Traffic-Aware UAV Placement Strategies for Load Balancing in 5G Cellular Hotspots

Byomakesh Mahapatra*, Anuradha Verma*, Deepika Gupta†, Pankaj K. Sharma*, and Ashok K. Turuk‡
*Department of Electronics and Communication Engineering,
National Institute of Technology Rourkela, Odisha, India
†Department of Electronics and Communication Engineering,
Dr. S. P. M. International Institute of Information Technology, Naya Raipur, Chhattisgarh, India
‡Department of Computer Science and Engineering,
National Institute of Technology Rourkela, Odisha, India
Email: byomakesh22@gmail.com, 519ec1026@nitrkl.ac.in, deepika@iiitnr.edu.in, sharmap@nitrkl.ac.in, akturuk@nitrkl.ac.in

***Abstract*—In the fifth-generation (5G) network, dependency on the cellular platforms increases due to an increase in the number of cellular and wireless devices. In such network, a hotspot situation arises when the user density goes beyond the threshold capacity. To reduce the load of this hotspot we have proposed a traffic-aware proactive load balancing (TPLBA) strategy. This strategy used a feedback approach to monitor and control the traffic load at the cellular base station or gNodeB. When the traffic load goes beyond a certain value, the main control unit (MCU) present at the base band unit (BBU) takes preventive actions by putting one or more number of F-RRHs at the probable hotspot. These F-RRH share the traffic load of the gNB to maintain the quality-of-service (QoS) of the cellular network. To implement the proposed strategy, we have used *Tu-Vienna LTE* simulator. Further, the simulation results show that the proposed TPLBA algorithm significantly improves the QoS by improving UE throughput, UE spectral efficiency, and blocking probability.**

I. Introduction

The advancement of technologies like internet-of-things (IoT), machine-to-machine (M2M), vehicle-to-vehicle (V2V) communications etc., increases the cellular and wireless traffic at the base station [1], [2]. An increase in this traffic increases the channel occupancy and computational load at the base station. A hotspot situation is observed at the corresponding base station when this cellular traffic goes beyond the threshold limit. In this hotspot situation, the base station restricted further user attachment. This restriction impacted various quality-of-service (QoS) parameters like UE throughput, UE spectral efficiency, and blocking probability [3].

A. 5G H-RAN Architecture

The 5G radio access network (RAN) architecture allows multiple radio access technologies (M-RAT) to process over a single radio platform [4]. The third-generation partnership project (3GPP) in Release-16 approved the 5G-RAN architecture, which is shown in Figure 1. This architecture consists of various functional units, which are as follows:

- ***gNodeB:*** The gNodeB (gNB) consists of a static remote radio head (S-RRH), and base band unit (BBU) connected through a fronthaul link. The RRH handles the uplink/downlink communications of RF signals and allocates channel bandwidth to the user equipment (UE) [5]. The BBU handle the user plane (UP) functions and performs packet processing at the baseband processors.
- ***5G core network:*** The 5G core network comprises of five main units:
  - **Access and Mobility Management Function (AMF):** The AMF in the 5G core transmits and receives all the initial UE related information through different interfaces. The AMF is responsible for UEs connections and mobility-related information.
  - **Unified Data Management (UDM):** The UDM in 5G is used for authentication of UEs and stores the identity of each UEs.
  - **Session Management Function (SMF):** SMF is responsible for selecting an appropriate user plane function during the setup of a protocol data unit (PDU) session between end devices.
  - **User Plane Function (UPF):** The UPF in the 5G core network is responsible for packet forwarding, packet routing, interconnection to the Internet Service Providers (ISP), IP and Multimedia Service (IMS) data network. The UPF also manages QoS and throttling of data or bandwidth based on the service level agreement (SLA) and load at the gNB of the 5G network.
  - **Policy Control Function (PCF):** The PCF is one of the control plane units responsible for QoS monitoring, making QoS policy, and implementing charging rules for associated UEs. The information regarding SLA between UE and operator is stored by PCF unit.
- ***Backhaul Link:*** The backhauls are the optical interconnection network between BBU and core network in a 5G architecture. All the user plane and control plane data are transferred through this network using different optical modulation techniques.
- ***Flying-RRH:*** The F-RRH are the UAV assisted access point embedded with radio transceiver units and lower-level signal processing units. In general, F-RRH resides at a gNB in ideal conditions, and is called upon based on

