ENHANCED APPROACH TO LATENCY MITIGATION IN 4G LTE MOBILE COMMUNICATION NETWORKS

R. O. Okeke
Electrical/Electronic Engineering Department,
University of Port Harcourt
Nigeria
remigius.okeke@uniport.edu.ng

I. B. Asianuba
Electrical/Electronic Engineering Department,
University of Port Harcourt
Nigeria
ifeoma.asianuba@uniport.edu.ng

ABSTRACT

4G networks currently provide the highest broadband internet connectivity in areas where 5G is not yet available. Studies on latency in 4G networks exist, but they often rely on analytical estimations without considering the factors that affect packet latency. Measurements of one-way packet latency in LTE networks are primarily conducted using ping programming, assuming symmetric latency in uplink and downlink directions, which may not reflect real-life scenarios accurately. In contrast to lab environments with low reported latency, a comprehensive study on the factors impacting packet latency in realistic 4G (LTE) networks is lacking. This paper aims to understand how packet latency is affected by various factors in a realistic environment and propose strategies for improving latency mitigation in 4G LTE networks. The paper further focuses on designing and simulating a latency mitigation technique for 4G networks. The interest is to maintain optimal Quality of Service (QoS) in the network, given the high traffic demand. Traffic scenarios were emulated by integrating five mobile users with varying demands and 20 wired node users in the radio network. Through simulations using Riverbed Modeler®, it was observed that without mitigation, latency increased steadily over time, becoming unmanageable. However, after implementing a differentiated scheduling algorithm, latency significantly decreased and remained constant, indicating success in mitigating latency in the 4G network.

Keywords: Latency mitigation, 4G LTE, QoS, Riverbed Modeler®.

INTRODUCTION

The 4G LTE (Long Term Evolution) network is the fourth generation in mobile telecommunication systems. It was developed to meet the increasing demands of mobile device users, AI-based gadgets, and applications with zero tolerance for lag, such as online gaming, VoIP, and video streaming. Over time, the network has seen improvements to address the growing bandwidth demand (Egena Onu, 2017).

The original 4G LTE build had data rates of 50Mbps and 100Mbps for uplink and downlink, respectively, which were insufficient for the rising bandwidth requirements. To tackle this issue, the LTE-Advanced network was introduced, offering data rates ranging from 100Mbps to 1Gbps (Egena Onu, 2017). The benefits of the 4G LTE network include improved data rates at the cell edge, enhanced spectral efficiency, scalable bandwidth, compatibility with earlier releases and different systems, reduced delays in connection establishment and transmission latency, seamless mobility between different radio-access technologies, and reasonable power consumption for mobile terminals (Christine, 2013; Isaam et al., 2009; Adebayo, 2014).

However, reducing network latency in wireless communication is a crucial requirement, especially for real-time applications such as remote surgery. Latency, also known as lag, can be defined as the time delay experienced in a communication network while transmitting data from the source to the destination. There is a common misconception that network speed is primarily determined by the network's bandwidth. While a higher bandwidth allows for greater traffic capacity, enabling the transmission of a large volume of data, it is important to note that if significant delays occur between data requests and acknowledgements within the network, the bandwidth becomes less effective in ensuring optimal performance. To efficiently improve latency in future LTE deployments, a comprehensive understanding of latency in current mobile networks is necessary. This involves analyzing the various network components that contribute to overall latency, investigating the reasons for latency persistence, and examining the effects of latency on critical network parameters such as network load, distance to mobile terminals and base stations, and packet rate. Efforts to reduce latency in LTE networks can be approached through informed analysis. Leveraging edge computing and differentiated scheduling techniques can be effective in achieving latency reduction.

LITERATURE REVIEW

Wylie-Green and Svensson (2010) conducted a study on latency in a 3GPP LTE system using the ping method. Their research aimed to evaluate LTE radio performance in various user scenarios. The experiments covered fixed and mobile throughput, multiple concurrent users, handover performance, and latency tests. They used a Category 2 LTE User Equipment (UE) with maximum speeds of 50 Mbps downlink and 25 Mbps uplink. The tests were carried out in a multi-site field test area, with Cluster A (2 GHz band) and Cluster B (Lower 700 MHz band) as depicted in Figure 1.

Figure 1: Multi-site test area of Wylie-Green, & Svensson, (2010)

Blajic, Noguli, and Družijani (2007) adopted an analytical approach to study latency in an LTE network. Their research primarily aimed to measure the current levels of latency and overall Quality of Service (QoS) in the network. The team focused on achieving specific performance goals such as improved peak data rates, reduced latency, increased spectrum efficiency, cost and complexity reduction. In their paper, they presented key strategies employed to mitigate latency in the advanced LTE framework, both in the control plane and client plane. The mitigation scope of their approach is depicted in Figure 2.

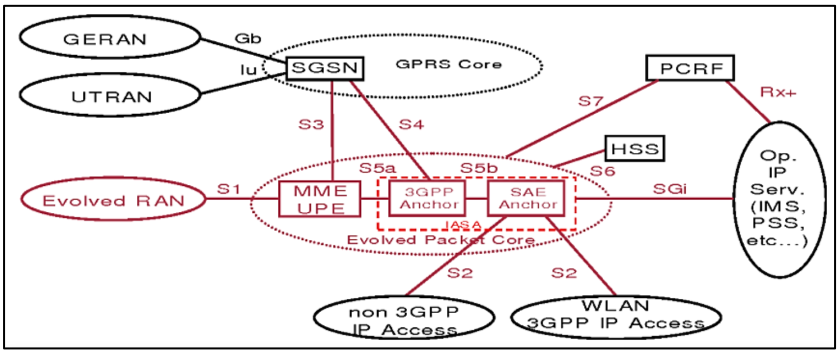

Figure 2: Logical high level architecture for the System Architecture Evolution (SAE) (Logical high level architecture for the System Architecture Evolution (SAE))

Arlos and Feidler (2010) conducted a study to evaluate the one-way delay in the downlink of an LTE network for packets with different sizes. The objective was to investigate how the packet size affects communication efficiency, specifically focusing on the one-way delay in the downlink of third-generation mobile networks. Since one-way delays play a crucial role in user perception, especially in interactive and real-time applications like VoIP, their research aimed to address this important aspect. However, challenges were encountered in accurately measuring the One-Way Delays due to the sensitivity of these measurements to clock synchronization. Consequently, their study focused primarily on the downlink, which carries the majority of traffic to the user and significantly impacts the quality of all IP-based services. The experimental setup utilized is depicted in Figure 3.

Xu and Fischione (2012) focused on studying one-way delay (OWD) in LTE networks, specifically for smart grid components. It was found that the existing LTE uplink scheduler did not consistently meet latency requirements, limiting component accommodation. To address this, they proposed a new scheduler based on LTE medium access control, designed for the smart grid. The approach considered both smart grid components and common user equipment, utilizing a mathematical linear optimization problem. They developed an algorithm to solve this problem, aiming to improve latency satisfaction. Also, Laner et al. (2012) conducted an analysis of latency in LTE and HSPA networks, employing a hybrid model combining active probing and white-box testing. The setup provided insights into latency characteristics of network components in both the radio access and core networks.

