MAUN Fen Bil. Dergi., 13, 1, 70-80 Araştırma Makalesi/ Research Article
MAUN J. of Sci., 13, 1, 70-80 DOI : 10.18586/msufbd.1672775

Integrating Queuing Theory and GBM for Dynamic Resource Allocation in 6G Networks

Tuğçe BİLEN
Yapay Zeka ve Veri Mühendisliği Bölümü, Bilgisayar ve Bilişim Fakültesi, İstanbul Teknik Üniversitesi
İstanbul, Türkiye

: bilent@itu.edu.tr 0000-0001-6680-8748

Received (Geliş): 09.04.2025 Revision (Düzeltme):02.05.2025 Accepted (Kabul): 20.05.2025

ÖZ

6G ağları, önceki nesillere kıyasla kesintisiz bağlantı ve önemli gelişmeler vaat etmektedir. Küçük hücreler, kapsama alanını artırarak, gecikmeyi azaltarak ve kapasiteyi yükselterek bu başarıda kritik bir rol oynar. Ancak, bu hücrelerin yoğun şekilde konuşlandırılması, büyük durum ve eylem uzayları nedeniyle kaynak tahsisi açısından zorluklar doğurur. Binlerce hücre için kararlar alınması gerekir ve kullanıcı hareketliliğinden kaynaklanan sık yeniden tahsisler algoritmik karmaşıklığı artırır. Bu durum gecikme, veri aktarım hızı ve paket kaybı gibi temel performans metriklerini olumsuz etkiler. Bu zorluklara çözüm olarak, kuyruk kuramına dayalı hücre durum tahmini ile Gradient Boosting Machine (GBM) öngörülerini birleştiren dinamik bir kaynak tahsis modeli öneriyoruz. Bu bütünleşik yapı, bant genişliği, hesaplama kaynakları ve enerji kullanımının zamanında ve uyarlanabilir şekilde tahsis edilmesini sağlar. Simülasyon sonuçları, modelin gecikmeyi %35, paket kaybını %42 oranında azalttığını ve veri aktarım hızını %28 artırdığını göstermektedir. Bu kazanımlar, öngörüye dayalı ve uyarlanabilir kaynak yönetiminin 6G ağlarındaki etkinliğini ortaya koymaktadır.

Anahtar Kelimeler: 6G, Gradient Boosting Makineleri, Kaynak Tahsisi, Kuyruk Teorisi, Küçük Hücreler

6G Ağlarında Dinamik Kaynak Tahsisi için Kuyruk Teorisi ve GBM'nin Entegrasyonu

ABSTRACT

6G networks promise seamless connectivity and major advancements over previous generations. Small cells play a critical role by enhancing coverage, reducing latency, and increasing capacity through proximity to users. However, their dense deployment introduces challenges in resource allocation due to vast state and action spaces. Decisions must be made for thousands of cells, and frequent reallocations from user mobility add algorithmic complexity. These factors degrade key performance metrics such as latency, throughput, and packet loss. To address this, we propose a dynamic resource allocation model combining queuing-theory-based cell state estimation with Gradient Boosting Machine (GBM) predictions. This integration enables timely and adaptive allocation of bandwidth, computing resources, and energy. Simulations show the proposed model reduces latency by up to 35%, packet loss by 42%, and improves throughput by 28% over static methods. These gains highlight the effectiveness of predictive, adaptive resource management for 6G networks.

Keywords: 6G, Gradient Boosting Machines, Queuing Theory, Resource Allocation, Small Cells

INTRODUCTION

6G can enable seamless connectivity by enabling significant improvements compared to previous generations. It aims to enhance connectivity and intelligence by utilizing advanced technologies such as terahertz (THz) spectrum and AI-driven algorithms [1]. Thanks to these technologies, 6G can achieve sub-millisecond latency and terabit per second (Tbps) data rates. Accordingly, it can support innovative applications like holograms, immersive virtual reality experiences, and fully automated systems [2]. 6G small cells play a key role in this success. Although they are short-range, they improve the network by providing better coverage and faster speeds. Additionally, small cells support flexible and incremental deployment. The 6G network can provide low delays, reliable service, and higher capacity by placing many small cells close to users [3]. This architecture enhances user experience by providing seamless connectivity and satisfies the needs of future applications.

