

Article



Convergence Bounds for Limited Processor Sharing Queue with Impatience for Analyzing Non-Stationary File Transfer in Wireless Network

Irina Kochetkova ^{1,2,*}, Yacov Satin ³, Ivan Kovalev ³, Elena Makeeva ¹, Alexander Chursin ¹, and Alexander Zeifman ^{2,3}

- ¹ Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St., 117198 Moscow, Russia; elena-makeeva-96@mail.ru (E.M.); chursin-aa@rudn.ru (A.C.)
- ² Institute of Informatics Problems, Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences, 44-2 Vavilova St., 119333 Moscow, Russia; a_zeifman@mail.ru
 ³ Department of Applied Mathematics, Valagda State University, Laping 15, 160000 Valagda, Russia;
 - Department of Applied Mathematics, Vologda State University, Lenina 15, 160000 Vologda, Russia; yacovi@mail.ru (Y.S.); kovalev.iv96@yandex.ru (I.K.)
- * Correspondence: gudkova-ia@rudn.ru

Abstract: The data transmission in wireless networks is usually analyzed under the assumption of non-stationary rates. Nevertheless, they strictly depend on the time of day, that is, the intensity of arrival and daily workload profiles confirm this fact. In this article, we consider the process of downloading a file within a single network segment and unsteady speeds—for arrivals, file sizes, and losses due to impatience. To simulate the scenario, a queuing system with elastic traffic with non-stationary intensity is used. Formulas are given for the main characteristics of the model: the probability of blocking a new user, the average number of users in service, and the queue. A method for calculating the boundaries of convergence of the model is proposed, which is based on the logarithmic norm of linear operators. The boundaries of the rate of convergence of the main limiting characteristics of the queue length process were also established. For clarity of the influence of the parameters, a numerical analysis was carried out and presented.

Keywords: queuing system; elastic traffic; inpatient claim; non-stationary intensity; convergence analysis; bounds on the rate of convergence; wireless network; file transfer; daily traffic profile; blocking probability

1. Introduction

The fifth generation (5G) networks will consist of different services with different specifications. Experts have identified network slicing as a key technology to enable 5G networks [1–4]. Within the framework of network slicing, many models are proposed with different principles for slicing radio resources [5]. In some cases, it is necessary to distribute slices between several users, and this distribution and the slices can be changed depending on time, for example, in [6–8], various options for re-slicing the network are shown. Earlier, we studied models for studying network slicing in the framework of papers [9,10], but we considered a model in the form of a queuing system with stationary intensities.

Given the need to re-slice the network and because all processes are non-stationary and depend on time. For example, a traffic profile—user activity is different depending on the time of day. Activity may also depend on the time of year; in the summer many users go on vacation and fly to other countries, and in winter most users are actively working and on weekends they sit at home watching movies, for example. Taking into account this dependence, it is necessary to consider models with non-stationary intensities. In our work, we consider an example of downloading user files depending on the time of day. For convenience, we consider one slice of the radio frequency channel taking into account the



Citation: Kochetkova, I.; Satin, Y.; Kovalev, I.; Makeeva, E.; Chursin, A.; Zeifman, A. Convergence Bounds for Limited Processor Sharing Queue with Impatience for Analyzing Non-Stationary File Transfer in Wireless Network. *Mathematics* **2022**, *10*, 30. https://doi.org/10.3390/ math10010030

Academic Editor: Daniel-Ioan Curiac

Received: 17 November 2021 Accepted: 17 December 2021 Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). non-stationary nature, and look at the behavior of the characteristics. As a mathematical model, we take a queuing system with a non-stationary arrival intensity of user arrival in the system, as in [11,12], where such systems with non-stationary intensities are used for the joint service of radio frequencies.

This article discusses a special, but rather non-standard, heterogeneous birth and death process, for which, in principle, one can apply the methods described, for example, in articles [13,14], but due to the specifics of the model, in order to obtain acceptable estimates, the authors had to "trick". In particular, the rate of convergence in the example under consideration actually turns out to be rather slow, so that the existing (and constructed) limiting regime begins to adequately show the situation at sufficiently large times Usually, uniformization is used as a method for calculating the transition probabilities for Markov chains, as in [15,16]. However, the methods based on it work very poorly in the case of slow convergence, which takes place for the considered models. In addition, without a prior understanding of when the limiting regime is reached, significant computational efforts are required in order to be at least to some extent confident that the obtained solution is the required one [17].

The goal of the paper is to explore the nonstationary queue-length process based on file transferring in the wireless network and analyze the performance measures of this system model. The remainder of the paper is organized as follows. In Section 2, the queueing system and its performance measures are presented. The convergence analysis for large service rates and arrival rates of the mentioned queueing system are described in Section 3. Application for file transfer in the wireless network and numerical analysis of the considered system model are discussed in Section 4, followed by the conclusions in Section 5.

2. Queuing System

2.1. Overview and Assumptions

In this paper, we will consider a queuing system with elastic traffic and inpatient claims. To describe the flow of requests with a variable number of users, a Poisson flow of the first kind with the following parameters is suitable: the arrival intensity $\lambda(t)$, the minimum requirement for the resource *b*, and the length of the transmitted data block $\theta(t)$. Table 1 reflects the main parameters that describe the system model and the corresponding terms of the mathematical model. In the system under consideration, there is a resource of volume *C*, a storage device with a finite capacity *r*. Requests also have the property of impatience—they leave the queue with intensity $\gamma(t)$. We will assume that the block length is equal to some value $\theta(t)$. The entire volume *C* is divided equally between the orders, that is, if the number of customers is 1, then the entire resource is consumed by this customer, and the service rate is $\theta(t)/C$; if the number of customers is 2, then the service rate is $2\theta(t)/C$ —the resource volume is divided in half. In the case when *C* cannot be divided equally between the customers with the provision of the minimum guaranteed threshold *b*, a new customer enters the queue.

