Resource Allocation and Sharing for Transmission of Batched NB IoT Traffic over 3GPP LTE

Sergey Stepanov*, Mikhail Stepanov*, Ariunaa Tsogbadrakh*, Juvent Ndayikunda*, Umer Andrabi†
* Moscow Technical University of Communications and Informatics, Moscow, Russia
† Moscow Institute of Physics and Technology (State University), Moscow, Russia
stpnvsrg@gmail.com, m.s.stepanov@mtuci.ru, ariunaa-ariun@mail.ru, juvndayi@mail.ru, umer.andrabi@rediffmail.com

***Abstract*—The model of resource allocation and sharing for conjoint servicing of real time and batched NB IoT traffic streams over LTE cell facilities is constructed. In the model three types of traffic flows are considered. The first two represent the flows of sessions coming from wireless video surveillance cameras. Depending on the number of traffic sources the flows are described by Poisson or Engset models. The resource of the cell is also used by flow of sessions coming from NB-IoT sensors. This flow follows Poisson model with batch arrivals and possibility of waiting if available radio resource is occupied. The number of waiting places and maximum allowed time of waiting are restricted. All random variables used in the model have exponential distribution with corresponding mean values. Using the model the main performance measures of interest are given with help of values of probabilities of model's stationary states. The model and derived algorithms of characteristics estimation can be used for study the scenarios of resource sharing between LTE and NB-IoT traffic flows.**

I. Introduction

The Internet of Things (IoT) has developed tremendously over the past few years. Billions of unattended devices (sensors, video surveillance cameras, etc) are collecting an enormous amount of data and send the collected data to analytical centers [1]–[8]. The necessity of processing large volumes of varied data (Big Data) has raised significant challenges for the researchers. The main problem is to organize the effective usage of restricted cell transmission resources for conjoint transmission of low-cost and low-traffic produced by smart meters and multimedia big data collected by video surveillance systems deployed for safety reasons [2], [3].

Technically the process of servicing the combination of two or more heterogeneous big data streams has been supported by activities of 3GPP after validation of a dedicated in-band deployment mode in LTE Rel. 13. This document describes the rules how LTE technology can be used for conjoint servicing the traffic streams produced by video surveillance cameras and sensor traffic streams generated by NB-IoT technology.

Unfortunately 3GPP just provides the flexibility that how the radio resources can be shared between LTE and NB-IoT technologies, but does not provide concrete solutions on how these resources should be shared. This problem should be solved by mathematical modeling with taking into account the peculiarity of traffic streams forming and characteristics of conjoint servicing the heterogeneous data streams.

The main aim of this paper is to address the above mentioned challenges by constructing an analytical framework for modeling the process of resource sharing for conjoint servicing of real time and batched NB IoT traffic streams over LTE cell facilities. LTE-based network used by surveillance operator to establish monitoring services by using video surveillance cameras (LTE-devices) and NB-IoT sensors (NB-IoT-devices). Individual data stream of one LTE or NB-IoT-device consist of sequence of sessions [2], [3]. The proposed model is generalized the results of [2], [3] by taking into account the dependence of sessions coming from LTE-devices on the number of traffic sources and the possibility of waiting for blocked sessions of NB-IoT-devices coming by batches. The possibility of waiting based on the property to tolerate in some degree the delay in servicing of sessions coming from NB-IoT-devices. Technically it can be provided by using the relay stations. It allows to decrease the necessary amount of resource needed to serve NB-IoT traffic with given values of performance measures. The model and derived algorithms of performance measures estimation are based on results of [9]–[15] and can be used for study the scenarios of resource sharing between LTE and NB-IoT traffic flows. Two scenarios will be considered: Static, when resources are strictly divided among LTE and NB-IoT traffic and Dynamic, when resources are fully shared.

The rest of the paper is organized as follows. In Section II the mathematical description of the model will be presented. Here the system of state equations that relates the model's stationary probabilities is outlined and main performance measures and algorithm of their estimation will be defined. Numerical assessment is performed in Section III. Conclusions are drawn in the last section. The process of sessions forming by LTE and of NB-IoT devices is shown at Fig. 1.

II. Mathematical description of the model

A. Structural components

Let us consider an LTE cell with a base station placed in its center. The total amount of available radio resource of LTE cell in uplink direction is measured in units of its smallest granularity. We call one of such units as a channel or resource unit. Let us suppose that total amount of resource units is a linear function of the number of resource blocks (RB). There is some number of LTE-devices, representing video surveillance cameras, and quite a big number of NB-IoT-devices, representing a great variety of sensors, actuators and smart meters, located in the cell and connected to the corresponding base station.

In the model we consider two flows of sessions representing the traffic of LTE-devices and one flow of sessions

ISSN 2305-7254

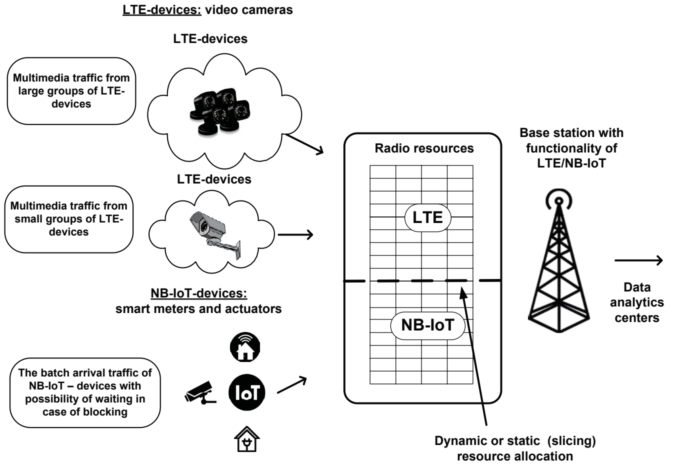

Fig. 1. The model of sessions forming by LTE and of NB-IoT devices

representing the traffic of NB-IoT-devices. Let us denote by v, the total number of resource units and by c denote the transmission speed provided by one unit. As it was supposed earlier the value of c usually equals to the minimum transmission requirement for servicing of one session of NB-IoT-devices. Let us suppose that the first flow of sessions is generated by large population of LTE-devices. Because of this assumption the changing of their number doesn't influence greatly the intensity of sessions coming. In this case we can use the poissonian model when describing the process of sessions coming. Let us denote for this type of flow by λ_1 the intensity of sessions arriving, α_1 denotes the parameter of exponentially distributed service time of one session, b_1 denotes the number of resource units used for servicing of one session and a_1 denotes the intensity of offered traffic expressed in potential number of connections (erlangs). The following relation is true $a_1 = \frac{\lambda_1}{\alpha_1}$.