Although small cells are critical for supporting the high capacity and low latency demands of 6G networks, their dense deployment introduces unique resource allocation challenges. Here, resource allocation decisions must be computed for thousands of cells due to the huge state and action spaces of the 6G network. Also, frequent resource reallocation due to high user mobility increases the algorithmic overhead. All of these make resource management challenging. This resource management challenge affects critical performance parameters, including latency, throughput and packet loss. Therefore, the rapid increase in 6G network traffic and the proliferation of small cells necessitates a proactive and dynamic resource allocation mechanism.

Based on these, we propose a dynamic resource allocation model based on queuing theory and GBM in this paper. More specifically, we aim to develop a proactive capacity planning model that combines queuing-theory-based small cell state analysis with GBM-driven predictions to ensure dynamic resource scaling. The main contributions of the paper could be summarized as follows:

- We introduce a new methodology that combines queuing- theory-based small cell state analysis with GBM-driven predictive modelling. Queuing theory identifies critical patterns, while GBM enables future resource requirement forecasting for adaptive adjustments.
- The queuing-theory-based small cell state analysis provides a deep understanding of system utilization and traffic dynamics. These insights are transformed into actionable features for accurate and reliable resource predictions.
- The GBM-driven allocation mechanism empowers real- time resource allocation by utilizing predictive analysis. This ensures the allocation of bandwidth, computing power, and energy usage.

The rest of the paper is organized as follows: we first explain the Related Work. Then, we describe the proposed topology. Then, we give the proposed approach. After these, we provide the performance with evaluation results. Finally, we conclude the paper by also giving future work.

RELATED WORK

The literature shows different works for resource allocation management in 6G networks. The resource allocation for 6G massive machine-type communications (mMTC) is optimized to meet diverse QoS needs in [4]. It proposes a multiple numerology framework with non-orthogonal multiple access and a semi-persistent algorithm to minimize bandwidth usage. They aim to enhance support for many low-power devices in 6G networks. Here, they adjust signal formats and use a scheduling algorithm to allocate resources efficiently. In this methodology, the signal settings, channel characteristics, and quality requirements are considered. Also, they measure the resource usage compared to traditional methods. The deep reinforcement algorithm learning for intra- and inter-slice management is proposed in [5]. This work aims to improve resource sharing among different network slices in 6G. With this adaptive learning approach, they aim to manage resources within and across slices. The signal configurations, resource units, and service quality needs are used as parameters in this methodology. Also, the service delays and overall resource consumption are considered to evaluate the efficiency of the proposed model. To address the dynamic user demands of 6G UAV networks, a resource allocation algorithm is proposed in [6]. This algorithm employs decision-making based on partial network information to plan spectrum and movement. Here, area coverage, drone settings, and signal conditions are considered as algorithm parameters. This methodology evaluates improvements in data transfer speeds and learning efficiency. The resource allocation for software-defined 6G networks is optimized by dynamically switching between centralized and distributed policies to adapt to network conditions in [7]. This work uses adaptive learning to switch between centralized and distributed resource control. To balance performance in software-controlled 6G networks, this work considers performance feedback, user counts, and system priorities. As a result of these operations, this work observes balanced performance and reduced system overhead. The study aims to optimize resource allocation in 6G integrated satellite-terrestrial networks by combining base station and satellite resources to increase user throughput in [8]. It proposes a single-stage user-centric approach using a reinforcement learning model to allocate physical resource blocks and satellite beam hopping strategies. Therefore, it integrates user needs with learning-based resource allocation to increase access rate. This model uses satellite coverage, resource units, and user demand profiles. Also, to evaluate the efficiency, the number of users served and data transfer rates are assessed. A study aims to optimize power control in dense 6G in-factory subnetworks to maximize spectral efficiency while minimizing sensing and signaling overhead in [9]. This work utilizes network modeling to manage interference and optimize power use. Accordingly, this approach aims to enhance efficiency in factory-based 6G networks. For this aim, it