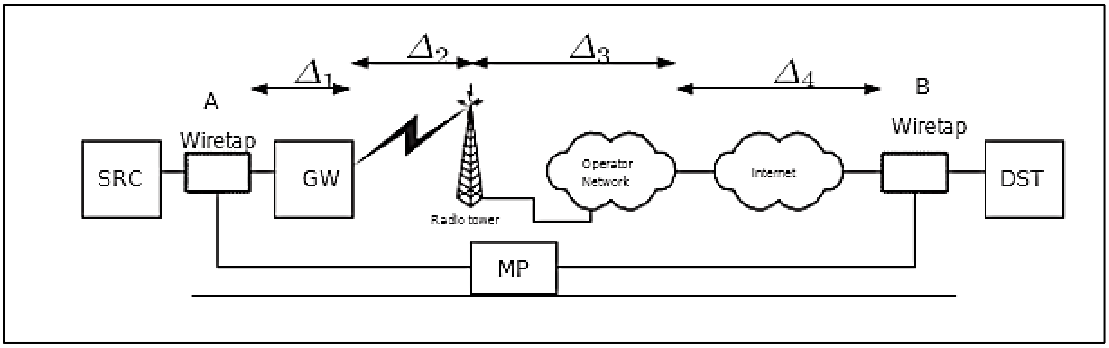

Figure 3: Experimental Setup of Arlos & Feidler (2010)

ISSN 2056-5860

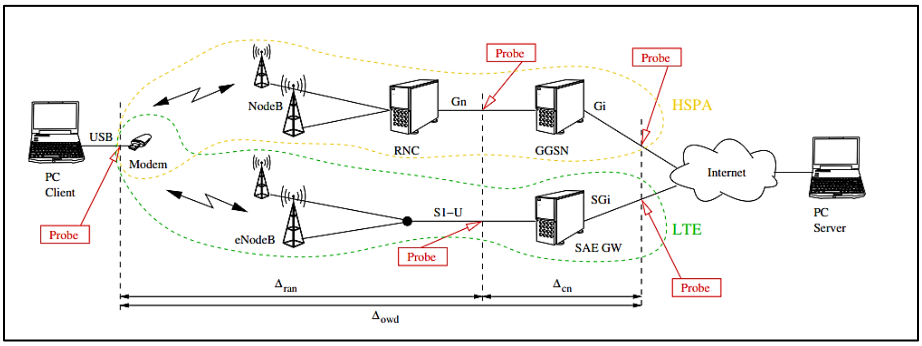

Figure 4: Experimental Setup of Laner et al (2012)

Blajic, Noguli, & Družijani (2007) addressed the issue of control plane and user plane latency; however, their emphasis was primarily on the control plane, and they relied on arbitrary parameters and estimations rather than actual latency measurements. Their approach to user plane latency only considered processing delays, which is not a comprehensive approach, leading to the assumption of symmetrical latency between uplink (UL) and downlink (DL) transmissions, which is not valid in real-time scenarios. Consequently, the latency values obtained from their experiment only provided lower limits based on different sub-layers of the LTE network.

Wylie-Green & Svensson (2010) and Xu & Fischione (2012) both focused on latency mitigation techniques based on round-trip time (RTT) latency. However, Wylie-Green & Svensson (2010) limited their investigation to two packet sizes, while Xu & Fischione (2012) conducted experiments with a wider range of packet sizes. Although Xu & Fischione (2012) had more extensive data, both teams still faced the limitation of predominantly focusing on the control plane, without achieving individual latency measurements for UL and DL signaling.
This limitation was addressed by Arlos & Feidler (2010) and Laner et al. (2012). Arlos & Feidler (2010) focused solely on DL, while Laner et al. (2012) investigated both DL and UL transmissions. They also considered packet sizing for both one-way delay (OWD) and RTT scenarios, resulting in more comprehensive results compared to previous attempts.

Evidently, gaps in knowledge still exist, and this paper aims to further enhance the work of Laner et al. (2012). by not only addressing latency in OWD and RTT scenarios and packet data sizing but also incorporating factors such as packet rate, network load, and user position. Both the Radio and Core networks will be considered, with the goal of achieving maximum latency reduction within the scope of this report.

METHODOLOGY

Design Overview

This section focuses on the development of an improved approach for packet latency mitigation in wireless networks. The objective is to utilize differentiated scheduling and edge computing techniques to prioritize users and reduce latency. The study aims to address packet latency, which impacts network speed and quality of service, an important factors for wireless network subscribers. Various latency scenarios, including packet size, packet rate, network load, and user position, are examined to provide comprehensive results. A realistic network packet latency measurement setup is crucial to ensure accurate evaluations, utilizing tools and settings that closely resemble real-world conditions.

Figure 5 shows a basic mobile network without the internal intricacies and additional complications that it entails. The term "User Equipment" (UE) is used to denote mobile devices.

It might be a cell phone, a laptop with a mobile broadband adapter, or any other wireless communication device that end-users utilize. The Evolved Packet System's (EPS) access network, known as the "Radio Network", is made up of UMTS-based telecommunication base stations (eNodeBs or eNBs) and Radio Network Controllers (RNC). The EPC, also known as the Core Network (CN), is in charge of the UE and the general control of the network. The Application Server (AS) is the component in the mobile network that receives the data packets delivered by UEs in the Up-Link (UL) direction, and also serves as a data packet generator, sending data packets to UEs on the Down-Link (UL) direction.

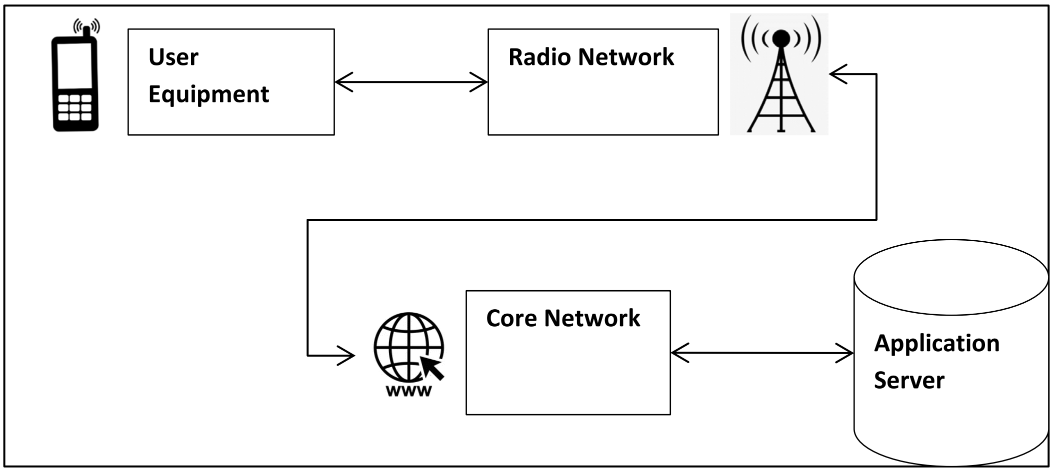

Figure 5: Basic Mobile Network

The network in this paper will rely on the real-time data obtained from Ashish Kurian's (2018) research, which involved conducting an experiment to measure and mitigate latency using a carefully designed setup. Kurian had direct access to network components and a deep understanding of the network architecture and system parameters such as available bandwidth. The measurement setup consist of the radio network, the core network, and user traffic characteristics, which are crucial elements to be extensively modeled in this research.

The User Equipment

LTE mobile devices, like smartphones, are extensively used in mobile networks. They possess the ability to send and receive data packets, enabling downlink acceptance and uplink transmission. These devices also leverage radio access network resources for data communication, facilitating network service access and traffic generation. To simulate a mobile device effectively, the simulation tool should support packet generation and reception, as well as traffic generation. In Ashish's (2018) experiment, a software-based packet generator and traffic generator were utilized. These tools operated on a Linux-based machine connected to a mobile device via USB. The mobile device received the generated packets, transmitted them through the radio access network using a SIM card, and their latency was measured through time-stamping with Wireshark. User Datagram Protocol (UDP) was chosen over Transmission Control Protocol (TCP) to minimize transmission time. The simulation tool must faithfully replicate the mobile device setup illustrated in figure 6.

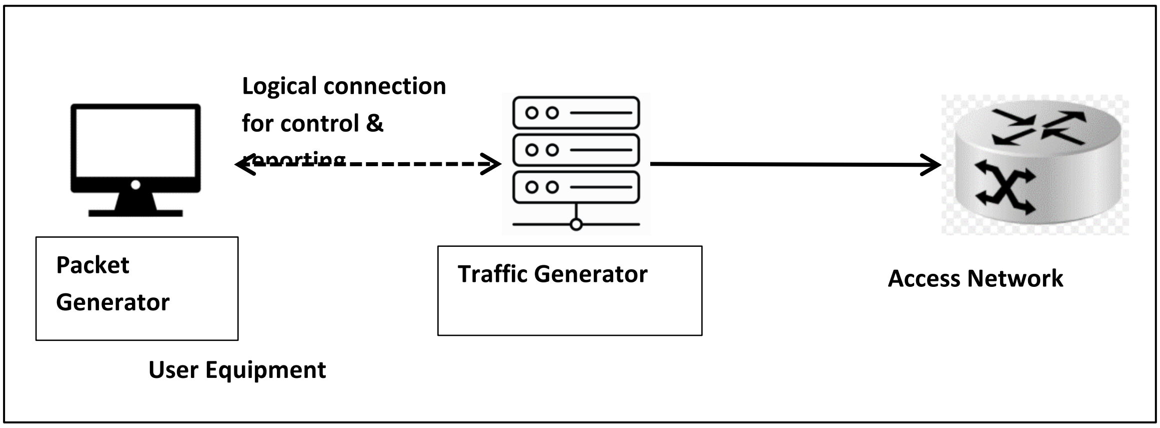

Figure 6: User Equipment Configuration

The Radio Network

To create a realistic experimental radio network, accessing existing eNodeBs is not feasible due to proprietary restrictions. Instead, a small-scale base station like the Ericsson Base Station Rrul 81 B38 can be used. By connecting this base station to network simulators and degraders, the required packet transmission patterns can be generated. Realistic radio channel conditions and diverse transmission patterns are crucial for simulating the network scenarios discussed, particularly in the radio network component. These scenarios are divided into Data-Driven conditions, focusing on data errors and signal quality, and Time-Driven conditions, concerning transmission delays. One important Data-Driven condition is Fading Channel, which includes Multipath Fading, path loss-based fading, and interference-based fading. Multipath fading occurs when the transmitted signal reflects off various obstacles along the signal path. The relationship between interference (I) and noise level (N) is described in equation 1.

