CELLULAR NETWORK RESOURCE DISTRIBUTION METHODS FOR THE JOINT SERVICING OF REAL-TIME MULTISERVICE TRAFFIC AND GROUPED IOT TRAFFIC

Umer Mukhtar Andrabi,

Moscow Institute of Physics and Technology (State University),
Moscow, Russia, umer.andrabi@phystech.edu

DOI: 10.36724/2072-8735-2020-14-10-61-69

Sergey N. Stepanov,

MTUCI, Moscow, Russia, stpnsrg@gmail.com

Manuscript received 03 August 2020;

Revised 07 September 2020;

Accepted 28 September 2020

Juvent Ndayikunda,

MTUCI, Moscow, Russia, juvndayi@mail.ru

Margarita G. Kanishcheva,

MTUCI, Moscow, Russia, juvndayi@mail.ru

Keywords: Narrowband Internet of Things (NB-IoT), Long Term Evolution (LTE), Radio Resource Management (RRM), IoT, LTE Radio Frame Structure, Network Slicing

Immense growth in the volumes and multiplicity of data to be collected in future Internet of Things (IoT) applications is one of the crucial challenges for the networking organizations as they develop from 4G+ to true 5G systems. Particularly bulk of this traffic includes complex, unstructured and varied data (Big Data) evolve from smart networking ecosystems (LTE-devices, NB-IoT devices). Although 5G offers many low power wide area technologies (Lora WAN, GSM and NB-IoT etc.), principally NB-IoT seems very promising addressing the problem because of its certain characteristics like high fault tolerance, delay tolerance, higher coverage area etc. However, due to the limited bandwidth (180 kHz) availability one of the challenges is how to efficiently use these resources to support and handle massive number of growing IoT devices, also resource management and allocation methodology between LTE and NB-IoT traffic flows. In this context, several key issues for IoT communications in 5G networks should be addressed to satisfy quality of service (QoS) provisioning. In this paper, we proposed a mathematical model for Operator Surveillance systems for sharing radio resources between LTE and NB-IoT. The model utilizes the technique of network slicing for resource management. The proposed techniques provide scenarios that aims to offer a trade-off between the two types of traffics by guaranteeing the network performance and avoiding unproductive utilization of available resources.

Information about authors:

Umer Mukhtar Andrabi, PhD student, Moscow Institute of Physics and Technology (State University), the chair of infocommunication networks and systems, Moscow, Russia.

Sergey N. Stepanov, professor, doctor of science, MTUCI, head of the chair of communication networks and commutation systems, Moscow, Russia

Juvent Ndayikunda, PhD student, MTUCI, the chair of communication networks and commutation systems, Moscow, Russia

Margarita G. Kanishcheva, Master degree student, MTUCI, the chair of communication networks and commutation systems, Moscow, Russia

Для цитирования:

Умер Мухтар Андраби, Степанов С.Н., Ндайикунда Ж., Канищева М.Г. Процедуры распределения ресурса сотовой сети при совместном обслуживании мультисервисного трафика реального времени и группового трафика интернета вещей // T-Comm: Телекоммуникации и транспорт. 2020. Том 14. №10. С. 61-69.

For citation:

Umer Mukhtar Andrabi, Stepanov S.N., Ndayikunda J., Kanishcheva M.G. (2020) Cellular network resource distribution methods for the joint servicing of real-time multiservice traffic and grouped IoT traffic. *T-Comm*, vol. 14, no.10, pp. 61-69. (in Russian)

Introduction

Over the last decade, the IoT technologies have advanced extensively and they have been merged in various domains. IoT has accomplished major improvement in big data processing, heterogeneity, and performance. Based on transmission rate, the communication services of IoT can be indelicately graded into two types: high-data-rate services (such as video service) and low-data-rate services (such as meter reading service) [1]. More than 67% of the overall IoT services are represented by low-data-rate services, which make these services appropriate for WAN technologies.

Due to the constant developments in IoT, these technologies have become advanced and widespread. Based on transmission distance, IoT can be categorized as Short -distance communication and WAN communication [2]. WAN communication technology is desirable in low-data-services such as smart monitoring and surveillance (Multimedia Big Data) [3], generally referred as Low-Power Wide Area Network (LPWAN) technology.

The Narrow-Band Internet of Things (NB-IoT) is a massive Low Power Wide Area (LPWA) technology recommended to handle low-data-rate devices (or Low-end traffic) applications. The typical applications are smart metering (meters, sensors etc.) and intelligent monitoring (Video Cameras) [4] [5].

System operators consider data transmission and resource management one of the crucial parts of surveillance systems. These systems include a number of cameras that implement the functions of video surveillance, a large number of sensors that perform the functions of monitoring temperature, pressure, intrusion, etc. The analytical centers, process the information based on the results retrieved from monitoring systems. Afterwards, management decisions are made to ensure the required level of safety and the efficiency of various technical systems. Information is transmitted using the resources of communication networks.

The functional model of an Observation (Surveillance and Monitoring) system is shown in Fig. 1. Often, surveillance devices are installed in places where availability of a fixed communication network is challenging and impractical. In such situation, the mobile network infrastructure is used to transmit the collected information. To realize this opportunity, the 3GPP (3rd Generation Partnership Project) consortium has developed the NB-IoT (Narrow Band Internet of Things) standard, where several scenarios of connecting various low-speed telemetry sensors to the LTE network are considered [6].