Table 1. Comparison of Resource Allocation Algorithms

Work	Aim	Approach	Parameters	Evaluation Results
Sadi et al. [4]	Optimize resources for machine-type communications in 6G	Uses multiple signal formats and allocation algorithm	Signal frequencies Channel size Quality needs	Reduced resource usage
Mhatre et al. [5]	Improve resource allocation for network slices in 6G	Uses reinforcement learning for slice management	Signal settings Resource blocks Service quality	Lower delays Less resource use
Zhou et al. [6]	Enhance resource use for UAVs in 6G	Uses belief-based learning for spectrum and path planning	Area size UAV settings Signal strength	Higher data rates Faster learning
Nouruzi et al. [7]	Optimize resources in software-defined 6G networks	Uses reinforcement learning with mode switching	Feedback data Performance weights User count	Better performance balance Lower overhead
Sun et al. [8]	Maximize user access in satellite-terrestrial 6G	Uses reinforcement learning for unified allocation	Satellite beams Resource blocks User demands	More users served Higher data flow
Abode et al. [9]	Improve power use in factory 6G networks	Uses neural networks for interference control	Network devices Signal data Power limits	Better efficiency Reliable performance
Wu et al. [10]	Enhance user experience in 6G	Uses quality mapping for diverse needs	Service quality User expectations Adjustment factors	Improved user satisfaction
Shen et al. [11]	Manage resources in dense IoT 6G networks	Uses learning for cell selection and interference	Signal Strength Device count Time slots	Stable performance Handles many devices
Zhang et al. [12]	Optimize resources and mobility in 6G	Uses sensing and access schemes for handover	Signal timing Movement data Service quality	Fewer connection issues Better service
Guo et al. [13]	Improve device-to-device resources in 6G	Uses distributed learning for allocation	Frequency bands Signal states Service needs	Higher data flow Less power use
Rezazadeh et al. [14]	Reduce conflicts in 6G network slices	Uses multi-agent learning for protocol design	Computing resources Traffic types	Lower delays Better resource use
Wang et al. [15]	Optimize resource allocation for 6G slices	Uses learning for slice admission and reuse	Computing units Request patterns	Better resource use Fewer failures
Zhao et al [16]	Optimize resources for user experience in 6G edge networks	Uses digital modeling and learning-based allocation	Network conditions, user demands, service quality	Improved user satisfaction, reduced congestion

focuses on device interactions, signal conditions, and power constraints. This work measures efficiency gains and reliable network performance to evaluate the proposed model. A study aims to enhance resource allocation in 6G by developing a uniform QoE-aware criterion to address multi-demand challenges in hyper-heterogeneous networks in [10]. It proposes a novel exponential mapping function integrating minimum and expected user demands to balance quality of experience and resource efficiency. More specifically, this approach maps user needs to resource decisions to meet diverse demands and improve user satisfaction in 6G networks. Here, the service quality, user expectations, and adjustment settings are used as method parameters. Also, the performance of the proposed model is evaluated by measuring enhancements in user experience and satisfaction. A study aims to optimize wireless resource management in 6G-enabled high-density IoT networks in [10]. This study supports many IoT devices in crowded 6G networks to improve system capacity and reduce interference. Specifically, it proposes a simulation-based framework using Q-learning for cell selection and static interference coordination to manage ultra-dense networks with femtocells and small cells. This framework includes signal strength, device counts, and scheduling slots. This study assesses the stable performance and ability to handle many devices. An optimization algorithm for 6G resource allocation and mobility management is proposed in [12]. It proposes a non-orthogonal multiple access scheme and a sensing-aided handover mechanism to dynamically adjust handover parameters and allocate resources. Therefore, it combines sensing and access methods to improve handovers and resource use. In this algorithm, signal timing, movement data, and service quality requirements are used as system parameters. Also, this work uses connection issues and service quality parameters for model evaluation. A study to optimize resource allocation in D2D-enabled 6G networks is proposed in [13]. It proposes a federated learning-aided deep reinforcement learning approach for decentralized resource allocation. Specifically, it applies distributed learning to allocate resources efficiently and enhance direct device-to-device connections in 6G. This learning methodology utilizes frequency settings, signal conditions, and service needs for resource allocation. Also, this work aims to improve data transfer, and reduce power consumption. A study aims to address 6G inter-slice resource contention by developing adaptive communication protocols in [14]. It uses coordinated learning to design efficient resource-sharing protocols and reduce conflicts in 6G network slices. This work also aims to reduce delays and improve resource utilization by using computing resources and network traffic type parameters. Similarly, a reinforcement learning-based slice admission control framework with periodic and shortage-triggered resource recycling is proposed in [15]. This study aims to optimize resource allocation and service provisioning in 6G network slicing to meet dynamic service demands and ensure security. For this aim, it focuses on computing resources and patterns of service requests to obtain better resource use and fewer service failures. To enhance user experience across diverse traffic types, resource allocation for 6G edge networks is optimized in [16]. The work proposes a digital twin edge network framework combined with a learning-based algorithm to manage communication resources efficiently. The aim is to support multiple user devices while ensuring high service quality. The approach integrates digital modeling and adaptive learning to adjust resource allocation dynamically. The methodology considers network conditions, user demands, and service quality needs. Performance is evaluated by measuring improved user satisfaction and reduced network congestion compared to traditional methods.