Second flow of sessions originated from LTE-devices require a large amount of cell transmission capacity. Usually these are sessions generated by video cameras of improved quality. Number of such kind of LTE-devices is comparatively small. In this case Engset model is used to describe the sessions coming. Let us denote for this type of flow by β_2, the parameter of exponentially distributed time between successive sessions arrival, s_2 denotes the number of users, α_2 denotes the parameter of exponentially distributed service time of one session, b_2 denotes the number of resource units used for servicing of one session and a_2 denotes the intensity of offered traffic expressed in potential number of connections. The following relation is true $a_2 = \frac{s_2\beta_2}{\alpha_2+\beta_2}$.

In order to simplify the model it is supposed that blocked LTE-device sessions are lost without repeating an attempt.

Sessions for servicing the traffic of NB-IoT-devices arrive in batches according to poissonian model with intensity λ_d. With probability f_j, $j = 1, \ldots, g$, the arriving batch has $j \leq v + w$ sessions, where w is the number of waiting positions for blocked NB-IoT-device sessions. The arriving batch having j sessions is taken for servicing entirely if $j \leq v - i$, where i is the number of resource units occupied for servicing the traffic of LTE-devices sessions. If $j > v - i$, then $(v - i)$ sessions are accepted for servicing, the rest $(j + i - v)$ sessions are accepted for waiting entirely if $j + i - v \leq w$. If $j + i - v > w$ then w sessions are accepted for waiting, the rest $(j + i - v - w)$ sessions are lost without resuming.

The volume of the transmitted file has exponential distribution with mean value of F bits. The time of servicing of one session from NB-IoT-devices by one resource unit has exponential distribution with parameter $\mu_d = \frac{F}{c}$. The maximum allowed waiting time for sessions occupying the waiting position is restricted by random time having exponential distribution with parameter equals to σ. Mathematical description of the model is shown in Fig. 2.

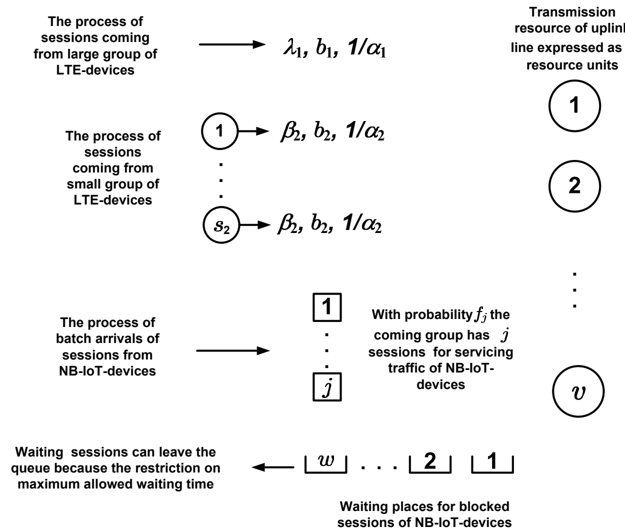

Fig. 2. The model of conjoint servicing of real time and batched NB IoT traffic streams over LTE cell

B. Marcov process

Let us denote by $i_1(t)$ and $i_2(t)$ correspondingly the number of active sessions of LTE-devices of the first and second flows being on servicing at time t, and by $d(t)$ we denote the number of active sessions of NB-IoT-devices being on servicing at time t. The dynamic of a model states changing is described by multidimensional Markov process with components

$$r(t) = (i_1(t), i_2(t), d(t)),$$

defined on the finite set of model's states S. Let us denote by (i_1, i_2, d) the state of $r(t)$, and by S denote the set of all possible states.

The vector (i_1, i_2, d) belongs to S when i_1, i_2, d varies as follows

$$i_1 = 0, 1, \ldots, \min(s_1, \left\lfloor \frac{v}{b_1} \right\rfloor); \quad (1)$$

$$i_2 = 0, 1, \ldots, \left\lfloor \frac{v - i_1 b_1}{b_2} \right\rfloor);$$
$$d = 0, 1, \ldots, v + w - i_1 b_1 - i_2 b_2.$$

Let us denote by $p(i_1, i_2, d)$ the value of stationary probability of state $(i_1, i_2, d) \in S$. It can be interpreted as portion of time the model stays in the state (i_1, i_2, d). This interpretation gives the possibility to use the values of $p(i_1, i_2, d)$ for estimation of model's main performance measures.

C. System of state equations

System of state equations is obtained after equating the intensity of transition $r(t)$ out of the arbitrary model's state (i_1, i_2, d) to the intensity of transition $r(t)$ into the state (i_1, i_2, d). Let us denote for the state $(i_1, i_2, d) \in S$ by i the total number of resource units occupied for transmission of traffic originated by LTE-devices $i = i_1 b_1 + i_2 b_2$. By using the indicator function we represent all equations of the system of state equations in one relation [9],[10]. It will be shown latter that this is very convenient for realizing the iterative algorithms of solving the corresponding system.

