© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710727-4766

Variability-Driven Iterative MLR-LP Approach to Resource Allocation in 5G Network Slicing: Minimizing Latency in Ride-Hailing Services

THEOPHILUS IREBHUDE AGHUGHU[1], BELLO O. LAWAL[2], BRAIMOH ABDULLAHI IKHARO[3]
[1,2,3] *Department of Computer Engineering, Edo State University Iyamho, Edo State, Nigeria*

Abstract- The advent of 5G networks has ushered in a new era of communication technology characterized by unprecedented speed, ultra-low latency, and higher reliability. However, intelligent and adaptive resource allocation in 5G network slicing is critical to meeting consistent sub-10 ms latency, a value that aligns with the performance benchmark of ultra-reliable low-latency communications such as ride-hailing. This study deployed a hybrid Multiple Linear Regression–Linear Programming (MLR-LP) framework for optimizing bandwidth, memory, and signal strength to achieve latency reduction. Real-time data were collected from ride-hailing sessions in 5G-covered areas of Benin City, Nigeria, capturing latency, bandwidth, memory, and signal strength. The MLR model established the predictive relationship between network resources and latency, achieving a strong R^2 value of 0.941. The regression equation was embedded as the objective function of an iterative LP model, which optimized bandwidth, memory, and signal strength allocations. The iteration process was guided by practical feasibility and variability analysis, particularly the unit step standard deviation, to progressively expand the bounds of resource variables in a controlled manner until feasible sub-10 ms latency was consistently obtained. The results demonstrate that the variability-driven iterative MLR-LP approach effectively minimizes latency to reliably support latency-sensitive services and enhancing 5G slicing performance. The study concludes that integrating predictive modeling with optimization techniques provides both theoretical and practical contributions, offering a possible solution for adaptive 5G resource management.

Index Terms: 5G Network Slicing, Latency, Linear Programming, Multiple Linear Regression, Resource Allocation.

I. INTRODUCTION

5G technology offers critical advancements over previous mobile network generations, particularly through its ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC) capabilities (1), (2). Among these, URLLC is especially suited for mission-critical and latency-sensitive applications (3). These features make 5G particularly promising for industries that depend on low latency such as ride-hailing services by offering significantly improved responsiveness compared to 4G networks (4).

Ride-hailing applications such as Bolt and Uber operate via GPS-enabled mobile platforms that connect passengers with nearby drivers in real time (5). The performance of such services depends critically on fast and stable network connectivity, with low latency being essential for functions like live tracking, dynamic pricing, and seamless ride dispatch (6). As a result, ride-hailing provides a practical and relevant use case for evaluating real-time 5G performance. Excessive latency in ride-hailing applications can cause service delays, mismatched driver assignments, and suboptimal user experience (5). While 5G network slicing offers the flexibility to create service-customized virtual networks, network providers often struggle with inefficient resource allocation within slices, undermining the full potential of low-latency performance (7), (8).

A growing body of literature has explored various approaches to address these challenges, including machine learning models for latency prediction, linear and mixed-integer programming for resource allocation, and network slicing strategies for service differentiation. (9) developed a standardized Integer Linear Programming (ILP) model to optimize mobile network resource allocation. They emphasized the diversity of emerging 5G use cases including enhanced Mobile Broadband (eMBB), industrial automation, and critical safety communications which impose varying requirements in terms of latency, throughput, and reliability. To accommodate these needs, the paper highlighted the role of end-to-end network slicing

© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710727-4766

and virtualization as essential technologies for enabling efficient service delivery over shared infrastructure. The proposed ILP model specifically tackled the offline network slice embedding problem by detailing how virtual nodes and links could be effectively mapped onto physical infrastructure. Although the paper successfully presented a clear and standardized formulation for resource allocation, its scope was limited to offline scenarios, and it did not integrate predictive mechanisms or real-time adjustments. (10) used Multiple Linear Regression with real xHaul latency data to model relationships between bandwidth and load. Their latency estimates were feasible, but lacked optimization. (11) in their paper, reviewed RL methodologies for slice orchestration. They identified a lack of hybrid models involving regression.

