RESEARCH ARTICLE

Energy-Efficient QoE-Driven Radio Resource Management Method for 5G and Beyond Networks

MYKOLA BESHLEY1,2, NATALIA KRYVINSKA2, AND HALYNA BESHLEY1
1Department of Telecommunications, Lviv Polytechnic National University, 79000 Lviv, Ukraine
2Department of Information Systems, Faculty of Management, Comenius University in Bratislava, 814 99 Bratislava, Slovakia
Corresponding author: Mykola Beshley (mykola.i.beshlei@lpnu.ua)
This work was supported by the Ukrainian Project “Development of the Novel Decentralized Mobile Network Based on Blockchain-Architecture and Artificial Intelligence for 5G/6G Development in Ukraine” under Grant 0120U100674; in part by the Ukrainian Project “Development the Methods and Unified Software-Hardware Means for the Deployment of the Energy-Efficient Intent-Based Multi-Purpose Information and Communication Networks” under Grant 0120U102201; and in part by the Faculty of Management, Comenius University in Bratislava, Slovakia.

ABSTRACT Energy-efficient Radio Resource Management (RRM) for 5G and beyond networks has become a key research challenge due to increasing Small Cells (SCs) densities and the high Quality of Experience (QoE) requirements of business users. Ensuring QoE and energy efficiency is essential in mobile networks, but these goals are often opposing and rarely addressed simultaneously in existing solutions. In this paper, we propose to include the QoE criterion in the RRM technique for 5G and beyond multi-layer networks, which will allow ordering individual QoE for business users. We developed a new radio resource allocation and optimization method to address changing user QoE requirements and reduce energy consumption in multi-layer 5G networks. The proposed method differs from the known ones in that it considers the QoE requirements of business users and load localization to optimally distribute the service process between Macro Cells (MCs) and SCs. This method uses a Voronoi diagram to energy-efficiently design the 5G Radio Access Network (RAN) by switching SCs to sleep mode when they are not serving active users. As a result, a balance is struck between user QoE requirements and network energy efficiency. Based on simulations, it is proved that the proposed method allowed more efficient use of accessible radio resources by 25% and reduced the energy consumption of the 5G RAN by 8.7% to provide the ordered QoE for users compared to traditional RRM methods.

INDEX TERMS Energy consumption, energy efficiency, heterogeneous networks (HetNet), quality of experience (QoE), radio resource management (RRM), radio access network (RAN), resource allocation, voronoi diagram.

I. INTRODUCTION

The traffic that users generate is constantly growing. This is due to the fact that the market is constantly being updated with modern devices, the servicing of which requires broadband [1]. Broadband provides high-speed data transmission and constant connection to the Internet, the ability to both receive and transmit information at high speeds. Today, broadband is considered to be more than 2 Mbit/s.

Unfortunately, it is not always possible to provide such speed to users today [2]. This is because the user load is heterogeneous and usually localized, at one time or another, in well-defined places, which is most typical for urban conditions. Therefore, a situation arises where some Base Stations (BSs) are overloaded, and the Quality of Service (QoS) to their users is unsatisfactory, while others remain virtually without users [3]. The current radio access systems are not focused on the user, they are focused on the coverage area.

The existing networks should be designed with a user-oriented design [4], [5], [6]. The main difference between
user-oriented design and other design methods is the attempt to optimize the system so that users use its resources the way they want, rather than forcing users to change their behavior to accommodate the network. One of the future approaches to such design is to modify the existing 5G architecture using an improved Radio Resource Management (RRM) method, considering the user’s individual Quality of Experience (QoE) requirements. The measurement of QoE can be done at different scales and include different units of metrics. It can be measured with a qualitative or quantitative scale. To calculate the QoE parameter “user satisfaction”, an example, an ordered (qualitative) scale, including ratings from 1 to 5, is used, where 1 represents bad quality and 5 represents excellent quality [7]. QoE provides a broader view of how quality is obtained from the end-user perspective, compared to the more specific view of individual network parameters (throughput, delay, loss) that QoS provides. Nevertheless, QoS has a high impact on QoE. To guarantee users a certain level of QoE, operators must allocate the necessary share of the radio resource. And the higher the QoE level, the more share of radio resources must be allocated. The RRM is a set of algorithms used to optimize the use of scarce licensed spectrum. The main task of RRM is to ensure QoS and optimal use of resources [8]. For both 4G and 5G networks, the main method of RRM is the use of spatial multiplexing, which makes it possible to achieve better network performance by introducing a high number of Small Cells (SCs), such as micro, pico, and femto cells, increasing the rate of frequency spectrum reuse and using effective resource scheduling algorithms [9]. However, the high density of SC leads to a significant increase in the network’s energy consumption. Energy efficiency is a major concern for current and future mobile networks [10], [11], [12]. The energy-optimized utilization is a key factor for efficient resource allocation planning and has a major impact on the user experience [13]. Network life depends on a well-organized energy balancing scheme in 5G networks.

Accordingly, the deployment of SCs should be aimed at improving the energy efficiency of the network by putting SCs in idle mode when they do not serve any active users. This requires solving the trade-off between providing the necessary network throughput to ensure the QoE requirements of the users per unit area and the energy efficiency of the network.

A. BACKGROUND

Currently, the deployment of SCs takes place using the deterministic (stationary) method. This method has been widely used for the coverage planning of a cellular system. These methods are effective for homogeneous network topology with fixed cell size, taking into account the interference between cells. However, they are inefficient for Heterogeneous Networks (HetNet) since it is only used to model coverage with specific cell size. It is known that network coverage in a city with dense buildings is difficult due to the non-uniformity of user load and non-uniform fading of the useful signal, which violates the fixed geometric structure of the network. Additional installation of the BSs on the hexagonal scheme doesn’t provide the optimal cell size to satisfy the capacity requirements of the system. Stochastic geometry is more suitable for planning HetNet cell structures with random parameters [14], [15], [16].

With this in mind, the challenge of deploying HetNet has created the need to consider a variety of user and base station (BS) configuration models for system realistic performance evaluation and design. In particular, it is worth focusing on the HetNet deployment models proposed by standardization bodies such as the Third Generation Partnership Project (3GPP). To deploy Macro Cells (MCs), 3GPP solutions rely either on a single MC or on grid-based models where a finite number of MCs are placed as evenly spaced points on a plane [17]. The deployment of SCs occurs randomly within an MC. It is assumed that they should operate at different frequencies to avoid interference between MCs and SCs.

Since the coverage area of SCs is not large, the same licensed frequency band can be reused, which greatly improves the spectral efficiency of the network. Traditional fractional frequency reuse and soft frequency reuse methods considered by 3GPP standards are used to reduce inter-cell interference [18]. To allocate radio resources between users in the 5G network, standardized algorithms Round Robin (RR), Maximum Channel State Information (CSI), and Proportional Fair (PF) are used [19].

B. MOTIVATION AND CONTRIBUTIONS

Most of the existing approaches on one goal of optimizing or improving the energy efficiency of the 5G network or enhancing the QoS. This is because there is some contradiction in the resource management process. Since improving the QoS requires additional radio resources and BSs for more coverage, this leads to an additional increase in network energy consumption and inter-cell interference.

However, with dynamic traffic demand and dense non-uniform distribution of user load, frequency reuse methods lead to low spectrum utilization in some cells and a deficit of spectrum in others. And standardized resource scheduling algorithms are ineffective in adapting to business users’ changing requirements for individual QoE.

Thus, there are currently no practical solutions to develop methods for deploying multi-layered 5G networks that simultaneously improve network energy efficiency and satisfy changing user requirements for ordered QoE levels. Accordingly, the problem of adaptively allocating 5G resources while minimizing energy savings and ensuring high QoE requirements remains unresolved. Therefore, further research is needed to develop a new energy-efficient RRM method for beyond 5G and 6G networks, considering user requirements for service individualization.

The novelty and the main contributions of the work are as follows:

1) We propose a new Fractional Frequency Reuse (FFR) scheme using the Voronoi diagram that tries to find the most
suitable subband for newly deployed SCs when it is turned on in any sector or zones of the MCs area.

2) We propose QoE-based traffic differentiation technique for the 5G and beyond network;

3) We propose a new energy-efficient QoE-driven RRM method for adaptive resource planning and formation of the 5G Radio Access Network (RAN). This idea is achieved by shifting the center of gravity of operation from rigid, always-on SCs focused on the BS to cells that adapt their on/off state based on the QoE requirements of the user;

4) We develop a simulation model that uses a systematic object approach to design functional blocks of a mobile communication network based on known 5G standards and the proposed RRM method;

5) We provide a more generalized view of quantifying 5G network performance quality to optimize radio resources and network energy consumption by proposing a QoE-based index of service-level quality;

6) We use simulation to demonstrate that our suggested method is better than the alternatives.