For practical applications, the NB-IoT implementation within the LTE frequency range is of interest. This scenario is called in band [7]. In this case, one of the 180 kHz LTE resource blocks is allocated for NB-IoT traffic. To transfer NB-IoT traffic, the resource block is divided into 12 subcarriers of 15 kHz (a solution has already been standardized in which the resource block is divided into 48 subcarriers of 3.75 kHz). A single subcarrier (single tone mode) or several subcarriers (multitone mode) can be used to service a single NB-IoT device.

The later scenario is applicable when there are free subcarriers. It should be noted that the adoption of the standard does not mean that the rules for its application are formulated. Such solutions need to be configured, which consists in choosing the values of the parameters governing the reception and servicing of incoming information flows in relation to the features of the communication under consideration.

Figure 1. Functional model of the formation of the input data flow IoT in the network of the operator of observation systems

Analysis of resource allocation scenarios in slices

A reasonable choice of the numerical values of the parameters is the main goal of scientific research in this area of telecommunications. Information from sensors and cameras are transmitted as a series of communication sessions [8]-[10].

Incoming informational messages differ in terms of the required resource and requirements for transmission conditions. So, for servicing communication sessions of high-speed video cameras, a large number of resources are required instantaneously and the corresponding information should be transmitted with a minimum delay. While for communication sessions of low-speed smart meters require a small number of resources and data transmission can afford some delay. Sharing the resources of a mobile network is limited for various reasons, as it leads to redistribution in favor of streams with small requests for transmission speed. This is pronounced, especially in overload conditions. Distribution of a telecommunication resource between incoming information flows without taking into account the peculiarities of their formation according to the principle of a uniform measure for all "one-size-fits-all" does not suit the communications market. To create conditions for servicing numerous communication applications such as M2M and high-quality delivery of multimedia content in wireless networks. It was necessary to change the approach to utilize the existing transmission resource. The corresponding concept was developed and is called "network slicing".

It consists of a series of procedures that allow division of the existing physical network architecture into several independent logical networks. Each of which is configured based on the quality of service requirements for communication applications using these networks. The constituent parts of the concept are: A set of network functions (network functions), virtualization method (virtualization), and the orchestration system. Network function include a load balancer, used as separate modules for link building. While, Virtualization used to create logically independent networks on one physical resource and orchestration system, which is a procedure for coordinating the components involved in the life cycle of each slice.

Communication session flow model parameters

We discussed the functional features of the communication system under study and formulate the main assumptions that will

be used in constructing its mathematical model. Let us start by defining the flows of requirements for the allocation of information transfer resources. The model considers the process of receipt and servicing of a certain number of streams of communication sessions from Internet of Things devices. These includes video cameras and other similar video surveillance systems, as well as NB-IoT devices consisting of various telemetry sensors. Among these devices are medical sensors, navigation sensors, temperature, sensors, alarm devices, fire sensors, etc.

Each stream is a superposition of communication sessions from a large number of independent devices of the same type. Therefore, based on the known principles of probability theory, it can be approximately considered Poisson.

The properties of the communication sessions from the cameras depend on a number of parameters. First of all, it is necessary to take into account the characteristics of the codec (for example, using the H.264 standard), method of buffering data before sending (for example, using traffic shaping methods). Depending on the requirements of the target application (in particular, playback delay), the entire communication session can be "sliced" using traffic shaping into separate sessions that require a certain minimum data rate. For example, in the case of H.264, when transmitting a stream with 1080P HD resolution with an average level of quality and without changing other settings, the average stream rate from one camera will be approximately 1.4 Mbit / s.

Of course, this value depends on the parameters of the video stream, for the account of which the mentioned calculator application can be used. Thus, a session is a fragment of a session with a constant speed generated by the target application after the rate adaptation procedure. Bitrate of cameras vary widely from hundreds of kilobits to several megabits per second. The time for transmitting video information is determined by the amount of resource used and can range from a few seconds to tens of seconds. The number of video cameras per cell can vary from several tens to several hundreds.

The intensity of applications for the transfer of sessions can be associated with the number of connected cameras. We denote their number by N_v and by Y_v we denote the intensity of sessions from one camera. Then, by virtue of the above assumptions, the receipt of applications for the transfer of sessions from video cameras obeys the Poisson law with an intensity of $\lambda_v = N_v Y_v$. Note that taking into account the number of cameras is very arbitrary and requires that the cameras form a homogeneous family with similar characteristics. If this cannot be done within the framework of a single stream, then it is possible to divide all cameras into several separate families, each of which generates its own Poisson stream with parameters for using the transmission resource that are different from other streams.

Consider the properties of communication sessions from NB-IoT devices. Since they use a low data rate, the number of such devices assigned to the LTE base station is very large. According to the 3GPP partnership project, a single LTE macrobase station with a range of 1 km will be able to serve information flows from about 50 thousand NB-IoT devices. The resource of such a base station is enough to serve consumers in dense urban conditions, when about 1,500 households per 1 sq. Km, each of which has about 40 sensors.