978-1-6654-2337-3/21/$31.00 © 2021 IEEE

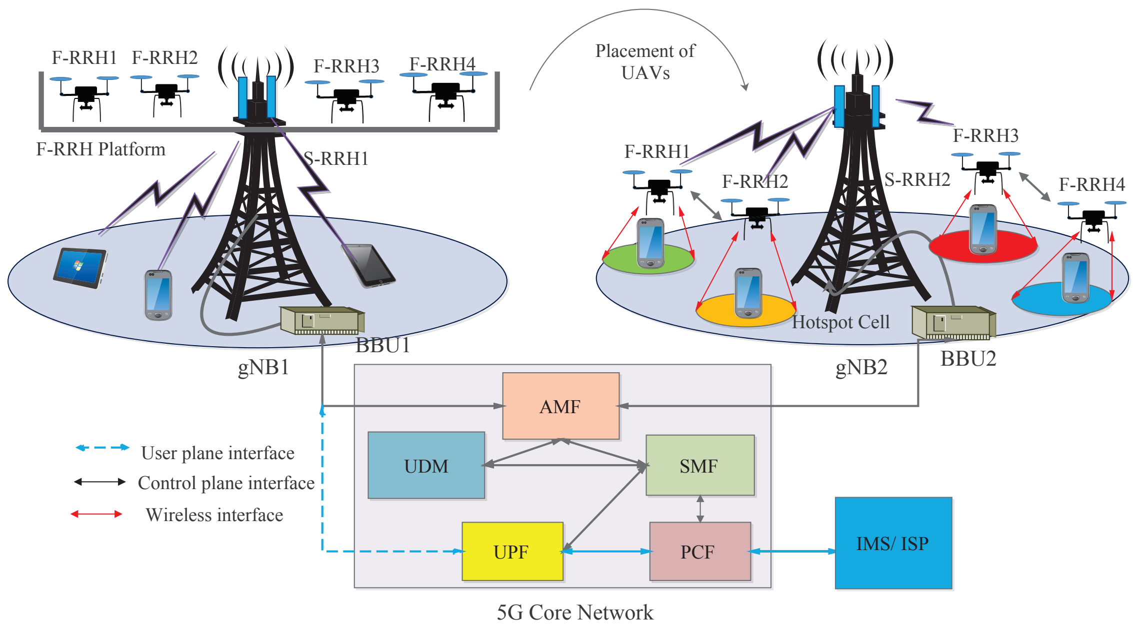

Fig. 1: 5G H-RAN with a Hotspot cell.

demand by the cellular cell.

B. Quality-of-Service Parameters in 5G H-RAN

In a cellular network, the QoS is defined as the method or mechanism which determines the performance of the different network's parameters based on a predefined SLA [6]. The QoS parameters which are considered for H-RAN are:

- ***Average UE Throughput:*** A UE throughput in cellular communication, is defined in term of the numbers of packets or bits successfully received by a UE, transmitted by the gNB within a specified period of time. The UE throughput in a cellular system depends on the characteristics of the transmission medium, network congestion and the number of UE attached to a cellular base station.
- ***Average UE Spectral Efficiency (SE):*** The spectral efficiency in a cellular network is defined as the number of bits transmitted to a UEs by maintaining the QoS. The spectral efficiency of a UE depends on the number of UEs attached to the gNB and the Channel Quality Index (CQI) of the system [7]. A higher CQI value provides better SE by reducing the probability of traffic congestion at the gNB.
- ***UE Blocking Probability (BP):*** According to the tele-traffic theory, the BP is the chance that a UE fails to attach to the cellular base station within a specified time period [8]. The blocking probability depends on various factors like channel condition, bandwidth availability, and UEs arrival rate.

C. Our Contributions

In this paper, we present the following contributions:

- Proposed a hybrid network architecture (H-RAN) for 5G cellular network.
- Proposed traffic aware proactive load balancing algorithm (TPLBA) to avoid/reduce traffic congestion and jamming at a cellular hotspot.
- Developed a traffic based feedback mechanism for deployment F-RRH at the hotspot cell to support higher traffic load.

The rest of the paper is organized as follows; related work is mentioned in section II, the proposed work is mentioned in section III. Section IV will give simulation framework and performance analysis, followed by conclusion in section V.