The remainder of the article is organized as follows. Section II describes the related work of the RRM methods in 5G and 6G networks. Section III addresses our proposed 5G multi-layer system architecture and energy-efficient QoE-driven RRM method. Section IV presents an evaluation of the effectiveness of the proposed solution through simulations and experiments. Section V discusses future work and some limitations of our contribution. Section VI concludes the paper.

II. RELATED WORK

The most common approach in RRM methods is to introduce flexible scheduling of resources, which distributes accessible resources dynamically, taking into account various limitations, such as system bandwidth, energy consumption or QoS. This section discusses the latest RRM methods for 5G and 6G networks.

In [9], authors reviewed in detail the recent works on RRM methods in 5G networks aimed at the deployment of “green” wireless technology. In particular, the review demonstrates the architecture of a heterogeneous multi-layer network, which is the basis of all works to achieve optimal RRM in terms of ensuring the QoS and 5G energy consumption. The proposed technique for spectrum allocation, resource allocation fairness, interference mitigation, QoS, and QoE are discussed. RRM schemes that optimize the 5G/6G network energy consumption are also investigated. The authors described the feasibility and complexity of implementing these methods in practice at the end of the paper. In the paper [20] the authors have developed a unique mathematical model based on stochastic geometry to calculate the coverage probability of a 5G network consisting of MCs and SCs. The downlink throughput under cognitive interference coordination and Signal to Interference plus Noise Ratio (SINR) threshold are analyzed.

Zhenni Pan and his colleagues [21] propose two specific methods to optimize 5G energy consumption for macro-level trade-offs between spectrum efficiency and QoS, that accommodate immediate traffic changes on the basis of their dynamic configuration of MC. In particular, the authors have developed an Active Cell Rotation (ACR) scheme for the optimal energy utilization of the BSs. The above-proposed approach is to automatically manage the number of active BSs using a different frequency scheduling scheme, thus guaranteeing the QoS requirements for the user’s equipment. An enhanced cooperative ACR scheme is also offered, allowing BSs to attain the best spectrum use with energy savings based on traffic utilization by Markov processes. Several experimental results demonstrate that the ACR techniques can impact energy and spectrum eco-efficiency solutions while satisfying future traffic-intensive and QoS requirements. Unfortunately, this work is not suitable for multi-layer HetNet with different cell sizes. And also, these methods do not consider the possibility of ordering individual QoE requirements of business users, which, even under poor radio conditions, should get a high-quality service. Solving these problems is the main goal for the development of future 6G networks.

In [22], the method of dynamic channel allocation based on the Context Multibranch Bandit (CMAB) theory is proposed. In particular, the algorithm is based on stochastic data about user behavior, namely the user’s location, and QoE. Users with the same conditions are formed into certain clusters for which resource allocation is planned. However, the work does not give a comprehensive combination of the mechanism of reasonable activation and deactivation of BSs with the proposed method of reducing the energy consumption of the 5G network.

The Energy-Efficient Deep Reinforcement Learning Resource Allocation (EE-DRL-RA) method for network slicing in 5G is proposed in [23]. This method is based on the use of integrated learning, considering Deep Reinforcement Learning (DRL) and Deep Learning (DL) techniques, to make resource allocation decisions in 5G. In addition, the authors calculate the optimal BS transmission power and resource blocks (RBs) for user throughput, solving the problem of energy-efficient resource allocation. The RRM technique proposed in this paper is unique due to the simultaneous distribution of power and RBs, providing insulation of slices with low processing and time difficulty. Based on this study, it is found that this approach provides better performance than the current published known method in relation to the speed of convergence, computational complexity, energy efficiency, the number of users served, and the degree of inter-slice isolation. The paper states that for the next work, it is proposed to develop this method to take into account the users’ intent to order the desired QoS delivery in the resource allocation process.

The article [24] provides exciting research on the business case for 6G. The authors provide a comprehensive overview of state-of-the-art trends, emerging technologies,
apps, user devices, and the structure of the 6G network. It notes that future 6G networks must adapt to users’ needs, taking into account their behavior and service quality desires. It emphasizes that the energy efficiency of networks is important in the 6G building process. It also argues that Artificial Intelligence (AI) will improve the efficiency of future networks.

Vibha Jain and his colleagues [25] present a cooperative resource allocation and computation offloading method that integrates deep learning techniques with reinforcement in 6G using Cybertwin. In order to guarantee QoS for end-users via the minimization of packet latency and total 5G energy consumption with more efficient cache management, the Multi-Agent Twin Delayed Deep Deterministic (MATD3) algorithm is used as the basis of the proposed system. Comparative benchmarking tests have shown that the proposed MATD3 system decreases end-to-end latency and energy consumption by 13.8% and 12.5% compared to Multi-Agent Deep Deterministic Policy Gradient (MADDPG), while increasing the task success rate by 4%.

In [26], a resource allocation method was developed with an algorithm for optimal user aggregation to reduce the service delay. Three approaches to QoS-aware User Aggregation (QoSA) are proposed: Block-Coordinate Descent (QoSA-BCD), the Alternating-Direction Method of Multipliers (QoSA-ADMM), and Multi-Flow (QoSA-MF). However, the work does not take into account the effect of this method on interlayer interference, which can cause problems in the multi-layer HetNet.

In [27] by analogy, the authors have developed a method for resource allocation and aggregation of users to reduce the total energy consumption and QoS control. The technique consists of two stages. In the first one, there is a merger of similar users and resource allocation using a cost-based algorithm. The power allocation problem for each device based on SINR and QoS was decided in the second stage by a deterministic method based on the decomposition structure. This algorithm showed good performance for a large network, providing fast convergence. Unfortunately, each user was allowed to use only one channel resource unit. Each MC and SC were not applied to more than one unit. And this work did not consider distributional fairness and individualization of service.

In another paper by Li et al. [28], the joint user aggregation and resource allocation methods were solved using a nonlinear mixed-integer fractional programming method to reduce energy consumption. The technique was improved by accounting for user QoS using the Dinkelbach method and dual Lagrange decomposition. The results show that the developed method provides lower network energy consumption than the user aggregation algorithm with maximum SINR. In addition, the convergence time of the proposed algorithm was optimal at low computing complexities, which increases the cost-effectiveness. However, when the number of users increased, the method became inefficient. While [28] proposed Active Antenna System (AAS) in a 5G heterogeneous network, Xu et al. [29] presented an energy-dependent user association in an environment with energy-saving interactions between BSs. The trade-off between offloading users and saving energy was an incentive for them. The primal-dual interior point method was used to construct a user association scheme. The result indicated an increase in energy for HetNet with energy cooperation compared to a network without energy cooperation.

The authors [30] developed a methodology to optimize energy consumption at the base station level, the interconnection of end devices, and resource and energy allocation. The main idea of the work was to minimize energy consumption without the need to distribute traffic. First, the method solved the problem of user aggregation and resource allocation. Then a fast algorithm was used to achieve optimal energy allocation, which did not require tuning. The optimization of BS coverage was done using a heuristic algorithm based on the greedy principle.

In [31], the authors provide primary considerations for green communications based on Machine Learning (ML). The authors describe how AI technologies control 5G resources and increase energy efficiency. Existing challenges are discussed, and questions for future research in green 6G are highlighted.

The paper [32] proposes interesting AI-based solutions that should be adopted in various characteristics of 6G networks during the deployment and management of the layered architecture. In addition, this paper presents AI technologies that will improve 6G network throughput in the future. Conceptual applications of ML in 6G network architecture are discussed to improve computational efficiency, communication reliability, energy management, scalability and resilience, and new channel estimation.

Salh A. and colleagues [33] proposed a new learning algorithm for optimal resource allocation, power minimization, and guaranteed solution for Ultra-reliable and Low-latency Communications (URLLC) scheduling in Beyond 5G networks. The authors improved Generative Adversarial Networks (GANs), which apply a sufficient number of extreme events in order for the deep-RL agent to generate synthetic data with high fidelity similar to real data, based on an adjustable number of extreme events in the data set. The proposed enhancements allow the deep-RL technique to generate large amounts of data that are practically used in real-time operations.

In [34], the authors developed a technique for intelligent packet transmission scheduling in cognitive networks. For this purpose, the authors applied a new Generative Adversarial Network and Deep Distribution Q Network (GAN-DDQN). The authors proposed several mathematical models to improve the intelligent packet data transmission in cognitive networks, such as: transmit packet rate model, transmission delay model, throughput and Mean Opinion Score (MOS) model to improve the QoE for users. Simulation results show that the proposed GAN-DDQN increases the throughput and packet rate while reducing the power.
consumption and transmission delay (TD) compared to the known ones.

