i \geq 0$, in the considered in our paper model of the queuing network with retrials is equal to $\bar{W} \sum_{n=0}^{N} T_n$. Therefore, the blocks of generator (1) are much larger than the blocks of the generator in Reference [10]. Thus, the algorithm from Reference [28], which was used for numerical work in Reference [10], is not very effective for solving system (13). To solve this system, we recommend to use a more effective numerically stable algorithm recently developed in Reference [29].

4. Performance Measures

The average number N_{orbit} of customers in the orbit is computed by

$$N_{orbit} = \sum_{i=1}^{\infty} i\boldsymbol{\pi}_i\mathbf{e}.$$

The average number $N_{network}$ of customers in the network at an arbitrary moment is computed by

$$N_{network} = \sum_{i=0}^{\infty} \sum_{n=1}^{N} n\boldsymbol{\pi}(i,n)\mathbf{e}.$$

The probability P_{imm} that an arbitrary customer is admitted to the network immediately upon arrival is computed by

$$P_{imm} = \frac{\sum\limits_{i=0}^{\infty} \sum\limits_{n=0}^{N-1} \boldsymbol{\pi}(i,n)(\hat{D} \otimes I_{T_n})\mathbf{e}}{\hat{\lambda}}$$

where the average arrival rate of primary customers $\hat{\lambda}$ is computed by

$$\hat{\lambda} = \sum_{i=0}^{\infty} \sum_{n=0}^{N} \boldsymbol{\pi}(i,n)(\hat{D} \otimes I_{T_n})\mathbf{e}$$

and $\hat{D} = \sum\limits_{r=1}^{L} D_r$.

Remark 1. *If all the matrices $D^{(i)}$ are the diagonal, that is, the underlying process ψ_t of arrivals cannot make the jumps into other states at the moments of retrials, then the average arrival rate of primary customers $\hat{\lambda}$ is equal to the arrival rate λ of the MMAP that arrives at time intervals when the orbit is empty. The value λ is given by $\lambda = \boldsymbol{\theta}\hat{D}\mathbf{e}$ where the row vector $\boldsymbol{\theta}$ is the unique solution of the system $\boldsymbol{\theta}(D_0 + \hat{D}) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. If some of the matrices $D^{(i)}$ are non-diagonal, then generally speaking $\hat{\lambda} \neq \lambda$.*

The average intensity $\lambda^{(l)}_{out-serve}$ of flow of customers who leave the network after successful service from the l-th node is computed by

$$\lambda^{(l)}_{out-serve} = \sum_{i=0}^{\infty} \sum_{n=1}^{N} \boldsymbol{\pi}(i,n)(I_{\bar{W}} \otimes S_{N-n}(\mathbf{q}^{(l)}))\mathbf{e},\ l = \overline{1,L},$$

where $\mathbf{q}^{(l)}$ is a column vector of size L with all zero components except the component $(\mathbf{q}^{(l)})_l = q_l$.

The average intensity $\lambda_{out-serve}$ of flow of customers who leave the network after successful service is computed by

$$\lambda_{out-serve} = \sum_{i=0}^{\infty} \sum_{n=1}^{N} \boldsymbol{\pi}(i,n)(I_{\bar{W}} \otimes S_{N-n}(\mathbf{q}))\mathbf{e}.$$

The average intensity $\lambda^{(l)}_{out-imp}$ of flow of customers who leave the network due to impatience from the l-th node is computed by

$$\lambda^{(l)}_{out-imp} = \sum_{i=0}^{\infty} \sum_{n=2}^{N} \boldsymbol{\pi}(i,n)(I_{\bar{W}} \otimes Z_n(\boldsymbol{\beta}^{(l)}))\mathbf{e},\ l = \overline{1,L},$$

where $\boldsymbol{\beta}^{(l)}$ is a column vector of size L with all zero components except the component $(\boldsymbol{\beta}^{(l)})_l = \beta_l$.

The average intensity $\lambda_{out-imp}$ of flow of customers who leave the network due to impatience is computed by

$$\lambda_{out-imp} = \sum_{i=0}^{\infty} \sum_{n=2}^{N} \boldsymbol{\pi}(i,n)(I_{\bar{W}} \otimes Z_n(\boldsymbol{\beta}))\mathbf{e}.$$

The probability of an arbitrary customer loss due to impatience from the orbit is computed by

$$P^{imp-orbit}_{loss} = \hat{\lambda}^{-1}\gamma N_{orbit}.$$

The probability of an arbitrary customer loss due to non-persistence from the orbit is computed by

$$P_{loss}^{nonpersist-orbit} = \hat{\lambda}^{-1} h \sum_{i=1}^{\infty} \boldsymbol{\pi}(i,N)(D^{(i)} \otimes I_{T_N})\mathbf{e}.$$