Unlike other studies, our model integrates queuing theory with GBM to create a unified framework for resource management. Most existing approaches consider traffic analysis and machine learning as independent processes. However, our model bridges this gap by directly utilizing insights from queuing theory as input features for GBM. This novel integration enables both proactive resource planning and real-time adjustments. Our model minimizes latency and packet loss by focusing on dynamic resource scaling. Therefore, it offers a scalable and robust solution for next-generation 6G networks. The strengths of our model compared to other studies can be listed as follows:

- **Integration of Mathematics and Predictive Modeling**: This study employs a traffic flow model in conjunction with a predictive tool, distinguishing it from most existing research that predominantly relies on learning-based methodologies.
- **Emphasis on Network Traffic Dynamics**: The analysis focuses on the movement of data within small cells to anticipate demand, rather than concentrating on individual devices or specific services.
- **Proactive Resource Optimization**: The approach involves preemptive planning to optimize bandwidth, computational resources, and energy consumption, in contrast to traditional reactive strategies.
- **Enhancement of User Experience**: The objective is to minimize delays and errors in congested networks, thereby providing a viable solution to the challenges faced in 6G environments.

THE PROPOSED SYSTEM ARCHITECTURE

This paper considers a knowledge-defined networking-based architecture for the dynamic 6G resource allocation. This architecture consists of three main components: data, control, and knowledge planes, as shown in Figure 1.

The Data Plane includes the physical elements of 6G small cell architecture, such as small cells, macro cells, and user equipment. It transfers the real-time data of these elements to the knowledge plane for dynamic resource allocation. To achieve this aim, the Knowledge Plane includes the Dynamic Resource Allocation Framework. This framework consists of two main stages: Queueing-Theory-Based Small Cell State Determination and GBM-based Prediction for Dynamic Resource Allocation. Here, the data received from the Data Plane goes through two main stages and turns into dynamic resource allocation decisions as actionable knowledge. We have a Control Plane between these two layers to manage dynamic resource management operations. Therefore, the Control Plane receives actionable knowledge (dynamic resource allocation decisions) from the Knowledge Plane. This knowledge is then used to allocate resources dynamically in the data plane.