An interesting solution is proposed in [35], where the authors propose a heterogeneous SpiderNet architecture for 6G. The main idea of this paper is to use an AI-based database for intelligent network management to maximize spectral efficiency, energy efficiency, and user QoE.

Based on the related works, we found that it is difficult to implement all methods simultaneously. This is due to the fact that there are certain contradictions between them. For example, improving the QoS/QoE requires additional radio resources and BSs for denser coverage, which causes an increase in the energy consumption of the network and inter-cell interference. As such, new RRM methods need to be developed that simultaneously address the problem of multi-criteria optimization, promoting a user-centric QoE and energy-efficient approach for 5G and 6G networks.

In this case, we propose an approach to frequency allocation and SCs deployment using the Voronoi diagram for the 5G network and beyond. We present block diagrams of an energy-efficient QoE-oriented RRM method for resource scheduling and 5G radio access layer structure. This method allows automated resource allocation adaptation to business users’ needs. In contrast to the well-known ones, a balance between user QoE requirements and network energy efficiency is achieved by switching SCs to sleep mode when they are not serving active users. Also, the uniqueness of the method is that users who want a high quality of service and are in overhead BS conditions will be able to get the ordered level of service at a higher price. And vice versa, users who do not require high quality of service will be able to get these services at a lower price.

### III. PROPOSED ENERGY-EFFICIENT QOE-DRIVEN RRM METHOD FOR 5G AND BEYOND NETWORKS

Planning of multi-layer 5G mobile networks is more complex than single-layer ones [36]. It is necessary to separately consider the density of BSs of each layer and then determine the superposition of all layers in calculating parameters for end users. When planning by the classical (stationary) method, such a heterogeneous network will not be effective due to the heterogeneity of the load in different territories. In the case of heterogeneous networks, the most preferred method is the integration of many SCs into the existing radio access network structure.

For this purpose, based on the Voronoi diagram, we propose a method of planning the operation of BSs depending on the location of the user load and QoE requirements of business users. This approach ensures high energy efficiency of the network by putting the BSs into inactive mode without degrading the QoE requirements of business users. The variables (symbols) which are used for problem formulation are shown in Table 1.

| Symbol | Definition |
|--------|------------|
| $S_e$  | system spectral efficiency |
| $C$    | channel throughput |
| $\Delta F$ | channel bandwidth |
| $S$    | coverage area |
| $\text{SNR}$ | signal to interference plus noise ratio |
| $i$    | base station (cell) index |
| $j$    | user equipment (UE) index |
| $r$    | resource block (RB) index |
| $t$    | subframe index |
| $I$    | time period |
| $I_{\text{d}}$ | duration in seconds |
| $\text{SNR}_{i,j}(t)$ | SINR of $j$ from $i$ over $r$ |
| $W_i^u$ | transmission power of base station |
| $R_{il}$ | channel gain between cell $i$ and user $j$ |
| $X$    | set of all base stations |
| $N_0$  | noise power spectral density |
| $F_r$  | channel bandwidth of the allocated resource $r$ |
| $\eta_i(r)$ | spectral efficiency of the UE $i$ of the $r$-th cell |
| $D_{\text{GRB}}$ | total load of GBR users with QoE demands in cell $i$ |
| $P_{i}(t)$ | association coefficient of $j$ with $i$ |
| $d_{\text{QoS}}(t)$ | throughput for QoS demand of GBR user $j$ at time $t$ |
| $C_{i}^d$ | throughput of cell $i$ at time $t$ for user $j$. |
| $C_{i}^c$ | throughput of cell $i$ at time $t$ for user $j$. |
| $r_{i,j}$ | load of GBR user with QoE demands $j$ on cell $i$. |
| $a_{i,j}$ | the number of accessible RBs in cell $i$. |
| $a_{i,j}$ | the number of accessible RBs in the 5G for user $j$. |
| $N_{\text{GRB}}$ | number of GBR users with QoE demands |
| $S_{i}$ | Jain fairness index of load distribution. |
| $N_{\text{NGBR}}$ | number of NGBR users with QoE demands |
| $R_{i}(t)$ | data rate of $j$ from $i$ |
| $U_{i}(t)$ | utility of NGBR user $j$ with QoE demands |
| $E_{\text{MAC}}(t)$ | overall energy consumption of the 5G RAL |
| $C_{\text{MCBS}}$ | macro cell (MC) base station |
| $E_{i}(t)$ | energy consumption of $i$-th MCBS, |
| $E_{i}(t)$ | small cell (SC) base station |
| $E_{i}(t)$ | energy consumption of $i$-th SCBS, |
| $E_{\text{NGBR}}(t)$ | total accessible RBs for user NGBR $j$ at time $t$ in $i$. |
| $Z$    | set of all MC base station |
| $y$    | set of all SC base station |
| $z$    | set of all users |
| $d_j$  | user demands for a certain type of data traffic |
| $r_j$  | required data traffic type and QoE of the $j$ user |
| $v_j$  | active state of base station $X_i$ |
| $s_j$  | sleep state of base station $X_i$ |
| $E_{SC}$ | fixed energy consumption of SC in active state |
| $E_{SC}$ | fixed energy consumption of SC in sleep state |
| $E_{SC}$ | energy consumption depending on the load of SC |
| $S_{\text{MCBS}}$ | slope of the load-dependent FC in SC |
| $S_{\text{MCBS}}$ | slope of the load-dependent FC in SC |
| $C_{\text{MCBS}}$ | maximum throughput of a SC |
| $C_{\text{MCBS}}$ | maximum throughput of a MC |
| $C_{\text{NGBR}}$ | throughput is needed to serve users with high QoE |
| $C_{\text{NGBR}}$ | throughput is needed to serve users with low QoE |

### A. MULTI-LAYER HETEROGENEOUS NETWORK ARCHITECTURE

This paper proposes the architecture of a beyond 5G heterogeneous network consisting of 5 layers, as shown in Fig. 1. The core layer implements the network management functions and is deployed as a software-driven SDN/NFV (Software Defined Networking/Network Functions Virtualization)
core, which uses a cloud environment to store and process data about the network. This layer has software and hardware that provides users localization, determines its velocity and movement direction, calculates the threshold power and SINR coefficient, and other necessary calculations. The following functions are performed at the core level:

- adaptive change of the SCs structure and frequency planning depending on the input load;
- analyzing the BSs status and ordered user-centric QoE requirement;
- continuous analysis of signal data received from the lower levels, changes in the distance to the BS for each user, the total number of active network users;
- loading map generation and modification;
- user distribution over the RAN structure;
- processing and storage of user data received by the 5G BSs;
- controlling and storing the data from lower layers, managing the aggregation process, and regulating the network's energy consumption.

In the proposed architecture, the MC is a gateway through which the cells of the lower level are connected to the network. The range can be up to 1-2 km.

The Pico Cells (PCs) layer or SCN adaptively in the mode of functioning changes the states depending on the load and serves the users with high requirements to the traffic. The range is 100-200 m. In this case will have two states: Idle, the BS listens to the environment, sends Broadcast_ID, defines the vector of distance change to the BS, and active, the BS is working in normal mode, contains information about the allowable and current users. Serves low-mobile users (low traffic vector) that cause high service requirements (high QoS_class, QoE-5). Connected to a higher level BS via optical fiber cable.

The Femto Cells (FCs) layer or SCN−1 is dynamically deployed in areas of high user congestion and performs similar functions as Layer 2 cells.

The User Equipment (UE) layer has QoS_class attributes and sends Channel State Information (CSI), SINR, and QoE on request.

Since the user load is localized irregularly, the placement of BSs should also be irregular. In places with a higher density of user load, there should be a higher density of SCs and vice versa. Stochastic geometry can be a solution to this problem. According to the proposed architecture of partitioning the coverage area of layer two between the BSs, we propose carrying out using the Voronoi method [37]. The essence of the method is to partition the plane so that each partitioning area consists of a set of points, each of which is closer to a certain object than to another.

**B. FREQUENCY ALLOCATION TECHNIQUE FOR THE 5G AND BEYOND NETWORKS USING THE VORONOI DIAGRAM**

According to the proposed network architecture, the coverage areas of SCs are significantly smaller than the coverage area of MCs. Therefore, SCs will be much more per unit area. The related work section found that traditional frequency planning methods are not designed for use due to the increasing level of intercellular interference. Therefore, there is a need to find new methods of frequency planning which minimize the number of interference at all layers. We consider the classic frequency reuse scheme of covering the area with hexagonal cells successfully used in modern mobile networks [38], [39].