The probability of an arbitrary customer loss due to impatience from the network is computed by

$$P_{loss}^{imp-net} = \hat{\lambda}^{-1} \lambda_{out-imp}.$$

The probability of an arbitrary customer loss due to impatience from the lth node of the network is computed by

$$P_{loss}^{imp-net,l} = \hat{\lambda}^{-1} \lambda_{out-imp}^{(l)},\ l = \overline{1,L}.$$

The probability of an arbitrary customer loss is computed by

$$P_{loss} = 1 - \frac{\lambda_{out-serve}}{\hat{\lambda}} = P_{loss}^{imp-orbit} + P_{loss}^{imp-net} + P_{loss}^{nonpersist-orbit}.$$

The average intensity $\mu_{out}^{(l)}$ of output flow of successfully served customers from the l-th node is computed by

$$\mu_{out}^{(l)} = \sum_{i=0}^{\infty} \sum_{n=1}^{N} \boldsymbol{\pi}(i,n)(I_{\bar{W}} \otimes S_{N-n}(\mathbf{m}^{(l)}))\mathbf{e},\ l = \overline{1,L},$$

where $\mathbf{m}^{(l)}$ is the column vector of size L with all zero components except the component $(\mathbf{m}^{(l)})_l = \mu_l$.

The load of the l-th node ρ_l can be found as follows:

$$\rho_l = \frac{\mu_{out}^{(l)}}{\mu_l(1 - P_{loss}^{imp-net,l})},\ l = \overline{1,L}.$$

This characteristic of the node operation is very important because knowledge of its value is helpful to recognize the so-called bottlenecks in the network and to make certain managerial updates.

5. Numerical Examples

We present the results of three numerical experiments. In the first experiment, we illustrate the importance on account of possible dependence of arrival processes of primary and retrial customers. The aim of the second experiment is the numerical investigation of the dependence of the main performance measures of the system on the threshold N and illustration of the importance of account the impatience of customers staying in the network. In the third experiment, we show how our results can be used for localization of the bottlenecks of the network and improvement of the performance of the network via upgrading the bottleneck node.

Remark 2. *The importance of correlation in the arrival process for a similar queuing network without retrials was shown in Reference [10]. In this paper, we do not present the results illustrating the importance of the correlation. We mention only that the correlation in the arrival process of primary customers has an essential impact on the networks with retrials as well.*

Example 1. *Let us consider a queuing network consisting of $L = 3$ nodes. The mean service rates at these nodes are $\mu_1 = 2$, $\mu_2 = 1.5$, $\mu_3 = 2$, respectively.*

The transition probabilities of the customers in the network $q_{l,k}$, $l = \overline{1,L}$, $k = \overline{0,L}$, $k \neq l$, are defined as the corresponding entries of Table 1.

The probabilities defining the choice of the node at the moment of admission of a customer from the orbit are chosen as: $p_1 = 0.2$, $p_2 = 0.3$, and $p_3 = 0.5$. The intensity of impatience of a customer from the orbit is assumed to be $\gamma = 0.02$. The probability of a customer departure from the orbit after an unsuccessful retrial is $h = 0.3$. The intensities of impatience of customers from the nodes are given as follows: $\beta_1 = 0.05$, $\beta_2 = 0.01$, $\beta_2 = 0.03$.

Table 1. Transition probabilities $q_{l,k}$.

$q_{l,k}$	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$l = 1$	$\frac{1}{2}$	-	$\frac{1}{4}$	$\frac{1}{4}$
$l = 2$	$\frac{1}{3}$	$\frac{2}{15}$	-	$\frac{8}{15}$
$l = 3$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	-

We assume that the $MMAP$ arrival flow of primary customers when the orbit is empty is defined by the following matrices:

$$D_0 = \begin{pmatrix} -1.764 & 0.014 \\ 0.07 & -0.42 \end{pmatrix}, D_1 = \begin{pmatrix} 0.07 & 0.007 \\ 0 & 0.14 \end{pmatrix},$$

$$D_2 = \begin{pmatrix} 0.028 & 0.035 \\ 0.0042 & 0.203 \end{pmatrix}, D_3 = \begin{pmatrix} 1.603 & 0.007 \\ 0.0021 & 0.0007 \end{pmatrix}.$$

The retrial rates of customers from the orbit when i customers are staying in the orbit are defined by the entries of the matrix

$$D^{(i)} = i\tilde{D},$$

where

$$\tilde{D} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.02 \end{pmatrix}.$$

Remark 3. *We intentionally choose the matrix $D^{(i)}$ in the simple diagonal form. This allows us to calculate the average individual retrial rate α of a customer from the orbit as*

$$\alpha = \boldsymbol{\sigma}\tilde{D}\mathbf{e},$$

where $\boldsymbol{\sigma}$ is the unique solution of the following system

$$\boldsymbol{\sigma}(D_0 + \sum_{l=1}^{L} D_l - \mathrm{diag}\{\tilde{D}\mathbf{e}\}) = \mathbf{0}, \quad \boldsymbol{\sigma}\mathbf{e} = 1.$$

In this example, α is equal to 0.1185929648.