Despite the significant body of work exploring network slicing and resource allocation for 5G networks, a notable gap persists across optimization techniques within the context of latency minimization in ride-hailing services. Current resource allocation models rarely prioritize real-time latency minimization in the context of ride-hailing's operational demands (12). Furthermore, many models overlook local performance conditions such as signal strength that directly impact end-to-end latency (13), (14).

In the absence of a context-specific optimization framework, network slicing remains largely theoretical and fails to meet the practical needs of mobility services in Nigerian urban settings. Therefore, there is a compelling need to develop an application and location-specific, regression-based optimization model.

The methodological framework of this study aims to address these gaps by adopting the usage of real-world primary data and recognizing that network latency in ride-hailing applications can be significantly influenced by three key 5G network resources: bandwidth, edge memory, and signal strength.

Efficient bandwidth allocation is fundamental in 5G network slicing, as bandwidth directly determines the data rate and service quality of each slice (15). Edge memory allocation is equally critical and justifiable because 5G slices rely on virtualized network functions (VNFs) and Multi-access Edge Computing (MEC), where sufficient RAM and buffer allocation ensures faster packet processing and reduced queueing delay (16). Furthermore, signal strength allocation is very vital can be managed via transmit power control, beamforming, and RB scheduling to enhance reliability and reduce retransmissions, thereby lowering latency (17). Considering bandwidth, memory, and signal strength as allocable resources is therefore consistent with 5G slicing principles, as each plays a direct role in shaping latency performance and overall Quality of Service (QoS).

II. MATERIALS AND METHODS

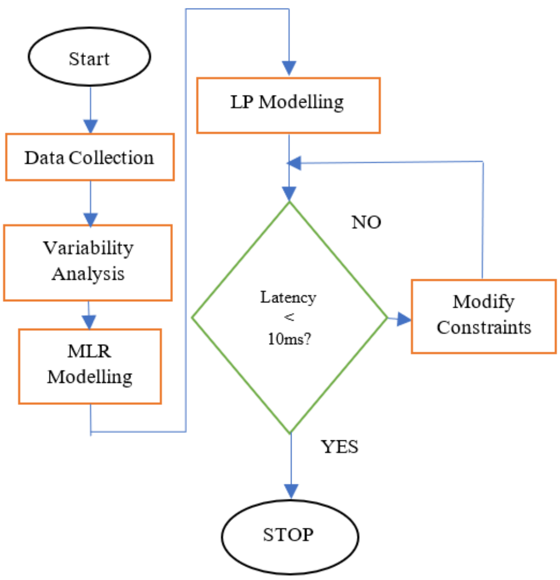

Figure 1: Research methodology flowchart

Figure 1 shows the flowchart of the complete research methodology. Variability analysis is first conducted on the primary data collected from ride-hailing sessions within 5G-covered areas of Benin City, Nigeria. This step provides insight into the distribution and fluctuation of each resources, laying the foundation for both accurate modeling and adaptive optimization. A Multiple Linear Regression (MLR) model is then applied to quantify the relationship between the network resources and latency. The MLR equation estimates the extent to which each network resource contributes to latency outcomes, thereby providing a statistically validated basis for optimization. The resulting regression model is then embedded within a Linear Programming (LP) framework, where it serves as the objective function. This LP model is formulated to minimize latency below 10 ms target by identifying optimal resource allocations under specified constraints controlled by practical feasibility and the standard deviation values of each resources.

A. Method for data collection

This study utilizes primary data. It relies on a combination of mobile and network diagnostic tools installed on an android smartphone for real-time observations and measurements within active 5G network zones in Benin City. The data focuses on four key network performance indicators critical to ride-hailing service efficiency: latency, bandwidth, memory, and signal strength. These variables are monitored during live sessions of a Bolt ride-hailing application. Table 1 shows the diagnostic software for the data collection process.

Table 1: Software for data collection.

Ride-hailing platform	Bolt App
ISP and Server information	Wi-Fi analyzer, Net analyzer
Latency Data	Fast.com, RF Benchmark
Bandwidth Data	Glasswire, Trepn profiler
Memory Data	Simple system monitor, Android system info
Signal Strength Data	G-CPU monitor, Traffic monitor

This study adopts a purposive (judgmental) sampling technique to select zones where data are collected. From 10 bolt ride sessions, 30 sample points are selected across different parts of Benin City with confirmed presence of 5G network signal. Figure 2 shows the hotspots of the locations where data are collected.