For this reason, we propose a slightly improved approach to frequency reuse for MC and SC. This allocation will ensure more efficient frequency reuse than the existing one. We propose a FFR technique that tries to find the most suitable subband for newly deployed SCs based on the Voronoi diagram when it is turned on in any sector or zones of the MCs area. If we draw imaginary lines from the centers of three neighboring MCs, then a triangle is formed, which will serve as the location of SCs. To minimize the number of interference on the edges of MCs, we propose to create another “triangle” in the middle of the formed “triangle”. This is done so that the frequencies inside the large “triangle” are shifted counterclockwise by one position relative to the MC.

Within the small “triangle” the same frequency subband as in the near zone of MCs will be used (Fig. 2 a), or this “triangle” will be divided into three equal parts in each of which the frequency subband of the inverse sector will be used (Fig. 2 b).

This allocation will provide a more efficient reuse of frequencies than the existing ones. The “triangles” formed as a result of such manipulations will serve as locations of SCs according to the method of placement of base stations based
FIGURE 2. Frequency allocation for the effective implementation of “SC” in the existing HetNets.

on Voronoi diagrams. This approach provides an opportunity to determine where a BS of a certain type should be placed and at what frequency it will operate (Fig. 3 a and Fig. 3 b). This will ensure that unnecessary base stations are not installed and the level of interference is reduced.

FIGURE 3. Proposed fractional frequency reuse based frequency allocation for 5G HetNet using Voronoi diagram.

The proposed approach will allow a more rational use of network frequency resources compared to the existing methods and increase the spectral efficiency of the network as a whole. The system spectral efficiency is generally defined as [40]:

\[ S_{ef} = \frac{C}{\Delta F \cdot S} \]  

(1)

where \( C \) is the channel throughput, \( \Delta F \) is the channel bandwidth, and \( S \) is the coverage area.

The throughput that must be provided for each user according is determined by Shannon’s theorem, as:

\[ C = \Delta F \cdot \log_2(1 + SINR) \]  

(2)

Each user has information about the received and transmitted signal level to the BS. The spectral efficiency of the network with an irregular structure of SCs depends only on the range of the BS of a certain layer and the value of SINR.

Using the new method of frequency planning, the accessible bandwidth for the user will increase almost three times. Accordingly, the spectral efficiency will grow in proportion to the number of users located within the triangle, formed by base stations of regular coverage.

C. QOE-BASED TRAFFIC DIFFERENTIATION TECHNIQUE FOR THE 5G AND BEYOND NETWORKS

According to QoE estimates, the controller makes their classification under the ordered QoS requirements during data flow processing. The main elements that will manage the process of traffic prioritization at this level are the Policy and Charging Rules Function (PCRF).

The information flows are classified into two groups to provide the QoS parameter, such as flows with a Minimum Guaranteed Bit Rate (GBR) and flows Non-Guaranteed Bit Rate (NGBR). Flows of GBR type have a predefined minimum bit rate set when creating or changing a flow. If there are free resources on the radio channel, transmitting data with a higher bitrate than the minimum bitrate set is possible. Maximum Bit Rate (MBR) limits can also be set. Flows of this type are used, for example, when transmitting Voice over IP (VoIP) traffic. Flows of NGBR type do not guarantee any minimum bit rate. We advise defining which class this or that traffic belongs to based on the known QoS Class Identifier (QCI) parameter. There are nine states of the QCI parameter, each of which is associated with a specific Type of Service (ToS) followed by a transmission channel, throughput, packet loss, and latency. A QCI is a tag in an IPv4 packet called a “channel identifier” (Fig.4).

The use of adaptive traffic prioritization to ensure QoE requirements will allow users using different types of mobile applications simultaneously, if necessary, to use different transmission channels, which are subordinate to BSs of different layers. And then provide high-speed broadband access for each user, or use a shared GBR & NGBR channel, which will allow quality with minimal delay to serve the large volume of data that users will generate. Prioritization will also offload the low-speed channels subordinate to the macro-base station by setting priorities and directing them to serve “volume” users at the FCs layer.

The proposed approach will create a clear, streamlined algorithm for serving different types of service requests at the appropriate levels. As a result, this orderliness will maximize the QoS provision and increase the number of users of such a network, which will lead to an overall increase in company profits.

FIGURE 4. Proposed QoE-based traffic differentiation technique.

The GBR&NGBR channel is formed as a result of the aggregation of frequencies to the user from MC and
SC simultaneously. The decision on aggregation is made at the controller layer. The SDN controller combines two BSs into one virtual station, thereby performing frequency aggregation to the user. For example, let the user is served by a macro BS, which can provide a throughput of 20 Mbit/s. However, this throughput is not enough for the user for a certain individual service. For example, the user requires a high QoE and is not enough. In this case, the controller decides to connect this user additionally to a lower layer station, which can provide 50 Mbit/s to the user. As a result, the frequency spectrum is aggregated according to the method described above, and the user will be able to get a throughput of 100 Mbit/s.

The planning of multi-layer heterogeneous mobile networks requires a network planning method that considers the feasibility and optimality of BSs placement depending on load localization, type of custom GBR and NGBR services [41], and QoE evaluation of users. And ensuring at the same time high energy efficiency of the network by transferring inactive BSs not used to serve users.

D. PROBLEM FORMULATION FOR QOE-DRIVEN RADIO RESOURCE ALLOCATION

We first formalize the SINR model to determine the load that the corresponding user GBR or NGBR contributes in the presence of free radio resources. In a heterogeneous network, the momentary SINR of an end user $j$ from cell $i$ for a dedicated resource block (RB) $r$ is given as [42]:

$$\text{SINR}_{i,j,r}(\tau) = \frac{W_{tx}^i(\tau) \cdot |g_{i,j,r}|^2}{\sum_{k \in X \setminus i} W_{tx}^k \cdot |g_{k,j,r}|^2 + F_r \cdot N_0},$$  

(3)

where $W_{tx}^i(\tau)$ is the cell transmission power and $g_{i,j,r}$ is the channel gain between cell $i$ and user $j$ for RB $r$. $\sum_{k \in X \setminus i} W_{tx}^k \cdot |g_{k,j,r}|^2$ is the interference from all other cells, except the serving cell, for UE $j$ (because the reuse frequency is equal to 1), $N_0$ is the noise power spectral density and $F_r$ is the channel bandwidth of the allocated resource RB blocks for a particular type of service. In case $\text{SINR}_{i,j,r}(\tau)$ determines the average SINR of all RBs in the time period $t$ within a subframe $(r \in (t - l, t))$, where $l$ is the duration in seconds. Further, the resulting spectral efficiency of the UE $j$ of the $i$-th cell on all subframes is equal to the $\eta_{i,j}(t)$ and formalized as $\log_2(1 + \text{SINR}_{i,j}(t))$, according to Shannon’s theorem.

The load that a generates of GBR UE is the ratio of the amount of occupied resources to the total amount of free resources on the network. Firstly, the total load $D_{\text{GBR}(QoE),i}$ caused by GBR UE per cell is determined as $\sum_{j \in \text{GBR}(QoE)} P_{i,j}(t) \cdot d_{G_{\text{GBR}(QoE)}}^j(t)$. Where $P_{i,j}(t)$ is coefficient, that meaning 1 if user $j$ is connected to the $i$-th BS at time $t$, otherwise, it is equal to 0. According to the ordered QoE estimates, the parameter $d_{G_{\text{GBR}(QoE)}}^j(t)$ is the throughput demand of user $j$ GBR at time $t$. In the cell $i$ the $\omega_{\text{GBR}(QoE),\text{used}}^j(t)$ determined the number of time-frequency resources (RBs) used at time $t$. The proportion of RBs used by user GBR $j$, serviced by cell $i$, define as:

$$\omega_{\text{GBR}(QoE),\text{used}}^j(t) = \frac{P_{i,j}(t) \cdot d_{G_{\text{GBR}(QoE)}}^j(t)}{\min\left(F_r \cdot \eta_{i,j}(t), C_{BH}^j(t)\right)}.$$  

(4)

Therefore, it follows that the throughput of the 5G radio access network is bounded by the accessible RBs and the spectral efficiency of the $i$-th cell. Therefore, it follows that the throughput of the 5G radio access network is bounded by the accessible RBs and the spectral efficiency of the $i$-th cell. The $\rho_{BH}^j(t)$ is the load of the $i$-th cell by GBR user $j$ is the ratio of the number of busy resources to the number of accessible resources and is represented as:

$$\rho_{BH}^j(t) = \frac{\omega_{\text{GBR}(QoE),\text{used}}^j(t)}{\min(\omega_{AC}^j(t), \omega_{BH}^j(t))}.$$  

(5)

where $\omega_{AC}^j(t)$ is the number of accessible RBs in cell $i$ and $\omega_{BH}^j(t)$ is the number of accessible RBs in the 5G network for user $j$, and is formed as $\left[\frac{C_{BH}^j(t)}{F_r \cdot \rho_{BH}^j(t)}\right]$, where $\lfloor \cdot \rfloor$ is the minimum integer, $C_{BH}^j(t)$ is the throughput of cell $i$ at time $t$.