$$Noise\ rise = {}^{(I+N)}/_{N} \qquad (1)$$

The quality of the received signal is determined by the signal to interference plus noise ratio (SINR) metric. A higher SINR indicates a better signal quality and reduces the likelihood of transmission errors. On the other hand, pathloss represents the fading of the signal over distance and is considered a large-scale characteristic of the channel. It takes into account the impact of signal fading as the distance between the transmitter and receiver increases. In this paper, two conditions are met to better account for the pathloss in the radio network:

1. The Hexagonal site layout is utilized for 12 sites, and a wraparound feature is applied during simulation to put the boundary effect of signal interference to minimal levels. The spacing configuration is shown in the Figure 7.
2. The radio environment in consideration is set to be a suburban setting with an average inter-site distance of 1.5 kilometers.

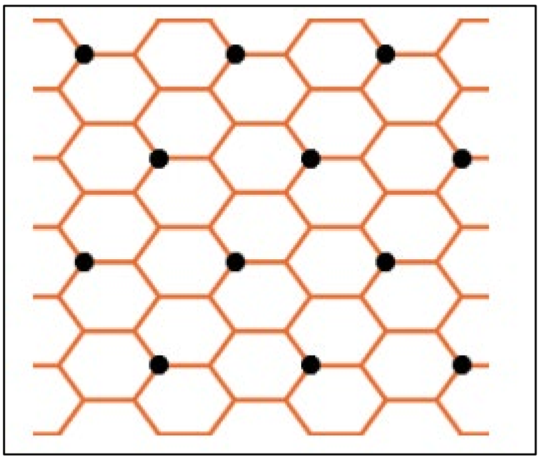

Figure 7: Network Site Layout Pattern

Having put these conditions in place, the pathloss is designed in the simulation tool to function based on the Okumura-Hata pathloss model. The model is suited for both point-to-point and broadcast communications, and covers mobile station antenna heights of 1–10 m, base station antenna heights of 30–200 m, and link distances from 1–10 km.

For Urban Environments:

$$L_U = 69.55 + 26.16\log_{10} f - 13.82\log_{10} h_B - C_H + [44.9 - 6.55\log_{10} h_B]\log_{10} d \quad (2)$$

For small or medium cities:

$$C_H = 0.8 + (1.11\log_{10} f - 0.7)h_M - 1.56\log_{10} f$$

For large cities:

$$C_H = \begin{cases} 8.29\,(\log_{10}(1.54h_M))^2 - 1.1\,, & if\ 150 \le f \le 200 \\ 3.2\,(\log_{10}(11.75h_M))^2 - 4.97\,, & if\ 200 < f \le 1500 \end{cases}$$

Where:

L_U = Path loss in urban areas. Unit: decibel (dB)
h_B = Height of base station antenna (m)
h_M = Height of mobile station antenna (m)
f = Frequency of transmission (MHz)
C_H = Antenna height correction factor
d = Distance between the base and mobile stations (km).

For Sub-Urban Environments: (for areas where the buildings are not as high or as densely packed as urban residences)

$$L_{SU} = L_U - 2\left(\log_{10}\frac{f}{28}\right)^2 - 5.4 \quad (3)$$

Where;

L_{SU} = Path loss in suburban areas (dB)
L_U = Path loss from the small city version of the Urban Environment model (dB)
f = Frequency of transmission (MHz)

The IMT Advanced multipath model for macro-cells in a suburban setting is studied with an estimated UE speed of 0.8 m/s to represent the impacts of signal reflections or obstructions that might occur in a suburban environment owing to e.g. moving cars and buildings. The Parameters are shown in Table 1 below.

Table 1: IMT-Advanced Macro-Cell Parameters

Parameters	Value
Base station transmitted power (dBm)	43
Minimum Coupling Loss (dB)	30
Base station antenna gain (dBi)	18
Base station antenna height (m)	30
Interference Limit Power (dBm)	-109

In the DL, inter-cell interference is simulated by assuming that all surrounding cells of the studied cell transmit at a prior fixed power, represented as a percentage of their maximum transmitting power, to account for it in the considered cell (cell that serves the UE under test). Different levels of noise gain are simulated in the UL.

Core Network & Application Server

This simulation design utilizes software functions within virtual machines (VMs) to implement various tasks of the Core Network. Virtualization involves running an operating system or software on a physical computer that previously had its own operating system and applications. Multiple VMs, referred to as guests, can operate concurrently on a single physical host, eliminating the need for separate hardware for each VM. In this setup, the virtual machine named "ctrl" performs Mobility Management entity (MME), Serving Gateway (SGW), and Packet Network Data Gateway (PGW) functionalities. The control VM connects to the Home Subscriber Server (HSS) VM, performing HSS functions for identity verification. The core network simulation includes a "backhaul delay emulation" VM that adds deterministic transport delay to both uplink and downlink packets. Refer to Figure 8 for a description of the core network simulation operation.

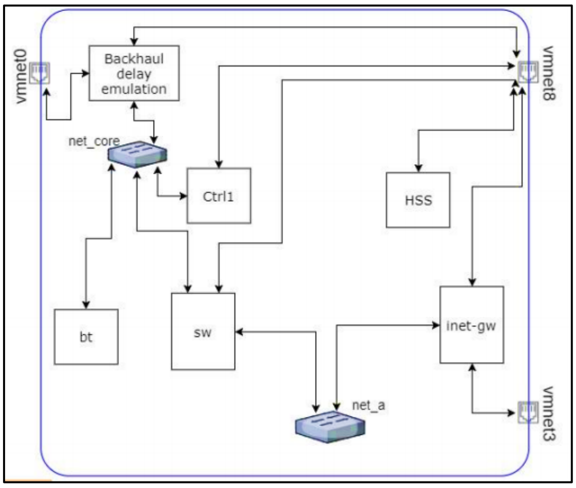

Figure 8: VMware Core Network simulation Block diagram

Network Calibration

In order to analyze the impact of six chosen aspects on packet latency, measurements were conducted by unilaterally varying the values of each aspect while keeping the rest unchanged. Separate measurement experiments were performed for each aspect to evaluate its impact on packet latency. To ensure statistical accuracy, the measurements were repeated five times within the context of each scenario. After analyzing the results, it was determined that five repetitions were sufficient to eliminate statistical errors. Regarding network load, four different levels were considered: low, medium, high, and very high load for both the wireless and core

ISSN 2056-5860

networks. The default inter-cell load for downlink (DL) was set to 50% of the adjacent cell's transmit power, while the default inter-cell load level for uplink (UL) was set to a 3 dB noise rise.

User distance was also taken into account, with different distances considered. The default distance from the base station was set at 0.90km, representing a cell boundary user in a hexagonal network design. The default packet size was set to 1200 bytes due to simulation software limitations, even though the common packet size in IP networks is typically 1500 bytes. For channel bit rate tracking, the radio network simulator was used to simulate varying user distances from the eNB and inter-cell load on the downlink. The average bit rate observed for cell edge users was 5.2 Mbps, and three users with a 1.7 Mbps bit rate were selected as the default for in-cell upload on the downlink. Table 2 displays the different values considered in the experiment for the six scenarios, with the default values highlighted in grey.