II. Related Work

Researchers have proposed various solutions to evenly distribute the traffic and improve QoS in a cellular cell. Authors in [9] proposed a Cell Zooming (CZ) based load balancing technique for LTE and LTE-A cellular cells. In this CZ approach, the effective coverage area of eNodeB (BS) dynamically changes w.r.t the UE density and available channel occupancy. In [10] authors proposed a Zone-based Load Balancing (ZLB) algorithm. In this work, an overloaded BS is selectively handover the overloaded UEs to the neighbouring base stations. In [11] authors have proposed a co-operative load balancing (CLB) approach to balance load within a cluster of the cellular cell. In this approach, load sharing between RRHs occurs based on fixed threshold values of available bandwidth under a Base Band Unit (BBU) control.

The authors in [12] proposed two proactive algorithms, i.e., the worst zone algorithm (WZA) and shift algorithm (SA) for UEs load balancing in an ultra-dense network (UDN). The proposed load balancing algorithms are run on a base station controller, which is always in ready mode to handle new UEs whenever the BS is overloaded. The mentioned WZA algorithm used adaptive threshold and Jain fairness index (JFI) to determine the worst zone and handover condition for a UE. At the same time, the mentioned SA algorithm used shifting parameters to decide the load sharing condition between adjacent cells within a small cell cluster. In [13] authors proposed a machine learning approach for deployment of access point enabled UAV at the dense small cell network to enhance the traffic handling capacity of the target base station. Here, the proposed Auto-regressive Integrated Moving Average (ARIMA) model used a regression model to predict a base station's future traffic. Based on these predicted values, the network controller deploys the required numbers of UAVs at the overloaded cell.

Techniques proposed in all the above literature effectively work when the adjacent base stations of a cluster are at distinct load conditions. But, all these mentioned techniques lack a solution to cluster overloading situations when all residing BSs are overloaded. As a solution to this issue, we propose a traffic-aware proactive load balancing strategy to improve QoS at the hotspots.

III. Proposed Work

In this section, various mathematical models and load balancing approaches for a cellular hotspot in a 5G network are presented.

A. System Model

The proposed system model considered M number of UEs connected to a cluster of N number of gNB in uplink and downlink connections. To formulate the UEs arrival and load model, we have only considered the uplink connection between UEs and gNB.

- ***Traffic Model:*** In a cellular network traffic load T_L at a gNB depends on two factors i.e number of attached UEs and the time period (T) for which a UEs hold the available channels.

$$T_L = M \times l_i \times T \tag{1}$$

Let traffic load l_i generated by UE_i increases the bandwidth utilization by ΔB_{ij} amount in a time period t at j^{th} gNB. So the total traffic load at gNB_j is calculated as:

$$T_{Lj} = \sum_{i \in |M|} l_i \tag{2}$$

In terms of bandwidth utilization the above equation can be expressed as:

$$T_{Lj} = \sum_{i \in |M|} \Delta B_{ij} \tag{3}$$

- ***Threshold Value Calculation:*** In a cellular system, the upper threshold T_{LU} determines the overloaded condition at the gNB, whereas a lower threshold T_{LL} determines underloaded conditions for the gNB. An underloaded gNB allows large numbers of handover UEs from an overloaded gNB to make it balanced. The gNB, in which the current load lies between T_{LU} and T_{LL} is considered as balanced. In this work, we have assumed that a balanced gNB allows the new UE request, if the average load $T_{L,avg}$ and usable capacity $\bar{C}_j$ at gNB_J is greater than l_i.

According to [14] the upper threshold T_{LU} and lower threshold T_{LL} depends on two factors i.e, load factor (α_f) and the average available capacity ($T_{L,avg}$). The upper threshold can be calculated as:

$$T_{LU} = T_{L,avg} + \alpha_f \times T_{L,avg} \tag{4}$$

The lower threshold value can be calculated as:

$$T_{LL} = T_{L,avg} - \alpha_f \times T_{L,avg} \tag{5}$$

The load factor α_f determines the load tolerance limit of a gNB and can be calculated as

$$\alpha_f = \frac{T_{L,max} - T_{L,avg}}{T_{L,avg}} \tag{6}$$

where $T_{L,max}$ is the maximum load that a gNB can sustain without damaging the system. The $T_{L,max}$ is always greater than the upper threshold load T_{LU}. Figure 2 shows the different load margin and threshold conditions for a gNB. Decision for UE handover and UAV placement are taken based on these threshold values.