At each point, network resource data (bandwidth, memory allocation, signal strength) and corresponding latency readings are collected to build a comprehensive dataset for regression and optimization modeling.

During a live session of Bolt app usage, measurements are taken in real time, under different network loads, resources and varying environmental conditions, to reflect true usage behavior. The collection and processing of user/driver data adheres to ethical standards to protect user privacy.

B. Method for data validation

Validity is ensured by using two independent tools to measure the same metric (e.g., Fast.com and RF Benchmark for latency, G-CPU and Traffic Monitor for signal strength). Average value is taken in cases of disparity.

C. Data preprocessing

The dataset undergoes systematic preprocessing to ensure quality and reliability before modeling. Obvious logging errors are identified and corrected by re-taking the affected ride-hailing sessions. Duplicate records generated by repeated logging during real-time collection are removed. Outlier capping is also applied to minimize the influence of unusual spikes in latency, bandwidth, or memory usage. Reasonable thresholds are enforced to prevent extreme values from distorting regression estimates. Signal strength values are constrained within the range of -120 dBm to -40dBm, consistent with 3GPP specifications. Sample points that fall outside this range are flagged as invalid, and the corresponding ride-hailing sessions are re-taken to ensure accuracy.

Through these steps, the dataset is maintained to be free of duplicates, corrected for measurement inconsistencies, and restricted to the practical operational ranges of 5G network parameters, thereby supporting valid regression and optimization analysis.

D. Variability analysis of network parameters

To understand the behavior of key 5G network parameters in real-world conditions, this study conducts a variability analysis on the collected primary data. The objective of this analysis is to assess the degree of fluctuation, dispersion, and consistency in the values of bandwidth, memory, and signal strength during ride-hailing requests.

The variability analysis involves computing key descriptive statistics for each of the three 5G network parameters: bandwidth (kbps), memory usage (MB), and signal strength (dBm). The primary metrics for the assessment includes the range, standard deviation and the standard deviation unit form. These metrics provides a comprehensive overview of the central tendency and dispersion of each variable. To visually support these findings, bar charts is generated to display the variability in magnitude and distribution across the observed data points. The final results of the variability analyses are compiled into summary tables, which guided the construction of feasible ranges and constraint update logic within the LP modeling phase.

E. MLR model formulation

To quantify the relationship between the 5G network resources and latency during ride-hailing operations, a Multiple Linear Regression (MLR) model is formulated. The purpose of this model is to estimate the extent to which each of the three key resources, bandwidth, memory, and signal strength influences the dependent variable, latency. This step establishes a statistically grounded basis for subsequent optimization via Linear Programming (LP).

F. Mathematical modelling of latency using MLR

Recognizing the three independent variables, the Multiple Linear Regression model can be expressed as

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 \pm \varepsilon \quad (1)$$

Where: Y = Estimated latency (ms),

X_1 = Bandwidth (kbps), X_2 = Memory (MB),

X_3 = Signal strength (dBm),

β_0 = Intercept (baseline latency when all predictors are zero),

β_1, β_2, β_3 = Regression coefficients representing the effect of each variable,

ε = Error term

Based on the real-time primary data collected, the fitted regression equation takes the form:

$$\text{Estimated latency} = \beta_0 + \beta_1(\text{Bandwidth}) + \beta_2(\text{Memory}) + \beta_3(\text{Signal strength}) \pm \varepsilon \quad (2)$$

G. Linear programming problem and model formulation

To achieve optimal 5G network performance for latency sensitive applications like ride-hailing, a Linear Programming (LP) model is developed using the Multiple Linear Regression (MLR) equation as the objective function. This model aims to identify the best combination of network resources; bandwidth, memory, and signal strength that minimizes latency, while adhering to real-world operational constraints and requirements. These specified boundaries define the feasible solution space for the LP optimization, ensuring that the model recommendations are constrained within realistic and possible limits.