The figures given correspond to the case when each device per hour transmits on average approximately 100 bytes of information. The approximate values in the specifications of the NB-IoT standard indicate the maximum values of the transmitted message of the order of 1280 bytes and the speed towards the base station of the order of 20 kbit / s for singletone mode and 250 kbit / s for multitone mode. With the development of technology, data transfer rates will increase. Suppose that the amount of information transmitted has an exponential distribution with an average value of F , expressed in seconds.

Similar to how this was done for the stream of sessions from video cameras, the frequency of applications for transferring information files from NB-IoT devices can be related to the number of such devices. We denote their number by N_d and by Y_d we denote the intensity of the receipt of files from one camera. It is assumed that the receipt of files from NB-IoT devices follows the Poisson law with intensity $\lambda_d = N_d Y_d$.

Resource Allocation Scenarios

Consider a separate cell of the LTE network, which is circular, with a base station located in the center. In a mobile networks based on LTE technology, the resource for servicing incoming requests is a variety of resource blocks (RB), which depends on the allocated frequency band, for example, for 1.4 MHz – 6 RB, for 5 MHz – 25 RB, for 10 MHz – 50 RB, for 15 MHz – 75 RB, and for 20 MHz – 100 RB. Using the LTE standard allows you to achieve impressive aggregate data rates. So in the 20 MHz band, speeds reach values of up to 50 Mbit / s for the uplink (from the subscriber to the base station) and up to 100 Mbit / s for the downlink (from the base station to the subscriber). With the further development of the technology, these values will reach upto 500 Mbit / s and 1 Gbit / s, respectively. Flexibility in the distribution of resources, the use of IP at the access level and the core of the network allows us to organize the servicing of traffic stream. It includes both real-time services and elastic data with the provision of specified quality indicators for receiving and transmitting information flows.

We will use the capabilities of the LTE network to organize the work of an operator of observing systems. Suppose that in a cell information flows are served from ordinary consumers of wireless services, and a resource is leased for transmitting the traffic of the observing systems operator. Resource sharing was performed using network slicing capabilities. We denote by r_0 the volume of the resource in the Republic of Belarus, which is allocated to the operator. Part of this resource is used to service the traffic of video cameras, and another part is used to service the traffic of IoT devices using NB IoT technology.

In the process of servicing incoming requests, a certain number of resources are allocated to support customer requests for information transfer. The distribution of resources is considered on a time scale determined by the flow of applications. Each request is assigned a random amount of transmitted user traffic, which has an exponential distribution with a known average value.

When constructing the model, resource granulation is used, which is performed on the basis of the minimum requirements for its value from incoming requests.

It is clear that in this model the minimum resource requirements have requests for information services from NB-IoT devices. The average value of the corresponding requirement will be a resource unit or virtual channel. Let us denote by ν the speed provided by one channel. Next, we will consider the solution to the problem of estimating the quantity ν the resource volume required in terms of load and quality of service in virtual channels. This estimation performed in the uplink (uplink direction) from users of the services of an operator of surveillance systems to the resource of an isolated cell of an LTE network. The value ν is an integer and is a function of the number of resource blocks allocated to form the slice. We will assume that this function uniquely allows us to determine the number of resource blocks by the value of ν and vice versa.

The network slicing involves the use of a resource for servicing homogeneous traffic generated by one type of communication application with similar transmission speed requirements. In the analyzed case, there are two types of requests for information services: from video cameras and from NB IoT devices. Both types have different resource requirements, so it makes sense to use the network slicing procedure again and split the resource between these two types of traffic. The analysis of the peculiarities of the resource occupation by each of these types of requests will be performed in Section 6 (traffic from NB IoT devices) and in Section 7 (traffic from video cameras). In this case, more general models will be investigated than those considered in this section. So it will be assumed that files from NB IoT devices will arrive in groups of random size, and the arrival of video camera sessions will be specified by an arbitrary number of streams. Each with individual requirements for the transmission resource and the general nature of the distribution function of the session service time.

A strict resource sharing allows to equalize the session service indicators, however, the efficiency of its use deteriorates. The negative consequences of this phenomenon can be partially mitigated if the available resource is shared, but at the same time access is limited for some sessions in order to achieve the required loss of incoming requests. To equalize the quality of service indicators, various forms of access restriction are applied, which make it possible to create conditions for the preferential use of the resource for a dedicated group of flows.

The choice of a specific procedure largely depends on the availability of an efficient and computationally stable algorithm for assessing the required resource volume and parameters of the access restriction procedure. The presence of the noted properties of the algorithm is a prerequisite for its use in software and analytical complexes for planning communication networks. In the paper, an algorithm with the listed properties will be considered in Section 7.

Analysis of resource allocation patterns in slices

The sufficiency of the resource in the slice for servicing the incoming flow of requests for information services is estimated by the share of lost requests, and the efficiency of resource use is estimated by the occupancy rate of one virtual channel. To evaluate the introduced indicators, it is necessary to construct a mathematical model of the resource distribution in slices and conduct its research using the theory of Markov processes and numerical methods of linear algebra.