Table 2: Scenario Aspect Parameter Calibration

<table>
<tr><th>Scenario</th><th colspan="3">Direction</th><th colspan="5">Values Considered</th></tr>
<tr><td rowspan="4">Network Load</td><td rowspan="3">Radio Network</td><td rowspan="2">Intercell (interference from neighboring cells)</td><td>UL (Increase in noise floor in dB)</td><td>0</td><td>1.5</td><td>3</td><td>6</td><td>-</td></tr>
<tr><td>DL (transmit power of neighboring cells, expressed as a percentage of the maximum transmit power)</td><td>0</td><td>25</td><td>50</td><td>100</td><td>-</td></tr>
<tr><td>Intracell (the number of other users in the cell)</td><td>UL / DL</td><td>0</td><td>1</td><td>2</td><td>6</td><td>-</td></tr>
<tr><td>Core network (total number of other users of the core network)</td><td colspan="2">UL / DL</td><td>0</td><td>12</td><td>25</td><td>50</td><td>-</td></tr>
<tr><td>The distance from the user to the base station (km)</td><td colspan="3">UL/DL</td><td>0.1</td><td>0.5</td><td>0.7</td><td>0.9</td><td>-</td></tr>
<tr><td>Size of Data Packet (bytes)</td><td colspan="3">UL/DL</td><td>100</td><td>500</td><td>1000</td><td>1100</td><td>1200</td></tr>
<tr><td>Packet rate (pps)</td><td colspan="3">UL/DL</td><td>100</td><td>110</td><td>120</td><td>130</td><td>180</td></tr>
</table>

Latency Mitigation Technique

In this research, the main latency mitigation technique has been chosen to be Differentiated Scheduling. In LTE, a scheduler is used to allocate and share limited network resources among network users. Based on the simulation tool in use, the differentiated scheduling will be based on Physical Resource Blocks (PRB). A PRB is the smallest element of resource allocation assigned by the eNB scheduler, consisting of 12 consecutive subcarriers for one slot (0.5ms). Relationship between frequency bandwidth and PRB is given as:

$$PRB = 5 \times f(MHz) \tag{4}$$

The developed radio network for this measurement setup has a channel bandwidth of 5 MHz which by equation (4) corresponds to a total of 25 PRBs. The proposed scheduler implements a round-robin scheme in the downlink time domain, where each Transmission Time Interval (TTI) assigns all 25 PRBs to one user while others wait. Users with higher priority receive customized scheduling, allowing them to clear/receive more data. In the uplink, scheduling is frequency domain-based, with each user programmed for at least one PRB per TTI. The total number of users and SINR determine the PRBs allocated to each user. Designated users in the uplink receive greater priority for frequency domain resources, receiving x times more resources than others, where x represents the priority weight.

Simulation Summary

The simulation tool employed in this research work is known as Riverbed Modeler. OPNET (Optimized Network Engineering Tool) Riverbed Modeler is the development environment of OPNET simulator for research and development. Since User distance is kept at a uniform value of 1km from access point to user equipment; data rate also kept at a constant of 54Mbps; and Data Packet size randomized between 500 to 1500 bytes uniformly among all users, the main variation is then on the different user traffic and application demand. Figures 9 and 10 show the simulation design of the baseline system.

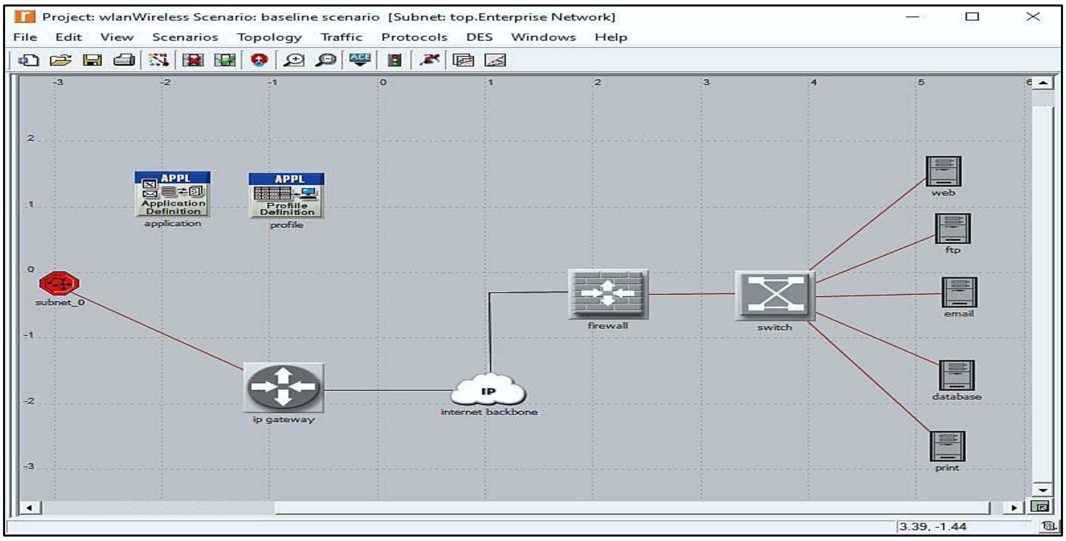

Figure 9: Network Main View

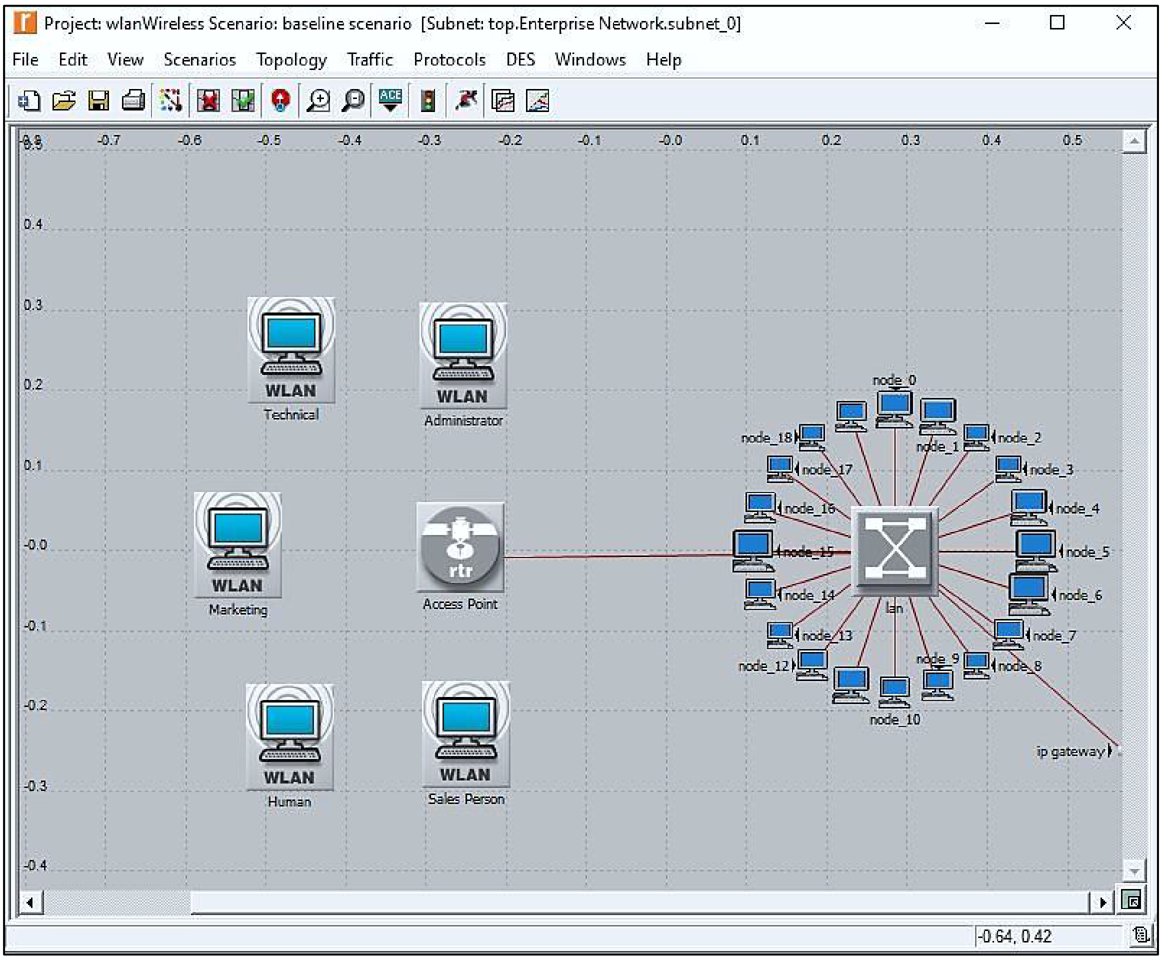

Figure 10: Expansion of Subnet_0 in Network Main View

RESULTS

Baseline Network results

The simulation was conducted over a duration of 30 minutes, during which random packets of varying sizes ranging from 500 bytes to 1500 bytes were sent and received. The obtained results were plotted on different graphs, with the x-axis representing the runtime from 0 to 30 minutes, and the y-axis representing the corresponding parameters. The separate graphs were created to display the latency statistics for the core network and the radio network. In order to avoid repetition, if a particular trend was observed across different scenario aspects in the results, an explanation for that trend was provided only at its initial occurrence. Subsequent occurrences of the trend were referred to the initial explanation. The delay or latency values were measured in seconds. The latency results for each of the profiles at the radio network level are given in Figures 11 to 16. Having seen the latency at the radio network end, the latency experienced at the core network is shown in Figure 17.