While the standard Linear Programming (LP) model offers a one-time optimal solution for minimizing latency based on fixed bounds and constraints, it may not always yield a feasible or practically deployable result within the dynamic conditions of real-time 5G networks. To address this, an iterative approach is introduced into the model to progressively adjust constraints until a latency (less than 10 ms) that meets ultra-reliable low-latency communication (URLLC) standards is achieved.

The LP formulation is structured around the principle that latency, the dependent variable, is a linear function of the three independent network variables as modeled and expressed in the MLR equation (2) earlier.

The LP problem formulation is expressed as follows:

Objective Function:

$$\text{Minimize Latency (L)} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 \quad (3)$$

Decision Variables: X_1 as bandwidth, X_2 as memory, X_3 as signal strength.

Latency Target: Latency (L) < 10 ms

Initial Subject (Iteration 1):

$$\begin{cases} x1min \leq x1 \leq x1max \\ x2min \leq x2 \leq x2max \\ x3min \leq x3 \leq x3max \end{cases}$$

Consequent Subject (Iteration 1+n):

$$\begin{cases} x1min \leq x1 \leq x1max + a \\ x2min \leq x2 \leq x2max + b \\ x3min \leq x3 \leq x3max + c \end{cases}$$

The values of a,b and c in each iteration is informed by insights gained during the variability analysis of the independent variables; bandwidth, memory, and signal strength. The variability metrics, particularly the unit step standard deviation, provide a rational basis for incrementally relaxing or tightening bounds in each iteration. The goal is to find the first feasible and optimal combination of network parameters that results in latency within the target threshold of < 10 ms.

III. RESULTS

This section presents the results of all the methods including details of the collected data and the outcomes of all data analysis. The findings are interpreted in line with the research aim which is to develop a model that minimizes latency through 5G network slicing (allocating available network resources such as bandwidth, memory, and signal strength).

H. Primary data overview

This study used primary data collected in real-time from ride-hailing sessions using the Bolt application in 5G-covered areas around Benin City. The dataset includes four key network data; latency (ms), bandwidth (Kbps), memory usage (MB), and signal strength (dBm), recorded during ride-hailing sessions in 5G-enabled areas of Benin City. Latency values varied across sessions, with the lowest readings indicating near-instantaneous response times and the highest reflecting network delays during periods of fluctuating connectivity. Bandwidth measurements also exhibited substantial variation, from modest data rates during congested periods to peak speeds exceeding 1 Mbps, highlighting the dynamic nature of available throughput in the study area. Memory values generally fell within a moderate range, with occasional spikes while signal strength, expressed in dBm, showed fluctuations linked to network coverage zones and possible interference effects, ranging from strong, stable connections to occasional weaker signals. The interplay of these metrics reflects the complex, real-world conditions experienced by ride-hailing applications, where performance depends not on a single factor but on the combined influence of multiple network resources. A summary statistic of the data is shown in Table 2. The values of these network variables form the foundation for the regression, and optimization analyses presented in the subsequent sections.

Table 2: Summary statistics of dataset

Metric	Latency (ms)	Bandwidth (kbps)	Memory (MB)	Signal Strength (dBm)
Count	30	30	30	30
Mean	40.23	1630.17	34.44	-79.93
Minimum	28.00	369.00	28.90	-110.00
Maximum	64.00	2387.00	39.60	-49.00

I. Variability analysis results

Table 3: Variability analysis results

	Minimum	Maximum	Range	Mean	Standard Deviation
Bandwidth	369	2387	2018	1630.167	539.7015
Memory	28.9	39.6	10.7	34.43667	3.359134
Signal Strength	-110	-49	61	-79.9333	16.21394

The results presented in Table 3, reveal distinct patterns of variability across the three resources. To visually reinforce these findings, Figure 3 presents a bar chart of the standard deviation values for each resource. This chart illustrates the pronounced variability in bandwidth compared to the other resources, highlighting the need to account for this factor in both the Multiple Linear Regression (MLR) and Linear Programming (LP) modeling phases.