Model description for Group arrival Files

Consider the process of servicing these NB IoT devices in the network of the observing systems operator. The operator uses the LTE network resource to transmit information to analytical centers that are involved in processing the results of observations. It is assumed that the NB IOT technology is implemented at the base station, which allows one of the resource blocks to be used to form data traffic channels and the network slicing procedure is performed, which allows the base station to be allocated for exclusively NB IOT traffic.

Parameters of input stream formation are given in section 3. Since we consider the process of servicing a single stream, in the designations of the parameters we will no longer use the index d , indicated in section 3 that the parameter belongs to the data transfer stream. Recall that one channel is required to transmit one information message of the NB IoT device. Files with the observation results coming from IoT sensors can be accumulated at traffic concentration points, then in groups of random volume they arrive at the base station of the wireless communication network and are forwarded to analytical centers for further processing. The functional model of the formation and maintenance of the data flow under consideration is shown in Fig. 2.

We will construct an input flow model that takes into account the noted features of the formation of incoming requests for information services.

In the Erlang's model, the time intervals between successive receipt of applications have an exponential distribution. The random nature of the moments of the appearance of applications means that at some time intervals they can arrive more often, and at some – less often. However, this feature of the Poisson stream is not enough to reflect the above-mentioned impulsive nature of the recipient file transfer applications, which is peculiar to the operation of NB IoT devices. We will construct a model that, on the one hand, reflects the group arrival of files, and on the other hand, saves the possibility of evaluating the characteristics of file transfer quality with computational complexity comparable to calculations using the Erlang model.

Figure 2. Group arrival of IoT data in the network of the operator of observation systems

Let us discuss the main stages of the mathematical model for the distribution of the channel resource, which takes into account the impulse nature of the appearance of files. In contrast to the Erlang model, a two-level scheme of reconstruction of the in-

coming traffic stream is considered here. The model has channels that serve the Poisson stream of individual groups of requests for transferring files of intensity λ . The arrival rate of groups of applications can be represented as a function of the number of NB IoT devices if we apply the input stream parametrization used in Section 3 (see also an example 1).

With probability f_s , the incoming group contains s files, $s = 1, 2, \dots, g$. we assume that $g = n$, i.e. the s index for f_s . Thus, the received group of files cannot be empty and the number of files in the group does not exceed the available channel resource.

It is clear that $\sum_{s=1}^v f_s = 1$. We will assume that the amount of information transmitted has an exponential distribution with an average value of F (in seconds).

From here it follows that the time of transmission of one file by one channel has an exponential distribution with the parameter μ . Denote by b the average number of applications in one group. The value of b is found from the expression.

$$b = \sum_{s=1}^g f_s s.$$

When a group of applications arrives, the following three development scenarios are possible.

1. ice has enough free resources to serve the entire group. Then all incoming applications simultaneously get service.
2. Slice does not have enough free resource to serve the entire group of applications. In this case, part of the applications equal to the number of free units of the resource gets the service, and the remaining ones are lost without renewal.
3. Slice is full occupied. Then the received group of applications is lost entirely and is not renewed.

The system of equilibrium equations and its solution

The quality of service of incoming applications and the process of occupying information transmission channels are determined by the following indicators: π_i (fraction of the time the slice stays in the busy state of all channels); π_c (share of the lost file transfer applications); π_l (the percentage of lost potential file transfers) and m is the average number of busy channels. The listed characteristics can be found if the values of the fraction of the time $p(i)$ of the stay of the model in a state with i occupied channels are known.

Let us denote by $S = \{i, i = 0, 1, \dots, v\}$, the state space of the model under study. Their change over time is described by a random process $r(t) = i(t)$, where $i(t)$ is the number of channels occupied at the time t by servicing applications.

A graphical illustration of the intensities and directions of transitions $r(t)$ from state (i) is shown in Fig. 3. Since all the random variables implemented in the model have an exponential distribution and are independent of each other, the introduced process $r(t)$ can be considered Markov, in accordance with the provisions of the constructive definition of the Markov process.

Let us denote by $p(i)$, the stationary probability of the state (i) . In order to estimate the values of $p(i)$, it is necessary to compile and solve a system of equilibrium equations. Acting in a standard way, we obtain the following system of equations:

$$p(0)\lambda = p(1)\mu, i=0 \quad (1)$$

$$p(i)(\lambda + i\mu) = (p(0)f_i + \dots + p(i-1)f_1)\lambda + p(i+1)(i+1)\mu, \\ i=1, 2, \dots, v-1;$$

$$p(v)v\mu = (p(0)f_v + p(1)(f_{v-1} + f_v) + \dots + p(v-1)(f_1 + f_2 + \dots + f_v))\lambda, \quad i=v.$$

Figure 3. Transition diagram for the random process $r(t)$ depending on the number of busy channels i and the number of received group of applications for transferring files k , determined by the implementation of the probabilities f_s , $s = 1, 2, \dots, v$

For the values of $p(i)$, the normalization condition will be:

$$\sum_{i=0}^v p(i) = 1.$$

Summarizing (1) with respect to i from 0 to $j-1$. After reducing such terms, we obtain the expression:

$$\lambda \sum_{i=0}^{j-1} p(i) + \mu \sum_{i=0}^{j-1} p(i)i = \lambda \sum_{i=1}^{j-1} \sum_{k=0}^{i-1} p(k) f_{i-k} + \mu \sum_{i=0}^j p(i)i. \quad (2)$$