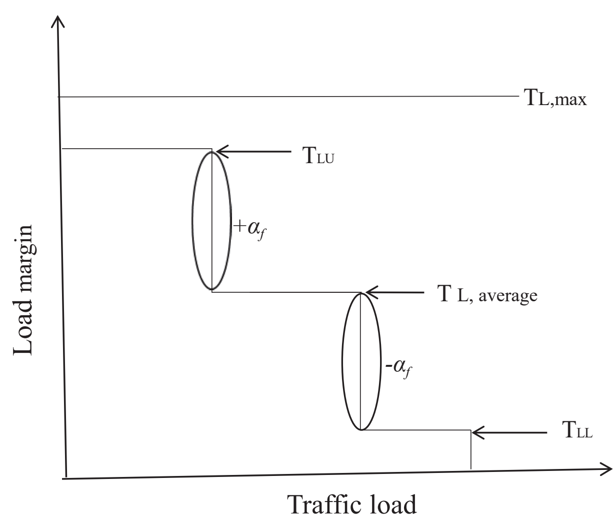

Fig. 2: Different load margin of gNB traffic.

B. Load Balancing Approaches

When a cellular cell is overcrowded, bandwidth and congestion increase and make it a hotspot cell. To avoid this hotspot situation, we have considered two approaches: reactive and proactive. Details on these approaches are discussed below:

1) Reactive Approach: In this approach, the service provider waits until the bandwidth utilization at the gNB exceeds its threshold value. When it exceeds this limit, control information is shared between gNB and UPF of the core network. After receiving this control information, the UPF instructs the gNB to start bandwidth throttling to decrease the number of attached UEs and reduce the coverage area of the BS. In between this process, the gNB sends another information to the core network to allocate more bandwidth for service resumption to provide new UEs. The UPF analysed the information and sends a command to gNB1 to release F-RRH towards the hotspot cell i.e., gNB2. When this F-RRH is placed at the specified place, it starts providing service to the overloaded users of gNB2.

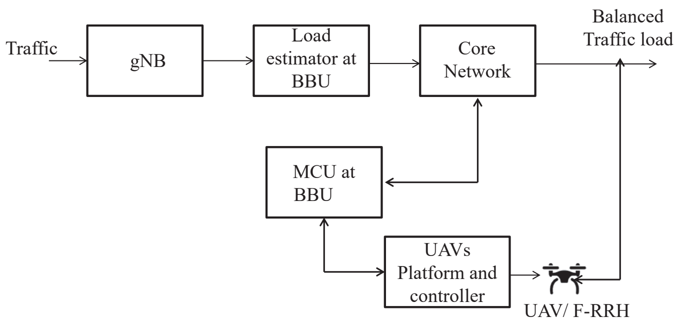

Fig. 3: Block diagram for reactive load balancing approach.

2) Proactive Approach: The proactive approach is a step-wise traffic control process. Here, the MCU continuously monitors the traffic and takes the necessary actions when congestion arises at the gNB. Instead of taking actions at the upper threshold, this technique takes proactive actions and incrementally places F-RRH at the hotspots.

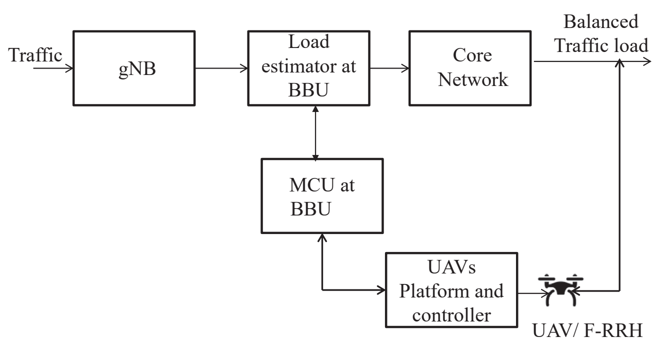

Fig. 4: Block diagram for proactive load balancing approach

Fig. 4 shows a block diagram for the proactive load balancing approach. In this approach, the F-RRH placement decision is taken directly by the MCU without the intervention of the core network. Traffic load at the gNB is continuously monitored by the MCU and regularly updated the information to the core network. If traffic at the gNB is increased, the MCU locally takes the placement decision and instruct the UAV platform to place F-RRHs as required at the hotspot cell. In this proactive approach, regular monitoring and incremental UAV placement help keep traffic below the threshold limit.

