Assessing Power Allocation Efficiency for RAN in Cloud-based Systems for 5G Networks

João Albuquerque, Glauco Gonçalves, Aldebaro Klautau

***Abstract*— Deploying edge and cloud computing architectures in 5G networks offers significant advantages, including reduced latency, lower core network traffic, and distributed processing capabilities. However, relocating latency-sensitive Radio Resource Management (RRM) functions, such as power allocation, to edge or cloud nodes may degrade the channel capacity and performance of User Equipments (UEs). This paper evaluates the impact of allocating a power allocation function at different network levels—Radio Access Network (RAN), Mobile Edge Computing (MEC), and cloud—using the ns-3 simulator with the 5G-Lena module. We implemented a simple, memoryless power allocation algorithm to assess channel capacity variations under varying cloud-distance latency conditions. We analyzed key performance indicators, including the Round-Trip Time (RTT) of control packets and channel capacity over time, to investigate the impact of latency on power allocation efficiency. The results reveal that allocating the power allocation function closer to the RAN achieves superior performance, with mean capacity values and variations meeting 3GPP standards. In contrast, placing the function at the MEC or cloud led to increased latency, insufficient capacity levels for some UEs, and more significant deviations from 3GPP requirements.**

***Keywords*— 5G, Radio Access Network, Radio Resource Management, Cloud, Mobile Edge Computing.**

I. Introduction

The advent of 5G networks has introduced transformative advancements in connectivity, data rates, and modularity. These improvements have enabled diverse and flexible architectures, allowing for dynamic allocation of network functions and optimized resource management within the core network [1]. As a result, 5G technology has become a cornerstone for modern communication systems, supporting a wide range of applications and services. One key shift facilitated by 5G networks is the migration of application allocation from localized servers to distributed cloud environments, including edge computing infrastructures [2]. This transition offers the potential to enhance scalability, flexibility, and resource utilization. Edge cloud systems, in particular, allow for remote management and orchestration of both software and hardware components, paving the way for innovative use cases and performance enhancements.

However, despite this migration's theoretical advantages, several practical challenges arise, particularly regarding latency and response time. Functionalities that demand ultra-low latency, such as Radio Resource Management (RRM) and dynamic power allocation, are highly sensitive to delays introduced by cloud-based processing [3]. Moreover, this paradigm shift is intrinsically tied to the disaggregation of the core network into various Virtualized Network Function (VNF), which can be arbitrarily distributed throughout the network, enabling a virtualized Next Generation Radio Access Network (vNG-RAN) on general-purpose, vendor-neutral hardware with virtualized functions [4]. Despite the potential for arbitrary distribution, the challenge lies in finding common ground regarding the performance of allocating delay-sensitive functions that require low response times and virtualizing these functions. This highlights the need to determine the most efficient approach for distributing them across the network. Evaluating these impacts through emulated or simulated environments is essential to understanding the feasibility and performance trade-offs associated with this paradigm shift.

João Albuquerque, Glauco Gonçalves, and Aldebaro Klautau are with LASSE - Telecommunications, Automation and Electronics Research and Development Center, Federal University of Pará (UFPA), Belém-PA, Brazil. E-mails: joao.barbosa.albuquerque@itec.ufpa.br; {glaucogoncalves, aldebaro}@ufpa.br;

II. Related Works

Previous research has acknowledged the challenges and potential issues of migrating RRM functions to cloud environments [5], [6]. These studies highlight concerns about performance evaluation without considering different possible geographical cloud positions and primarily focus on theoretical models without addressing the practical implications of implementing power allocation functions in real, emulated, or simulated 5G network stacks. Moreover, other works have predominantly explored algorithmic approaches, such as optimization techniques and machine learning models, to improve power allocation efficiency [7]. While these approaches provide valuable insights, they are not tested in comprehensive end-to-end 5G network environments.

Moreover, other recent studies, such as PlaceRAN [4], have delved into optimizing the placement of virtualized Radio Acess Network (vRAN) functions within disaggregated network environments. PlaceRAN presents a comprehensive optimization model that balances computational resource usage and aggregation of radio functions. Evaluating realistic topologies highlights the trade-offs in computational efficiency and virtualization. However, it does not compute performance indicators in an end-to-end simulation scenario, such as capacity for the end users.