Let $N(t) \in \{1, ..., |C/b|\}$ be the number of processed orders at the moment $t \ge 0$. Hence, the number |C/b| = N is the maximum number of requests that the device can process simultaneously. The state-space of the system looks like this:

$$X := \{ n \in 0, ..., N, ..., N + r : c(n) \le C \}.$$
(1)

Table 1. System model parameters.

Parameter	Description
$\lambda(t)$	Intensity of the flow of requests for the transfer of elastic data
b	Minimum guaranteed elastic data block transfer rate
$\theta(t)$	Average value of data block length
r	The queue of applications for the transfer of a block of elastic data
$\gamma(t)$	The rate of loss of impatient claims
Ĉ	Network bandwidth (service speed)

2.2. Continuous-Time Markov Chain

It is easy to see that the given model can be described by Markov process X(t), t > 0 where X(t) denotes the number of customers in the system at time t (queue-length process). Denote by $p_n(t) = P(X(t) = n)$, n = 0, 1, 2, 3, ...

From the above assumptions, the resulting behavior of the state probabilities are described by a forward Kolmogorov system:

$$p_0'(t) = -\lambda(t)p_0(t) + \frac{C}{\theta(t)}p_1(t)$$
(2)

$$p_n'(t) = \lambda(t)p_{n-1}(t) - \left(\frac{C}{\theta(t)} + \lambda(t)\right)p_n(t) + \frac{C}{\theta(t)}p_{n+1}(t), 1 \le n < N$$
(3)

$$p_N'(t) = \lambda(t)p_{n-1}(t) - \left(\frac{C}{\theta(t)} + \lambda(t)\right)p_n(t) + \left(\frac{C}{\theta(t)} + \gamma(t)\right)p_{n+1}(t)$$
(4)

$$p'_{n}(t) = \lambda(t)p_{n-1}(t) - \left(\frac{C}{\theta(t)} + (n-N)\gamma(t) + \lambda(t)\right)p_{n}(t) + \left(\frac{C}{\theta(t)} + (n+1-N)\gamma(t)\right)p_{n+1}(t),$$

$$N < n < N+r$$
(5)

$$p_{N+r}'(t) = \lambda(t)p_{N+r-1}(t) - \left(\frac{C}{\theta(t)} + r\gamma(t) + \lambda(t)\right)p_{N+r}(t).$$
(6)

Now, we consider the corresponding nonstationary situation. Namely, we suppose that the queue-length process $\{X(t), t \ge 0\}$ is an inhomogeneous continuous-time Markov chain. All possible transition intensities say $q_{ij}(t)$, are supposed to be non-random functions of time. We suppose that all intensity functions are nonnegative and locally integrable on $[0, \infty)$.

Denote by $\mathbf{p}(t) = (p_0(t), p_1(t), p_2(t), \dots, p_{N+r}(t))^T$ the vector of state probabilities at the moment *t*. Put $a_{ij}(t) = q_{ji}(t)$ for $j \neq i$ and $a_{ii}(t) = -\sum_{j\neq i} a_{ji}(t) = -\sum_{j\neq i} q_{ij}(t)$.

We can consider the forward Kolmogorov system (2)–(6) as a differential equation

$$\frac{d\mathbf{p}(t)}{dt} = A(t)\mathbf{p}(t),\tag{7}$$

in the space of sequences l_1 , where A(t) is a bounded for almost all $t \ge 0$ linear operator from l_1 to itself and it is generated be the corresponding transposed intensity matrix:

($-\lambda(t)$	$\frac{C}{\theta(t)}$	0	-	0	0	0		0	0	
	$\lambda(t)$	$-\left(\frac{C}{\theta(t)} + \lambda(t)\right)$	$\frac{C}{\theta(t)}$		0	0	0		0	0	
	0	$\lambda(t)$	$-\left(\frac{C}{\theta(t)} + \lambda(t)\right)$		0	0	0		0	0	
	:	•	•	:	•	-	•	÷	:	•	
	0	0	0		$-\left(\frac{C}{\theta(t)} + \lambda(t)\right)$	$\frac{C}{\theta(t)}$	0		0	0	(8)
	0	0	0		$\lambda(t)$	$-\left(\frac{C}{\theta(t)} + \lambda(t)\right)$	$rac{C}{ heta(t)} + \gamma(t)$		0	0	(0)
	0	0	0		0	$\lambda(t)$	$-\left(\frac{C}{\theta(t)} + \gamma(t) + \lambda(t)\right)$		0	0	
	:	- - -	-	:	•			÷	: : :		
	0	0	0		0	0	0		$-\left(\frac{C}{\theta(t)} + (r-1)\gamma(t) + \lambda(t)\right)$	$\frac{C}{\theta(t)} + \gamma(t)r$	
	0	0	0		0	0	0		$\lambda(t)$	$-\left(\frac{C}{\theta(t)}+\gamma(t)r\right)$	

2.3. Performance Measures

For analysis of the system let us consider some characteristics of the devoted model. First, the probability of blocking an incoming application:

$$P_{block}(t) = p_{N+r}(t).$$
(9)

The average number of serviced applications:

$$\bar{C}(t) = \sum_{i=1}^{N} i p_i(t) + N \cdot \sum_{i=1}^{r} p_{N+i}(t).$$
(10)

The average number of applications in the queue:

$$Q(t) = \sum_{i=N+1}^{N+r} (i-N)p_i(t).$$
(11)

3. Convergence Analysis

3.1. Definitions of Terms

We denote the mathematical expectation (the mean) of X(t) at the moment *t* if X(0) = k as $E(t,k) = E\{X(t)|X(0) = k\}$.