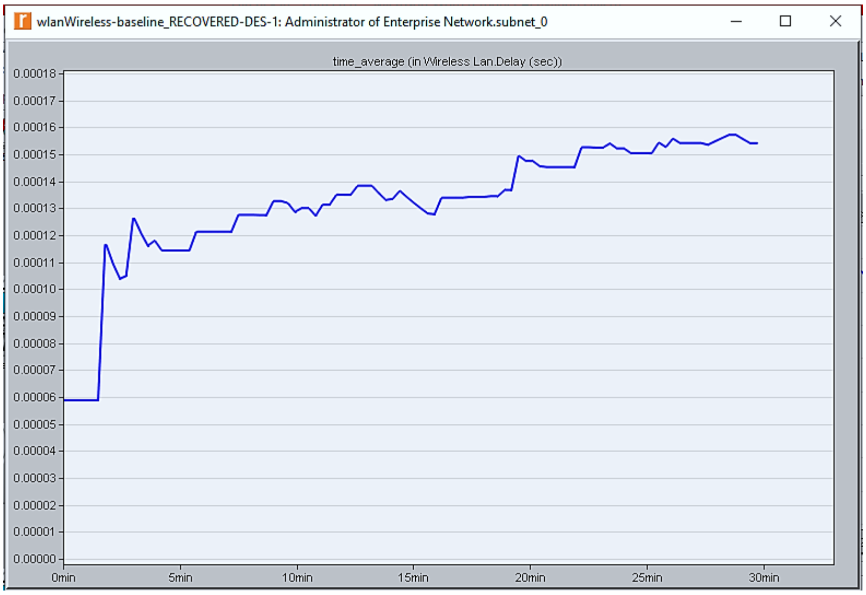

Figure 11: Latency at "Administrator" node

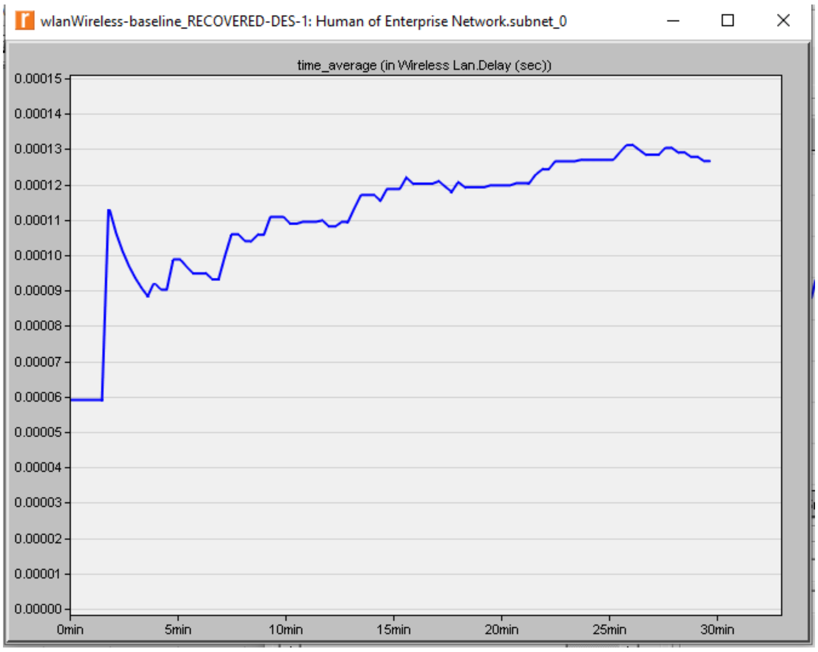

Figure 12: Latency at "Human" node

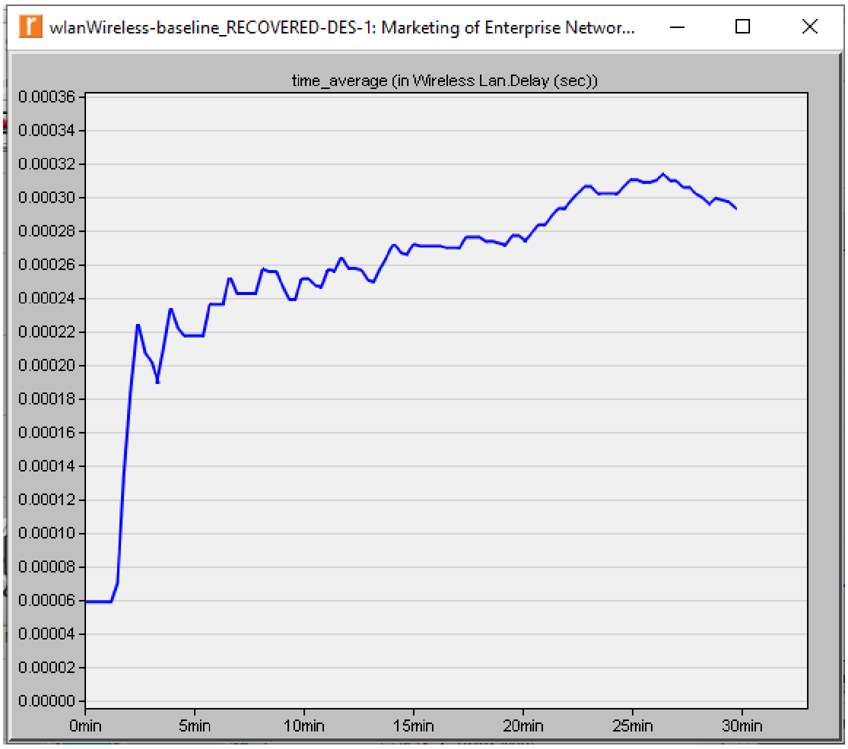

Figure 13: Latency at "Marketing" node

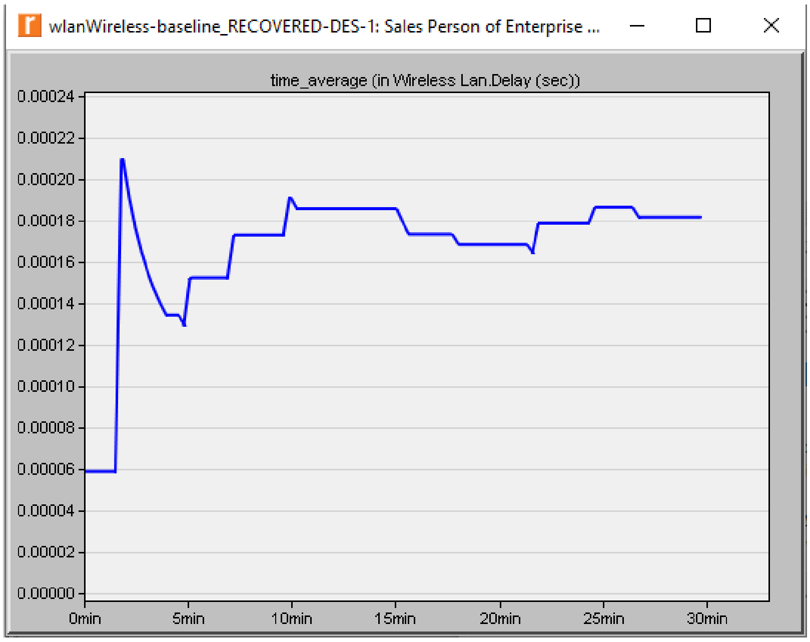

Figure 14: Latency at "Sales Person" node

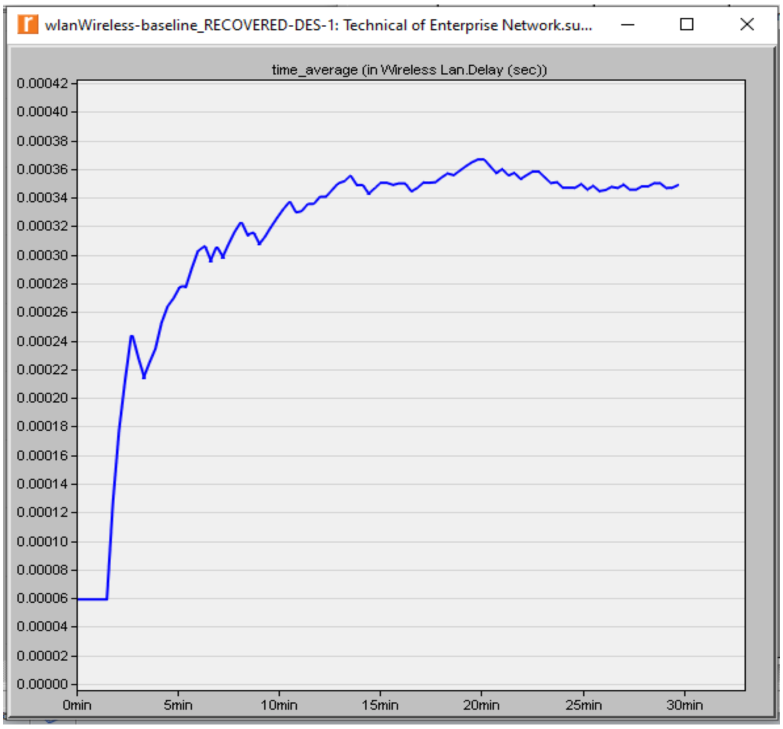

Figure 15: Latency at "Technical" node

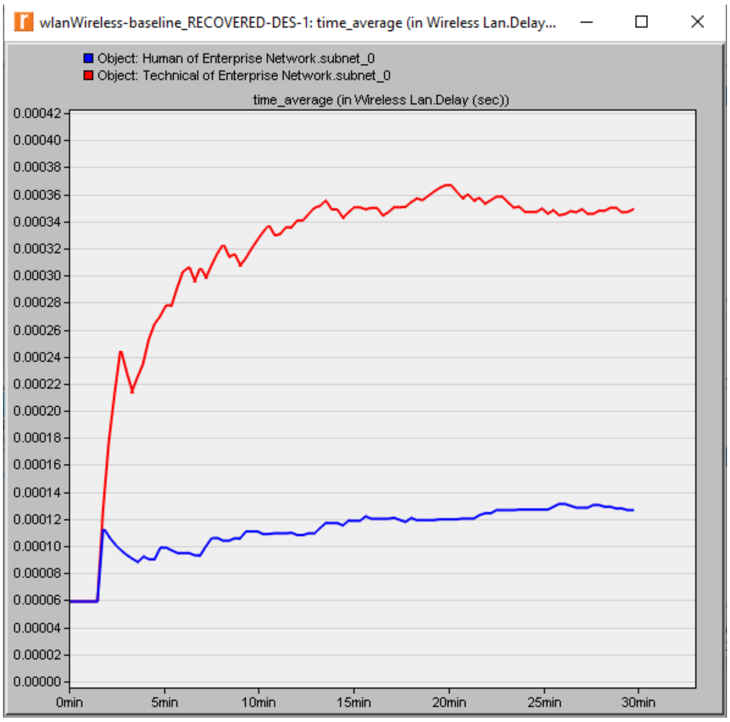

Figure 16: Latency/Delay of Heaviest and Lightest traffic plot against each other

ISSN 2056-5860

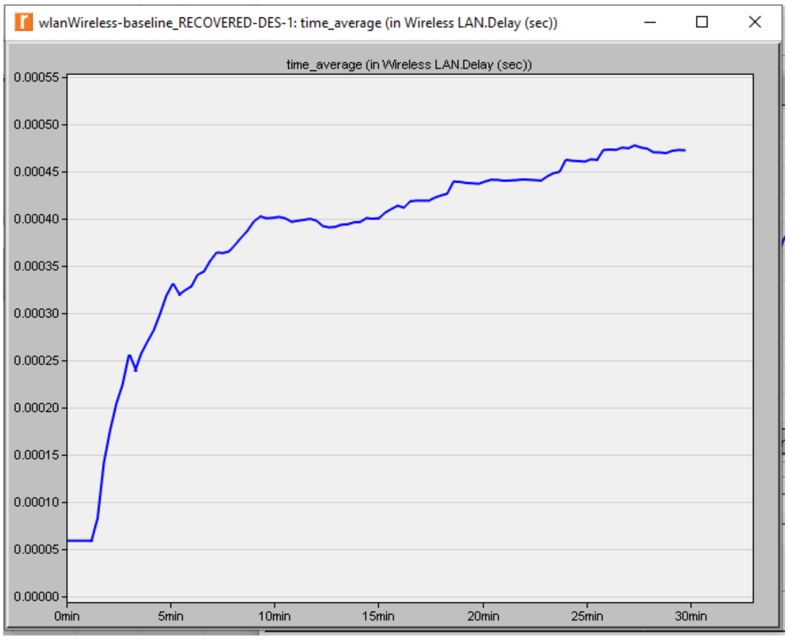

Figure 17: Latency at core network

Differentiated Scheduling Latency Mitigation

As mentioned in the earlier sections, differentiated scheduling prioritizes specific network users based on criteria established by the network designer. This prioritization can be implemented using various methods. In this report, the chosen algorithm for differentiated scheduling is the Low Latency Queuing (LLQ) Algorithm. The LLQ algorithm enables the utilization of a single priority queue that accommodates different classes of traffic. By employing a strict priority queuing scheme, LLQ ensures that delay-sensitive traffic, such as voice, is processed ahead of other