This paper seeks to bridge these gaps by simulating the migration of RRM functions, focusing on power allocation for User Equipment (UE), to different geographically positioned clouds. By incorporating a realistic 5G simulated environment using ns-3[1] and the 5G-Lena module[2], it provides a more accurate end-to-end assessment of the feasibility and performance of cloud-based RRM implementations, providing insights into their ability to meet stringent functionality latency requirements while ensuring optimal resource allocation for each UE, computing several performance metrics, such as capacity and Round-Trip Time (RTT) for the end users.

III. System Model and Challenges

5G networks are designed to support diverse communication categories, each with specific service requirements and performance demands. The three primary categories are Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), and Massive Machine-Type Communication (mMTC). eMBB is optimized for high data-rate applications, such as ultra-high-definition video streaming and virtual reality, requiring high throughput, low latency, and seamless mobility. URLLC addresses mission-critical applications, including autonomous vehicles and remote surgery, which demand ultra-low latency, high reliability, and minimal jitter. mMTC targets massive Internet of Things (IoT) deployments, prioritizing scalability, high connection density, and low power consumption, often tolerating higher delays for non-time-sensitive transmissions.

These categories emphasize the need for adaptable and efficient resource allocation mechanisms to meet their varied performance requirements. In particular, eMBB poses a unique challenge due to its demand for high throughput, which is directly influenced by noise at the receiver and the gain of the frequency-selective fading channel.

The maximum achievable throughput (for eMBB and other use cases) is theoretically governed by the Shannon capacity of the Gaussian channel, expressed as:

$$C = BW \cdot \log_2(1 + SNR), \tag{1}$$

where BW is the bandwidth, and SNR is the signal-to-noise ratio in linear scale. The allocated SNR depends heavily on the transmitted power since:

$$SNR = \frac{P_t \cdot |H(w)|^2}{N0 \cdot BW}, \tag{2}$$

where P_t is the transmitted power, $|H(w)|^2$ represents the channel's power, and $N0$ is the noise Power Spectral Density (PSD). According to the 3GPP requirements [8], the minimum value permitted for eMBB is around 50 Mbps in UMa (UMa) scenarios. To achieve this high value of channel capacity, the power of the transmission must be modified according to the noise and the channel gain, i.e., to the SNR's current state. As previewed in the requirements, to achieve the 50 Mbps, the minimum desired SNR^* for each UE is according to (1):

$$SNR^* = 2^{\frac{C}{BW}} - 1. \tag{3}$$

Since the SNR is measured at the receiver and the transmitter is responsible for power allocation, obtaining real-time channel state information poses a challenge. To address this, UE periodically reports a Channel Quality Indicator (CQI) to the Radio Access Network (RAN). The CQI can be mapped to an SNR value based on predefined lookup tables [9]. This mapping allows the RAN to adjust transmission power to meet throughput requirements dynamically.

Although this process appears straightforward, several challenges arise due to the latency introduced by cloud-based RRM processing, which can degrade the accuracy of SNR estimations. These challenges are particularly significant given the rapid variations in channel conditions, determined by the temporal stability of a wireless channel, characterized by its coherence time, T_{C}. Coherence time represents the interval during which the wireless channel remains with no relative variance. Specifically, coherence time is defined as the duration within which the channel's time correlation exceeds 0.5. It can be expressed mathematically using the following equation [10]:

$$T_{\mathrm{C}} = \frac{9c}{16\pi v f}, \tag{4}$$

where c is the speed of the radio wave, f is the transmission frequency, and v denotes the relative velocity between the receiver and transmitter. A high relative velocity v results in faster channel variations and a shorter coherence time, whereas stationary scenarios yield longer coherence times due to minimal relative motion.

Short coherence times or slow updates of channel state information can lead to delayed power adjustments, resource underutilization, and potential capacity degradation. Moreover, migrating RRM functions to remote cloud servers or edge clouds further exacerbates latency issues, hindering real-time decision-making. This delay is particularly critical for eMBB applications, which demand highly accurate and responsive resource allocation to maintain performance and reliability.