In the considered model, the arrival flows of primary and retrials customers are dependent. In the existing literature, such dependence was not analyzed. In this example, we clarify whether or not this dependence is essential. To this end, we also consider the case when customers in the orbit retry to obtain access to the network independently on primary customers as assumed in the existing literature. We suppose that each customer from the orbit makes repeated attempts with the intensity $\alpha = 0.1185929648$.

It is easy to see that the results for the model with independent flows of primary and retrial customers are obtained as the particular case of our results if we assume that the matrix $\tilde{D}$ has the following form:

$$\tilde{D} = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}.$$

Let us vary the number N of customers, which can be serviced in the network simultaneously, over the interval $[1, 15]$. The computations were performed on a PC with an Intel Core i7-8700 CPU and 16 GB RAM. The computation time for all N from 1 to 15 is about 4 min and 50 s.

Figures 2–4 illustrate the dependence on N of the probability of an arbitrary customer loss due to non-persistence from the orbit $P_{loss}^{nonpersist-orbit}$, due to impatience from the orbit $P_{loss}^{imp-orbit}$ and the loss probability of an arbitrary customer P_{loss} for the systems with dependent and independent arrivals of primary and retrial customers.

As it is seen from these figures, an account of the dependence of arrivals of primary and retrial customers is important for the precious prediction of the system performance measures. Under the chosen values of the network parameters, this dependence deteriorates the performance of the network. All loss probabilities are greater when the flows are dependent. In this example, the error in the estimation the loss probabilities can be up to 20% on their real values.

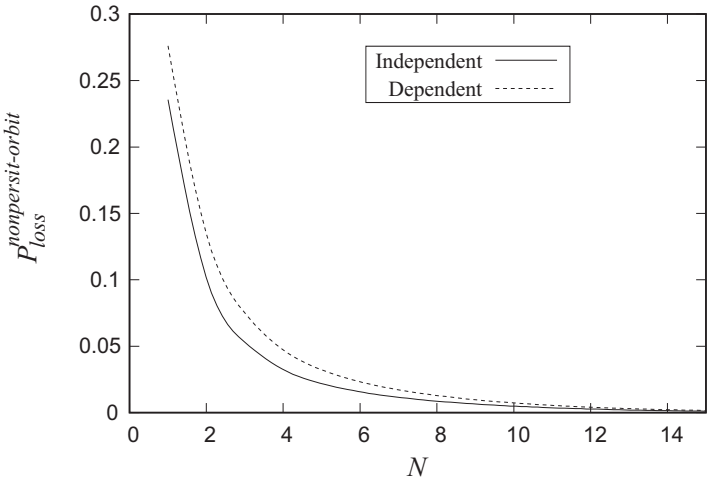

Figure 2. Dependence of the probability $P_{loss}^{nonpersist-orbit}$ on the number N.

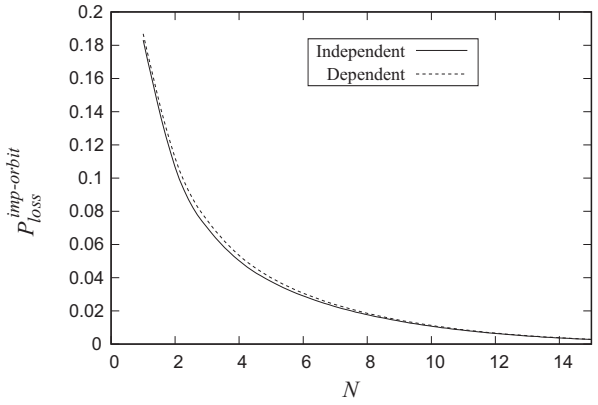

Figure 3. Dependence of the probability $P_{loss}^{imp-orbit}$ on the number N.