Usage Equation 5, the general load from GBR UE for BS $i$ at period of time $t$ is $\sum_{j \in NGBR(QoE)} P_{i,j}(t) \cdot \rho_{BH}^j(t)$. So, the average load through users’ GBR is defined as:

$$\overline{\rho}_{\text{GBR}(QoE)}^i(t) = \sum_{i \in X} \rho_{\text{GBR}(QoE),i}^j(t).$$  

(6)

where $X$ is the set of all BSs.

In order to ensure that all users achieve an equivalent level of QoE, they must be considered with fairness according to their needs and limitations. To determine the fairness of the distribution of user load between cells, we use the Jain’s fairness index [43]:

$$\Omega(t) = \frac{\left(\sum_{i \in X} \omega_{\text{GBR}(QoE),i}^j(t)\right)^2}{|X| \cdot \sum_{i \in X} \left(\omega_{\text{GBR}(QoE),i}^j(t)\right)^2},$$  

(7)

where $\Omega(t)$ ranges from the $\left[\frac{1}{|X|}, 1\right]$ higher it is, the more well-balanced the distribution of users among the BSs will be.

To the users of NGBR, the purpose is to choose a destination cell that maximizes the effective use of the network. In the case of NGBR users, this is to determine how efficiently the network resources can be utilized to increase the possible rates of all NGBR users. The achievable data rate under proportional fair scheduling is defined as:

$$R_{i,j}(t) = \eta_{i,j}(t) \cdot F_r \cdot \omega_{\text{NGBR}(QoE),\text{used}}^j(t) \left[\frac{|N_{\text{NGBR}(QoE)}^i|}{\omega_{\text{GBR}(QoE),\text{used}}^j(t)}\right],$$  

(8)

where $\lfloor \cdot \rfloor$ is the maximum integer and $\omega_{\text{NGBR}(QoE),\text{used}}^j(t) = \omega_{\text{GBR}(QoE),\text{used}}^j(t) - \omega_{\text{NGBR}(QoE),\text{used}}^j(t)$ is the total accessible RBs for users NGBR $j$ at time $t$ and $|N_{\text{NGBR}(QoE)}^i|$ is the number of NGBR users with QoE demands served by BS $i$. 

131608 VOLUME 10, 2022
Fairness maximization for the GBR users balances resource usage between cells, but such aggregation is not aware of the users’ channel status. This reduces the overall performance of the network, due, for example, to the fact that a user may be connected to a BS that provides weak signal strength, reducing the spectral efficiency as a result. The distribution of resource consumption between different BSs as well as the actual resources consumed by the BSs in the network should be sought to be taken into account. Further, the channel state of the user is fixed, because the GBR user (with fixed speed requirements to provide a certain level of QoE) with a better Channel Quality Index (CQI) is allocated less time and frequency resources to meet its speed relative to the GBR user with a bad channel. By switching users to other BSs, the resource usage allocation between different BSs not only changes but also impacts on the resource consumption of various BSs in the network.

The target function of the problem of effective allocation of radio resources for the GBR group is formalized in the form:

$$\max_{P_{i,j}(t)} \sum_{i \in X} P_{i,j}(t) \cdot \min(R_{i,j}(t), C_{ij}^{BH}(t))$$

subject to

$$\frac{1}{2} \sum_{j \in J} P_{i,j}(t) \cdot \min(R_{i,j}(t), C_{ij}^{BH}(t)) \geq \frac{1}{2} \sum_{j \in J} P_{i,j}(t) \cdot \min(w_{AC}(t), w_{BH}(t)), \forall i \in X$$

Fairness maximization for the GBR users balances resource usage between cells, but such aggregation is not aware of the users’ channel status. This reduces the overall performance of the network, due, for example, to the fact that a user may be connected to a BS that provides weak signal strength, reducing the spectral efficiency as a result. The distribution of resource consumption between different BSs as well as the actual resources consumed by the BSs in the network should be sought to be taken into account. Further, the channel state of the user is fixed, because the GBR user (with fixed speed requirements to provide a certain level of QoE) with a better Channel Quality Index (CQI) is allocated less time and frequency resources to meet its speed relative to the GBR user with a bad channel. By switching users to other BSs, the resource usage allocation between different BSs not only changes but also impacts on the resource consumption of various BSs in the network.

The target function of the problem of effective allocation of radio resources for the GBR group is formalized in the form:

$$\max_{P_{i,j}(t)} \sum_{i \in X} P_{i,j}(t) \cdot \min(R_{i,j}(t), C_{ij}^{BH}(t))$$

subject to

$$\frac{1}{2} \sum_{j \in J} P_{i,j}(t) \cdot \min(R_{i,j}(t), C_{ij}^{BH}(t)) \geq \frac{1}{2} \sum_{j \in J} P_{i,j}(t) \cdot \min(w_{AC}(t), w_{BH}(t)), \forall i \in X$$

The overall energy consumption of the proposed 5G radio access layer is estimated as:

$$E_{total5G} = \sum_{i=0}^{z-1} E(i)_{MC} + \sum_{i=0}^{y-1} E(i)_{SC},$$

where $E(i)_{MC}$ is the energy consumption $MCBS_{i}$ i-th macro cell, $z$ is the number of macro-level BSs in the 5G network, $E(i)_{SC}$ is the energy consumption $SCBS_{i}$ i-th femto and pico cells, $y$ is the number of SCs BS in the 5G network.

The set of all base stations $X$ in a 5G network consists of the sum of $z$ and $y$. It is assumed that all UE users $j$ in the coverage area of base station $X_i$ can be served. The establishment of communication between $MCBS_{i}$, $SCBS_{i}$ and UE is indicated as $a_{i,j}$ and determined according to equation (16), where $MCBS_{i}(0 \leq i < z - 1)$ and $SCBS_{i}(0 \leq i < y - 1)$.

$$a_{i,j} = \begin{cases} 1, & \text{if } u_j \text{ connecting to } X_i \\ 0, & \text{otherwise, } (0 \leq i < X_i + 1, 0 \leq j < u) \end{cases}$$

The demand for a certain type of data traffic by the user is designated as $d_j$, and its value is determined as:

$$d_j = \begin{cases} 1, & \text{if } u_j \text{ required data service} \\ 0, & \text{otherwise} (0 \leq j < u) \end{cases}$$

The required data traffic type and QoE of the UE is defined as:

$$r_j = \begin{cases} 1, & \text{if } u_j \text{ demanded high throughput QoE} \text{high}, \\ 0, & \text{if } u_j \text{ demanded low throughput QoE} \text{low} \end{cases}$$
In Equation (19) \( v_i \) is related to whether it is in the active state \( X_i \) or not, and \( s_i \) is related to whether it is energy-saving or not (20):

\[
v_i = \begin{cases} 
1, & \text{if } X_i \text{ is in on state, } (0 \leq i < X + 1) \\
0, & \text{otherwise}
\end{cases}
\]

(19)

\[
s_i = \begin{cases} 
1, & \text{if } X_i \text{ is in sleep state, } (0 \leq i < X) \\
0, & \text{otherwise}
\end{cases}
\]

(20)

Using the above equations, we obtain the energy consumption \( E(i)_{SC} \) of small cells \( SCBS_i \), as:

\[
E(i)_{SC} = v_i \cdot (1 - s_i) \cdot (E_{SC}^v + \rho_{SC} \cdot E_{SC}^s(i)) + s_i \cdot (1 - v_i) \cdot E_{SC}^s,
\]

where, if \( SCBS_i \) it is in an active state, the energy consumption is defined as the sum of the fixed energy consumption \( E_{SC}^v \) and the energy consumption \( E_{SC}^s(i) \), which depends on the load of the SC \( \rho_{SC} \). The energy consumption is calculated for the sleep state in energy-saving mode. Depending on the user load, the energy consumption \( E_{SC}^s(i) \) is defined as:

\[
E_{SC}^s(i) = E_{SC}^s, (21)
\]

where \( E_{SC}^{s, \text{max}} \) is the maximum throughput of a SC, \( C_{QoEh} \) is the throughput is needed to serve users with high quality of service, such as high-resolution video, the user QoE score is 5), is \( C_{QoEl} \) the throughput is needed to serve users with low quality of service, such as low-resolution video, the user QoE score is 2).

Similarly, the energy consumption \( E_{MC} \) for macro cells \( MCBS_i \) is defined by the Equation (23), as the sum of the fixed energy consumption and the energy consumption depending on the user load.

\[
E(i)_{MC} = E_{MC}^f + \rho_{MC} \cdot E_{MC}^s(i),
\]

(23)

where \( E_{MC}^{s, \text{max}} \) is defined as (24):

\[
E_{MC}^s(i) = E_{MC}^{s, \text{max}}, \sum_{j=0}^{u-1} \left( d_j \cdot a_{j,i} \cdot \frac{(r_j \cdot C_{QoEh} + (1 - r_j) \cdot C_{QoEl})}{C_{MC}^{s, \text{max}}} \right),
\]

(24)

where \( C_{MC}^{s, \text{max}} \) is the maximum throughput of a macro cell.