In relation (2), we write in more detail the first sum on the right. It takes the following form:

$$p(i)f_1 + (p(0)f_2 + p(1)f_1) + (p(0)f_3 + p(1)f_2 + p(2)f_1) + \dots + \\ + (p(0)f_{j-1} + p(1)f_{j-2} + \dots + p(j-2)f_1) = \\ = p(0)(f_1 + f_2 + \dots + f_{j-1}) + p(1)(f_1 + f_2 + \dots + f_{j-2}) + \dots + \\ + p(j-2)f_1 = \\ = p(0) \sum_{i=1}^{j-1} f_i + p(1) \sum_{i=1}^{j-2} f_i + \dots + p(j-2) \sum_{i=1}^1 f_i.$$

Using the expression obtained, the equality $\sum_{i=1}^v f_i = 1$ and relation (2), we find the recursive formula

$$j\mu p(j) = \lambda \left(p(0) \sum_{i=j}^v f_i + p(1) \sum_{i=j-1}^v f_i + \dots + p(j-1) \sum_{i=1}^v f_i \right) \quad (3)$$

connecting sequential values of $P(f)$, $j = 1, 2, \dots, v$. The calculation algorithm $P(j)$ consists of the following steps.

1. P the value of $P(0) = 1$.
2. for values of J varying successively from 1 to v , we find the unnormalized probabilities $P(j)$, using recursion following from (3).

$$p(j) = \frac{\lambda}{j\mu} \left(p(0) \sum_{i=j}^v f_i + p(1) \sum_{i=j-1}^v f_i + \dots + p(j-1) \sum_{i=1}^v f_i \right)$$

3. Determine the normalization constant.

$$N = \sum_{j=0}^v p(j).$$

4. Calculate the normalized probabilities $p(j)$, using the relation.

$$p(j) = \frac{p(j)}{N}, \quad j = 0, 1, \dots, v.$$

Thus, to estimate the stationary probabilities $p(j)$, it is sufficient to use recursion (4). From a computational point of view, this procedure is not difficult to carry out. Let's move on to the definition and calculation of indicators of the quality of service for incoming applications.

Application service indicators and their assessment

The fraction of the busy time of all channels of the slice π_t is found from the relation $\pi_t = p(v)$. The share of π_c of lost file transfer requests is from the expression

$$\pi_c = \frac{\text{Rate of lost orders.}}{\text{Intensity of the applications received.}}$$

To estimate the intensity of blocked requests for file transfer, let us consider an arbitrary state (i) . While in this state, the slice has $v-i$ free channels. Then the arrival of the group containing $v-i+1, v-i+2, \dots, v$, files, will lead to the loss of 1, 2, ..., i of requests for file transfer, respectively (see Fig. 3). Thus, the average number of claims lost in state (i) is determined from the relation:

$$f_{v-i+1} \cdot 1 + f_{v-i+2} \cdot 2 + \dots + f_v \cdot i,$$

equals to $\sum_{k=0}^{i-1} p(i) f_{v-k} (i-k)$, and the intensity of the lost requests λ_b is found by averaging the above expression over all states of the slice :

$$\lambda_b = \lambda \sum_{i=1}^v p(i) \sum_{k=0}^{i-1} f_{v-k} (i-k). \quad (5)$$

The intensity of the received applications λ is found from the equality $\lambda = \lambda b$. The final expression for evaluating π_c is written as the ratio:

$$\pi_c = \frac{1}{b} \sum_{i=1}^v p(i) \sum_{k=0}^{i-1} f_{v-k} (i-k). \quad (6)$$

The average number m of busy slice channels and the percentage of π_l lost potential connections are determined in a standard way:

$$m = \sum_{i=0}^v p(i) i, \quad \pi_l = \frac{\lambda b - m\mu}{\lambda b}. \quad (7)$$

Let us prove that the equality $\pi_c = \pi_l$. To do this, it is enough to establish the relationship between the intensities of the flow of applications received for service, refused and completed service.

Let us add (3) for all j . Let us add (3) over all j from 1 to v . After reducing similar terms and using the previously introduced notation, we obtain the equality,

$$\lambda b = \lambda b \pi_c + m\mu,$$

which proves the formulated statement. Note that relation (7) is an analogue of the conservation laws in the slice of the intensities of incoming and serviced requests. The parameter $a = \frac{\lambda b}{\mu}$

determines the average number of potential connections for the studied slice model. Then equality (7) can be rewritten in the form

$$a = a \pi_c + m. \quad (8)$$

From the recursive expression (4), we signify that the characteristics of the model does not depend on the specific values of λ and μ , but on their relationship.