The Markov chain X(t) is called *weakly ergodic*, if $\lim_{t\to\infty} ||\mathbf{p}^1(t) - \mathbf{p}^2(t)|| = 0$ for any initial conditions $\mathbf{p}^1(0) = \mathbf{p}^1 \in \Omega$, $\mathbf{p}^2(0) = \mathbf{p}^2 \in \Omega$. For our situation *any* $\mathbf{p}^1(t)$ is considered as a *quasi-stationary distribution* of the chain X(t).

The mentioned Markov chain X(t) also has the limiting mean $\phi(t)$, if $|E(t;k) - \phi(t)| \rightarrow 0$ as $t \rightarrow \infty$ for any k.

We recall the logarithmic norm of operator function from l_1 to itself is calculated as (12):

$$\gamma(B(t))_{1} = \sup_{i} \left(b_{ii}(t) + \sum_{j \neq i} |b_{ji}(t)| \right),$$
(12)

and the bound

$$\|U(t,s)\| \le e^{\int_s^t \gamma(B(\tau)) d\tau},\tag{13}$$

is valid for the Cauchy operator of the corresponding differential equation

$$\frac{d\mathbf{x}}{dt} = B(t)\mathbf{x}.$$
(14)

3.2. Preliminary Considerations

$$A(t) = \begin{pmatrix} -\lambda(t) & \mu_{1}(t) & 0 & \cdots & 0 & 0 \\ \lambda(t) & -(\mu_{1}(t) + \lambda(t)) & \mu_{2}(t) & \cdots & 0 & 0 \\ 0 & \lambda(t) & -(\mu_{2}(t) + \lambda(t)) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\mu_{N+r-1}(t) + \lambda(t)) & \mu_{N+r}(t) \\ 0 & 0 & 0 & \cdots & \lambda(t) & -\mu_{N+r}(t). \end{pmatrix}$$
(15)

As indicated above, the considered method is based on the concept of the logarithmic norm and the corresponding estimates for the Cauchy operator.

Assuming that $p_0 = 1 - \sum_{i=1}^{N+r} p_i(t)$, then from (7) we get the following equation:

$$\frac{d\mathbf{p}}{dt} = B(t)\mathbf{p} + \mathbf{g}(t), \quad t \ge 0,$$
(16)

where $\mathbf{g}(t) = (\lambda(t), 0, 0, ..., 0)^T$ and B(t) equals (17).

$$B(t) = \begin{pmatrix} -(\mu_{1}(t) + 2\lambda(t)) & \mu_{2}(t) - \lambda(t) & -\lambda(t) & -\lambda(t) & \cdots & -\lambda(t) & -\lambda(t) \\ \lambda(t) & -(\mu_{2}(t) + \lambda(t)) & \mu_{3}(t) & 0 & \cdots & 0 & 0 \\ 0 & \lambda(t) & -(\mu_{3}(t) + \lambda(t)) & \mu_{4}(t) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -(\mu_{N+r-1}(t) + \lambda(t)) & \mu_{N+r}(t) \\ 0 & 0 & 0 & 0 & \cdots & \lambda(t) & -\mu_{N+r}(t). \end{pmatrix}$$
(17)

The solution to this equation can be represented in the following form:

$$\mathbf{p}(t) = U^*(t,0)\mathbf{p}(0) + \int_0^t U^*(t,\tau)\mathbf{g}(\tau) \,d\tau,$$
(18)

where $U^*(t, s)$ is the Cauchy operator of the corresponding homogeneous equation:

$$\frac{d\mathbf{x}}{dt} = B(t)\mathbf{x}.$$
(19)

Next, we will consider estimates in "weighted" norms. Suppose $d_1, d_2, ..., d_{N+r}$ are positive numbers. Then let

$$D = \begin{pmatrix} d_1 & d_1 & \cdots & d_1 \\ 0 & d_2 & \cdots & d_2 \\ & & & \\ & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{N+r} \end{pmatrix}$$
(20)

We will denote by $\|\mathbf{z}\|_{1\mathbf{D}} = \|\mathbf{D}\mathbf{z}\|_1$. Note that B(t) is essentially non-negative, that is, all off-diagonal elements of B(t) are non-negative for any $t \ge 0$. Then we get:

$$B^{**}(t) = DB(t)D^{-1} = \begin{pmatrix} -(\mu_1(t) + \lambda(t)) & \frac{d_1}{d_2}\mu_1(t) & 0 & \cdots & 0 & 0 \\ \frac{d_2}{d_1}\lambda(t) & -(\mu_2(t) + \lambda(t)) & \frac{d_2}{d_3}\mu_2(t) & \cdots & 0 & 0 \\ 0 & \frac{d_3}{d_2}\lambda(t) & -(\mu_3(t) + \lambda(t)) & \cdots & 0 & 0 \\ 0 & 0 & \frac{d_4}{d_3}\lambda(t) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\mu_{N+r-1}(t) + \lambda(t)) & \frac{d_{N+r-1}}{d_{N+r-1}}\mu_{N+r-1}(t) \\ 0 & 0 & 0 & \cdots & \frac{d_{N+r-1}}{d_{N+r-1}}\lambda(t) & -(\mu_{N+r}(t) + \lambda(t)). \end{pmatrix}$$
(21)

Let us put the following:

$$\gamma_{**}(t) = \inf_{i} \left(|b_{ii}(t)| - \sum_{j \neq i} \frac{d_j}{d_i} b_{ji}(t) \right).$$

$$(22)$$

Then we will get:

$$\gamma(B(t))_{1\mathsf{D}} = \gamma\left(\mathsf{D}B(t)\mathsf{D}^{-1}\right) = \sup_{i} \left(b_{ii}(t) + \sum_{j \neq i} \frac{d_j}{d_i} b_{ji}(t)\right) = -\gamma_{**}(t).$$
(23)

For some positive δ we put $d_1 = 1$, $d_{k+1} = \delta d_k$, $k \ge 1$.