The optimization problem of the proposed energy-saving scheme is formulated as follows, taking into account the above constraints of problem (9) and (13):

\[
\min_{E_{\text{total}5G}, v_i} \text{subject to}
\]

\[
v_i + s_i \leq 1, \quad (0 \leq i \leq y - 1)
\]

(25)

(26)

Since \( SCBS_i \) it can be in one of the active, energy-saving and non-active states, the sum \( v_i \) for the active state and for the non-active state must be less than or equal to 1.

Thus, according to [44], the energy efficiency of a heterogeneous 5G network is defined as the average data rate of MCs and SCs divided by the total energy consumption of MCs and SCs.

Accordingly, to improve the energy efficiency of 5G networks, it is necessary to solve the problem (Equation 9 and Equation 13) of optimal allocation of the necessary radio resources to provide the throughput for the user’s QoE requirements, so as to minimize the energy consumption of the 5G radio access network by switching the SCs BS to an energy-saving state (Equation 25).

The optimization problem formulated in this article is a mixed integer linear programming since some variables in the proposed optimization are constrained to some discrete integer values.

\section{F. ENERGY-EFFICIENT QOE-DRIVEN RRM METHOD FOR RESOURCE PLANNING AND FORMATION OF THE 5G RADIO ACCESS LAYER STRUCTURE}

To solve the above problem of effective allocation of radio resources, taking into account QoE requirements of users, we proposed a block scheme of QoE-oriented method of resource allocation and formation of the structure of 5G RAN (Fig.6).

![Figure 6](https://example.com/f6.png)

\textbf{FIGURE 6.} A general block scheme of the energy-efficient resource allocation and formation of the 5G coverage based on QoE requirement and Voronoi diagram.

This method is programmatically automated in a simulation model, using ready-made program classes that solve integer linear and nonlinear programming problems. The developed method is based on three main steps. In the first stage, the SDN network controller, based on the exchange of signaling data between BSs and users, collects statistics on the time-space localization of user load in order to create a specific load map for a particular coverage area. It also considers the possibility for users to order a certain level of QoE from the network operator by means of a mobile application through QoE estimation. In the second stage of the method, the system analysis of the collected QoE requirements of the users and the received information on the load localization for the decision-making on the users’ individual service
takes place. At the next stage, a decision on modification and formation of the RAN structure (if not created so far) based on the intuitive control logic of the SDN network controller is made automatically.

We will consider the working principle of this method in detail. During the first stage (Fig. 7a), the users of 5G network are connecting to the MC $M(S) = \{MC_1, MC_2, \ldots, MC_i\}$. The following network describes all users attributes: threshold transmission power $P$ from the UE to the BSs; SINR; QoS_class of services GBR/NGBR, and the ordered QoE score of the quality of service. Based on these attributes, the UE initiates a request for service (QoE on request) sent through the “up” channel to the BS. The SCs in energy-saving mode are characterized by the disabled channel “down” (Down-link), only the channel “up” is enabled. The MC base station, having received signaling messages from users, gives instructions to SCs through channel UL, to determine Channel State Information (CSI) that characterizes the distance change vector to the BS by changing the signal power ($P_{UL_{UE_j}}$), and then for each SC a matrix of “permissible users” $K_j$, containing statistics on all users that a particular BS can serve. The formation of the load map is carried out by the SDN controller, which analyzes the matrices $\{K_j\}^{S_j}$, where $\{K_j\}$ is the matrix of UE that can be served by the BS, $\{S_j\}$ is the matrix UE, which are already served) the BS and “discard” users who can not be served due to incoherence of radio channel CSI parameters.

$$K_j = \{UE_{res}, UE_{res}, \ldots, UE_{res}\}$$
$$S_j = \{UE_{cur}, UE_{cur}, \ldots, UE_{cur}\}$$

(27)

where $UE_{res}$ is the user equipment can be serviced by SCN, SCN−1, $UE_{current}$ is the user device currently serviced by SCN, SCN−1.

Consequently, the matrices $\{K'\}$ are formed for each BS of the level SCN, SCN−1. This matrix is for storing information about UE, which will be serviced by specific BSs. This information is sent to the lower level, and a general map of the present user load $R_j$ is formed.

$$R_j = \sum K_j + \sum S_j,$$

(28)

For each of the SCs, a different matrix of users $\{K'\}$ (29) is formed, which can be served by a particular BS.

$$K'_j = \sum K_j - \sum UE_{notserviced},$$

(29)

$UE_{notserviced}$ a user device that a particular BS cannot serve due to low signal strength.

A matrix $\{K'_j\}$ (30) is generated based on the control information on the analysis of the network state and the active load map. This matrix includes information on users already served by the radio access layer structure and information about users who potentially need to be serviced by a certain BS.

$$R'_j = \sum K'_j - \sum UE_{notserviced},$$

(30)

The second stage of the block scheme, which is responsible for analyzing the state of the radio access layer, is one of the main and is shown in Fig. 7b. The controller decides how to provide service to the UE user with a specific QoE score within each accessible BS. Three rules are used to make the selection:

1) If a high distance-vector change value characterizes the UE to the BS, the users will be served by the $MC_{DL}$. Similarly, the $R_j$ will serve users requiring any other QoS_class, particularly those requiring low QoE, in case it is not possible to connect it to an alternative SC.
2) If the user UE is characterized by the low distance-vector change value to the BS and high traffic demands QoS_class_high, which corresponds to estimates QoE (4-5) and the connection of it to the RAN active structural elements will not lead to overload, then SDN controller to connect to the corresponding SC with parallel channel aggregation of MC_DL. Thus, by aggregating frequencies, a better quality of service is provided. As a result, the matrices \( \{K_j\} \) and \( \{S_j\} \) are analyzed for the necessity to modify the structure of the 5G RAN by moving the SCs to an inactive state to reduce the network energy consumption. The management is transferred to the third block if there is such a need, otherwise the current load map is restructured, and the control is transferred to the first stage of the general block diagram of the method.

3) If the UE user is characterized by the low distance-vector change value to the BS and high traffic demands, which corresponds to estimates QoE (4-5) and there is no possibility of connecting BSs, a decision is made to use a larger number of SCs and the verification proceeds to the third stage of the general block-schematic method.

By obtaining the necessary information about UE and matrix \( \{S_j\} \), which contains information for active users of SCs, as well as using data about the location of BSs and their coverage area, the controller decides which BS and their number are needed to serve the current load. The radio resource allocation in the 5G network is based on the analysis of user QoE requirements. The essence of this is to allocate this share of resources to provide throughput for a service that requires a specific level of QoE. This leads to the restructuring of the 5G RAN based on Voronoi diagram (Fig. 8 a). This is achieved by shifting the center of gravity of operation from rigid, always-on SCs focused on the BS to cells that adapt their on/off state based on the QoE requirements of the user. Then the definition and verification of DL parameters for users of UE and identification of service structural elements of the RAN for each UE according to the algorithm shown in Fig. 8 b.

Each of the SC creates a UE set for which the power level value of the received signal is greater, compared to a certain limit value \( P_{max} \). For this purpose, each SC sends a broadcast echo request which includes the identifier of the given SC to all UE in SC coverage zone. After receiving the request, each user UE responds by giving the received value \( (P_j, SINR_j) \). Then the measured data are transmitted to the SDN controller. According to the algorithms proposed above, the controller blocks users whose BS signal strength values are below the set threshold. On this basis, the controller forms for each user from the received set the list of BSs, whose SINR exceeds the allowable value. In this case, the controller generates a set of potential BS, capable of serving this user. From the set of potential BS is selected BS with the highest SINR value and checked to see if it is not overloaded. The overload of the selected BS is removed from the set of potential BSs, and the process is repeated until a BS capable of serving the user is selected. The controller then commands the user to switch to the selected cell. In case all the SCs are overloaded, the user will be served by the MCs. The system then switches back to step 3 (display block execution).