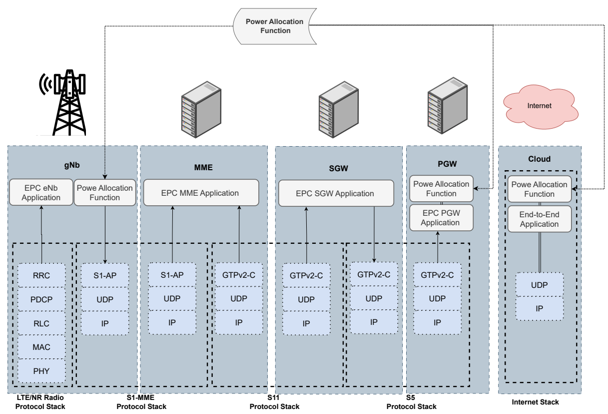

Fig. 1. Power allocation function placements on top of the interfaces and protocols stack.

In our context, the power allocation function is implemented as an algorithm within an application. This function is designed to ensure that the processing delay incurred in generating the appropriate power allocation remains fixed at one millisecond, regardless of the location of the power allocation function. The placement of the power allocation function follows the Long-Term Evolution (LTE) 4G Non-

[1]https://www.nsnam.org/
[2]https://5g-lena.cttc.es/

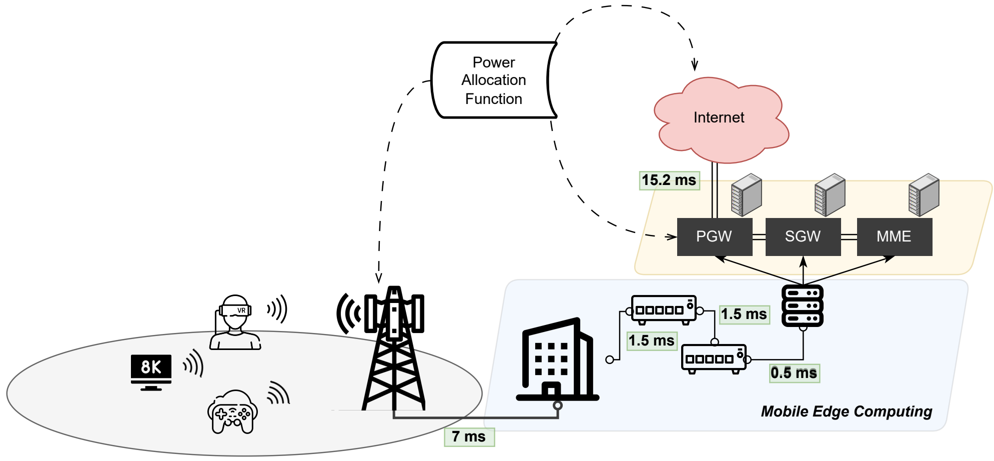

Fig. 2. All entities used in the simulated 5G environment.

Standalone (NSA) architecture, also employed in the 5G-Lena module [11].

According to this setup, the power allocation function is integrated into the application layer, as shown in Fig. 1, without altering the existing protocol stack layers, which include Radio Resource Control (RRC), Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC), and Physical (PHY) layers (the LTE/New Radio (NR) stack). All nodes' S1-AP interface, GTPv2, Internet Protocol (IP), and transport layer stacks remain unchanged. The only modification is the addition of the new power allocation application to the application layer in the Packet Data Network Gateway (PGW), Next Generation Node B (gNB), and the Internet node, alongside the existing Evolved Packet Core (EPC) application in the PGW and the end-to-end application in the Internet node. Consequently, all interfaces, including S1-MME, S11, and S5, are unaffected. In the following sections, we demonstrate the power allocation strategy and the simulated environment used to deploy the simulation. We also analyze the impact of latency on power allocation efficiency in cloud-based environments.

IV. Algorithm and Power Allocation Strategy

As discussed in Section III, dynamic power allocation enhances the feasibility of eMBB by providing better stability in maintaining channel capacity and, consequently, the throughput for each UE. An algorithm was developed to manage power allocation in a scenario with a single RAN. This algorithm evaluates the current channel state, mapped using the CQI report values received from each UE. However, as noted earlier, the reported CQI values may not significantly reflect the most up-to-date channel conditions when latency increases between the RAN and the node executing the functionality.

To address this issue, the proposed algorithm operates memoryless, considering only the current channel state without accounting for its historical variations. Algorithm 1 illustrates this approach. The algorithm assumes a maximum allocated transmission power of 40 dBm per UE. Using Eq. (3) and a bandwidth of 100 MHz per UE, the target SNR^* in dB scale is calculated as -3.83.