Algorithm 1: Traffic-aware Proactive Load Balancing Algorithm (TPLBA)

Input : Number users (*M*), Traffic generated by i^{th} UE (l_i)
Output: Usable capacity at gNB_j ($\bar{C}_j$), and balanced load at gNB

$\bar{C}_j \leftarrow$ Total capacity of gNB_j – Current load at gNB_j
for *each* UE_i **do**
  Locate a gNB say gNB_j nearest to the UE_i, having usable capacity $\bar{C}_j$
  if *($T_{LUj} > l_i$) and ($\bar{C}_j > l_i$)* **then**
    assign UE_i to gNB_j
    Update the usable capacity of gNB_j , $\bar{C}_j \leftarrow \bar{C}_j - l_i$
    Update the current load T_L
  end
  else
    Find a gNB say $gNB_{j'}$ within the same cluster, and having a usable capacity $\bar{C}_j$
    if *($T_{LUj'} > l_i$) and ($\bar{C}_{j'} > l_i$)* **then**
      Assign UE_i to $gNB_{j'}$
    end
    else
      Send a control information to MCU to initiate F-RRH deployments
      Relocate the UE_i to $F{-}RRH_k$
      Continue till gNB_j and $gNB_{j'}$ are balanced
    end
  end
end

IV. Simulations Framework and Performance Analysis

This section describes the framework for performance analysis of different QoS parameters mentioned in Section I. The simulation is carried out in two phases, i.e., simulation setup phase and experimental analysis phase. Details of this are described as below:

A. Simulation Set-Up

To simulate a hotspot cell, we have used the Tu-Vienna LTE simulator [15]. The gNBs are placed at the center of the cell, and UEs are assigned in increasing order and analyze the behaviours of different performance parameters. Table I list the system parameters used for simulation, along with their ranges. For simulation, we have taken 20 MHz LTE-A bandwidth and 100 physical resource blocks (PRB) [16]. Further, we have taken a cluster of three cells, each having a gNB with a coverage area of 0.5 km^2. Further, we have considered a low power access point (AP) as F-RRH with a transmission range of 200 meters with an altitude of 100 meters.

TABLE I: Simulation Parameters

Parameters	Values
Number of cell in the cluster	3
Number of gNB per cell	1
Cellular area of each cell	500 m^2
Number of sector antennas per gNB	3
TX range of gNB	750 m
Bandwidth availability	20 MHz
Number of PRB	100
Maximum no. of PRB per UE	2
Simulation Time	300 seconds
Number of cells in a cluster	3
Number of UEs	10-100
UE placement strategy	Random
Upper threshold load margin	80%
Maximum number of UAV (F-RRH) placed per cell	2
TX range of UAV	250 m
Altitude of UAV	100 m

B. Performance Analysis

To analyze the performance, number of UEs is varied from 10 to 100. Fig. 5 shows a grid of 19 cells, each having a gNB at its center position. We have only considered cell numbers 7, 11, and 12 as a cluster for traffic analysis and assigned the UEs to these selected gNBs. To make a hotspot situation larger number of UEs are initially assigned to gNB 11 and observed the different performance parameters.

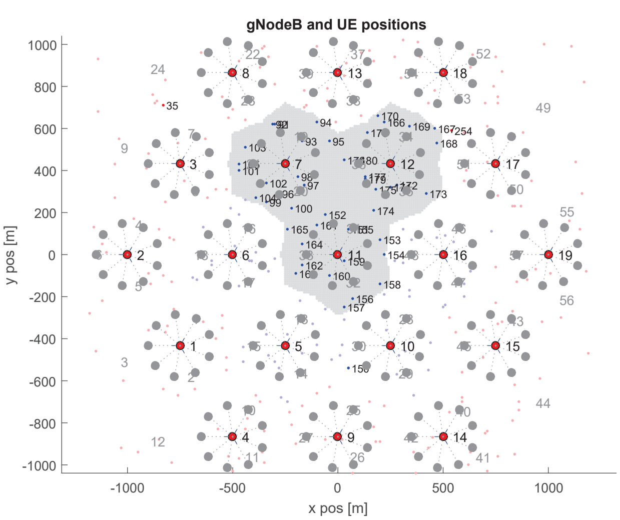

Fig. 5: Initial UEs and gNBs placement within a cellular cluster of three cell.