**IV. EVALUATING THE EFFICIENCY OF THE PROPOSED SOLUTION BY SIMULATION AND EXPERIMENT**

A simulation model of 5G network was created to evaluate the effectiveness of the proposed solutions to optimize the network energy consumption by the criterion of the ordered QoE by business users. The model takes into account the main technical parameters of 5G to create full-scale research conditions, based on Matlab 5G Toolbox [45]. 5G Toolbox provides fully 3GPP-compliant functions and reference designs for modeling, simulation and verification [46], [47], [48]. The simulation model also automates the proposed QoE-driven RRM method by writing its own software code. The proposed method is implemented using the Discrete-Event Simulation and Modelling in Java (DESMO-J) library.
This tool supports both event-driven and process-driven simulation representations. DESMO-J provides access to ready-to-use classes such as queues with bounded or infinite space, random number generators based on different random number distributions, various statistical data collection tools, and their graphical visualization. We also used an off-the-shelf Java library to create two-dimensional Voronoi diagrams using the Fortune algorithm to simulate adaptive 5G coverage formation by SCs.

The model consists of a working field on which stations of 3 levels are placed: MC, PC, and FC. In the simulations, it is assumed that the placement of MC leads to relatively long transmission distances. In this case, the transmission channels from BS to UE and interference from BS to BS in the external environment are modeled only as Non-Line of Sight (NLoS) propagation. The dense deployment of SCs in HetNets leads to use Line of Sight (LoS) transmissions in the outdoor environment, in particular this case is typical for FCs. Because the transmitting distances between UEs and femto BSs are dramatically reduced. The UEs may located a few meters to BSs. For a realistic scenario, the transmission channels for PCs are modeled only as NLoS propagation by generating random cell models based on Voronoi diagrams. Because within the FC coverage there may be obstacles in the signal propagation process, such as buildings, trees and uneven terrain. Within the operating field, the simulation of user traffic takes place. With a certain periodicity, the image on the working field is updated, and the graphs, which are the final goal of the model development, are displayed (Fig. 9).

The MCs are placed in forming an equilateral triangle, which is the study area of the model and the design area of small cells. When calculating the simulation results, only the users of this area will be taken into account.

SCs within the triangle are placed according to the following principle: FCs are installed, if necessary, in places of the most significant accumulation of user load. PCs are set according to the parametric point process with the minimum distance between the BSs, followed by the Voronoi tessellation.

The simulation model considers user parameters such as mobility, which means the speed and direction of movement, the throughput required by the user, the probability of activity, and the non-activity of user parameters.

Our simulator uses the system parameters listed in Table 2 and Table 3.

The generation of traffic by users takes place in several stages:

- Generate a session and select its parameters, such as session usage time, type of service, and required QoE score (from 1 to 5);
- Generate traffic for the session until it finishes;
- After it ends, generate new traffic with new parameters.

In the simulation model, each BS includes a list of users, a list of active and passive users, a node name identifier, and a load-sharing table. The BS also has appropriate methods to calculate the zone in which the user is located. This is necessary for the further algorithm, which will be to analyze the user zone and resource requirements. There is a calculation of the number of active users in different zones of the BS. Each iteration of the generated data users sends to the corresponding serving BS. After reading the data, the possible free resources are analyzed. The BS in the proposed model allows distribution with a channel width from 0.2 to 20 MHz [49]. Network testing and performance analysis were conducted over a 12-hour period. During this, the load graphs

![Figure 9](image-url)
The total load graph is similar to the one observed in real networks because there are similar load fluctuations and variations within a reasonably wide range. The maximum load on the MC level is observed only during the highest load hours, and at an average load on the network, the MC level is loaded only by 40-70%. This is due to the fact that the SCs take most of the user load, due to which the macro coverage has enough resources to serve users who are outside the coverage of SCs, as well as users with high mobility. This ensures that users with high mobility are served with minimal handover. The lost load is negligible and occurs when users are very densely located in a small area. Then the maximum load of the SCs adjacent to this area is observed, and the users who could not be served by the SCs due to insufficient resources the SCs are served by the MC. There is a lost user load if the MC is also overloaded.

Switching to the energy-saving mode of SCs is reasonable because at low load on the network, most SCs are idle, so the load on them is insignificant, or there is no load at all. Switching to the active mode of most SCs in the hour of the highest load is also reasonable, as users generate a significant amount of traffic, and we must ensure the requirement of the ordered level of QoE.

As a result of 20 series of simulations, a graph of the average share of active SCs was obtained (Fig. 15). Thus, the proposed method improves the energy efficiency of the network by ensuring that, on average, only 60% of all cells are active. Accordingly, the remaining 40% of the cells are switched to energy-saving mode. Assuming that the average number of active SCs is 60%, it is no longer necessary for all BS to be activated.

Since FCs have a short range and are installed mainly indoors, they are more often idle than PCs. The PCs which are installed to cover an area with a radius of 200-300m, are active for a longer period of time than FCs because due to the greater radius of action, there is a greater probability of appearance of the user in the area of the picocell than in the area of PC action.

The proposed method permits only a certain fraction of the BSs, providing an energy-saving opportunity. Suppose \( n\% \) is the fraction of active SCs at a fixed moment. This SC consumes 15 Wh of energy according to paper [50], and the total SCs energy consumption we define, as \( \sum E \). Then the fraction of non-active cells defines as \( 100\% - n\% \), and the fraction of cells defines as \( 100\% - n\% \). Each of these non-active cells consumes 8 Wh of energy. The average energy consumption of a cell defines as:

\[
E_{cell} = 15Wh \cdot (n\%/100\%) + 8Wh \cdot ((100\% - n\%) /100\%)
\]

\[ (31) \]
According to the simulation results, the increase in energy efficiency was 9-42%. Experimental studies show that the developed method is effective both at low and high network load, but at high load the increase in energy efficiency is somewhat less than at low load. This is due to the fact that at high load the share of active SCs is larger, which leads to an increase in the total energy consumption by the base stations of the network. We also conclude that network goals should depend on the context of operation, particularly network load and QoE requirements. In the context of low load on the network cells, guaranteeing high QoE is easily achieved by a significant excess of available radio resources. The focus should be on minimizing energy consumption rather than system throughput. Conversely, under high load conditions, the limited available radio resources require that resource allocation algorithms focus on system throughput to guarantee the QoE requirements of users, with energy consumption becoming less critical.

To compare and evaluate the effectiveness of implementing the new method with the known, we used the criterion of probability that the user will receive a certain throughput channel. For this purpose, we used the Cumulative Distribution Function (CDF) of the probability of allocation of the average throughput for users when using the different RRM methods. The CDF of average throughput using the existing RR and PF methods [19] was compared with the proposed method relative to the etalon CDF of throughput based on the ordered users’ QoE requirements that characterize a certain level of service quality. Namely, allocating the necessary throughput to view the video service in the ordered quality.

Accordingly, the resource allocation method will be more effective, whose cumulative function is close to the etalon one. The comparison was carried out for two cases of user localization: under the conditions of preferential localization in the edge zone of the cell (low values of SINR Fig. 16 a, blue curve and the central zone of the cell high values of SINR Fig. 16 a, red curve). The CQI value corresponding to the measured SNR is calculated using the mapping technique shown in Fig. 16 b.

The percentage of energy saved is calculated using the formula:

\[
E_{\text{Saved}} = 100\% \left(1 - \frac{E_{\text{cell}}}{\sum E}\right)
\]  

(32)
As a result of the comparison, it was found that the proposed method provides better adaptability of resource allocation to ensure the ordered QoE compared to the known ones. In particular, the proposed method conducts rational resource allocation by analyzing the users QoE scores by solving the problem of flexible redistribution of radio resources between different user requirements by adapting the function to the etalon, as shown in Fig.17.

We investigated the possibility of using the proposed method for the adaptive formation of the 5G RAN structure to reduce the network’s energy consumption. We proposed that when the BSs do not serve any users, put them in energy-saving mode and thus form the SC structure of RAN, which will adjust to the needs of users. To fully increase the energy efficiency of 5G network, it is necessary to provide a similar procedure both at the core layer in the process of Virtual Machine (VM) server organization and at the SDN network (Fig.18 a), using the minimum number of network nodes with the provision of requirements on the ordered QoE for users. The representation of the obtained results in the form of CDF allows for visually evaluating the adaptability of a specific method for allocated throughput to provide the QoE requirements of users. And also to estimate if there is an excess of the allocated resources that leads to giving higher throughput to the users than they require. If such redundancy exists, our method will reduce the energy consumption of the network by deactivating the SCs, while guaranteeing the throughput for a certain level of user QoE, as shown in Fig.18 b.

We determined that the proposed method reduces the energy consumption of the network by 8.7%.