The system of state equations can be written as follows

$$P(i_1, i_2, d)\Big\{ \Big(\lambda_1 I(i + b_1 \le v) + i_1 \alpha_1 I(i_1 > 0)\Big) + \quad (2)$$
$$+ \Big((s_2 - i_2)\beta_2 I(i + b_2 \le v) + i_2 \alpha_2 I(i_2 > 0)\Big) +$$
$$+ \lambda_d \Big(I(i + d + 1 \le v) + I(i + d + 1 > v, i + d - v < w)\Big) +$$
$$+ \alpha_d \Big(d I(d > 0, i + d \le v) + (v - i) I(d > 0, i + d > v)\Big) +$$
$$+ (i + d - v)\sigma I(i + d - v > 0)\Big\} =$$
$$= P(i_1 - 1, i_2, d)\lambda_1 I(i_1 > 0, i + d \le v) +$$
$$+ P(i_1, i_2 - 1, d)(s_2 - i_2 + 1)\beta_2 I(i_2 > 0, i + d \le v) +$$
$$+ \sum_{\ell=1}^{d} P(i_1, i_2, d - \ell)\lambda_d \times$$
$$\times \left(f_\ell + I\Big(i + d = v + w\Big) \sum_{j=\ell+1}^{g} f_j \right) +$$
$$+ P(i_1 + 1, i_2, d)(i_1 + 1)\alpha_1 \Big(I(i + d + b_1 \le v) +$$
$$+ I(i + d + b_1 > v, i + b_1 \le v)\Big) +$$
$$+ P(i_1, i_2 + 1, d)(i_2 + 1)\alpha_2 \times$$
$$\times \Big(I(i + d + b_2 \le v, i_2 + 1 \le s_2) +$$
$$+ I(i + d + b_2 > v, i_2 + 1 \le s_2, i + b_2 \le v)\Big) +$$
$$+ P(i_1, i_2, d + 1)\Big((d + 1)\alpha_d I(i + d + 1 \le v) +$$
$$+ \big((v - i)\alpha_d + (i + d + 1 - v)\sigma\big) I(i + d + 1 > v, i + d + 1 - v \le w)\Big).$$

Values $P(i_1, i_2, d)$ satisfy the normalizing condition

$$\sum_{(i_1, i_2, d) \in S} P(i_1, i_2, d) = 1.$$

D. Performance measures

The model performance measures depend on the type of the session considered and can be defined by summing probabilities $p(i_1, i_2, d)$ over corresponding subsets of S. Let us denote for the state (i_1, i_2, d) by i the total number of resource units occupied by active sessions of LTE-devices $i = i_1 b_1 + i_2 b_2$. Let us define for the k-th flow of sessions of LTE-devices $k = 1, 2$ by $\pi_{t,k}$ the portion of time when available number of free resource units are insufficient for excepting of a session of k-th flow, $\pi_{c,k}$ denotes the portion of lost sessions of k-th flow, $\pi_{\ell,k}$ denotes the portion of lost traffic of k-th flow, Λ_k denotes the intensity of coming sessions of k-th flow, m_k denotes the mean number of resource units occupied by servicing the sessions of k-th flow and y_k denotes the mean number of sessions of k-th flow being on servicing. The introduced characteristics are define in the following way

$$\pi_{t,k} = \sum_{\{(i_1, i_2, d) \in S \mid i + d + b_k > v\}} p(i_1, i_2, d); \quad k = 1, 2;$$
$$\pi_{c,1} = \pi_{t,1};$$
$$\pi_{c,2} = \frac{\sum_{\{(i_1, i_2, d) \in S \mid i + d + b_2 > v\}} p(i_1, i_2, d)(s_2 - i_2)\beta_2}{\sum_{\{(i_1, i_2, d) \in S\}} p(i_1, i_2, d)(s_2 - i_2)\beta_2};$$
$$\Lambda_1 = \lambda_1;$$
$$\Lambda_2 = \sum_{\{(i_1, i_2, d) \in S\}} p(i_1, i_2, d)(s_2 - i_2)\beta_2;$$
$$m_k = \sum_{(i_1, \ldots, i_n, d) \in S} p(i_1, \ldots, i_n, d) i_k b_k; \quad k = 1, 2;$$
$$y_k = \sum_{(i_1, \ldots, i_n, d) \in S} p(i_1, \ldots, i_n, d) i_k; \quad k = 1, 2;$$
$$\pi_{\ell,1} = \pi_{t,1};$$
$$\pi_{\ell,2} = \frac{a_2 - y_2}{a_2}.$$

Let us introduce the definitions of performance measures of NB-IoT-device sessions servicing. The mean number b_d of sessions in one group is defined by relation

$$b_d = \sum_{j=1}^{v+w} f_j \, j.$$

The portion $\pi_{d,1}$ represents the sessions lost because at the moment of session coming the available number of free resource units and waiting positions is insufficient for excepting of a session is defined by relation

$$\pi_{d,1} = \frac{1}{\lambda_d b_d} \times \left(\sum_{\{(i_1, i_2, d) \in S \mid i + d = v + w\}} \sum_{\ell=0}^{d} p(i_1, i_2, d - \ell) \times \right.$$
$$\left. \times \lambda_d \sum_{j=\ell+1}^{g} f_j \, (j - \ell) \right).$$

The portion $\pi_{d,2}$ of sessions lost after the excess of maximum allowed waiting time is defined by relation

$$\pi_{d,2} = \frac{1}{\lambda_d b_d} \sum_{\left\{(i_1,i_2,d)\in S \mid i+d>v\right\}} p(i_1,i_2,d)(i+d-v)\sigma.$$