Given the known SNR^*, the newly allocated transmission power for a UE is computed based on the difference between the desired and current SNR values. According to Eq. (2), the updated transmission power is expressed as:

$$P_t = P_t + (SNR^* - SNR), \tag{5}$$

on a logarithmic scale.

Because fading channel gain can significantly affect the SNR, the allocated power may sometimes exceed the allowable transmission limit. To prevent this, the algorithm enforces an upper bound on transmission power, ensuring compliance with realistic power constraints. Similarly, a lower bound is implemented to guarantee that the allocated power does not fall below acceptable levels. This dual-bound approach maintains the allocated power within operational limits, enabling practical deployment in real-life transmission scenarios.

Algorithm 1 Power Allocation Algorithm

Require: $maxTxPower = 40dBm$ ▷ Maximum transmission power
Require: $minTxPower = 0dBm$ ▷ Minimum transmission power
Require: $SNR^* = -3,83dB$ ▷ Desired SNR
  if $SNR \neq SNR^*$ **then**
    $UeTxPower \leftarrow currentTxPowerx$
     $UeTxPower = UeTxPower + (SNR^* - SNR)$
    if $UeTxPower > maxTxPower$ **then** ▷ Upper bound
     return $maxTxPower$
    else if $UeTxPower < minTxPower$ **then** ▷ Lower bound
     return $minTxPower$
    end if
  end if
     return $UeTxPower$

V. Simulated Environment

To evaluate the challenges of power allocation functionality, we developed a 5G simulation scenario as illustrated in Fig. 2.

Nine UEs under the same eMBB category were deployed in this environment using Time-Division Duplexing (TDD). Communication occurs in a Single-Stream Multiple-Inputs and Multiple-Outputs (MIMO) configuration, where only one UE interacts with the RAN at a given time. The core network employs an NSA architecture, comprising a Mobility Management Entity (MME), a PGW, and a Serving Gateway (SGW). These core functions were migrated to a Mobile Edge Computing (MEC) infrastructure. The connection between gNB and the core network traverses two switches, each introducing delays as specified in Fig. 2. These delay values were derived from measurements based on [12].

The PGW connects to an external Internet node that generates UE traffic and will also work as the cloud. The traffic source utilizes User Datagram Protocol (UDP) with fixed packet sizes, transmitting one packet per 100 milliseconds. CQI reports are generated during the uplink transmission phase as part of the power adaptation process. These reports provide feedback to the RAN about the quality of the radio channel. Each UE periodically sends CQI feedback, subsequently mapped to an SNR value.

The power allocation process can occur at different nodes, including the RAN, MEC, or the Internet (cloud). The gNB receives CQI reports from all UEs, and maps them in SNR values, then sends a packet to the next hop if power allocation is not performed locally at the RAN. The server computes power allocation and returns the results if configured for MEC processing. Otherwise, the request is forwarded to the Internet node for processing. Once power allocation is determined, the updated configuration flows back through the network to the gNB, as shown in Fig. 3.

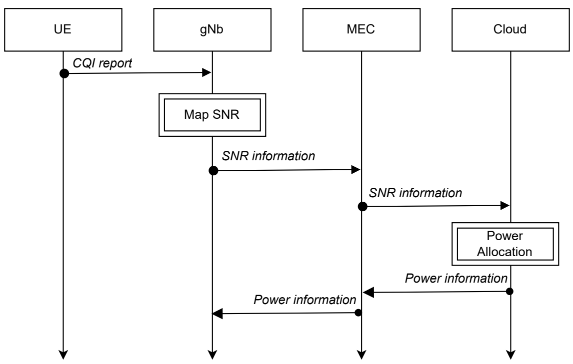

Fig. 3. The flow diagram of the packets' information through the network in the case of the power allocation function in the cloud.

VI. RESULTS

To evaluate the proposed scenario, we implemented the network setup depicted in Fig. 2 using ns-3, leveraging the 5G-Lena module to deploy 5G features in three distinct simulations, each differing in the geographical placement of the power allocation function. The simulation environment, as discussed previously, follows a drop-based approach, where UEs positions were randomly assigned but remained consistent across all simulations. Notably, UEs 3, 4, and 5 were positioned closer to the gNB, while the remaining UEs were placed at greater distances to prevent significantly higher SNR values. To ensure symmetry, these UEs were equidistant from the gNB. All UEs were associated with the same application, as described in Section V, without any prioritization.