Example 1: For a model with a group arrival of requests for file transfer, let us calculate the characteristics π_t , π_c , m and compare them with similar indicators calculated for the Erlang model for the same values of a , v . We find their values from the following considerations. The volume of the transferred file has an exponential distribution with an average value of $F = 200$ bytes. The speed and data transfer provided by one NB-IoT channel will be taken equal to and $u = 20$ kbit / s (see Section 3). We will set the parameters of the group receipt of applications by the relations: $f_1 = 0.5$; $f_5 = 0.5$. Hence, the average number of files in the group is $b = 3$. The potential number of connections a (i.e., the average number of connections in the absence of losses) is determined from the expression $a = \frac{\lambda b}{\mu} = 1$. From the assumptions made, it follows that the average file transfer time per channel NB IoT is determined from the expression

$$\frac{1}{\mu} = \frac{F}{u} = \frac{200.8}{20000} = 0.08 \quad c = 80 MC. \quad \text{Suppose that one sensor sends applications with an intensity, equal to 10 applications per minute. Then } N \text{ is the average number of sensors that create the specified file stream with the listed characteristics, is deter-}$$

mined from the expression: $N = \frac{\lambda b}{\gamma} = 75$. Next, we calculate

the unnormalized values of $p(i)$. For these purposes, recursion (4) is used.

Finally, we find:

$$\pi_i = 0.0336; \quad \pi_c = 0.1477; \quad m = a(1 - \pi_c) = 0.8523$$

For the Erlang model with the same parameter values, similar indicators have the form:

$$\pi_i = \pi_c = 0.0031; \quad m = a(1 - \pi_c) = 0.9969.$$

A comparison of the characteristics shows that despite the equal number of channels and the same proposed traffic, the model with group file arrival has significantly worse service rates than the Erlang model. This circumstance must be taken into account when planning communication systems with the pulsed nature of arrival of applications.

9. Plannin a Slice Channel Resource

The main provisions: Let us consider the features of using the model with the group receipt of applications for transferring NB IoT files to estimate the required channel resource of the slice. Although the model under study does not belong to the class of models described by the process of birth and death, the resource planning problem is solved by methods similar to those considered for the Erlang and Engset models.

Parameters and characteristics calculator: The main task of the calculator is to implement the function of selecting some structural parameters and characteristics of the slice model based on others, known from the measurement results or used as standard indicators.

In the model with group arrival of applications, the input parameters are v , f_s , $s = 1, \dots, g$ and a , and the calculated characteristics are π_c and m . To simplify the situation, we will further assume that the maximum group size g and the probability distribution of the number applications in the group f_s , $s = 1, \dots, g$, are known and do not change during the counting process. Evaluation of parameters and quality indicators of service of incoming applications is carried out by analogy with how this problem was solved for Erlang and Engset models.

It is necessary to find the remaining two values from two of the four known values of the parameters and characteristics of the model. The direct problem does not cause any difficulties when the values of π_c and m are determined from the known input parameters of the model. Software implementations of the calculation algorithms equipped with a suitable interface make it possible to build a model parameters and characteristics calculator with group applications and use it to solve the problem of assessing the required network infrastructure.

Channel number planning: As a criterion of sufficiency of the resource, we choose the fulfillment of the given limit for the number of lost applications. This is one of the easiest ways to plan the number of link resource units.

Let us denote the normative value of the share of lost applications by π_{norm} . π_{norm} is fixed first. Further, the minimum number of channel resource units for which the percentage of lost requests does not exceed π_{norm} is determined by brute force or some other algorithm. This will be the solution to the formulated problem. Let us study the effect that group arrivals of applications has on its decision. The calculation results are presented in

Fig. 4, which shows the dependence of the share of lost π_c applications on the number of channels for different scenarios of the formation of groups of incoming applications.

Three scenarios are considered. In the first – the arrival of applications is subject to the rules formulated for the Erlang model. There is one entry in the group with probability one. For 2nd and 3rd scenarios, the maximum number of applications in a group is 10, but groups are formed according to different probabilistic laws. For the 2nd scenario: $g = 10$; $f_s = \frac{1}{10} = s = 1, \dots, 10$. For 3rd scenario: $b = 10$; $f_1 = 0.5$; $f_{10} = 0.5$. Note that in the last scenario, the pulsed nature of the arriving applications is most pronounced. The average number of potential connections for all scenarios is the same and is determined from the equality $a = 30$ Erl. The curve numbers related to the implementation of a particular scenario are indicated in the figure by the corresponding numbers.

The desired value was taken equal to the minimum number of channels at which $\pi_c < \pi_{norm} = 0.03$. For the 1st, 2nd and 3rd scenarios, the required number of channels is given by the numbers 38, 53 and 57, respectively.

The data obtained show a noticeable effect of the pulsed nature of the receipt of applications on the results of estimating the number of channels. This fact must be considered when planning a slice resource. Estimation of the required number of channels using the principle of rational distribution of the information transfer resource is carried out by analogy with the solution of a similar problem investigated for the Erlang model, and will not be considered here.

Example 2: For a model with a group arrival of requests for file transfer, there are two types of solutions to handle the traffic:

1. Estimating the volume of a slice resource sufficient to service a given amount of NB IoT traffic with the required quality.
2. Estimating the maximum possible number of NB IoT devices that can be served on a given slice with the required quality.

Let's start by solving the type one solution. Let's set the quality of file service as $\pi_c = 0.03$. The size of the transferred file has an exponential distribution with an average value of $F = 100$ bytes. The rate of information transfer provided by one NB IoT channel is assumed to be $u = 100$ kbit / s (see Section 3.2). Let us set the parameters of the group receipt of files by the os: $f_i = \frac{1}{5}$, $i = 1, 2, 3, 4, 5$.