The portion π_d of sessions lost due the all reasons considered in the model is defined by relation

$$\pi_d = \pi_{d,1} + \pi_{d,2}.$$

The mean number y_s of sessions being on service is defined by the following expression

$$y_s = \sum_{\left\{(i_1,i_2,d)\in S \mid i+d\le v, d>0\right\}} p(i_1,i_2,d)d + \sum_{\left\{(i_1,i_2,d)\in S \mid i+d>v\right\}} p(i_1,i_2,d)(v-i).$$

The mean number y_w of sessions being on waiting is defined by relation

$$y_w = \sum_{\left\{(i_1,i_2,d)\in S \mid i+d>v\right\}} p(i_1,i_2,d)(i+d-v).$$

The mean number y_d of sessions being in the system on servicing or waiting is defined by formula

$$y_d = y_s + y_w.$$

The mean time T_d of session being in the system on servicing or waiting is defined by relation

$$T_d = \frac{y_d}{\lambda_d b_d (1-\pi_{d,1})}.$$

The introduced performance measures are related by conservation laws given below

$$\lambda_1 = \lambda_1 \pi_{t,1} + y_1\alpha_1; \qquad (3)$$

$$\Lambda_2 = \Lambda_2 \pi_{c,2} + y_2\alpha_2;$$

$$\lambda_d b_d = \lambda_d \pi_{d,1} b_d + y_s\alpha_d + y_w\sigma.$$

They can be proved by algebraic transforms of the system of state equations (2) or by using the Little formula. To calculate the performance measures according formulated definitions it is necessary to solve the system of state equations (2).

E. Solution of the system of state equations: general case

In general case the system of state equations (2) doesn't have any special features that simplify the solution. Because of this reason the values of $p(i_1,i_2,d)$ can be found by Gauss-Zeidel iterative algorithm [9],[10]. Let us describe main steps of the algorithm. Let us define by $p^{(j)}(i_1,i_2,d)$ approximation number j for unnormalized values of $p(i_1,i_2,d)$ obtained by Gauss-Zeidel iterative algorithm, $(i_1,i_2,d)\in S$. Let us denote by $L(i_1,i_2,d)$ the coefficient of $p^{(j)}(i_1,i_2,d)$ in the left part of (2). For each state $(i_1,i_2,d)\in S$ it defines in the following way

$$L(i_1,i_2,d) = \Big\{\Big(\lambda_1 I(i+b_1\le v) + i_1\alpha_1 I(i_1>0)\Big) + \qquad (4)$$

$$+\Big((s_2-i_2)\beta_2 I(i+b_2\le v) + i_2\alpha_2 I(i_2>0)\Big) +$$

$$+\lambda_d\Big(I(i+d+1\le v) + I(i+d+1>v, i+d-v<w)\Big) +$$

$$+\alpha_d\Big(dI(d>0, i+d\le v) + (v-i)I(d>0, i+d>v)\Big) +$$

$$+(i+d-v)\sigma I(i+d-v>0)\Big\}.$$

In accordance with definition of Gauss-Zeidel iterative algorithm [9],[10] the components of approximation number $(j+1)$ for $p^{(j+1)}(i_1,i_2,d)$ are found from known components of approximations number (j) and $(j+1)$ with following relations

$$p^{(j+1)}(i_1,i_2,d) = \frac{1}{L(i_1,i_2,d)}\times \qquad (5)$$

$$\times\Bigg\{P^{(j,j+1)}(i_1-1,i_2,d)\lambda_1 I(i_1>0, i+d\le v) +$$

$$+P^{(j,j+1)}(i_1,i_2-1,d)(s_2-i_2+1)\beta_2 I(i_2>0, i+d\le v) +$$

$$+\sum_{\ell=1}^{d} P(i_1,i_2,d-\ell)\lambda_d \times$$

$$\times\left(f_\ell + I\Big(i+d=v+w\Big)\sum_{j=\ell+1}^{g} f_j\right) +$$

$$+P^{(j,j+1)}(i_1+1,i_2,d)(i_1+1)\alpha_1\Big(I(i+d+b_1\le v) +$$

$$+I(i+d+b_1>v, i+b_1\le v)\Big) +$$

$$+P^{(j,j+1)}(i_1,i_2+1,d)(i_2+1)\alpha_2\times$$

$$\times\Big(I(i+d+b_2\le v, i_2+1\le s_2) +$$

$$+I(i+d+b_2>v, i_2+1\le s_2, i+b_2\le v)\Big) +$$

$$+P^{(j,j+1)}(i_1,i_2,d+1)\Big((d+1)\alpha_d I(i+d+1\le v) +$$

$$+\Big((v-i)\alpha_d + (i+d+1-v)\sigma\Big)I(i+d+1>v, i+d+1-v\le w)\Big)\Bigg\}.$$

Upper index in $P^{(j,j+1)}(i_1,i_2,d)$ means usage in process of calculation the already found approximation $P^{(j+1)}(i_1,i_2,d)$ and if it unknown to this moment then the usage of known approximation $P^{(j)}(i_1,i_2,d)$. The initial approximation can be found from relations

$$P^{(0)}(i_1,i_2,d) = 1, \qquad (i_1,i_2,d)\in S.$$

At each step of iterative algorithm the convergence is checked with help of relation

$$\frac{\sum\limits_{(i_1,i_2,d)\in S}\Big(|P^{(j+1)}(i_1,i_2,d)-P^{(j)}(i_1,i_2,d)|\Big)}{\sum\limits_{(i_1,i_2,d)\in S}P^{(j+1)}(i_1,i_2,d)}\le\varepsilon,$$

where ε is taken from interval $10^{-6}\dots10^{-10}$.

The formulated version of algorithm in some cases (very rarely) may lose the convergence property. In this case it is necessary to construct the convergent version of Gauss-Zeidel iterative algorithm. For this purpose it is sufficient to put one of unknown probabilities $P(i_1,i_2,d)$ in (2) to 1 and after to solve obtained heterogeneous system of linear equations by ordinary Gauss-Zeidel iterative algorithm (4), (5). It is necessary to mention that in this case the number of iterations required for convergence is strongly decreased compare to (4), (5).