packets in different queues (Semeria, 2001). In other words, priority is given to delay-sensitive traffic over other types of traffic. Multiple classes can be assigned priority status, and all traffic from these classes are directed to a single, high-priority queue when a policy map with several priority classes is set. The latency correction configuration specifies a single Priority Queue with a capacity of 1 Mbps, which is time-shared between two applications using an implicit policer. Voice and interactive video traffic are placed in a high-priority queue and are granted higher precedence over data traffic. Although these classes are in the same queue, they are rate-limited separately: audio traffic is limited to 540 kbps, while video traffic is limited to 460 kbps. After implementing the LLQ algorithm in the simulation tool's application configuration, the entire simulation was rerun within the same 30-minute timeframe, yielding new results for the same categories as the baseline simulation. The obtained results, compared with the baseline test, are presented in graph format of Figures 18 to 23.

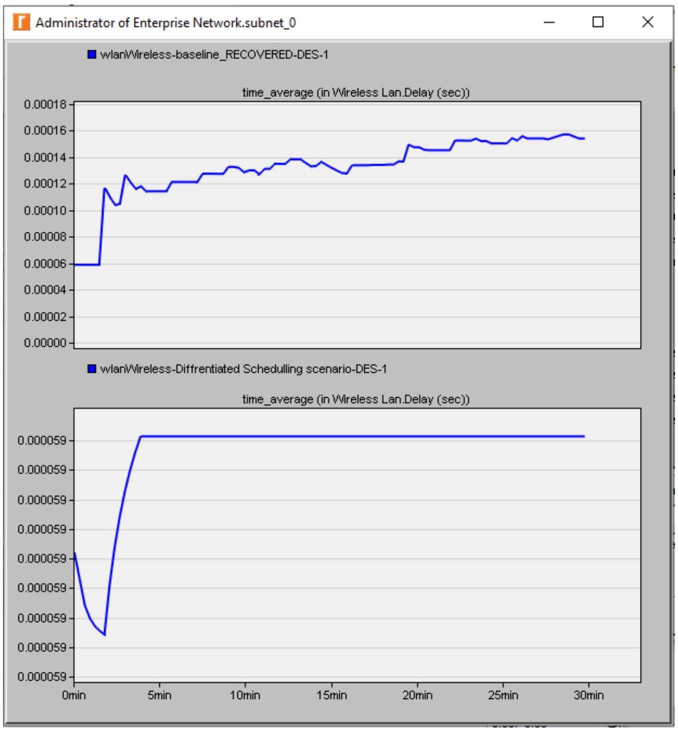

Figure 18: Administrator node showing latency before correction (top) and after correction (bottom)