Suppose that one sensor sends files with intensity γ equal to 1 file per minute. we will assume the number of NB IoT devices, equal to 37500. From the data given, it follows that the average file transfer time by one NB IoT channel is determined from the expression $\frac{1}{\mu} = \frac{1}{u} = 0.008$ s = 8 ms. We define the intensity of the proposed traffic a by the equality $a = 5$ Earl.

From here we find the value $\lambda = \frac{a\mu}{ba} = 208.33$ filegroups per second. Determination of the required value of the information transmission resource is performed by sequentially reading the number of channels v and checking the condition $\pi_c \leq 0.03$. At the first execution of the inequality, the search for the resource value ends. The results of solving the problem are shown in Fig. 5. The sought value of the number of channels is $v = 14$. Therefore, the slice resource must provide a speed of 1.4 Mbit / s.

Now let's turn to the solution of the second. Let's set the quality of file service as $\pi_c = 0.03$. We take the values of the parame-

ters of the receipt of files that were used to solve the first problem: $F = 100$ bytes; $u = 100$ kbps; $f_i = \frac{1}{5}$, $i = 1, 2, 3, 4, 5$; $\frac{1}{\mu} = 8$ ms; $\gamma = 1$ file / min.

The N_d number of NB IoT devices will be variable. The arrival rate of file groups $\lambda = \frac{N_d \gamma}{bd}$. The estimation of the maximum possible number of NB IoT devices that can be served on a given slice with the required quality is performed by sequentially increasing the number of NB IoT devices N_d and checking the condition $\pi_c \leq 0.03$. At the first execution of the opposite inequality, the search for N_d ends. The results of solving the problem are shown in Figure 6. The sought value for the number of NB IoT devices should not exceed $N_d < 43000$.

Figure 4. Dependence of the estimation of the number of slice channels on the group receipt of applications for a given value of losses $\pi_{norm} = 0.03$

Figure 5. Estimation of the slice resource volume sufficient to service a given amount of NB IoT traffic with the required quality

Figure 6. An estimate of the maximum possible number of NB IoT devices that can be served on a given slice with the required quality

Conclusion

In this paper firstly we discussed about how NB-IoT has a huge role to play in the sphere of Internet of things and how this technology can be used to address the problem of growing data and devices tremendously in our communication networks.

Then we discussed our mathematical model for Operator Surveillance systems for resource allocation and sharing for joint servicing of real time video traffic of surveillance cameras and NB-IoT data traffic of smart meters and actuators over LTE cell facilities. In the model the access control is utilized to create the environment for distinguished servicing of incoming sessions. All random variables applied in the model have exponential distribution with equivalent mean values but the results attained are effective for models with arbitrary distribution of service times. In the model, the key performance measures of interest are given with help of values of probabilities of model's stationary states. The recursive algorithm of performance measures estimation is recommended. The constructed analytical framework additionally offers the possibility to find the volume of resource units and access control parameters required for serving incoming traffic with given values of performance indicators. Future models can be further advanced to incorporate the possibility of reservation and using the processor sharing discipline for serving NB-IoT sessions traffic.

References

1. Y. Li and M. Chen, "Software-defined network function virtualization: A survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.
2. F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, "Understanding mobile traffic patterns of large-scale cellular towers in urban environment," *IEEE/ACM Trans. Netw.*, vol. 25, no. 2, pp. 1147–1161, 2015.
3. S. Stepanov, M. Stepanov, A. Tsogbadrakh, J. Ndayikunda and U. Andrabi, "Resource Allocation and Sharing for Transmission of Batched NB IoT Traffic Over 3GPP LTE," *2019 24th Conference of Open Innovations Association (FRUCT)*, Moscow, Russia, 2019, pp. 422-429. doi: 10.23919/FRUCT.2019.8711920.
4. Cellular System Support for Ultra-Low Complexity and Low Throughput Cellular Internet of Things, document 3GPP TR 45.820, 2015.
5. E-UTRA Physical channels and modulation—Chap.10 Narrowband IoT, document 3GPP TS 36.211, 2016.
6. Official website 3rd Generation Partnership Project (3GPP). www.3gpp.org.
7. Stepanov S. N., Stepanov M. S. Planning transmission resource at joint servicing of the multiservice real time and elastic data traffics // *Automation and Remote Control*. 2017, vol. 78. no. 11, pp. 2004-2015.
8. Stepanov, S.N., Stepanov, M.S. Efficient Algorithm for Evaluating the Required Volume of Resource in Wireless Communication Systems under Joint Servicing of Heterogeneous Traffic for the Internet of Things // *Automation and Remote Control*, 2019, vol.80, no.11, pp. 1970–1985.
9. Stepanov, S.N., Andrabi, U.M., Stepanov, M.S., Ndayikunda, J. Reservation Based Joint Servicing of Real Time and Batched Traffic in Inter Satellite Link // *Proc. of 2020 Systems of Signals Generating and Processing in the Field of on Board Communications*. Moscow, Russia, 2020. pp.1-5,
10. Stepanov S.N., Stepanov M.S. 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)*, vol 919. pp. 264-275. Springer, Cham.