3.3. Bounds on the Rate of Convergence for Large Service Rates

First, we let $\delta > 1$. Then we will get the following:

$$\alpha_{1}(t) = \mu_{1}(t) + \lambda(t) - \delta\lambda(t) = \mu(t) - (\delta - 1)\lambda(t),$$

$$\alpha_{k}(t) = \mu_{k}(t) + \lambda(t) - \frac{1}{\delta}\mu_{k-1}(t) - \delta\lambda(t) \ge \left(1 - \frac{1}{\delta}\right)\mu_{k-1}(t) - (\delta - 1)\lambda(t) \ge \left(1 - \frac{1}{\delta}\right)(\mu(t) - \delta\lambda(t)), \ 2 \le k < N + r,$$
(24)

$$\alpha_{N+r}(t) = \mu_{N+r}(t) - \frac{1}{\delta}\mu_{N+r-1}(t) + \lambda(t) \ge \left(1 - \frac{1}{\delta}\right)(\mu(t) - \delta\lambda(t)).$$

Therefore,

$$\gamma_{**}(t) = \min(\alpha_i(t)) = \left(1 - \frac{1}{\delta}\right) (\mu(t) - \delta\lambda(t)).$$
(25)

Theorem 1. Let there be a positive number delta > 1 such that,

$$\int_0^\infty (\mu(t) - \delta\lambda(t))dt = +\infty.$$
 (26)

Then the Markov chain X(t) is weakly ergodic and has the following convergence rate bounds:

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\|_{1D} \le e^{-\int_{0}^{t} (1 - \frac{1}{\delta})(\mu(t) - \delta\lambda(t)) d\tau} \|\mathbf{p}^{*}(0) - \mathbf{p}^{**}(0)\|_{1D},$$
(27)

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\| \leq 4\delta^{N+r}e^{-\int_{0}^{t} (1 - \frac{1}{\delta})(\mu(t) - \delta\lambda(t)) d\tau} \|\mathbf{p}^{*}(0) - \mathbf{p}^{**}(0)\| \leq 8\delta^{N+r}e^{-\int_{0}^{t} (1 - \frac{1}{\delta})(\mu(t) - \delta\lambda(t)) d\tau},$$
(28)

for any initial conditions $\mathbf{p}^*(0)$, $\mathbf{p}^{**}(0)$ and any $t \ge 0$.

We let that $W = \min_{k \ge 1} \frac{d_k}{k} = \min_{k \ge 0} \frac{\delta^k}{k+1}$. Then we will get $W \|\mathbf{p}\|_{1E} \le \|\mathbf{p}\|_{1D}$.

Corollary 1. From the conditions of Theorem 1, X(t) has a limit mean, then we say $\phi(t) = E(t, 0)$, and we obtain that the following estimate is true for any j and any $t \ge 0$:

$$|E(t,j) - E(t,0)| \le \frac{1 + \delta^{j-1}}{W} e^{-\int_{0}^{t} \left(1 - \frac{1}{\delta}\right)(\mu(t) - \delta\lambda(t)) \, d\tau}.$$
(29)

3.4. Bounds on the Rate of Convergence for Large Arrival Rates

Now consider the case $\delta < 1$ and assume that $\delta \in [\frac{r-1}{r}, 1)$. In this case, we have:

$$\alpha_1(t) \ge \mu(t) + (1 - \delta)\lambda(t),$$

$$\begin{aligned}
\alpha_k(t) &\geq (1-\delta)\lambda(t) + \mu_k(t) - \frac{1}{\delta}\mu_{k-1}(t) \geq \\
(1-\delta)\lambda(t) - \mu(t)\left(\frac{1}{\delta} - 1\right), \quad 2 \leq k \leq N+r-1 \\
\alpha_{N+r}(t) &\geq \lambda(t) - \left(\frac{1}{\delta} - 1\right)\mu(t).
\end{aligned}$$
(30)

Then it follows that we have the same:

$$\gamma_{**}(t) = \min(\alpha_i(t)) = \left(\frac{1}{\delta} - 1\right)(\delta\lambda(t) - \mu(t)).$$
(31)

Theorem 2. Let

$$\int_0^\infty (\delta\lambda(t) - \mu(t)) \, dt = +\infty,\tag{32}$$

for some $\delta \in [\frac{r-1}{r}, 1)$. Then, the Markov chain X(t) is weakly ergodic and has the following convergence rate bounds:

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\|_{1D} \le e^{-\int_{0}^{t} \left(\frac{1}{\delta} - 1\right)(\delta\lambda(t) - \mu(t)) d\tau} \|\mathbf{p}^{*}(0) - \mathbf{p}^{**}(0)\|_{1D'}$$
(33)

and

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\| \le 8\delta^{N+r}e^{-\int_{0}^{t} \left(\frac{1}{\delta} - 1\right)(\delta\lambda(t) - \mu(t))\,d\tau},$$
(34)

for any initial conditions $\mathbf{p}^*(0)$, $\mathbf{p}^{**}(0)$ and any $t \ge 0$. In addition, there is also a marginal mean and bound (29).

In addition, note the following: if the process is homogeneous (i.e., all intensities are constant), then the conditions of Theorems 1 and 2 are equivalent to the inequalities $\mu > \lambda$ and $\mu < \lambda$, respectively.