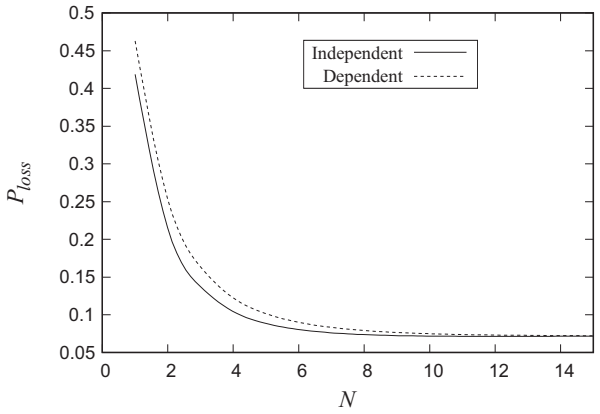

Figure 4. Dependence of the probability P_{loss} on the number N.

Example 2. *In this example, we numerically investigate the dependence of the main performance measures of the system on the threshold N. We also investigate the importance of accounting the impatience of customers in the nodes of the network. Let us assume that all system parameters are the same as in the previous example except for the matrix*

$$\tilde{D} = \begin{pmatrix} 0.2 & 0.002 \\ 0.001 & 0.02 \end{pmatrix}.$$

In the previous example, we chose relatively small intensities of impatience of customers from the nodes: $\beta_1 = 0.05$, $\beta_2 = 0.01$, $\beta_2 = 0.03$. The question arises, whether or not it is possible to ignore such small intensities of impatience and assume that the customers are patient, that is, $\beta_l = 0$, $l = \overline{1, L}$.

Figures 5–7 illustrate the dependence on N of the average number N_{orbit} of customers in the orbit, the probability P_{imm} that an arbitrary customer is admitted to the network immediately upon arrival and the probability of an arbitrary customer loss due to impatience from the network $P_{loss}^{imp-net}$ for the cases of the patient and impatient customers in the nodes of the network.

The probability of an arbitrary customer loss due to impatience from the network $P_{loss}^{imp-net}$ is equal to 0 when the customers in the nodes are patient and essentially increases with the growth of N when the customers are impatient. It is worth to note that the loss of the customers due to impatience in the nodes causes the reduction of the average number of customers in the orbit and the increase of the probability P_{imm}.

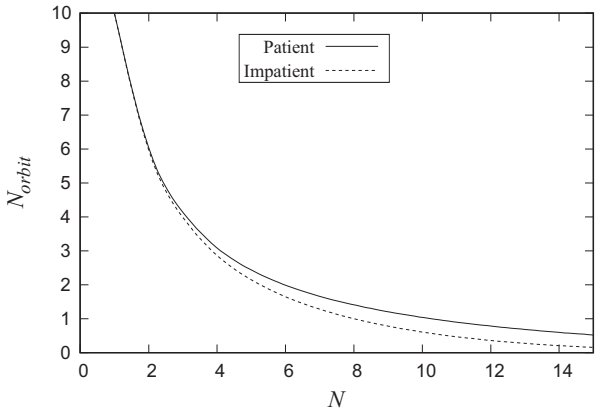

Figure 5. Dependence of the average number N_{orbit} on the number N.

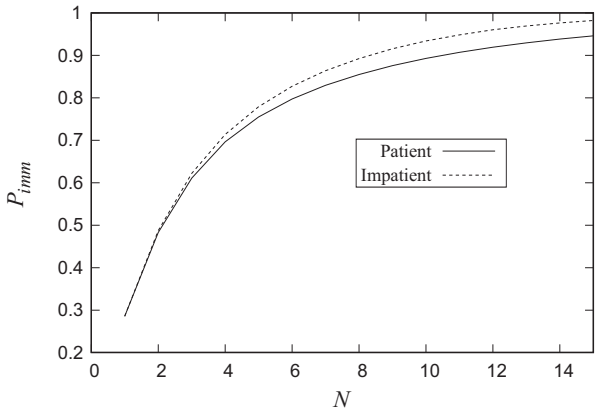

Figure 6. Dependence of the probability P_{imm} on the number N.

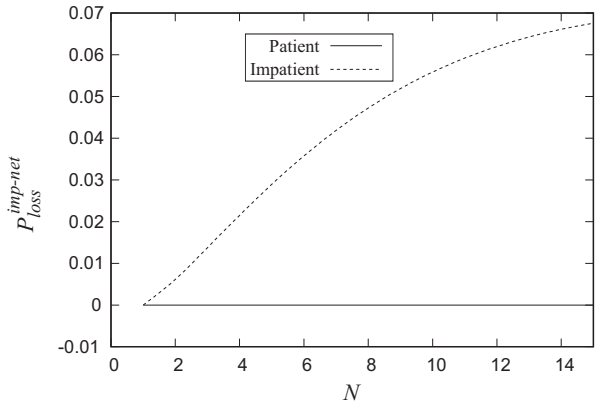

Figure 7. Dependence of the probability $P_{loss}^{imp-net}$ on the number N.