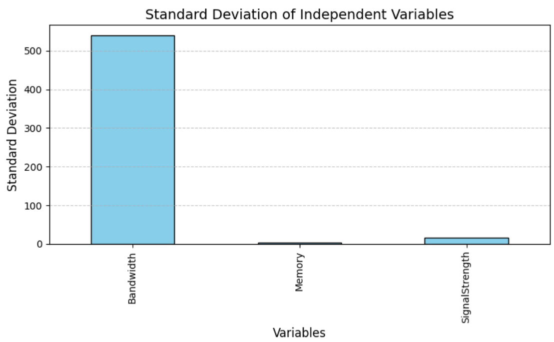

Figure 3: Bar chart of standard deviation for network resources.

In addition to the raw variability measures, the standard deviations for bandwidth, memory, and signal strength were converted into unit-form to enable direct comparison on a uniform scale. The smallest observed standard deviation was for memory, which was assigned a value of 1 unit and used as the normalization baseline. Dividing the

© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710727-4766

standard deviations of bandwidth and signal strength

Variable	SD (σ)	Normalized Step Size (SD-ratio relative to memory)
Bandwidth	539.70	539.70 ÷ 3.36 ≈ 160.6 Kbps
Memory	3.36	1.00 MB (baseline)
Signal Strength	16.21	16.21 ÷ 3.36 ≈ 4.82 dBm

by this baseline yielded unit-form values as presented in Table 4

Table 4: Unit-form Standard deviation.

J. Multiple linear regression model results

Multiple Linear Regression (MLR) model was implemented to examine the relationship between latency (dependent variable) and the three network resources; bandwidth, memory, and signal strength (independent variables).

Figure 4,5 and 6 presents scatter plots showing the distribution of each independent variable against latency. These visualizations provide an initial indication of negative trends, suggesting that increases in bandwidth, memory, or signal strength are associated with reductions in latency.

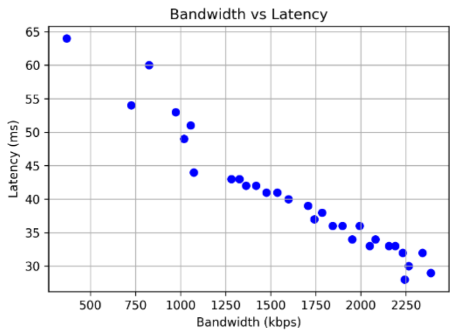

Figure 4: Scatter plots of bandwidth vs latency.

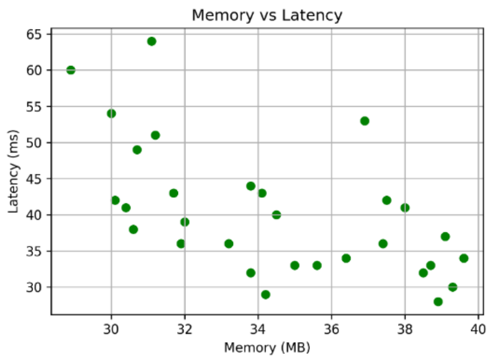

Figure 5: Scatter plots of memory vs latency.

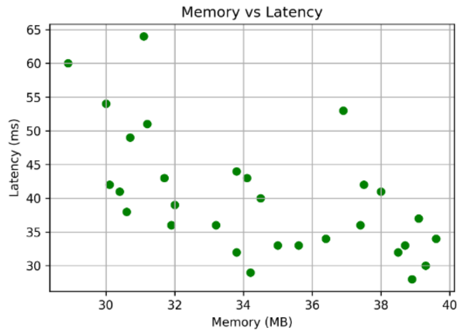

Figure 6: Scatter plots of signal strength vs latency.

Based on these observed relationships, the final regression equation derived from the model is:

Latency = $69.4548 - 0.0148X_1 - 0.2078X_2 - 0.0258X_3$ (4)

K. Interpretation of coefficients

i. *Bandwidth:* A unit increase in bandwidth (Kbps) results in an average reduction of 0.0148 ms in latency, assuming other factors are constant.
ii. *Memory:* A unit increase in available memory (MB) decreases latency by approximately 0.2078 ms.
iii. *Signal Strength:* Because signal strength is negative (in dBm), an increase (i.e., less negative) improves latency. Each 1 dBm improvement leads to a 0.0258 ms decrease in latency.

L. Model performance and validation

The statistical performance of the multiple linear regression (MLR) model was evaluated using both coefficient specific and collective metrics. The performance metrics are summarized in Table 5 and 6.