And it is also worth noting that when all intensities of the process are 1-periodic, then there is a weak ergodicity X(t) and the estimates of Theorem 1 or Theorem 2 if:

$$\int_0^1 \lambda(t) dt \neq \int_0^1 \mu(t) dt.$$
(35)

3.5. Perturbed CTMC and Bounds

In this subsection, we will consider the application of the general perturbation bounding in the same way as in the work [13]) for the models under study. We will consider a "perturbed" queue-length process $\bar{X}(t), t \ge 0$ with the corresponding transposed intensity matrix $\bar{A}(t)$, where the "perturbing" matrix $\hat{A}(t) = A(t) - \bar{A}(t)$ is small. That is, we assume that the perturbed queue is of the same nature as the original one. Then the perturbed intensity matrix also has the same structure with the corresponding perturbed intensities $\bar{\theta}(t), \bar{\gamma}(t), \bar{\lambda}(t)$. We assume that $\bar{\mu}_n(t) = \frac{C}{\bar{\theta}(t)} + \max(0, n - N)\bar{\gamma}(t)$ for $n \ge 1$.

We suppose that:

$$\left|\frac{1}{\theta(t)} - \frac{1}{\theta(t)}\right| = \left|\frac{1}{\dot{\theta}(t)}\right| \le \hat{\epsilon}, |\gamma(t) - \bar{\gamma}(t)| = |\hat{\gamma}(t)| \le \hat{\epsilon}, |\lambda(t) - \bar{\lambda}(t)| = |\hat{\lambda}(t)| \le \hat{\epsilon}.$$
(36)

Hence, we will get the following:

$$|\mu_{n}(t) - \bar{\mu}_{n}(t)| = |\hat{\mu}_{n}(t)| = |\frac{C}{\theta(t)} + \max(0, n - N)\gamma(t) - \frac{C}{\bar{\theta}(t)} - \max(0, n - N)\bar{\gamma}(t)|$$

$$\leq \left|\frac{C}{\theta(t)} - \frac{C}{\bar{\theta}(t)}\right| + \max(0, n - N)|(\gamma(t) - \bar{\gamma}(t))|$$

$$\leq C\hat{\epsilon} + r\hat{\epsilon} = (C + r)\hat{\epsilon}.$$
(37)

Then, from (8) we obtain the following bound:

$$\|\hat{A}(t)\| = 2\sup_{k} |\hat{a}_{kk}(t)| = 2\max(|\hat{\lambda}(t)|, |\hat{\lambda}(t)| + |\hat{\mu}_{n}(t)|, |\hat{\mu}_{N+r}(t)|) \le 2(C+r+1)\hat{\epsilon}.$$
 (38)

Now from Theorem 1 and Corollary 1 in the paper [13] the following bounds of the perturbation follow.

Theorem 3. Suppose that under the conditions of Theorem 1 or Theorem 2 the Markov chain X(t) is exponentially ergodic, that is,

$$e^{-\int_{s}^{\cdot} \gamma_{**}(\tau) d\tau} \le K e^{-\gamma_0(t-s)},$$
 (39)

for some positive K, γ_0 . Then the following bounds of the perturbation take place:

$$\limsup_{t \to \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \le \frac{2\hat{\epsilon}(C+r+1)(1+\log(4K)+(N+r)|\log(\delta)|)}{\gamma_0},\tag{40}$$

and

$$\limsup_{t \to \infty} |E(t,0) - \bar{E}(t,0)| \le \frac{2(N+r)\hat{\epsilon}(C+r+1)(1 + \log(4K) + (N+r)|\log(\delta)|)}{\gamma_0}, \quad (41)$$

for any perturbed queue with the respectively closed intensities satisfying to (36).

4. File Transfer in Wireless Network

4.1. Multi-Service Network

The system model of this work has the form of a cell, in the coverage area of which mobile devices are located (Figure 1). Each of them has its MSISDN (Mobile Subscriber Integrated Services Digital Number)—the mobile subscriber number of digital network with the integration of services. Each user behaves as follows: he sends a request to download a file, then downloads it, and also the user can disappear from the system. Disappearance can be associated with leaving the coverage area of the cell, or with a change in the type of service, or with the end of the service. The conclusion follows from the description of the system model: if we sum up the flow from all users, then the duration of the intervals between requests will not depend on the number of users, which is described by the Poisson flow of the first kind.



Figure 1. Scheme of System Model.

The users are provided with various services that belong to different categories of data transmission. A more detailed overview of the categories is illustrated in the Table 2. However, of these, we consider only those that can be described by elastic traffic, such as email, file transfer, and others.

N.	Service Name	Description
1	DB Transactions	Transaction in a database—for example, trans- fer of funds from a credit card through a mobile application.
2	File Systems	Work with remote file systems. This is how ATMs are arranged, for example. There is a machine with an operating system, and ATMs are connected to it and all information is stored there.
3	File Transfer	File transfer via FTP.
4	Games	Online game traffic.
5	Instant Messaging Applications	Instant messaging systems - messengers What- sApp, Viber, Telegram, etc.
6	Legacy Protocols	Deprecated protocols.
7	Mail	Email.
8	Music Streaming	Streaming services for listening to music.
9	Network Operation	Network services.
10	Others	Others
11	P2P Applications	Applications where data transmission is based on the principles of peer-to-peer networks. The simplest example is Bittorrent applications.
12	Security	Online video cameras, data from alarm sensors, etc.
13	Streaming Applications	Streaming services for watching movies and video-chats.
14	Terminals	Mobile terminals for credit card payments.
15	VoIP	IP-telephony. For example, calls via WhatsApp or Skype.
16	Web Applications	Web applications—client-server applications in which the client interacts with the server us- ing a browser (Microsoft Office Online, Google Documents).

Table 2. Description of service types.