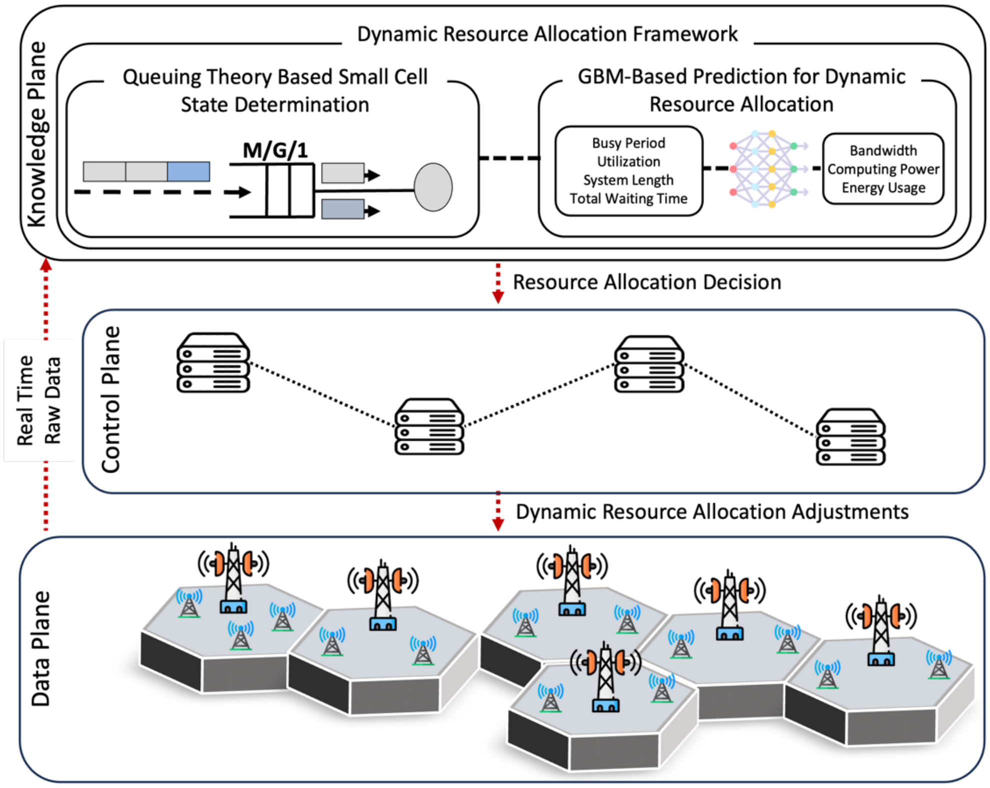


Figure 1. The Proposed System Architecture

THE PROPOSED APPROACH

This paper proposes a dynamic resource allocation model based on queuing theory and Gradient Boosting Machines (GBM). Our proposed model is executed in two main phases. We will conduct a queuing-theory-based small cell state determination in the first phase. In the second phase, we will incorporate small cell state metrics with GBM for dynamic resource allocation. Based on these results, we will dynamically adjust three primary resources (bandwidth allocation, computational power, and energy usage). The details of these phases will be summarized in the following subsections.

Queuing-Theory-Based Small Cell State Determination

In this phase, we aim to analyze network traffic patterns (busy period) and examine multiple influencing factors (utilization, system length, and total waiting time) with a queuing theory-based approach. These patterns and influencing factors are the real-time indicators of a small cell state.

To represent the operation of a small cell in a 6G network, we will utilize an *M/G/1* queue. User requests come to the small cell according to the Poisson process since it is memoryless and represents the random nature of mobile users [17], [18]. Here, we use λ_i to show the arrival rate of user requests to small cell *i*.

Additionally, service times for these requests follow a general distribution. Accordingly, the service rate ($\mu_i = \frac{1}{E[S_i]}$) shows the small cell's capability to process incoming requests. Here, $E[S_i] = \int_0^{inf} t.g(t)dt$ is the expected service time. Also, $G(s_i)$ and $g(s_i)$ represent the cumulative distribution function (CDF) and probability density function (PDF) of the service times, respectively. The small cell acts as a single server to represent its limited resources. These models and definitions form the basis for the following small cell state determination.

As explained above, we utilize busy period, utilization, system length, and total waiting time as real-time indicators to represent the small cell state. The busy period in an *M/G/1* queue refers to the continuous interval during which a small cell processes requests without idle time. Accordingly, the busy period begins when a request arrives at an idle server and terminates when the server becomes idle again. Also, new arrivals may extend the server's busy period during the service of one request.

This paper defines the busy period *B* with a recursive definition as given in Eq. 1. In this equation, *S* is the service time of the first request that initiates the busy period. *N(S)* is the number of arrivals during the service time *S*. *Bi* represents independent, identically distributed copies of the busy period initiated by those additional arrivals. Also, $\stackrel{d}{=}$ denotes equality in terms of distribution. This recursive definition provides a probabilistic framework to understand the busy period by considering the initial service time and additional arrivals.