To quantify the effectiveness of the use of the developed method of resource allocation in 5G network, we formed QoE-based index of service-level quality $Q$, which we also interpret as a normalized QoE criterion for network optimization:

$$Q = \frac{\sum_{QoE=1}^{5} \sum_{j=1}^{N_{users,QoE}} Y_{jQoE,\text{current}}(P_{\text{price}_{jQoE,\text{current}}})}{\sum_{QoE=1}^{5} \sum_{j=1}^{N_{users,QoE}} Y_{jQoE,\text{ordered}}(P_{\text{price}_{jQoE,\text{max}}})}, \quad (33)$$
where \( Y_{QoE_{\text{current}}} \) is the throughput provided to the j-th user of the QoE category, \( N_{\text{users}_{\text{QoE}}} \) is the number of user’s with QoE requirements, \( P_{\text{price}_{\text{current}}} \) is the price of the user’s service depending on the throughput provided for a particular QoE category, \( P_{\text{price}_{\text{max}}} \) is the maximum possible price of user service, \( Y_{\text{QoE}_{\text{ordered}}} \) is the throughput required by the i-th user for the QoE category.

\[
Y_{QoE_{\text{current}}} = \left( N_{RB_j} \cdot k \cdot l \cdot \text{MIMO} \cdot \text{Kod}_{\text{RATE}} \cdot \log_2(M) \right) \cdot S
\]  

(34)

where \( S \) is the percent of the user data in the frame (\( S = 0.75 \)), the quantity of allocated RBs \( N_{RB_j} \) per second for user \( j \), the quantity of REs (\( k = 12 \) is the quantity of subcarriers, \( l = 7 \) is the number of symbols), the quantity of antennas, code rate \( \text{Kod}_{\text{RATE}} \), modulation positioning \( \log_2(M) \).

The normalized value of the QoE criterion \( Q \) varies from 0 to 1, where a higher value means better quality of service and is calculated as the average value for each zone of 8 cells (from 1 to 16), characterizing a certain level of CQI at different bandwidths from 0.2 to 20 MHz (Fig.19).

Fig. 20 shows that to ensure a high level of ordered QoE in the network with eight BSs, which corresponds to the normalized QoE criterion \( Q = 1 \), the network operator must have a channel bandwidth of 20 MHz with the existing methods of RRM (dashed rectangle of blue color) and 15 MHz with the proposed RRM method (Fig. 21).
V. FUTURE RESEARCH DIRECTIONS AND DISCUSSIONS

According to the proposed method, switching to the energy-saving mode of BSs will reduce the number of active transmitters of BSs, thereby significantly reducing the energy consumption of SCs, the frequency of handover for users, and the cross-interference effects between SCs. In addition to the above, for the operator, it will mean a reduction in operating costs for the network and, consequently, an increase in profits. Using the proposed RAN construction method, the operator can introduce such a service as “RAN as a Service”. That is, to rent the RAN to other operators, that is, in periods when SCs do not serve the user load, to serve users of other operators, and at the same time to get a significant profit.

In particular, there are some limitations to the implementation of the proposed method on real 5G networks. Firstly, this approach is focused on business users who are ready to pay more money for high-quality service in moments of high network load in order to guarantee the quality of provision of a particular service. We believe that for practical implementation of such approach, mobile operators need to develop a special mobile QoE application for individual business users. This application will allow ordering a certain level of QoE for services at a certain fee price under different load conditions. Thus, the operator will charge an additional fee to the tariff plan for business users who want guaranteed service quality during high loads.

In future work, we will improve the proposed method of RRM using AI technology to ensure QoE and reduce energy consumption in 6G networks.

Using AI technologies, 6G networks will be fully automatic, not needing manual control. This means that 6G networks will not be proprietary, but self-managed, leading to the concept of “Individual Mobile Network on Demand”. Furthermore, our future research direction is to propose a novel AI-based resource management model for more useful QoE detection and energy optimization in 6G networks.

The proposed model will involve a two-stage training process. First, initial 5G/6G network statistics regarding user mobility and traffic demand are sent to the training block. The model will then find the best network configuration to maximize key performance indicators before the first network deployment. In the second stage, the monitoring system will collect statistics on the performance of the network already running, transferring this data to the neural network plane for further analytics. Such a solution would offer unlimited possibilities to improve the efficiency of the resource allocation process and minimize energy consumption, as all bottlenecks and failures would be reported to the intelligent control plane so that they could be eliminated in the future.

To predict the time load on a cell, in future work we will develop a recurrence neural network model using Long Short-Term Memory (LSTM) cells to predict user traffic on short and long-term scales. Also, this neural network model will allow predicting the business demands of the users with high accuracy based on the collected statistics about the individual QoE requirements of the users from the mobile application. In this case, there will be no need to constantly order their QoE requirements for services through the mobile app, making such processes self-managing.

VI. CONCLUSION

We provided a more generalized view of quantifying the quality of 5G/6G network performance by including a QoE-based index of service-level quality. This index takes advantage of the fact that quality can be viewed as an indirect measure of utility. Based on the assumption that not all user sessions need to have the same QoE weighting, we aim to provide a new radio resource scheduling approach. This approach focuses on allocating the throughput to obtain an individual QoE for the user at a certain fee price under different load conditions. This can be used by the mobile operator used to quantify the overall utility of providing system services to optimize radio resources and network energy consumption. For this reason, we developed a new RRM method taking into account the QoE criterion to improve energy efficiency and maximize individual user QoE in 5G HetNet. This method dynamically balances energy consumption and throughput according to system load and user QoE requirements filling the gap between existing approaches. The simulation results showed that the proposed method allowed to use the available radio resources more efficiently by 25% and reduce the energy consumption of 5G RAN by 8.7% to provide the requested QoE for users compared to traditional RRM methods, such as the PF scheduling algorithm. We proved that with the same radio resources the proposed method provides the best average index of service level quality on demand, which we interpret as a normalized QoE criterion. In particular, with the traditional method, the normalized QoE criterion \( Q = 0.7 \), and with the proposed method \( Q = 0.99 \). Furthermore, the future goal of the work is to propose a new AI-based framework for more useful QoE determination and energy efficiency improvement in future 6G networks.
[42] F. H. Khan and M. Portmann, “Joint QoS-control and handover optimization in backhaul aware SDN-based LTE networks,” Wireless Netw., vol. 26, no. 4, pp. 2707–2729, May 2019.

[43] O. Abuajwa, M. Roslee, Z. B. Yusoff, L. L. Chuan, and P. W. Leong, “Resource allocation for throughput versus fairness trade-offs under user data rate fairness in NOMA systems in 5G networks,” Appt. Sci., vol. 12, no. 7, p. 3226, Mar. 2022.

[44] Z. Albataineh, A. Andrawes, N. F. Abdullah, and R. Nordin, “Energy-efficient beyond 5G multiple access technique with simultaneous wireless information and power transfer for the factory of the future,” Energies, vol. 15, no. 16, p. 6059, Aug. 2022.

[45] 5G Toolbox 5G Toolbox Documentation. Accessed: Nov. 18, 2022. [Online]. Available: https://www.mathworks.com/help/5g/index.html?_tid=CRUX_lftnav

[46] Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16), document TR 38.901 V16.0.0, 3GPP, Oct. 2019.

[47] Technical Specification Group Radio Access Network; Study on Self evaluation Towards IMT-2020 Submission (Release 16), document TR 37.901, V16.0.0, 3GPP, Jun. 2019.

[48] Y. Xue, Y. Zhai, E. Sousa, W. Li, L. Zhang, Z. Hong, and Y. Wu, “5G system level simulation calibration using MATLAB 5G toolbox,” in Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB), Jun. 2022, pp. 01–06.

[49] H. Beshley, M. Beshley, M. Medvetskyi, and J. Pyrih, “QoS-aware optimal radio resource allocation method for machine-type communications in 5G LTE and beyond cellular networks,” Wireless Commun. Mobile Comput., vol. 2021, pp. 1–18, May 2021.

[50] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, “How much energy is needed to run a wireless network?” IEEE Wireless Commun., vol. 18, no. 5, pp. 40–49, Oct. 2011.

MYKOLA BESHLEY received the Ph.D. degree in telecommunication systems and networks from Lviv Polytechnic National University, in 2015, and the Dr.Sc. degree in 2021. He is currently an Associate Professor with the Telecommunications Department, Lviv Polytechnic National University. His current research interests include the next-generation Internet of Things, cloud computing, big data, software-defined networks, intent-based networks, and 5G heterogeneous networks.

NATALIA KRYVINSKA received the Ph.D. degree in electrical and IT engineering from the Vienna University of Technology, Austria, and the Habilitation degree (Docent title) in management information systems from Comenius University in Bratislava, Slovakia. She is a Full Professor at the Information Systems Department, Faculty of Management, Comenius University in Bratislava, and a University Lecturer at the E-Business Research Group, University of Vienna’s School of Business, Economics and Statistics. Her research interests include service management and systems engineering.

HALYNA BESHLEY received the Ph.D. degree in telecommunication systems and nets from Lviv Polytechnic National University, in 2021. She is currently an Assistant at the Telecommunications Department, Lviv Polytechnic National University. Her research interests include 5G, cloud computing, SDN, M2M, and the IoT.