4.2. Dataset Structure

The analysis of real traffic is of great importance, because the identification of the patterns of its arrival can make it possible to carry out studies of the system model described in the previous section in a non-stationary mode. Therefore, a task of this paper is formulated as follows: to analyze the traffic on one of the cell tower. There is a monitoring component to collect information about the system. Every hour, for each active device, the number of bits sent and received is added up. At the beginning of the next hour, the amount for the previous one hour is summed to the registration file. The principle of filling data during the monitoring is shown in Figure 2. There is a following designation $S_{d,t}^{up}$ —the sum of bits sent by the device per day d, at hour t. The part of the log file is shown in the Table 3. There are the column START_HOUR contains the full date and hour when the data was transferred; and the column MASKED_MSISDN—masked device identifier of the user; APP_CLASS—the class of the application that transfers data; UPLOAD—the number

of bits, which were sent as we denoted $S_{d,t}^{up}$; and the last DOWNLOAD—the number of received bits as we denoted $S_{d,t}^{down}$.



Figure 2. The principle of data accumulation during monitoring.

Table 3. Part of the log file.

START_HOUR	MASKED_MSISDN	APP_CLASS	UPLOAD	DOWNLOAD
12.02.2018 10:00	9BFA3001DE8C58C2453B6 9CE2E4A3704	Web Applications	111,339	78,208
12.02.2018 8:00	0C2C88351A593CD02727A 6207EB85E9E	Instant Messaging Applications	11,946	6741
12.02.2018 10:00	B9F45E1542162096408D9 4DD94499B89	Web Applications	67,143,832	4,230,675
12.02.2018 10:00	3CF6B81BC186E4C188D69 E1FE8919BB6	Web Applications	18,844	9769
11.02.2018 18:00	7EE1FCE60945D869A14EF 30E896E9131	Streaming Applications	17,219,335	403,361
11.02.2018 17:00	A746CCDCAC507B95C83 A3475C63C6BFD	Web Applications	235,455	131,712
11.02.2018 16:00	EA9D75DEB103F439A8C2 8300C2CC1757	Streaming Applications	136,767,000	4,078,817
12.02.2018 9:00	610FF30F1C9EC8939B4F BD259255BD63	Streaming Applications	3,330,234	112,395
12.02.2018 10:00	6DFD256C104B394A884E0 6A6EF7BE156	File Transfer	26,804,578	2,111,659
11.02.2018 17:00	6C9A6A5D444E7416C0C43 10E617EB611	Web Applications	587,786	87,897

4.3. Daily Traffic Profile

We consider elastic traffic, which is characterized by such a parameter as the length of the elastic data block. For the considered traffic model, monitoring data were taken with a class "File Transfer", which corresponds to the transfer of data using the FTP protocol.

To get the number of requests per second at a particular hour of the day for the selected application class, we perform aggregation of the form:

$$\lambda_{up}(t) = \frac{\sum_{d}^{\infty} S_{d,t}^{up}}{8 \cdot 1024 \cdot 1024 \cdot l} \tag{42}$$

$$\lambda_{down}(t) = \frac{\sum_{d}^{\infty} S_{d,t}^{down}}{8 \cdot 1024 \cdot 1024 \cdot l'},\tag{43}$$

where the received sum is determined in bits, d—date and l—average packet size, equal to 2 MB. Then, we perform the calculation on Formulas (42) and (43) and the results are shown in Table 4.

Times of Day, t	$\lambda_{up}(t)$	$\lambda_{down}(t)$
0	0.484847494	0.040711044
1	0.341865943	0.031079608
2	0.190823262	0.018468431
3	0.132143012	0.013982313
4	0.118697844	0.011683847
5	0.105803565	0.009923712
6	0.116474526	0.012585431
7	0.242942093	0.019329354
8	0.548677837	0.045828017
9	0.882484275	0.062668489
10	0.916328538	0.091028611
11	0.920243822	0.087240118
12	0.814816002	0.08172834
13	0.710077612	0.0853086
14	0.820103586	0.095698016
15	0.879327048	0.090266848
16	0.951204391	0.098226292
17	0.945908433	0.09180442
18	0.990004516	0.082148071
19	0.991890388	0.074312104
20	0.907643325	0.067218977
21	0.957584488	0.066221158
22	0.926637206	0.112652817
23	0.755175635	0.067874847

Table 4. Dependence of the number of requests on the time of day.

4.4. Fourier Series Approximation

From the found values it is necessary to obtain some continuous function of the dependence of the fluctuation of the upward and downward traffic flows on the time of day. To do this, we perform approximation by the Fourier series. For clarity, we will gradually increase the number of conditions in a row, which is to choose the most optimal option.

Consider the dependence of the intensity of the upward data flow λ_{up} from time, and we will carry out the approximation by the Fourier series with one, two, and three conditions. As a result, we get a functions of the form (44)–(46) respectively and for our example the parameters will take values, which are shown in Table 5:

$$a(t) = a_0 + a_1 \cos(wt) + b_1 \sin(wt)$$
(44)

$$a(t) = a_0 + a_1 \cos(wt) + b_1 \sin(wt) + a_2 \cos(2wt) + b_2 \sin(2wt)$$
(45)

$$a(t) = a_0 + a_1 cos(wt) + b_1 sin(wt) + a_2 cos(2wt) + b_2 sin(2wt) + a_3 cos(3wt) + b_3 sin(3wt).$$
(46)

Table 5. Parameters of Fourier Series Approximation for the upward data flow.