At the start of the simulation, the RAN was initialized with the same transmission power for all UEs, maintaining uniform conditions for evaluating power allocation performance. All results were measured over a 50-second simulation duration, focusing on their temporal variability. The fading characteristics and channel conditions—whether Line-of-Sight (LoS) or Non-Line-of-Sight (NLoS)—were updated according to Eq. 4.

TABLE I
RTT PERFORMANCE METRICS FOR CLOUD AND MEC NODES

	RTT (ms)		
Node	Minimum	Maximum	Average
MEC	18.2	32.3	25.48
Cloud	46.4	59.7	52.59

Table I presents the RTT measurements for power control packets in two deployment scenarios: cloud and MEC. The RTT was computed as the time difference between a control packet's transmission from the RAN and its return. In both cases, RTT was influenced by each network topology switch link latency along the path, while in the cloud scenario, it was further affected by internet link latency. The RTT for the RAN-only scenario was not computed, as the link latency between the UE and gNB was negligible.

In the cloud-based deployment, the evaluated RTT had an average value of 52.59 milliseconds, fluctuating between 46.4 and 59.7 milliseconds. In contrast, since MEC deployment is not subject to internet link latency, it exhibited a significantly lower RTT. The average RTT in MEC-based scenario was 25.48 milliseconds, ranging from 18.2 to 32.3 milliseconds. In both cases—MEC and cloud—the observed RTT variations were primarily influenced by buffer queuing and processing delays within the power allocation function.

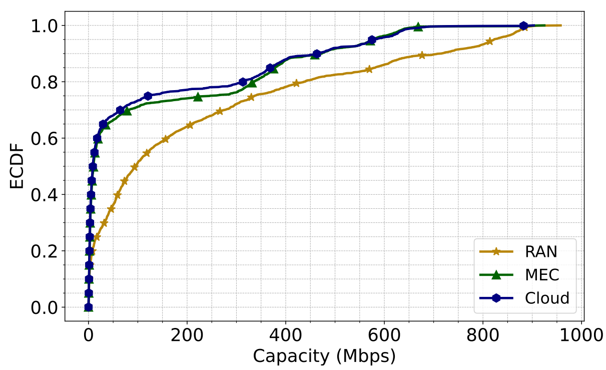

Fig. 4. The ECDF considering all measured UEs channel capacity values during the simulation.

The RTT measured values will significantly impact the channel capacity measurements. The channel capacity exhibits frequent variations due to the immediate application of power adjustments whenever a new CQI report is received. This rapid adaptation attempts to mitigate low SNR values, preserving channel capacity in scenarios with sudden degradation. As depicted in Fig. 4, considering all UEs in this measurement, the power allocation function, when allocated closer to the receivers, maintains a high percentage of channel capacity values that fit the 3GPP standards. I.e., when in RAN, 65% of the evaluated samples are in the requested range, as discussed in Section III. However, due to high latency in the transport network, the channel capacity drops and leads to almost 70% of samples being assessed out of the requisition in cloud and MEC scenarios, where, when placed in MEC, the overall ECDF result barely improves.

When assessing the channel capacity per UE, as shown in Fig. 5, some devices exhibit more consistent and higher capacity values. This is primarily due to higher channel gain and reduced fading variation, often resulting from shorter RAN-to-UE distances, which increase the likelihood of a LoS condition. Devices 3, 4, and 5 consistently maintain optimal capacity levels across all three placements, with average values exceeding 50 Mbps throughout the simulation. In the MEC scenario, UEs 1 and 8 achieved average capacity values aligned with the desired requirements, though they recorded lower minimum capacities even in the most favorable RAN configuration.

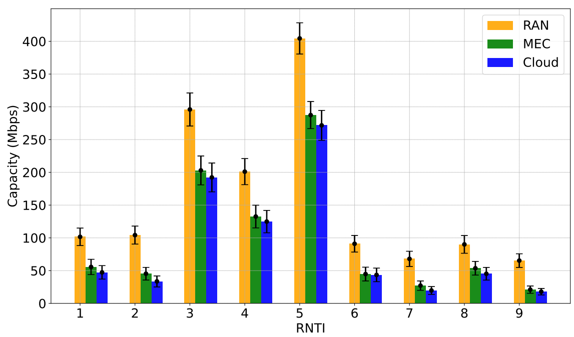

Fig. 5. The average channel capacity with confidence interval for UEs with power allocation for each possible placement.