F. Solution of the system of state equations: particular cases

1) Absence of NB-IoT traffic: In the initial model this situation corresponds to the case when $\lambda_d=0$. Let us denote by $p(i_1,i_2)$ the stationary probability of the state (i_1,i_2). Product form solution is valid for $p(i_1,i_2)$. It means that the following relation is true

$$p(i_1,i_2)=\frac{1}{N}\times\frac{a_1^{i_1}}{i_1!}\,\frac{\prod\limits_{j=0}^{i_2-1}(s_2-j)\gamma_2^{i_2}}{i_2!},\qquad(6)$$

where $a_1=\frac{\lambda_1}{\alpha_1}$ is offered load of the first flow, $\gamma_2=\frac{\beta_2}{\gamma_2}$ is the mean number of sessions of second flow coming during the mean time of serving of one session of the second flow and N is the normalizing constant

$$N=\sum_{(i_1,i_2)\in S}\frac{a_1^{i_1}}{i_1!}\,\frac{\prod\limits_{j=0}^{i_2-1}(s_2-j)\gamma_2^{i_2}}{i_2!}.$$

The product form (6) can be used for estimation of performance measures for small values of v. In general case the recursive algorithm is used.

The recursive algorithm of estimation the performance measures can be described as follows. Let us introduce the auxiliary variables $p(i)$ and $y_2(i)$

$$p(i)=\sum_{\{(i_1,i_2)\in S\ |\ i_1b_1+i_2b_2=i\}}p(i_1,i_2),$$

$$y_k(i)=\sum_{\{(i_1,i_2)\in S\ |\ i_1b_1+i_2b_2=i\}}p(i_1,i_2)i_k,\quad k=1,2.$$

By definition

$$\pi_{t,k}=\sum_{i=v-b_k+1}^{v}p(i);\quad k=1,2;\qquad(7)$$

$$\pi_{c,2}=\frac{\sum\limits_{i=v-b_2+1}^{v}\big(p(i)\,s_2-y_2(i)\big)}{\sum\limits_{i=0}^{v}\big(p(i)\,s_2-y_2(i)\big)}.$$

The recursion for estimation of performance measures is as follows

- Let $P(0)=1$. By definition $Y_1(0)=Y_2(0)=0$.
- Express values $Y_2(i)$, and $P(i)$ through $P(0)$, using relations
$$Y_1(i)=P(i-b_1)\,a_1;$$
$$Y_2(i)=P(i-b_2)\,s_2\,\gamma_2-Y_2(i-b_2)\,\gamma_2;$$
$$P(i)=\frac{1}{i}\,(b_1\,Y_1(i)+b_2Y_2(i))$$
for i from 1 to v.
- Find normalizing constant
$$N=\sum_{i=0}^{v}P(i).$$
- Find normalized values $p(i)$, $y_k(i)$
$$p(i)=\frac{P(i)}{N},\quad y_k(i)=\frac{Y_k(i)}{N},\quad k=1,2.$$
- Find values $\pi_{t,k}$ and $\pi_{c,2}$, using (7) and values of other characteristics using their definitions and (3).

2) Absence of LTE traffic: In the initial model this situation corresponds to the case when $\lambda_1=\lambda_2=0$. Let us denote by $p(d)$ the stationary probability of the state (d). The system of state equations will look like, as given below

$$P(0)\lambda_d=P(1)\alpha_d,$$
$$P(1)(\lambda_d+\alpha_d)=P(0)\lambda_d f_1+P(2)2\alpha_d,$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$P(v)(\lambda_d+v\alpha_d)=P(0)\lambda_d f_v+P(1)\lambda_d f_{v-1}+$$
$$+\ldots+P(v-1)\lambda_d f_1++P(v+1)(v\alpha_d+\sigma),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$P(v+w-1)(\lambda_d+v\alpha_d+(w-1)\sigma)=$$
$$=P(0)\lambda_d f_{v+w-1}+P(1)\lambda_d f_{v+w-2}+\ldots+$$
$$+P(v+w-2)\lambda_d f_1+P(v+w)(v\alpha_d+w\sigma),$$
$$P(v+w)(v\alpha_d+w\sigma)=P(0)\lambda_d f_{v+w}+$$
$$+P(1)\lambda_d(f_{v+w-1}+f_{v+w})+\ldots+$$
$$+P(v+w-1)\lambda_d(f_1+f_2+\ldots+f_{v+w}),$$

For values $P(d)$, $d=0,1,\ldots,v+w$ the normalizing condition is fulfilled: $P(0)+P(1)+\ldots+P(v+w)=1$. After algebraic transformation of the system of state equations the following recursive formula that relates $P(j)$, $j=1,2,\ldots,v+w$ can be obtained

$$(j\alpha_d+(j-v)\sigma I(j>v))P(j)=\qquad(8)$$
$$=\lambda_d\Bigg(P(0)\sum_{d=j}^{v+w}f_d+P(1)\sum_{d=j-1}^{v+w}f_d+\ldots+$$
$$+P(j-1)\sum_{d=1}^{v+w}f_d\Bigg),$$