Table 5: Metrics of individual predictors in the MLR model.

	Standard Error	t-value	p-value	VIF
Bandwidth	0.002	-9.538	0.000	3.797
Memory	0.154	-1.352	0.188	1.444
Signal Strength	0.050	-0.519	0.608	3.513

The standard errors values indicate the average amount by which the estimated coefficients deviate from the true population values, with lower values suggesting more precise estimates. In this case, bandwidth had the smallest standard error,

© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710727-4766

indicating a high level of precision in estimating its coefficient, while memory and signal strength had comparatively larger errors.

The p-values describes how significant the independent variables are to latency. At the conventional 5% significance level, only bandwidth showed a statistically significant relationship with latency. However, despite the non-significance of two predictors, their inclusion in the model can still be justified if they contribute to overall explanatory power and better prediction accuracy when considered jointly.

The variance inflation factor (VIF) values for Bandwidth, Memory, and Signal Strength were all below the conventional threshold of 5, indicating the absence of serious multicollinearity among the independent variables. This validates the stability of the regression coefficients and supports the inclusion of all three predictors in the multiple linear regression model.

Table 6: Metrics of the overall MLR model.

Metric	Value	Interpretation
R^2	0.941	Strong predictive power; 94.1% of latency variation explained
Adj-R^2	0.934	The model maintains a high explanatory power even after adjusting for possible overfitting
F-statistic	138.1	Substantially large, indicating that the model as a whole is statistically significant
RMSE (ms)	2.1547	Low prediction error in relative to latency target, good fit
MAE (ms)	1.6332	Small average absolute error
Durbin-Watson	1.535	Mild positive autocorrelation in residuals
Jarque-Bera	3.085	Residuals approximately normally distributed

The regression diagnostics results confirm the robustness of the multiple linear regression (MLR) model. The model achieved a high coefficient of determination, indicating strong explanatory power, while the root mean squared error (RMSE) and mean absolute error (MAE) values demonstrate low prediction errors. Given the sub-10 ms latency optimization target, these low error margins validate the reliability of the regression equation as an objective function in the iterative LP framework. The Durbin–Watson statistic suggests mild positive autocorrelation, but since the regression is embedded in an iterative optimization framework rather than used for hypothesis testing, the presence of autocorrelation does not compromise its validity. Also, residual normality is largely upheld as reflected by the Jarque-Bera value. Collectively, these results validate the stability and predictive accuracy of the regression equation, justifying its integration into the optimization framework.

M. Linear programming model results

Following the development of the MLR model, the derived regression equation was integrated into a Linear Programming (LP) framework to optimize network resource allocation for achieving sub-10 ms latency. The LP model treats the MLR equation as the objective function to be minimized resulting to the following:

Minimize: $\text{Latency} = 69.4548 - 0.0148X_1 - 0.2078X_2 - 0.0258X_3$

Based on the unit-form standard deviations observed in Table 4, A 1 MB increase in memory corresponds statistically to a 160.6 Kbps bandwidth increase and a 4.82 dBm improvement in signal strength. However, due to physical and system constraints, especially for signal strength, the model adopts more conservative and realistic step sizes of 150 Kbps for bandwidth, 1MB for memory and 1 dBm for signal strength. This hybrid approach maintains statistical integrity while ensuring practical feasibility.

Table 7 presents the iteration results, showing for each step, the bandwidth, memory, and signal strength combination, and the corresponding estimated latency.

Iteration	Bandwidth	Memory	Signal Strength	Estimated Latency	Status
1	2387	39.6	-49	27.16252	Infeasible
2	2537	40.6	-48	24.70892	Infeasible
3	2687	41.6	-47	22.25532	Infeasible
4	2837	42.6	-46	19.80172	Infeasible

5	2987	43.6	-45	17.34812	Infeasible
6	3137	44.6	-44	14.89452	Infeasible
7	3287	45.6	-43	12.44092	Infeasible
8	3437	46.6	-42	9.98732	feasible

Table 7: Iterative LP model results across iterations.

From the table, it is evident that latency decreases consistently with each iteration as resources remain near their upper feasible limits. The target latency was first achieved at iteration 8 with 3487 Kbps as optimal bandwidth, 46.6 MB as optimal memory, -42 dBm as optimal signal strength and estimated latency of 9.99 ms.