Analogously, Figures 8–10 illustrate the dependence on N of the probability of an arbitrary customer loss due to impatience from the orbit $P_{loss}^{imp-orbit}$, the probability of an arbitrary customer loss due to non-persistence from the orbit $P_{loss}^{nonpersist-orbit}$, and the loss probability of an arbitrary customer P_{loss} for the case of the patient and impatient customers in the network.

It can be seen that the impatience in the nodes reduced the probabilities of customers loss from the orbit (due to impatience or non-persistence). However, the total loss probability P_{loss} is essentially higher when the customers in the nodes are impatient. With the growth of N, the difference of values of this probability for the cases of the impatient and patient customers becomes more significant.

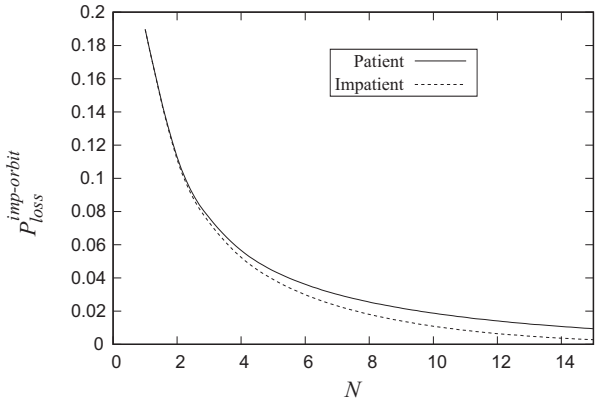

Figure 8. Dependence of the probability $P_{loss}^{imp-orbit}$ on the number N.

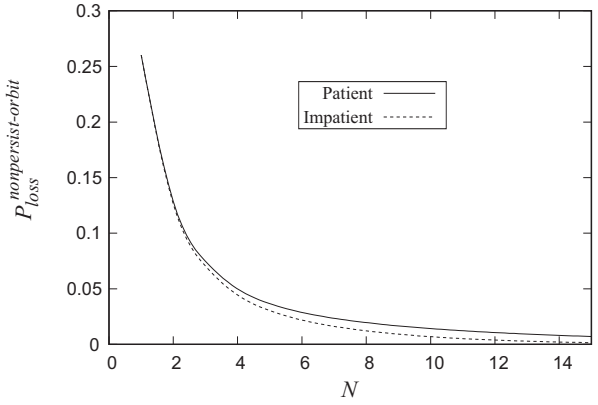

Figure 9. Dependence of the probability $P_{loss}^{nonpersist-orbit}$ on the number N.

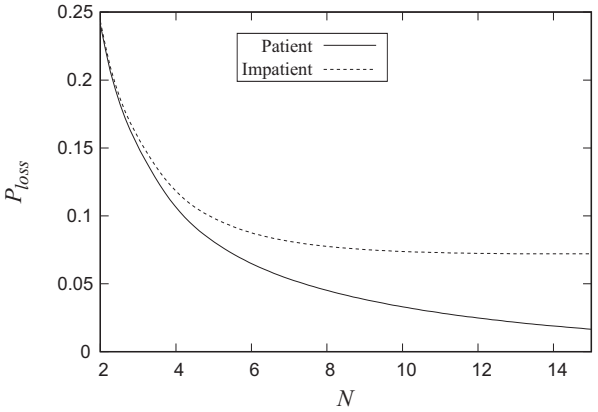

Figure 10. Dependence of the probability P_{loss} on the number N.

Let us introduce the following economical criterion of the quality of the network operation:

$$E(N) = \hat{\lambda}(a(P_{loss}^{imp-orbit} + P_{loss}^{nonpersist-orbit}) + bP_{loss}^{imp-net}),$$

where a is a charge paid for the loss of one customer from the orbit and b is a charge paid for the loss of one customer from the network per unit time. It is evident that the loss of a customer from the network is more painful than the loss of a customer from the orbit because the lost from the network customer possibly have already been serviced in some nodes, that is, the system has spent some resources for providing service to such a customer. Thus, in this example, we assume that $a = 1$ and $b = 3$.

Figure 11 illustrates the dependence of the economic criterion $E(N)$ on N for the cases of the patient and impatient customers in the network.