Fig. 5 shows a grid of 19 cells, each having a gNB at its center position. We have only considered cell numbers 7, 11 and 12 as a cluster for analysis of traffic and assignment of UEs to the selected gNBs. To analyze performance of simulation parameters number of UEs are varied from 10 to 100. To make a hotspot situation, a larger number of UEs are initially assigned to gNB 11, and different performance

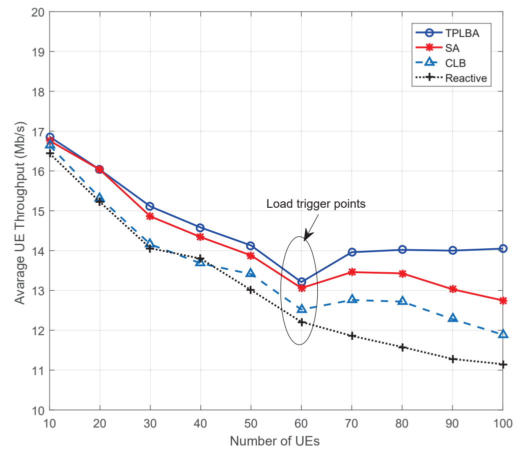

Fig. 6: Average UE throughput *vs* number of UEs.

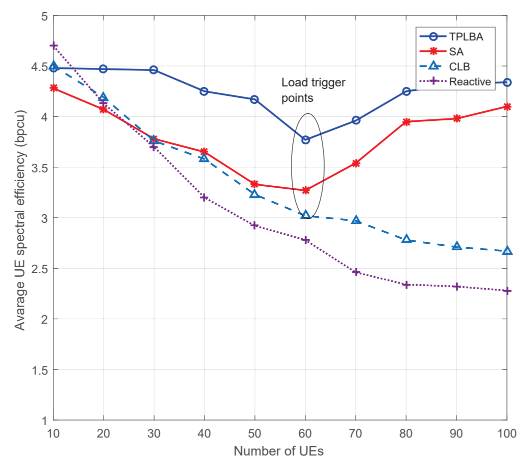

Fig. 7: Average UE spectral efficiency *vs* number of UEs.

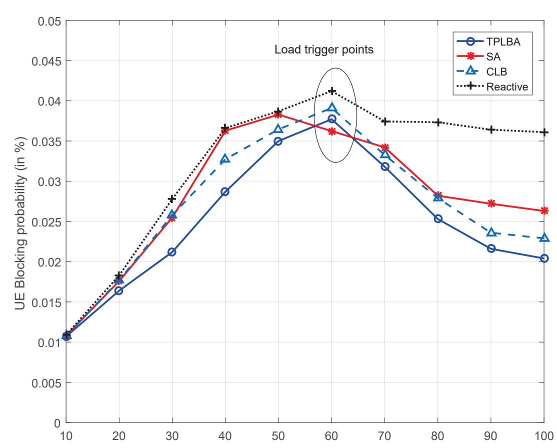

Fig. 8: UE blocking probability *vs* number of UEs.

parameters are observed.To analyze the performance of the proposed TPLBA algorithm, it is compared to other algorithms such as CLB [11], SA [12], and reactive [14] algorithms under the same traffic load and channel bandwidth.

Fig. 6 shows the impact of user arrival on the average UE throughput and depicts that, when the number of users increases, the UE throughput is decreased up to the trigger point. This inverse relationship is observed due to the exploitation of available channel bandwidth. At this trigger point, the MCU start the TPLBA algorithm and calls the required numbers F-RRHs at the cellular hotspot. Placement of this F-RRH at the hotspot enhances the available bandwidth at the corresponding gNB. Moreover, placement of F-RRH at the hotspot cell significantly improves the UE throughput by providing required amount of channel bandwidth to all the attached UEs. The comparison result between different load balancing algorithm shows that the TPLBA and SA algorithm provides more stable throughput as compared to CLB and reactive load balancing algorithm.

The impact of change in number of UEs on UE spectral efficiency is plotted in Fig. 7. It is observed that an increase in the number of UEs decreases the UE spectral efficiency due to increased channel occupancy up to the trigger point. But use of F-RRH at the load trigger point significantly improves the UE spectral efficiency for the TPLBA and SA algorithm. In contrast, the CLB and reactive strategies show a limited improvement in UE spectral efficiency. This distinct behavior of different algorithms occurs due to their load monitoring and control mechanisms. However, due to the limited availability of bandwidth, the UE spectral efficiency is reduced after the trigger point. Whereas, the TPLBA and SA algorithm shows a significant improvement in spectral efficiency after the trigger points.