To analyze the properties of *B* in Eq. 1, we apply the Laplace-Stieltjes Transform (LST) to the recursive equation. Here, $\tilde{B}(s)$ denotes the LST of *B*, and $\tilde{G}(s)$ denote the LST of the service time distribution G(s). Here, $\tilde{G}(s) = \int_0^{inf} e^{-sx}.g(x)dx$. Therefore, we can obtain Eq. 2. According to our small cell model, μ_i represents the processing capacity of small cell *i*, determined by the mean service time *E[Si]*. Also, *G(s)* and $G(s)$ are used to analyze system performance during different operational scenarios. This equation enables analytical solutions for properties like the mean and variance of the busy period. By using Eq. 2 and Eq. 3, we can derive the mean busy period *E[B]*, as given in Eq. 4.

$$B \stackrel{d}{=} S + \sum_{i=1}^{N(S)} B_i \qquad (1)$$

$$\tilde{B}(s) = \tilde{G}\left(s + \lambda\left(1 - \tilde{B}(s)\right)\right) \qquad (2)$$

As explained above, *E[S]* is the mean service time and $\rho = \lambda \cdot \mathbb{E}[S]$ is the utilization factor, representing the fraction of time the server is busy. Therefore, Eq. 4 represents the mean busy period by considering the impact of arrival and service rates on system performance.

$$\mathbb{E}[B] = -\frac{d\tilde{B}(s)}{ds}\Big|_{s=0} \qquad (3)$$

$$\mathbb{E}[B] = \frac{\mathbb{E}[S]}{1-\rho} \qquad (4)$$

In addition to the busy period, we have three additional metrics for the small cell state: Utilization (ρ), Expected System Length (*E[L]*), and Expected Total Waiting Time (*E[W]*) as given in Eq. 5. Here, the total waiting time includes the time spent in the queue and being serviced. Similarly, the total system length includes requests waiting in the queue and currently being served [19].
These metrics improve the overall understanding of a small cell state. More specifically, the busy period gives information about the base station's continuous active operational time. Additionally, we can assess the user experience regarding latency and the number of waiting requests by using other metrics.

$$\begin{aligned} E[L] &= \frac{\lambda \cdot E[S^2]}{2(1-\rho)} + \rho \\ E[W] &= \frac{\rho E[S]}{1-\rho} + \frac{\lambda \cdot \mathrm{Var}[S]}{2\mu(1-\rho)} \\ \rho &= \frac{\lambda}{\mu} \end{aligned} \qquad (5)$$

Therefore, when the small cell values go to the Knowledge Plane, they are modelled as *M/G/1* queues, and their states are determined. The GBM in the knowledge plane uses these states to obtain predictions for dynamic resource allocation.

GBM-Based Prediction for Dynamic Resource Allocation

GBM uses the small cell states to forecast resource demands. These demands are used to make decisions about resource scaling. Thus, in this phase, we leverage predictions generated by the GBM model to allocate resources dynamically. Here, we select busy period (Eq. 4), utilization (Eq. 5), system length (Eq. 5), and waiting time (Eq. 5) as input features for GBM. Therefore, we train the GBM to predict the allocations of our primary resources based on these features. Here, we define bandwidth allocation, computing power, and energy usage as primary resources.
In the beginning, our model is trained with the values of input features and corresponding resource allocation decisions. Here, GBM learns patterns from the data to dynamically predict the optimal allocation of resources. Accordingly, our model learns how the input metrics influence the resource allocation decisions by building a series of decision trees. These trees can learn the following patterns about small cells.

- Relationships between individual small cell state metrics and resource allocations. As an example, a high value of $E[B]$ indicates increased user demand. Accordingly, we should increase bandwidth to satisfy this demand.
- Interactions between small cell state metrics and resource allocations. For example, a higher value of $E[L]$ with lower ρ states that the server is not processing requests rapidly despite the high system length. Accordingly, we should increase the bandwidth to process the queued user requests faster.
- Nonlinear behavior states of metrics. For example, a value of $\rho \approx 1$ indicates that the system is approaching saturation in terms of load. So, we should increase the computing power. Table 1 gives a detailed version of these patterns.