Condition	w	<i>a</i> ₀	<i>a</i> ₁	b_1	<i>a</i> ₂	b_2	<i>a</i> ₃	<i>b</i> ₃
one	0.2519	0.6472	-0.168	-0.3675	_	_	_	_
two	0.2617	0.6521	-0.07656	-0.3992	0.1479	-0.1348	_	_
three	0.2603	0.6521	-0.08888	-0.3963	0.1384	-0.1411	-0.01446	0.05875

Using Matlab, we have built the plots of the considered approximation. It is easy to see that the plot of the obtained approximation for the one condition (Figure 3a) differs significantly from the initial data, therefore, so that is why we increased the number of conditions to two and the graph of the resulting function for two conditions (Figure 3b) significantly reflects more faithfully the relationship between real data. Then, according to the plot of the approximation for the three conditions (Figure 3c), it could be seen that, now, in the intervals with the highest network load, the function has become closer to the real data points. However, in general, the graph did not receive significant changes, and the variant of the approximation with two conditions can be considered the most optimal.



Figure 3. (a) Graph of the approximation function a(t) for upward flow with one condition; (b) Graph of the approximation function a(t) for upward flow with two conditions; (c) Graph of the approximation function a(t) for upward flow with three conditions.

4.5. Numerical Analysis

For the numerical analysis, we will consider the following example: the volume of the resource block is C = 100 Mbps, and the finite capacity drive is r = 100. The size of the transferred file equals to $\theta(t) = \theta = 10$ MB, that is 80 Mb, and the minimum transfer rate is b = 1 Mbps. The arriving intensity of device equals to $\lambda(t) = \lambda \cdot a(t)$, where according to Section 4.4 a(t) equals to (46), where parameters are indicated in the Table 5 on the line for the third condition. The intensity of the flow of leaving requests for the transfer of a block of elastic data due to "impatience" is $\gamma(t) = \gamma = 10^{-2}$.

Apply all our bounds for this specific situation.

For applying of Theorems 2 and 3 we put $\delta = 0.99$, $d_0 = 1$ and $d_{k+1} = \delta d_k$ for $k \ge 0$. Then, we will get:

$$\gamma_{**}(t) = \frac{1}{99} (2.97(a_0 + a_1 \cos(wt) + b_1 \sin(wt) + a_2 \cos(2wt) + b_2 \sin(2wt) + a_3 \cos(3wt) + b_3 \sin(3wt)) - 1.25), \quad (47)$$

$$e^{-\int_{s}^{t} \gamma_{**}(\tau) \, d\tau} \le e^{-0.016(t-s)}; \tag{48}$$

therefore, one can get $K = 2 \cdot 10^6$ and $\gamma_0 = 0.016$ in (39).

Now we obtain the following bounds on the rate of convergence:

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\| \le 3 \cdot 10^{6} \cdot e^{-0.016t} \|\mathbf{p}^{*}(0) - \mathbf{p}^{**}(0)\|,$$
(49)

from Theorem 2;

$$\|\mathbf{p}^{*}(t) - \mathbf{p}^{**}(t)\|_{1\mathsf{D}} \le 2 \cdot 10^{6} \cdot e^{-0.016t} \|\mathbf{p}^{*}(0) - \mathbf{p}^{**}(0)\|_{1\mathsf{D}},$$
(50)

$$|E(t,j) - E(t,0)| \le \frac{1 + 0.99^{j-1}}{W} e^{-0.016t},$$
(51)

from Theorem 2 and Corollary 1.

The corresponding perturbation bounds are:

$$\limsup_{t \to \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \le 5 \cdot 10^5 \hat{\epsilon},\tag{52}$$

and

$$\limsup_{t \to \infty} |E(t,0) - \bar{E}(t,0)| \le 10^8 \hat{\epsilon},\tag{53}$$

from Theorem 3.

To solve the Cauchy problem, the 4th order Adams-Multon method was used with the use of IntelliJ IDEA software, JDK and the JFreeChart library, the functionality of which is used for plotting. The convergence plots were built for the characteristics specified in Section 2.3, shown in Figures 4–6. To analyze the characteristics, two scenarios were chosen: 1. At the initial moment of time, our system is empty: X(0) = 0; 2. At the initial moment, the system is completely occupied, that is, all devices are occupied and there are no free places in the queue X(0) = 200. Figure 4 shows graphs for the blocking probability. Note that, since we are considering two scenarios, then according to Figure 4a. for the first scenario, the starting value of the probability is 0, since the system is empty and it is logical that there will be no locks, and for the second scenario, the starting value is 1 since the system is completely busy. We can notice that the blocking probability for the two scenarios converges at time t = 500 and according to Figure 4b. We see that its values fluctuate within the range of 0.000–0.013, and the period of fluctuation is 24. Further, in Figure 5, we see that the average number of service requests converges at 100; this may indicate a high load of devices in our system. However, the probability of blocking is small. Let us look at the average number of applications in the queue shown in Figure 6. As shown in Figure 6a. for the two scenarios, the graphs converge and the mean value is no less than 64 and no more than 75, as we can see in Figure 6b.



Figure 4. (a) Probability of blocking for $t \in [0, 1304]$. (b) Approximation of the limiting probability of blocking for $t \in [1304, 1329]$.



Figure 5. (a) The mean of applications served E(t,k) for $t \in [0, 1304]$. (b) Approximation of the limiting mean of applications served E(t,k) for $t \in [1304, 1329]$.



Figure 6. (a) The mean of applications in the queue Q(t) for $t \in [0, 1304]$. (b) Approximation of the limiting mean of applications in the queue Q(t) for $t \in [1304, 1329]$.

5. Conclusions

In this paper, we have investigated the process of downloading user files depending on the time in the form of a queuing system with elastic traffic and non-stationary intensities. We have developed a method for calculating the limits of convergence of such a model. Estimates of the rate of convergence are obtained based on the logarithmic norm of linear operators. As a result, it was found that the rate of convergence turned out to be low enough for an adequate representation of the situation in the limiting mode to begin at sufficiently long times. We evaluated the characteristics of such a model, namely, the probability of blocking new users, the average number of users downloading data, and the average number of users waiting for the download to start. We obtained the upper and lower bounds of their values.