Fig. 8 shows the relation between UE blocking probability *vs.* the number of UEs. The figure shows an initial increase in BP with the increase in the number of UEs upto the load trigger point. This increase in blocking probability occurs due to increase in bandwidth utilization and network congestion. But the blocking probability decreases due to increased available bandwidth with the use of different load balancing strategies at the trigger points. From the figure, it was also observed that the use of the TPLBA strategy at the trigger point significantly reduces the BP due to an increase in the bandwidth available with the placement of F-RRH at hotspot cell. In the reactive approach, however, the availability of limited bandwidth restricted any improvement in the likelihood of blocking probability.

V. Conclusion

An increase in dependency on the cellular network increases the load on cellular base station. The hotspot situation is found at a gNB, when the traffic load exceeds the threshold capacity. To avoid this hotspot situation, we have proposed the traffic-aware load balancing approach. The proposed load balancing strategy at the gNB is continuously monitored by a MCU unit using the TPLBA algorithm. The UAV controller gets the instruction to place F-RRH when traffic goes beyond a certain threshold limit. Further, the simulation results show that the proposed technique helps enhance QoS at the cellular hotspot by improving the average UE throughput, spectral efficiency, and reducing UE blocking probability.

VI. Acknowledgement

This work was supported by the Science and Engineering Research Board, Department of Science and Technology, Govt. of India (Project no. SRG/2019/000979).

References

[1] M. M. Mowla, I. Ahmad, D. Habibi, and Q. V. Phung, "A green communication model for 5G systems," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 264–280, 2017.

[2] P. K. Sharma and D. I. Kim, "Random 3d mobile uav networks: Mobility modeling and coverage probability," *IEEE Transactions on Wireless Communications*, vol. 18, no. 5, pp. 2527–2538, 2019.

[3] I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, 2014.

[4] S. Matoussi, I. Fajjari, S. Costanzo, N. Aitsaadi, and R. Langar, "5G RAN: Functional split orchestration optimization," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1448–1463, 2020.

[5] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent O-RAN for beyond 5G and 6G wireless networks," *arXiv preprint arXiv:2005.08374*, 2020.

[6] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "Fluidnet: A flexible cloud-based radio access network for small cells," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 915–928, 2016.

[7] Q. Zhang, J. Chen, L. Ji, Z. Feng, Z. Han, and Z. Chen, "Response delay optimization in mobile edge computing enabled UAV swarm," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3280–3295, 2020.

[8] M. Zafer and E. Modiano, "Blocking probability and channel assignment in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 869–879, 2006.

[9] H. Jiang, S. Yi, L. Wu, H. Leung, Y. Wang, X. Zhou, Y. Chen, and L. Yang, "Data-driven cell zooming for large-scale mobile networks," *IEEE Trans. Netw Serv. Manag.*, vol. 15, no. 1, pp. 156–168, 2018.

[10] M. Sheng, C. Yang, Y. Zhang, and J. Li, "Zone-based load balancing in LTE self-optimizing networks: A game-theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2916–2925, 2013.

[11] B. Mahapatra, R. Kumar, A. Turuk, and S. Patra, "CLB: a multilevel co-operative load balancing algorithm for C-RAN architecture," *Digit. Commun. Netw.*, vol. 5, no. 4, pp. 308–316, 2019.

[12] M. Salhani, "Comparison of the proactive and reactive algorithms for load balancing in UDN networks," *J. Commun.*, vol. 14, no. 12, pp. 1119–1126, 2019.

[13] J. Hu, H. Zhang, Y. Liu, X. Li, and H. Ji, "An intelligent UAV deployment scheme for load balance in small cell networks using machine learning," in *2019 IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–6.

[14] M. Salhani and M. Liinaharja, "Load balancing algorithm within the small cells of heterogeneous UDN networks: Mathematical proofs," *J. Commun.*, vol. 13, pp. 627–634, 2018.

[15] N. Mohsen and K. S. Hassan, "C-RAN simulator: A tool for evaluating 5G cloud-based networks system-level performance," in *IEEE 11th Int. Conf. on Wireless Mobile Comput. Netw. Commun. (WiMob)*, 2015, pp. 302–309.

[16] M. Elsaadany, A. Ali, and W. Hamouda, "Cellular LTE-A technologies for the future internet-of-things: Physical layer features and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2544–2572, 2017.