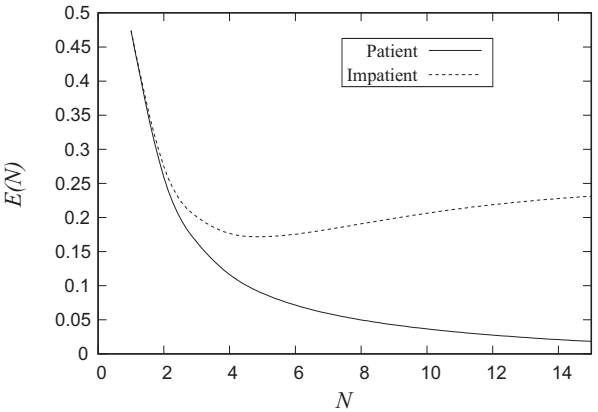

Figure 11. Dependence of the economical criterion $E(N)$ on the number N.

The optimal value for the case of patient customers is $E(N) = 0.0184$ when $N = 15$. This means that in the case of patient customers it is reasonable to accept as many customers as possible. The optimal value for the case of impatient in the network customers is $E(N) = 0.171865$ when $N = 5$. Presented in this paper results can be helpful for the optimal choice of the limit N imposed on the number of customers that can be admitted to the network simultaneously.

Example 3. *In this experiment, we show how our results can be used for identification of the bottleneck of the network and further improving the performance of the network via the proper upgrade of the bottleneck node. Let us choose the parameters of the network the same as in the previous example in the case of impatient customers. One can see that the loss probability of customers is quite high even for a large value of N ($N = 15$). About 7.2% of customers are lost due to different reasons. To understand the reason for such not very good operation of the network, let us compute the average load of each network's node.*

Figure 12 illustrates the dependence of loads of the nodes on the parameter N.

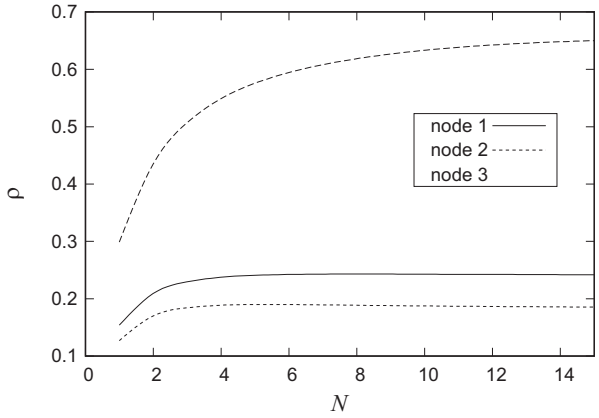

Figure 12. Dependence of the load of the nodes on the parameter N.

As it is seen from this Figure, the load of the third node is much higher than loads of other nodes. Let us make an upgrade of the third node in such a way that after upgrade the service rate in this node increases from 2 to 4 and compute the main performance measures of the network.

Figures 13 and 14 illustrate the dependence of the probability P_{imm} that an arbitrary customer is admitted to the network immediately upon arrival and the loss probability of an arbitrary customer P_{loss} on the parameter N before and after upgrade.

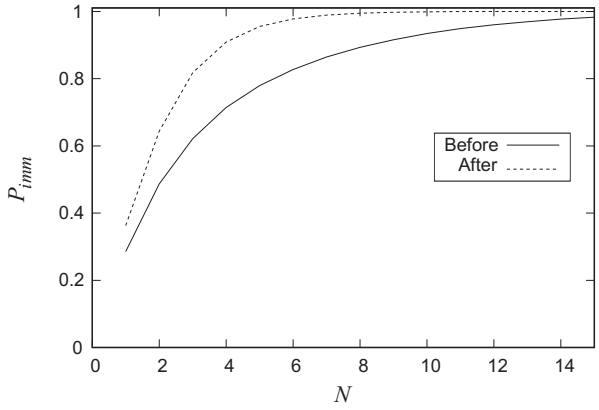

Figure 13. Dependence of the probability P_{imm} on the number N.

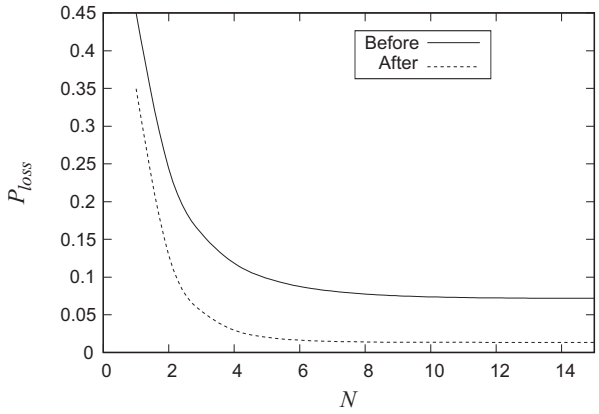

Figure 14. Dependence of the probability P_{loss} on the number N.