IV. DISCUSSIONS

This section provides an interpretation of the study's findings in relation to the objectives of resource allocation optimization and latency reduction in 5G network slicing. The data collection exercise provided real-time ride-hailing service measurements of latency, bandwidth, memory usage, and signal strength within 5G-covered areas of Benin City. Descriptive analysis revealed that latency values fluctuated significantly depending on network resource availability, with all instances exceeding the desired sub-10 ms threshold. This confirmed the necessity of a systematic optimization framework capable of aligning resource allocation with service-level requirements. The variability across bandwidth, memory, and signal strength also demonstrated that no single resource dimension alone could guarantee reduced latency, reinforcing the need for a combined approach.

Variability analysis showed that bandwidth exhibits the highest standard deviation, indicating greater inconsistency in available throughput values during observations. Memory usage showed comparatively lower variability, suggesting that its allocation remained relatively stable. Signal strength demonstrated moderate variability, which could be attributed to environmental and infrastructural factors affecting network reception.

The MLR model results showed that bandwidth, memory, and signal strength all had negative coefficients, indicating that increases in these resources were associated with reductions in latency. Among them, memory usage exhibited the strongest effect per unit, suggesting that efficient memory allocation plays a critical role in latency reduction. The model's predictive accuracy confirmed its suitability for embedding within the LP framework as an objective function, ensuring that the optimization process remained grounded in empirical data.

The LP model was infeasible under maximum observed resources as bounds, incapable of yielding latency below 10 ms. The results revealed that, starting from the maximum observed resource levels, an additional 1050 kbps of bandwidth, 7 MB of memory, and 7 dBm of signal strength were required to meet the latency threshold. This demonstrated that bandwidth is a critical determinant of 5G latency in ride-hailing applications and that multi-dimensional resource balancing, rather than reliance on a single parameter, is also essential for achieving latency levels supportive of ride-hailing and other low latency services.

N. Limitation of the Study

While this study provides practical insights into optimizing 5G network resources for latency-sensitive ride-hailing services, it was subjected to several limitations that should be acknowledged. The data used for model development was collected only in Benin City, Edo state. As a result, the model's applicability may not generalize across other cities or rural areas with different infrastructural or environmental conditions. Also, all measurements were conducted using Android-based mobile devices, and diagnostic tools. Variability in hardware capabilities, software optimization, or OS behavior could influence recorded values, potentially limiting cross-device consistency. Lastly, the model focused solely on three network resource variables; bandwidth, memory, and signal strength as predictors of latency. Other possible influential factors, such as CPU performance, transmission power, or backhaul delay, were not explicitly included due to data accessibility and model simplification.

V. CONCLUSION

This research successfully developed a hybrid MLR–LP framework for reducing network latency to below 10 ms in 5G network slicing specifically for low sensitive services like ride-hailing applications, where real-time responsiveness is critical to user satisfaction and operational efficiency. The results achieved feasible solutions with sub-10 ms latency, demonstrating its suitability for supporting latency-sensitive services and enhancing 5G slicing performance. The study concludes that integrating predictive modeling with optimization techniques provides both theoretical and practical contributions and that an iterative linear programming approach, guided by resource variability, is a viable method to adjust resource caps dynamically until a feasible latency solution is found. This offers a possible solution for adaptive 5G resource management and can serve as an effective decision-support tool for telecom engineers, enabling targeted resource upgrades to meet stringent latency demands while bridging a critical gap in empirical 5G latency optimization for mobility services in emerging markets.

ACKNOWLEDGMENT

I wish to express my gratitude to God Almighty for the gift of life, knowledge and provision of all the needed resources to undertake the research work. I extend my gratitude to my academic supervisor Dr. Bello O. Lawal and my co-supervisor Dr. Braimoh Abdullahi Ikharo for their strategic support and contribution to this work. Special thanks to my family and friends who have supported me financially and emotionally throughout the course of my academic journey.