By contrast, the remaining devices were unable to sustain acceptable average capacities in scenarios where the power allocation was not centered on the RAN, achieving notable capacity only at peak moments during the simulation. Specifically, UEs 1, 2, 6, 7, 8, and 9—all located farther from gNB-1—consistently demonstrated lower capacity values.

VII. Conclusion

Our study highlights the critical importance of appropriately placing latency-sensitive application components within the network hierarchy. By evaluating the deployment of a memoryless power allocation function at the RAN, MEC, and cloud levels, we demonstrated that increased distance from the terminal devices—and the resulting latency—can significantly degrade the channel capacity experienced by UEs. Through detailed simulation using ns-3 and the 5G-Lena module, we measured the RTT, tracked channel capacity variation over time, and provided a comprehensive statistical analysis for each UE.

The results indicate that while edge computing brings several advantages, not all functions are suitable for offloading beyond the RAN. When the power allocation function was executed at the MEC, some UEs failed to achieve capacity levels compliant with 3GPP standards. This issue was further exacerbated at the cloud level, where both average capacity and stability deteriorated significantly. These findings emphasize that deploying latency-sensitive RRM functions at higher network layers—especially without memory or prediction mechanisms—can undermine system performance. Therefore, such functions are more effectively hosted at the RAN, where lower latency allows for more responsive and reliable operation, preserving the quality of experience for end users.

Acknowledgements

This work was partially financed by the Innovation Center, Ericsson Telecomunicações S.A., Brazil; Brasil 6G project (RNP/MCTI grant 01245.010604/2020-14); OpenRAN Brazil - Phase 2 project (MCTI grant Nº A01245.014203/2021-14); Universal (CNPq grant 405111/2021-5); Project Smart 5G Core And MUltiRAn Integration (SAMURAI) (MCTIC/CGI.br/FAPESP under Grant 2020/05127-2).

References

[1] J. P. Esper, A. S. Abdallah, S. Clayman, W. Moreira, A. Oliveira, S. L. Correa, and K. V. Cardoso, "eXP-RAN—an emulator for gaining experience with radio access networks, edge computing, and slicing," *IEEE Access*, vol. 8, pp. 152975–152989, 2020.

[2] P. Y. Zhou, S. Fu, B. Finley, X. Li, S. Tarkoma, J. Kangasharju, M. Ammar, and P. Hui, "5G MEC computation handoff for mobile augmented reality," in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, pp. 129–136, IEEE, 2024.

[3] T. Akhtar, C. Tselios, and I. Politis, "Radio resource management: approaches and implementations from 4G to 5G and beyond," *Wireless Networks*, vol. 27, pp. 693–734, 2021.

[4] F. Zanferrari Morais, G. Matheus de Almeida, L. Pinto, K. Vieira Cardoso, L. M. Contreras, R. da Rosa Righi, and C. Bonato Both, "PlaceRAN: Optimal placement of virtualized network functions in the next-generation radio access networks," *arXiv e-prints*, pp. arXiv–2102, 2021.

[5] N. Kumar and A. Ahmad, "Quality of service-aware adaptive radio resource management based on deep federated Q-learning for multi-access edge computing in beyond 5G cloud-radio access network," *Transactions on Emerging Telecommunications Technologies*, vol. 34, no. 6, p. e4762, 2023.

[6] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: Conventional and new approaches," *Sensors*, vol. 20, no. 9, p. 2708, 2020.

[7] Z. Chen and Q. Liang, "Power allocation in 5G wireless communication," *Ieee Access*, vol. 7, pp. 60785–60792, 2019.

[8] T. GPP, "Service requirements for next generation new services and markets," *Technical Specification (TS)-38.900, Version*, vol. 16, no. 0, 2018.

[9] J. Fan, Q. Yin, G. Y. Li, B. Peng, and X. Zhu, "Mcs selection for throughput improvement in downlink lte systems," in *2011 Proceedings of 20th international conference on computer communications and networks (ICCCN)*, pp. 1–5, IEEE, 2011.

[10] T. S. Rappaport, *Wireless communications: principles and practice*. Cambridge University Press, 2024.

[11] G. Piro, N. Baldo, and M. Miozzo, "An LTE module for the ns-3 network simulator.," in *SimuTools*, pp. 415–422, 2011.

[12] B. Charyyev, E. Arslan, and M. H. Gunes, "Latency comparison of cloud datacenters and edge servers," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.