We considered the model in the form of a single network slice and, in the framework of further tasks, we can consider the model in the form of several slices. If we want to scale our system, for example, to add another incoming stream, it will be necessary to solve the problem of redistribution; we will have to adapt our method for the new case. In addition, in our case, we have a one-dimensional random process, and for a new system where it will be multidimensional, we will first need to identify a function—a mapping, to go to the one-dimensional case to apply our method. This complexity can be viewed as a challenge for future research.

Author Contributions: Conceptualization, supervision, A.Z. and I.K. (Irina Kochetkova); methodology, Y.S. and I.K. (Irina Kochetkova); software, validation, visualization I.K. (Ivan Kovalev) and E.M.; investigation, writing, I.K. (Irina Kochetkova), Y.S., Ivan Kovalev, E.M., A.C., A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Russian Science Foundation, grant number 19-11-00020 (recipients I.K. (Irina Kochetkova), Y.S., I.K. (Ivan Kovalev), A.Z., Sections 3–5). This paper has been supported by the RUDN University Strategic Academic Leadership Program (recipients A.C., E.M., Sections 1 and 2).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are sincerely grateful to Luis M. Correia (IST/INESC-ID, University of Lisbon) for providing the dataset for Section 4.

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

References

- 1. Kochetkov, D.; Almaganbetov, M. Using patent landscapes for technology benchmarking: A case of 5g networks. *Adv. Syst. Sci. Appl.* **2021**, *21*, 20–28. [CrossRef]
- Kochetkov, D.; Vuković, D.; Sadekov, N.; Levkiv, H. Smart cities and 5G networks: An emerging technological area? J. Geogr. Inst. Jovan Cvijic SASA 2019, 69, 289–295. [CrossRef]
- 3. Barakabitze, A.A.; Ahmad, A.; Mijumbi, R.; Hines, A. 5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges. *Comput. Netw.* **2020**, *167*, 106984. [CrossRef]
- 4. Khan, L.U.; Yaqoob, I.; Tran, N.H.; Han, Z.; Hong, C.S. Network Slicing: Recent Advances, Taxonomy, Requirements, and Open Research Challenges. *IEEE Access* 2020, *8*, 36009–36028. [CrossRef]
- Muhizi, S.; Ateya, A.A.; Muthanna, A.; Kirichek, R.; Koucheryavy, A. A novel slice-oriented network model. *Commun. Comput. Inf. Sci.* 2018, 919, 421–431. [CrossRef]
- 6. Ageev, K.A.; Sopin, E.S.; Yarkina, N.V.; Samouylov, K.E.; Shorgin, S.Y. Analysis of the network slicing mechanisms with guaranteed allocated resources for various traffic types. *Inform. Primen.* **2020**, *14*, 94–100. [CrossRef]
- 7. Yarkina, N.; Gaidamaka, Y.; Correia, L.M.; Samouylov, K. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation. *Mathematics* **2020**, *8*, 1177. [CrossRef]
- 8. Kochetkova, I.; Vlaskina, A.; Burtseva, S.; Savich, V.; Hosek, J. Analyzing the effectiveness of dynamic network slicing procedure in 5g network by queuing and simulation models. *Lect. Notes Comput. Sci.* **2020**, *12525*, 71–85. [CrossRef]
- 9. Vlaskina, A.; Polyakov, N.; Gudkova, I. Modeling and Performance Analysis of Elastic Traffic with Minimum Rate Guarantee Transmission under Network Slicing. *Lect. Notes Comput. Sci.* 2019, 11660, 621–634. [CrossRef]

- Vlaskina, A.S.; Polyakov, N.A.; Gudkova, I.A.; Gaidamaka, Y.V. Performance analysis of elastic traffic with minimum bit rate guarantee transmission in wireless network under network slicing, Izvestiya of Saratov University. *Informatics* 2020, 20, 378–387. [CrossRef]
- Gudkova, I.; Korotysheva, A.; Zeifman, A.; Shilova, G.; Korolev, V.; Shorgin, S.; Razumchik, R. Modeling and analyzing licensed shared access operation for 5G network as an inhomogeneous queue with catastrophes. In Proceedings of the International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, Lisbon, Portugal, 18–20 October 2016; pp. 282–287. [CrossRef]
- 12. Markova, E.; Satin, Y.; Kochetkova, I.; Zeifman, A.; Sinitcina, A. Queuing system with unreliable servers and inhomogeneous intensities for analyzing the impact of non-stationarity toperformance measures of wireless network under licensed shared access. *Mathematics* **2020**, *8*, 800. [CrossRef]
- 13. Zeifman, A.; Korolev, V.; Satin, Y. Two approaches to the construction of perturbation bounds for continuous-time Markov chains. *Mathematics* **2020**, *8*, 253. [CrossRef]
- 14. Zeifman, A.; Satin, Y.; Kryukova, A.; Razumchik, R.; Kiseleva, K.; Shilova, G. On the Three Methods for Bounding the Rate of Convergence for some Continuous-time Markov Chains. *arXiv* 2020, arXiv:1911.04086.
- 15. Arns, M.; Buchholz, P.; Panchenko, A. On the numerical analysis of inhomogeneous continuous-time Markov chains. *Informs J. Comput.* 2010, 22, 416–432. [CrossRef]
- 16. Andreychenko, A.; Sandmann, W.; Wolf, V. Approximate adaptive uniformization of continuous-time Markov chains. *Appl. Math. Model.* **2018**, *61*, 561–576. [CrossRef]
- 17. Katoen, J.-P.; Zapreev, I.S. Safe on-the-fly steady-state detection for time-bounded reachability. In Proceedings of the 3rd International Conference on the Quantitative Evaluation of Systems, California, CA, USA, 11–14 September 2016.