REFERENCES

[1] Kamal, M.A., Raza, H.W., Alam, M.M., Su'ud, M.M. and Sajak, A.B.A.B. (2021). Resource allocation schemes for 5G network: A review. Journal of Communications and Networks. pmc.ncbi.nlm.nih.gov/articles/PMC8512213

[2] Aliu, O. G., Imran, A., Imran, M. A. and Evans, B. (2020). A survey of self organisation in future cellular networks. *IEEE Communications Surveys & Tutorials, 19*(4), pp. 297–329. doi.org/10.1109/COMST.2017.2705683

[3] 3GPP. (2020). IMT-2020: The 5G system – enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). 3GPP.

[4] Liu, K., Chen, Z., Yamamoto, T. and Tuo, L. (2022). Exploring the impact of spatiotemporal granularity on the demand prediction of dynamic ride-hailing. *ArXiv preprint*. doi.org/10.48550/arXiv.2203.10301

[5] Olawole, M. O. (2022). Ride-Hailing services in Nigeria: Adoption, insights and implications. in B.S. Farah and J.A. Odeleye (eds.), Transport Technology and Innovations in Nigeria: Policy Guides to A Sustainable Future (pp. 121–147). Nigerian Institute of Transport Technology.

[6] Ojekere, S., Ojo, O. and Mkpandiok, D. (2022). Sustainable e-hailing mobility services in Nigerian cities: Issues and policy direction. Iconic Research and Engineering Journals, 6(5), pp. 90–96. www.irejournals.com/paper-details/1703897

[7] Cats, O., Kucharski, R., Danda, S. R. and Yap, M. (2021). Beyond the dichotomy: How ride-hailing competes with and complements public transport. *Transportation Research Record*. doi.org/10.1177/03611981231171155

[8] Afolabi, I., Taleb, T., Frangoudis, P. A., Bagaa, M. and Ksentini, A. (2022). Network slicing-based customization of 5G mobile services. ArXiv preprint. arxiv.org/abs/2201.07187

[9] Fendt, A., Lohmuller, S., Schmelz, L.C. and Bauer, B. (2023) A Network Slice Resource Allocation and Optimization Model for End-to-End Mobile Networks. *IEEE 5G World Forum (5GWF)*. doi.org/10.1109/5GWF.2018.8517075

[10] Klinkowski, M., Perelló, J. and Careglio, D. (2023). Application of linear regression in latency estimation in packet-switched 5G xHaul networks. *International Conference on Transparent Optical Networks (ICTON)*. doi.org/10.1109/ICTON59386.2023.10207222

[11] Hurtado-Sánchez, J. A., Casilimas, K. and Caicedo Rendon, O. M. (2022). Deep reinforcement learning for resource management on network slicing: A survey. Sensors, 22(8), pp. 3031. doi.org/10.3390/s22083031

[12] Ejaz, N. and Choudhury, S. (2025). A comprehensive survey of linear, integer, and mixed-integer programming approaches for optimizing resource allocation in 5G and beyond networks. *ArXiv preprint*. doi.org/10.48550/arXiv.2502.15585

[13] Bikkasani, D.C. and Yerabolu, M. (2024). AI driven 5G network optimization: A

comprehensive review of resource allocation, traffic management, and dynamic network slicing. American Journal of Artificial Intelligence, 8(2), pp. 55-62. doi:10.11648/j.ajai.20240802.14

[14] Lin, J.-Y., Chou, P.-H. and Hwang, R.-H. (2023). Dynamic Resource Allocation for Network Slicing with Multi-Tenants in 5G Two-Tier Networks. *Sensors*, *23*(10), pp. 4698. doi.org/10.3390/s23104698

[15] Zhou, J., Guo, C. and Wang, J. (2021). COQRA: A correlated Q-learning based resource allocation for multi-slice wireless networks. arXiv preprint. doi.org/10.48550/arXiv.2103.05888

[16] Li, Z., Liu, C., Li, H., Li, Y. and Wen, Y. (2019). Latency optimization for VNF chains with efficient VNF placement in 5G core networks. Applied Sciences, 9(21), pp. 4540. doi.org/10.3390/app9214540

[17] Boutiba, S. and Ksentini, A. (2023). Toward a DRL-LP solution for optimizing energy consumption in 5G networks. Energies, 16(15), pp. 5799. doi.org/10.3390/en16155799