The recursion for estimation of $P(j)$, $j = 1, 2, \ldots, v + w$ is as follows

- Let $P(0) = 1$.
- Express values $P(j)$ through $P(0)$ using relation

$$P(j) = \frac{\lambda_d}{j\alpha_d + (j - v)\sigma I(j > v)} \times \tag{9}$$

$$\times \left(P(0) \sum_{d=j}^{v+w} f_d + P(1) \sum_{d=j-1}^{v+w} f_d + \ldots + \right.$$

$$\left. + P(j-1) \sum_{d=1}^{v+w} f_d \right).$$

- Find normalizing constant

$$N = \sum_{j=0}^{v+w} P(j).$$

- Find normalized values $p(j)$

$$p(j) = \frac{P(j)}{N}, \quad j = 0, 1, \ldots, v + w.$$

- Find performance measures using the relations

$$\pi_{d,1} = \frac{1}{\lambda_d b_d} \sum_{d=0}^{v+w} p(v + w - d) \sum_{j=d+1}^{g} f_j (j - d)\lambda_d;$$

$$\pi_{d,2} = \frac{1}{\lambda_d b_d} \sum_{d=v+1}^{v+w} p(d)(d - v)\sigma;$$

$$\pi_d = \pi_{d,1} + \pi_{d,2};$$

$$y_s = \sum_{d=1}^{v} p(d)d + \sum_{d=v+1}^{v+w} p(d)(d - w);$$

$$y_w = \sum_{d=v+1}^{v+w} p(d)(d - v);$$

$$y_d = y_s + y_w;$$

$$T_d = \frac{y_d}{\lambda_d b_d (1 - \pi_{d,1})}.$$

III. NUMERICAL ASSESSMENT

Elaborated mathematical model can be used for analyzing the dependence of model's performance measures on the values of input parameters and the features of resource allocation strategies. Another important area of application is the estimation of the volume of resource units and the size of buffer required for serving incoming traffic with given values of performance indicators. Let us consider the numerical examples that illustrate the solution of the listed problems.

The level of traffic load will be characterized by the value of ρ the offered load of one resource unit. To define ρ it necessary to find the offered load of each traffic stream considered in the model. Let us denote by A_1 the offered load expressed in resource units for Poisson flow of sessions for servicing the traffic of LTE-devices

$$A_1 = \frac{\lambda_1}{\alpha_1} b_1. \tag{10}$$

Let us denote by A_2 the offered load expressed in resource units for Engset flow of sessions for servicing the traffic of LTE-devices

$$A_2 = \frac{s_2 \beta_2}{\alpha_2 + \beta_2} \frac{1}{\alpha_2} b_2 = \frac{s_2 \gamma_2}{1 + \gamma_2} b_2. \tag{11}$$

Let us denote by A_3 the offered load expressed in resource units for batched flow of sessions for servicing the traffic of NB-IoT-devices

$$A_3 = \lambda_d \sum_{j=0}^{g} f_j j \frac{F}{c} = \lambda_d b_d \frac{1}{\alpha_d}. \tag{12}$$

The value of ρ can be defined from relation

$$\rho = \frac{A_1 + A_2 + A_3}{v}.$$

Let us suppose that all traffic flows considered in the model generate the same offered load

$$A_1 = A_2 = A_3 = \frac{v\rho}{3}. \tag{13}$$

Relation (13) and expressions (10)–(12) for A_k, $k = 1, 2, 3$ give formulas for estimating the intensities λ_1, β_2, λ_d of sessions coming for each flow considered in the model from known values of ρ.

We begin the model's numerical assessment with Fig 3 that presents the mean values of occupied resource units m_1, m_2 and y_s vs. the value of ρ the offered load of one resource unit. The values of performance measures are obtained after solving the system of state equations (2) and using the definitions of characteristics through values of stationary probabilities formulated in subsection II-D. Model fixed parameters are $v = 120$ resource units (r.u.), $F = 200$ kbit. Transmission rate that is provided by one resource unit is taken from the relation $c = 200$ kbit/c. So mean time of file transmission is 1 sek. Let us suppose that $\alpha_1 = \alpha_2 = \alpha_d = 1$; $b_1 = 5$ r.u., $b_2 = 30$ r.u., $s_2 = \lfloor A_2/b_2 \rfloor + 10$. The size of the batch has probabilities $f_1 = 0{,}5$, $f_{19} = 0{,}5$ and demonstrate it's impulse character. Other fixed parameters: $w = 20$, $\sigma = 0{,}1$.

Presented results demonstrate that despite equal offered traffic for all flows in overload conditions, when $\rho > 1$, NB-IoT-device sessions obtain priority in occupying the transmission resource that doesn't included in SLA (Service Level Agreement). The only way to overcome mentioned difficulties is to divide the resource and used the obtained segments for separate servicing traffic of LTE-device sessions and NB-IoT-device sessions.

The results of usage the suggested approach are presented in Fig 4 and Fig 5. The model input parameters are the same as used in Fig 2 except $\lambda_1 = 0$; $w = 0$; $A_2 = A_3 = \frac{v\rho}{2}$; $f_1 = 0{,}5$, $f_9 = 0{,}5$. The values of characteristics for conjoint servicing are obtained after solving the system (2). The values of characteristics for separate servicing (they are marked by asterisk)

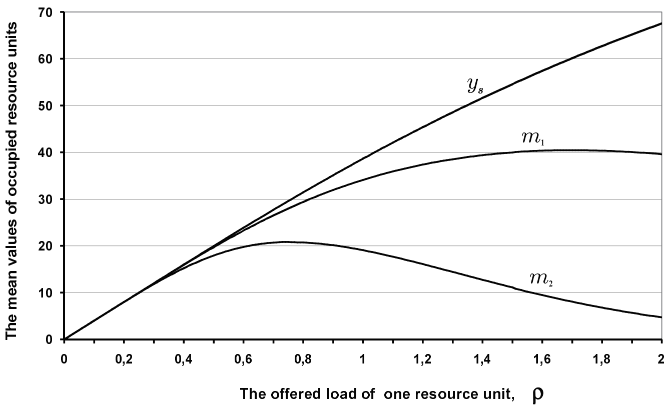

Fig. 3. Mean values of occupied resource units by traffic of LTE-device sessions and NB-IoT-device sessions vs. the offered load of one resource unit