ПРОЦЕДУРЫ РАСПРЕДЕЛЕНИЯ РЕСУРСА СОВОЙ СЕТИ ПРИ СОВМЕСТНОМ ОБСЛУЖИВАНИИ МУЛЬТИСЕРВИСНОГО ТРАФИКА РЕАЛЬНОГО ВРЕМЕНИ И ГРУППОВОГО ТРАФИКА ИНТЕРНЕТА ВЕЩЕЙ

Умэр Мухтар Андраби, Московского физико-технического института (Государственный университет), Москва, Россия, umer.andrabi@phystech.edu

Степанов Сергей Николаевич, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия, stpnsrg@gmail.com

Ндайикунда Жувен, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия, juvndayi@mail.ru

Канищева Маргарита Геннадьевна, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия

Аннотация

Увеличение объема и неоднородности трафика данных, которые будут собираться в будущих приложениях Интернета вещей (IoT), является одной из важнейших проблем для сетей связи по мере их перехода от 4G + к системам 5G. Большая часть этого трафика включает сложные, неструктурированные и разнообразные данные (большие данные), полученные из интеллектуальных сетевых экосистем (устройства LTE, устройства NB-IoT). Хотя 5G предлагает множество технологий с низким энергопотреблением (LoRa WAN, GSM и NB-IoT и т. д.), преимущество следует отдать NB-IoT вследствие таких ее достоинств как высокая отказоустойчивость, терпимость к задержкам, большая зона покрытия и т. д. Однако из-за ограниченной доступности полосы пропускания (180 кГц) возникает задача организации эффективного использования этого ресурса для поддержки и обработки огромного количества растущих устройств IoT, а также методологии управления ресурсами и их распределения между потоками трафика LTE и NB-IoT. В этом контексте необходимо решить несколько ключевых задач для IoT-коммуникаций в сетях 5G, чтобы обеспечить требуемое качество обслуживания (QoS). Предложена математическая модель систем наблюдения при совместном занятии радиоресурса трафиками LTE и NB-IoT. Модель использует технику разделения сети для управления ресурсами. В работе рассмотрены сценарии, цель которых – предложить компромисс между двумя типами трафика, гарантируя производительность сети и избегая непродуктивного использования доступных ресурсов.

Ключевые слова: Узкополосный Интернет вещей (NB-IoT), долгосрочное развитие (LTE), управление радиоресурсами (RRM), IoT, структура радиокадра LTE, сегментирование сети.

Литература

1. Y. Li and M. Chen, "Software-defined network function virtualization: A survey," IEEE Access, vol. 3, pp. 2542-2553, 2015.
2. F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, "Understanding mobile traffic patterns of large-scale cellular towers in urban environment," IEEE/ACM Trans. Netw., vol. 25, no. 2, pp. 1147-1161, 2015.
3. Степанов С.Н., Степанов М.С., Цогбадрах А., Ндайикунда Ж. Андраби О. Распределение ресурса и его совместное использование при передаче группового трафика в сети 3GPP LTE // The Proc of the 24th Conference of Open Innovation Association, FRUCT. 2019. Vol. 2019-April. pp. 422-429.
4. Cellular System Support for Ultra-Low Complexity and Low Throughput Cellular Internet of Things, document 3GPP TR 45.820, 2015.
5. E-UTRA Physical channels and modulation-Chap.10 Narrowband IoT, document 3GPP TS 36.211, 2016.
6. Official website 3rd Generation Partnership Project (3GPP). www.3gpp.org.
7. Степанов С.Н., Степанов М.С. Планирование ресурса передачи при совместном обслуживании мультисервисного трафика реального времени и эластичного трафика данных // Автоматика и телемеханика. 2017. № 11. С. 79-93.
8. Степанов С.Н., Степанов М.С. Эффективный алгоритм оценки требуемого объема ресурса беспроводных систем связи при совместном обслуживании гетерогенного трафика устройств интернета вещей // Автоматика и телемеханика. 2019. № 11. С. 108-126.
9. Степанов С.Н., Андраби О., Степанов М.С., Ндайикунда Ж. Использование резервирования при совместном обслуживании трафика реального времени и группового трафика данных на межспутниковой линии связи // Proc. of 2020 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia, 2020. pp.1-5.
10. Степанов С.Н., Степанов М.С. Модель и алгоритмы оценки характеристик пропускной способности узла доступа при совместном обслуживании трафика реального времени и эластичного трафика данных. In: Vishnevskiy V., Kozyrev D. (eds) Distributed Computer and Communication Networks. DCCN 2018. Communications in Computer and Information Science (CCIS), vol 919. pp.264-275. Springer, Cham.

Информация об авторах:

Умэр Мухтар Андраби, аспирант кафедры инфокоммуникационные сети и системы, Московского физико-технического института (Государственный университет), Москва, Россия

Степанов Сергей Николаевич, профессор, д.т.н., заведующий кафедрой сети связи и системы коммутации, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия

Ндайикунда Жувен, аспирант кафедры сети связи и системы коммутации, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия

Канищева Маргарита Геннадьевна, студент магистратуры кафедры сети связи и системы коммутации, Московский Университет Связи и Информатики (МТУСИ), Москва, Россия