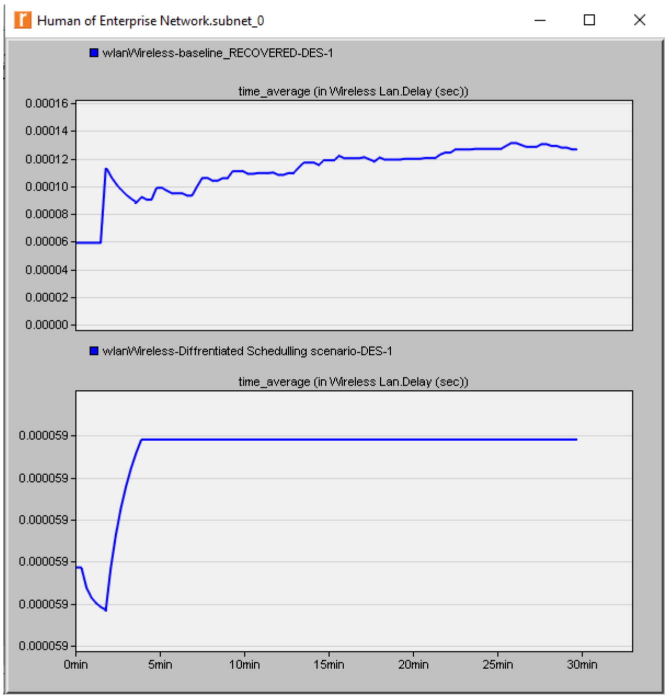

Figure 19: Human node showing latency before correction (top) and after correction (bottom)

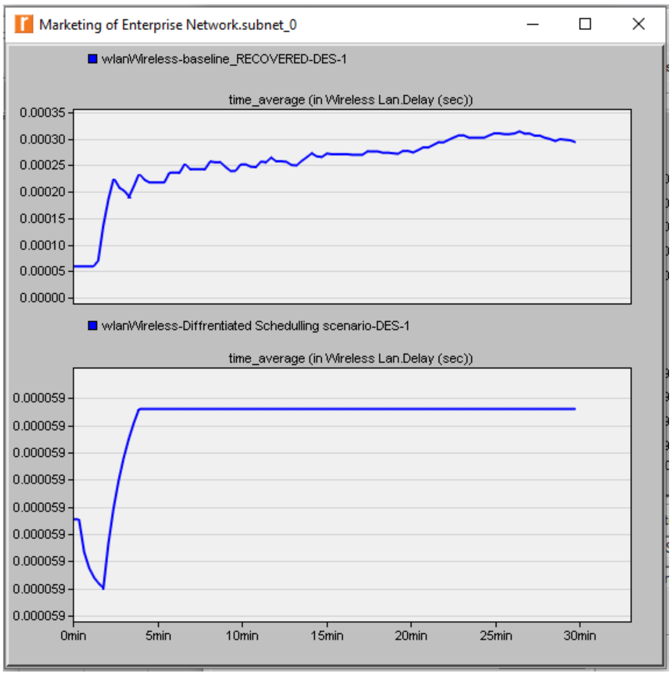

Figure 20: Marketing node showing latency before correction (top) and after correction (bottom)

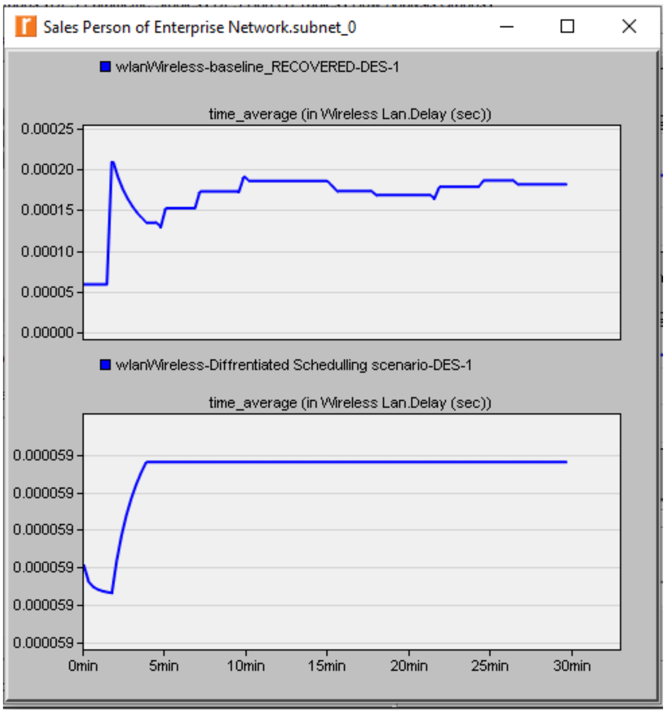

Figure 21: Sales Person node showing latency before correction (top) and after correction (bottom)

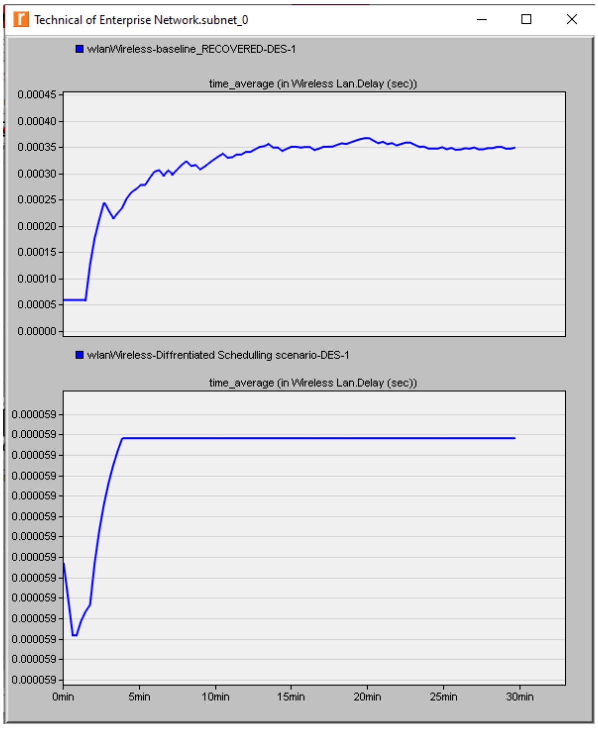

Figure 22: Sales Person node showing latency before correction (top) and after correction (bottom)

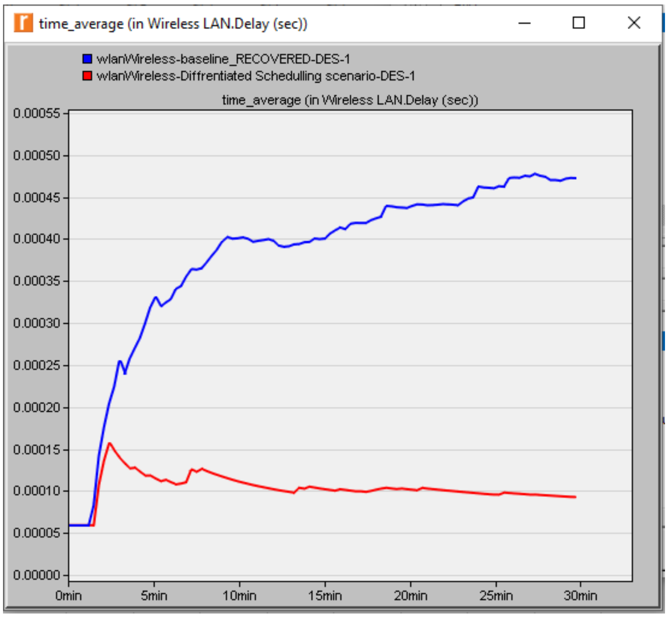

Figure 23: Comparison of Latency at core network

DISCUSSION

Upon examining the different graphs from Figure 10 to 16, it is evident that when maintaining a constant distance of 1km, data rate of 54Mbps, and data size uniformly distributed between 500-1500 bytes, the latency experienced in the network varies across the five user profiles with different levels of load and traffic. The user profile characterized by the highest traffic, known as the technical profile, exhibits a maximum latency of 0.00036 seconds, while the user profile with the lightest load, referred to as the human profile experiences a lower delay of 0.00013 seconds. It is important to note that these results were obtained without the implementation of the latency correction algorithm. The graph illustrating the latency of the heaviest and lightest load profiles is depicted in Figure 17. In summary, the latency within the network gradually increased as the communication run-time progressed from node to node. This indicates that as the network operates for a longer duration, the latency experienced also rises, which is undesirable for optimal network performance. Consequently, it becomes imperative to take necessary measures to mitigate this escalation of latency.