One can see from these Figures that after upgrade the loss probability P_{loss} essentially decreases and the probability P_{imm} of immediate access essentially increases. Figure 15 illustrates the dependence of the economic criterion $E(N)$ on N before and after the upgrade.

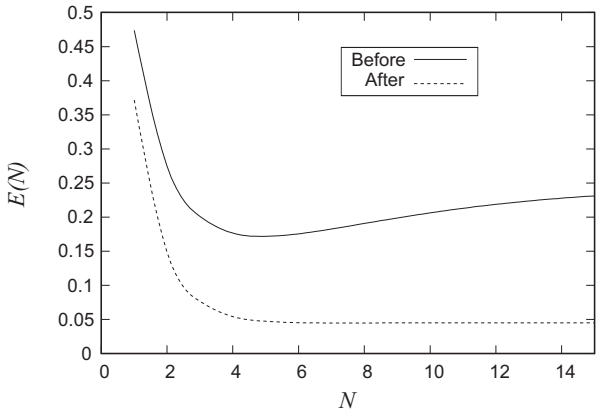

Figure 15. Dependence of the economic criterion $E(N)$ on the number N before and after upgrade.

The optimal value $E(N)$ after upgrade is $E(N) = 0.0448422$ when $N = 7$ what is almost four times smaller than the optimal value $E(N) = 0.171865$ before the upgrade (that was achieved when $N = 5$). Definitely, the increase of the service rate in the third node can cost some money. However, it allows increasing the optimal number of the limit N from 5 to 7 with the essential improvement of the quality of operation of the network.

It is worth to note that the improvement of the quality of operation of the network via the account of loads of the nodes can be achieved also via the modification of the routing of the customers in the node and the choices of the target node by the primary and retrial customers.

6. Conclusions

In this paper, we analyzed a semi-open queuing network with customers retrial. The number of customers in the network must not exceed the fixed threshold (capacity of the network). An arriving primary customer is admitted to the network and starts processing only if the current number of customers in the network is less than the network capacity. Otherwise, the customer moves to the orbit having an infinite capacity and tries to enter the network after random time intervals. The arrivals of primary customers and retrials of customers from the orbit depend on the same underlying process. This allows more adequately model real-world arrival processes than it can be done using known in the literature models of the arrival and retrial processes. Customers are impatient both in the orbit and infinite buffers of the nodes of the network. The behavior of the network is described by a multidimensional Markov chain. The generator of this chain is derived. The problem of computing

the stationary state distribution is discussed. The expressions for computing the main performance measures of the network are derived. Numerical results illustrate the importance of the account of the dependency of arrivals of the primary and retrial customers as well as the importance of account of impatience of customers. Possibilities of optimization of the quality of the network operation by means of the optimal choice of the network capacity and identification and elimination of bottlenecks in the network are demonstrated.

Author Contributions: Conceptualization C.K. and A.D.; methodology C.K. and S.D.; software S.D.; validation C.K., A.D. and K.S.; formal analysis C.K., S.D. and A.D.; investigation C.K., K.S. and A.D.; Writing—Original draft preparation, C.K. and S.D.; Writing—Review and editing, K.S. and A.D.; supervision, C.K.; project administration C.K. and K.S.