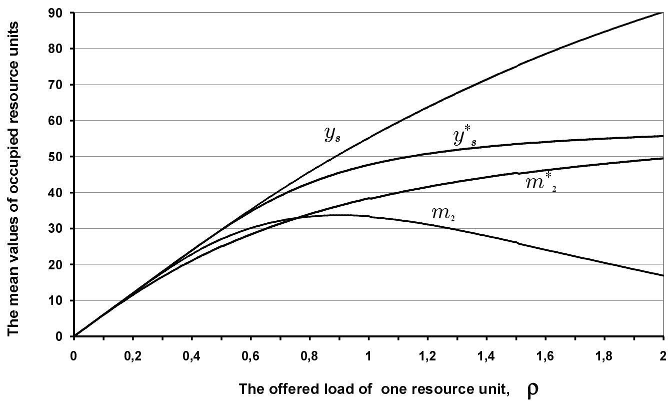

Fig. 5. Mean values of occupied resource units by traffic of LTE-device sessions and NB-IoT-device sessions vs. the offered load of one resource unit for conjoint and separate (marked by asterisk) usage of resource units

are obtained with help of recursive algorithms presented in subsections II-F1 and II-F2 correspondingly. The resource is divided into two equal parts according to the equality of offered loads. It is clearly seen that separate usage of resource helps to diminish the negative consequences of conjoint servicing the traffic streams of LTE-device sessions and NB-IoT-device sessions by LTE cell resource in overload conditions.

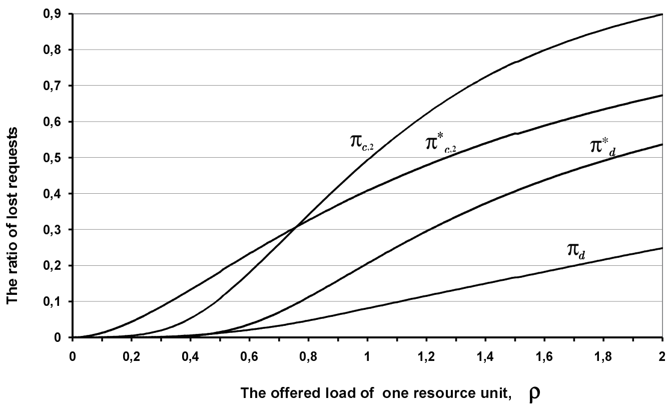

Fig. 4. Probability of losses of LTE-device sessions and NB-IoT-device sessions vs. the offered load of one resource unit for conjoint and separate (marked by asterisk) usage of resource units

Let us consider the usage of the model for estimation of the volume of resource units required for serving incoming traffic with given values of performance indicators. The model input parameters are the same as used in Fig 2 except $b_2 = 10$; $\rho = 1$. The required value of v is obtained by successively increasing of v and checking the criteria of sufficiency of the volume of resource in the form $\max(\pi_{c,1}, \pi_{c,2}, \pi_d) \leq 0{,}05$. The intermediate results of estimation of $\pi_{c,1}, \pi_{c,2}, \pi_d$ are presented in the Fig 6. The values of characteristics are obtained after solving the system (2). For all values of $v \geq 169$ $\max(\pi_{c,1}, \pi_{c,2}, \pi_d) \leq 0{,}05$. So the answer is $v \geq 169$.

Next problem is concerned with the estimation of buffer

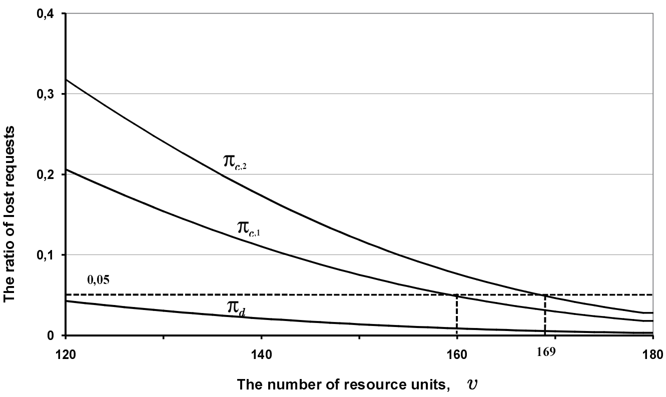

Fig. 6. The intermediate results of estimation of $\pi_{c,1}, \pi_{c,2}, \pi_d$ vs volume of resource v

size required for serving incoming traffic with given values of performance indicators. The model input parameters are the same as used in Fig 2 except $\lambda_1 = \beta_2 = 0$; $v = 60$ r.u.; $\rho = 0{,}9$; $A_3 = v\rho$. The required value of w is obtained by successive increasing of w and checking the criteria of sufficiency of the buffer size in the form $\pi_d \leq 0{,}05$ and $T_d \leq 1{,}5$. The intermediate results of estimation of π_d are presented in the Fig 7 and intermediate results of estimation of T_d are presented in the Fig 8. The values of characteristics are obtained with help of recursive algorithms presented in subsection II-F2. The minimal value of w when $\pi_d \leq 0{,}05$ and $T_d \leq 1{,}5$ is $w = 118$.

IV. Conclusion

In this paper we studied resource allocation and sharing for conjoint servicing of real time and batched NB IoT traffic streams over LTE cell transmission resources. In the model three types of traffic flows are considered. The first two represent the flows of sessions coming from wireless video surveillance cameras. Depending on the number of traffic sources the flows are described by Poisson or Engset models.