Upon analyzing the results of the radio network latencies following the implementation of the mitigation algorithm, as shown in Figures 18 to 23, a significant decrease in time delay across all five nodes becomes apparent. The latency experienced by each node is now approximately 0.00006 sec. This reduction can be primarily attributed to the prioritization of heavier traffic, which allows the network to efficiently address resource allocation. By handling the heavier traffic first and promptly, the network gains additional resources to allocate once the priority requirements are met. This noteworthy decrease in latency is not limited to the radio network alone; it also extends to the core network. A comparison between the latency results before and after mitigation in the core network is presented in Figure 23. The endeavor can be considered successful. The observed results vividly demonstrate the effectiveness of differentiated scheduling in any network. This is why it serves as one of the numerous mitigation tools in the advancement of the 4G network.

CONCLUSIONS

The objective of this report was to design and simulate a latency mitigation technique for 4G Networks. The 4G LTE network operates primarily on IP-based frameworks but also supports existing communication technologies. With the increasing demand for network traffic on the 4G network, maintaining optimal Quality of Service (QoS) is crucial. Latency mitigation is one approach to achieving this, and it serves as the focus of this paper. While the simulation tool approach may not be as effective as the experimental method, it provides a cost-effective means and establishes a foundation for future hybridization. The goal of this report was to simulate latency in an environment that closely resembles real-time conditions within the limitations of the simulation tool's capabilities and research scope.

Five mobile users with varying traffic demands and twenty wired node users were integrated into the radio network to simulate realistic traffic scenarios. The data rate, user distance, and data size were maintained at constant values.

Results obtained from the latency simulation without any mitigation after a 30-minute run revealed that latency in the radio network progressively increased with the amount of traffic in the network. This increase was continuous, indicating that over a longer runtime, latency could reach an unmanageable level. This was addressed as a differentiated scheduling algorithm was implemented in the network application configuration, and the system simulation was repeated to obtain new results. The new delays were significantly lower than the previous ones, and there was no further increase in latency after a certain period, resulting in a constant and optimal latency level. Thus, this analysis can be considered successful.

REFERENCES

3GPP (2010) Specification #: 25.913: Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN). Retrieved from:

https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1342 22 May 2021

Adebayo Emmanuel Abejide (2014) Performance Analysis of an LTE-4G Network Running Multimedia Applications. *Eastern Mediterranean University*, Gazimağusa, North Cyprus. Retrieved from: http://i-rep.emu.edu.tr:8080/jspui/bitstream/11129/1679/1/AbejideAdebayo.pdf 21 May 2021

Arlos, P. & Fiedler, M., (2010), Influence of the Packet Size on the One-Way Delay on the Down-link in 3G Networks. 10.1109/ISWPC.2010.5483732.

Ashish Kurian (2018) Latency Analysis and Reduction in a 4G Network. *Delft University of Technology*. Retrieved from: https://repository.tudelft.nl/islandora/object/uuid:e1badd8d-a384-49a1-b958-a0c1e499c539/datastream/OBJ/download 21 May 2021

Bamidele Moses Kuboye (2018) Performance Evaluation of Scheduling Algorithms for 4G (LTE). *Communications and Network*, 10, 152-163 DOI: 10.4236/cn.2018.104013

Blaji, T., Noguli, D., & Družijani, M. (2007). Latency Improvements in 3 G Long Term Evolution.

Burakov, Michael & Eldstål-Damlin, Albin. (2012). Master's Thesis an LTE Random Access Channel Model for Wireless Sensor Network Applications. Retrieved from: https://www.researchgate.net/publication/267841753_MASTER'S_THESIS_An_LTE_Random_Access_Channel_Model_for_Wireless_Sensor_Network_Applications 21 May 2021

Chadchan S. M. and Akki C. B. (2011) Priority-Scaled Preemption of Radio Resources for 3GPP. *International Journal of Computer Theory and Engineering*, *3*(6), pp 743-749 Retrieved from: http://www.ijcte.org/papers/403-G1159.pdf 21 May 2021

Christine De Monfreid (2013.) The LTE Network Architecture, *Alcatel Lucent Strategic White Paper*, Retrieved from: http://www.cse.unt.edu 21 May 2021

CompTIA, (n.d) What Is a Network Protocol, and How Does It Work? Retrieved from: https://www.comptia.org/content/guides/what-is-a-network-protocolb 21 May 2021

Egena Onu (2017) Planning and Optimisation of 4g/5g Mobile Networks and Beyond. University of Salford, Salford, UK. Retrieved from: http://usir.salford.ac.uk/id/eprint/45123/1/Egena%20Onu%20PhD%202017.pdf 21 May 2021

Ericsson (2017) 5G Systems: enabling the transformation of industry and society. Retrieved from: https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-systems.pdf. 22 May 2021

Firmin Frédéric (n.d) *The Evolved Packet Core*. 3GPP, Retrieved from: https://www.3gpp.org/technologies/keywords-acronyms/100-the-evolved-packet-core 21 May 2021

Hattachi Rachid El (2015) NGMN 5G White Paper. NGMN Ltd. Retrieved from: https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf 22 May 2021

Hossen Showket & Islam Neharul (2012) QoS Performance Evaluation of Video Conferencing over LTE. School of Computing Blekinge Institute of Technology 371 79 Karlskrona Sweden. Retrived from: https://www.diva-portal.org/smash/get/diva2:832428/FULLTEXT01.pdf 21 May 2021

Isaam Toufik, Mathew Baker, Stefania Sesia (2009) *LTE-The UMTS Long Term Evolution-From Theory or Practice*. John Wiley.

Ivan Penalba (2015) What differentiates 5G from LTE networks? *The Free Android*. Retrieved from: https://www.elespanol.com/elandroidelibre/20160223/diferencia-redes-lte/104489949_0.html 21 May 2021

ISSN 2056-5860

Laner M., Svoboda P., Romirer-Maierhofer P., Nikaein N., Ricciato F. and Rupp M.,(2012), A Comparison Between One-way Delays in Operating HSPA and LTE Networks. *10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, Paderborn.

Ocean News (2018). Tampnet Announces the Completion of the 4G LTE network Coverage in GoMRetrieved from: https://www.oceannews.com/news/subsea-cable/tampnet-announces-the-completion-of-the-4g-lte-network-coverage-in-gom 21 May 2021

Oludayo John Oguntoyinbo (2013) The Future of LTE: The Femtocells perspective. Aalto University School of Electrical Engineering. Retrieved from: https://core.ac.uk/download/pdf/80710971.pdf 21 May 2021

Riihonen, T., Wichman, R., & Werner, S. (2012). Evaluation of OFDM(A) Relaying Protocols: Capacity Analysis in Infrastructure Framework. *IEEE Transactions on Vehicular Technology*, *61*(1), 360-374.

Taranetz, Martin & Blazek, Thomas & Kropfreiter, Thomas & Muller, Martin & Schwarz, Stefan & Rupp, Markus. (2015). Runtime Precoding: Enabling Multipoint Transmission in LTE-Advanced System-Level Simulations. *Access*, IEEE. 3, pp 725-736. 10.1109/ACCESS.2015.2437903.

Wylie-Green, M., & Svensson, T. (2010). Throughput, Capacity, Handover and Latency Performance in a 3GPP LTE FDD Field Trial. 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, 1-6. DOI:10.1109/GLOCOM.2010.5683398. Retrieved at: https://www.semanticscholar.org/paper/Throughput%2C-Capacity%2C-Handover-and-Latency-in-a-LTE-Wylie-Green-Svensson/f969b56567f983c7fe17a94915b450cf614b2720 30 June 21

Xu, Yuzhe & Fischione, Carlo. (2012). Real-time scheduling in LTE for smart grids. 10.1109/ISCCSP.2012.6217872.