Funding: The work by Chesoong Kim has been supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2017R1D1A3A03000523). The work by S. Dudin, A. Dudin and K. Samouylov has been supported by RUDN University Program 5-100.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bolch, G.; Greiner, S.; De Meer, H.; Trivedi, K.S. *Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
2. Boucherie, R.J.; van Dijk, N.M. *Queuing Networks*; A Fundamental Approach; Springer: Cham, Switzerland, 2011.
3. Shortle, J.F.; Thompson, J.M.; Gross, D.; Harris, C.M. *Fundamentals of Queuing Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
4. Smith, J.M. *Introduction to Queuing Networks: Theory & Practice*; Springer: Cham, Switzerland, 2018.
5. Roy, D. Semi-open queuing networks: A review of stochastic models, solution methods and new research areas. *Int. J. Prod. Res.* **2016**, *54*, 1735–1752. [CrossRef]
6. Dallery, Y. Approximate analysis of general open queuing networks with restricted capacity. *Perform. Eval.* **1990**, *11*, 209–222. [CrossRef]
7. Dhingra, V.; Kumawat, G.L.; Roy, D.; de Koster, R. Solving semi-open queuing networks with time-varying arrivals: An application in container terminal landside operations. *Eur. J. Oper. Res.* **2018**, *267*, 855–876. [CrossRef]
8. Ekren, B.Y.; Heragu, S.S.; Krishnamurthy, A.; Malmborg, C.J. Matrix-geometric solution for semi-open queuing network model of autonomous vehicle storage and retrieval system. *Comput. Ind. Eng.* **2014**, *68*, 78–86. [CrossRef]
9. Jia, J.; Heragu, S.S. Solving semi-open queuing networks. *Oper. Res.* **2009**, *57*, 391–401. [CrossRef]
10. Kim, J.; Dudin, A.; Dudin, S.; Kim, C. Analysis of a Semi-Open queuing Network with Markovian Arrival Process. *Perform. Eval.* **2018**, *120*, 1–19. [CrossRef]
11. Palmer, G.I.; Harper, P.R.; Knight, V.A. Modelling deadlock in open restricted queuing networks. *Eur. J. Oper. Res.* **2018**, *266*, 609–621. [CrossRef]
12. Artalejo, J.R.; Gomez-Corral, A. *Retrial Queuing Systems: A Computational Approach*; Springer: Berlin/Heidelberg, Germany, 2008.
13. Falin, G.I.; Templeton, J.G.C. *Retrial Queues*; Chapman&Hall: London, UK, 1997.
14. Kim, C.S.; Dudin, S. Analysis of Semi-Open queuing Network with Customer Retrials. *J. Korean Inst. Ind.* **2019**, *45*, 193–202.
15. Mandelbaum, A.; Massey, W.A.; Reiman, M.I. Strong approximations for Markovian service networks. *Queuing Syst.* **1998**, *30*, 149–201. [CrossRef]
16. Chakravarthy, S.R. The batch Markovian arrival process: A review and future work. In *Advances in Probability Theory and Stochastic Processes*; Krishnamoorthy, A., Raju, N., Ramaswami, V., Eds.; Notable Publications Inc.: Branchburg, NJ, USA, 2001; pp. 21–29.
17. Lucantoni, D. New results on the single server queue with a batch Markovian arrival process. *Commun. Stat. Stoch. Models* **1991**, *7*, 1–46. [CrossRef]
18. Vishnevski, V.M.; Dudin, A.N. queuing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Autom. Remote Control* **2017**, *78*, 1361–1403. [CrossRef]

19. Strelen, J.C. Approximate analysis of queuing networks with Markovian arrival processes and phase type service times. In *Modellierung und Bewertung von Rechen- und Kommunikationssystemen*; Irmscher, K., Mittasch, C., Richter, K., Eds.; VDE- Verlag GmbH.: Berlin, Germany; Offenbach, Germany, 1997; pp. 55–70.
20. Gomez-Corral, A. A tandem queue with blocking and Markovian arrival process. *Queuing Syst.* **2002**, *41*, 343–370. [CrossRef]
21. Kim, C.S.; Dudin, A.; Dudin, S.; Dudina, O. Tandem queuing system with impatient customers as a model of call center with Interactive Voice Response. *Perform. Eval.* **2013**, *70*, 440–453. [CrossRef]
22. Kim, C.S.; Klimenok, V.; Taramin, O. A tandem retrial queuing system with two Markovian flows and reservation of channels. *Comput. Oper. Res.* **2010**, *37*, 1238–1246. [CrossRef]
23. Kim, C.S.; Park, S.H.; Dudin, A.; Klimenok, V.; Tsarenkov, G. Investigaton of the $BMAP/G/1 \rightarrow \bullet/PH/1/M$ tandem queue with retrials and losses. *Appl. Math. Model.* **2010**, *34*, 2926–2940. [CrossRef]
24. He, Q.M. Queues with marked calls. *Adv. Appl. Probab.* **1996**, *28*, 567–587. [CrossRef]
25. Dudin, A.N.; Klimenok, V.I. $BMAP/SM/1$ model with Markov modulated retrials. *Top* **1999**, *7*, 267–278. [CrossRef]
26. Graham, A. *Kronecker Products and Matrix Calculus with Applications*; Ellis Horwood: Cichester, UK, 1981.
27. Klimenok, V.I.; Dudin, A.N. Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queuing theory. *Queuing Syst.* **2006**, *54*, 245–259. [CrossRef]
28. Dudina, O.; Kim, C.; Dudin, S. Retrial queuing system with Markovian arrival flow and phase-type service time distribution. *Comput. Ind. Eng.* **2013**, *66*, 360–373. [CrossRef]
29. Dudin, S.; Dudina, O. Retrial multi-server queuing system with PHF service time distribution as a model of a channel with unreliable transmission of information. *Appl. Math. Model.* **2019**, *65*, 676–695. [CrossRef]