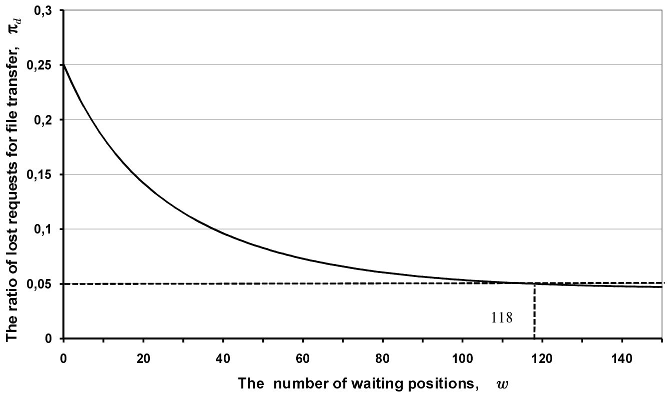

Fig. 7. The intermediate results of estimation of π_d vs buffer size w

Fig. 8. The intermediate results of estimation of T_d vs buffer size w

The transmission resource of the cell is also used by sessions for transmission of traffic of NB-IoT sensors. The corresponding flow of sessions follows Poisson model with batch arrivals and possibility of waiting if all resource units are occupied. The number of waiting positions and maximum allowed time of waiting are restricted. All random variables used in the model have exponential distribution with corresponding mean values. Using the model the main performance measures of interest are defined with help of values of probabilities of model's stationary states.

The model and derived algorithms of performance measures estimation can be used to study the scenarios of resource sharing between LTE and NB-IoT traffic flows. Two scenarios were considered: Static, when all available resources are divided among LTE and NB-IoT traffic and Dynamic, when all available resources are fully shared between LTE and NB-IoT traffic. It is shown that Static scenario can be used when it is necessary to guarantee the desired values of performance indicators especially when resource requirements of LTE and NB-IoT traffic flows significantly differs and traffic servicing is considered in overload conditions.

The constructed analytical framework additionally offers the possibility to find the volume of resource units required for serving incoming traffic with given values of performance indicators and to find the size of buffer that allows to serve NB-IoT traffic flows with acceptable value of delay instead of increasing the restricted volume of resource units. Proposed model can be further developed to include the possibility of reservation, to study the dependence of performance measures on the variance of NB-IoT sessions coming.

Acknowledgment

The research is partly supported by the Russian Foundation for Basic Research, project no. 16-29- 09497ofi-m.

References

[1] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, S. Guizani. "Internet-of- Things-Based smart cities: recent advances and challenges", *IEEE Commun. Mag.*, vol. 55, No 9. 2017. pp. 16-24.

[2] V. Begishev, V. Petrov, A. Samuylov, D. Moltchanov, S. Andreev, Y. Koucheryavy, K. Samouylov. "Resource Allocation and Sharing for Heterogeneous Data Collection over Conventional 3GPP LTE and Emerging NB-IoT Technologies", *Computer Communications*, vol.120, No 2. 2018, pp. 93-101.

[3] I. Gudkova, K. Samouylov, I. Buturlin, V. Borodakiy, M. Gerasimenko, O. Galinina, S. Andreev. "Analyzing Impacts of Coexistence between M2M and H2H Communication on 3GPP LTE System", *In: A. Mellouk et al. (Eds.). Wired/Wiless Internet Communications, WWIC 2014*. Springer, LNCS, vol 8458. 2014, pp. 162-174.

[4] A. Rico-Alvarino, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bern, T. Tirronen, E. Yavuz. "An overview of 3GPP enhancements on machine to machine communications", *IEEE Commun. Mag.*, vol. 54, No 6. 2016, pp. 1421.

[5] W. Zhu, P. Cui, Z. Wang, G. Hua. " Multimedia big data computing", *IEEE Multimedia*, vol. 22, No 3. 2015, pp. 96105.

[6] G. Margelis, R. Piechocki, D. Kaleshi, P. Thomas. "Low throughput networks for the IoT: lessons learned from industrial implementations", *Proc. of the IEEE 2nd World Forum on Internet of Things (WF-IoT)*, 2015, pp. 181186.

[7] D. Che, M. Safran, Z. Peng. "From big data to big data mining: challenges, issues, and opportunities", *Proc. of the 18th International Conference on Database Systems for Advanced Applications*, 2013, pp. 115.

[8] Nokia, Dynamic end-to-end network slicing for 5G, White Paper, 2017.

[9] S.N. Stepanov. *Osnovy teletraffika multiservisnykh setei (Fundamentals of Multiservice Networks)*. Eqo-Trends. Moscow, 2010, (in Russian).

[10] S.N. Stepanov. *Teoriya teletraffika: kontseptsii, modeli, prilozheniya (Theory of Teletraffic: Concepts, Models, Applications)*. Goryachaya Liniya-Telekom. Moscow, 2015, (in Russian).

[11] S.N. Stepanov "Model of Joint Servicing of Real-Time Service Traffic and Data Traffic. I", *Automation and Remote Control*, vol. 72. No 4. 2011, pp. 787–797.

[12] S.N. Stepanov "Model of Joint Servicing of Real-Time Service Traffic and Data Traffic. II", *Automation and Remote Control*, vol. 72. No 5. 2011, pp. 1028–1035.

[13] S.N. Stepanov, M.S. Stepanov. "Planning Transmission Resource at Joint Servicing of the Multiservice Real Time and Elastic Data Traffics", *Automation and Remote Control*, v. 78. 2017. pp. 2004–2015.

[14] S.N. Stepanov, M.S. Stepanov. "Planning the Resource of Information Transmission for Connection Lines of Multiservice Hierarchical Access Networks", *Automation and Remote Control*, vol. 79. No 8. 2018, pp. 1422–1433.

[15] S.N. Stepanov, M.S. Stepanov. "The Model and Algorithms for Estimation the Performance Measures of Access Node Serving the Mixture of Real Time and Elastic Data", *In: Vishnevskiy V., Kozyrev D. (eds) Distributed Computer and Communication Networks. DCCN 2018. Communications in Computer and Information Science (CCIS)*. Springer, LNCS, vol 919. 2018. pp. 264-275.