Therefore, as given in Table 2, we aim to dynamically determine the adjustments of three primary resources: Bandwidth Allocation, Compute Power, and Energy Usage. To achieve this, we give small cell state parameters to our GBM model after training. At that point, GBM uses the learned relationships to predict optimal values for resource allocations, as shown in Table 1. At that point, the Knowledge Plane transfers the actionable knowledge (dynamic resource allocation decisions) to the Control Plane. The Control Plane uses this knowledge to allocate resources dynamically in the data plane. Therefore, based on the predictions, the small cell dynamically adjusts its resources [20], [21].
In the following part, we will evaluate the performance of our proposed model in terms of latency, throughput, and packet loss.

Table 2. Conditions Learned by the GBM and Corresponding Resource Allocation Decisions

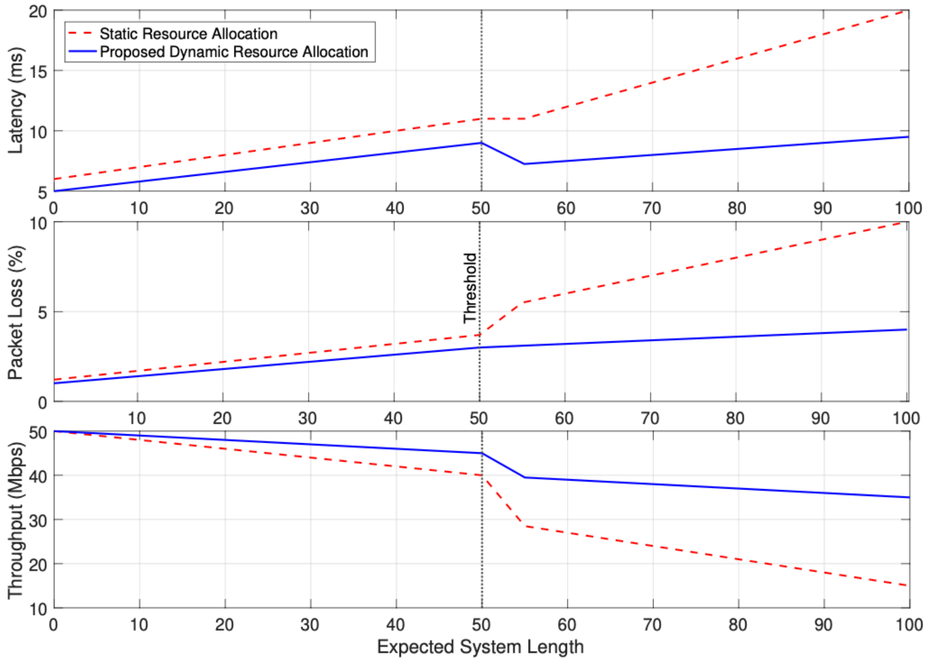
Condition Type	Small Cell State Metrics	Description	Resouce Allocation Decision
Individual Metric	High $E[B]$	Continuous activity, high demand	Increase bandwidth
	High $E[W]$	Long delays, high latency	Increase bandwidth
	High $E[L]$	Large queue, congestion	Increase compute power
	Low ρ	Underutilized server	Reduce energy usage
Metric Interaction	High $E[L]$ Low ρ	Queued requests, slow processing	Increase bandwidth Compute power
	Low $E[B]$ Low ρ	Low demand, idle system	Reduce energy usage
	High $E[W]$ High ρ	Delays, utilization near overload	Increase bandwidth Compute power
	Moderate $E[B]$ High ρ	Consistently busy state	Bandwidth Compute power increase
	High $E[B]$ Moderate $E[W]$	Stable workload, increasing activity	Increase bandwidth
Nonlinear Behavior	$\rho \approx 1$	Saturation, maximum capacity	Rapid compute power increase
	High $E[W]$ $\rho \approx 1$	Long delays, overload	Increase bandwidth Compute power
	Spike in $E[B]$	Abrupt demand surge	Temporarily boost bandwidth
	Drop in ρ High $E[L]$	Idle processing with queue backlog	Increase compute power
	Low ρ Low $E[L]$ High $E[B]$	Short bursts of activity	Maintain moderate bandwidth Reduce energy slightly

PERFORMANCE EVALUATION

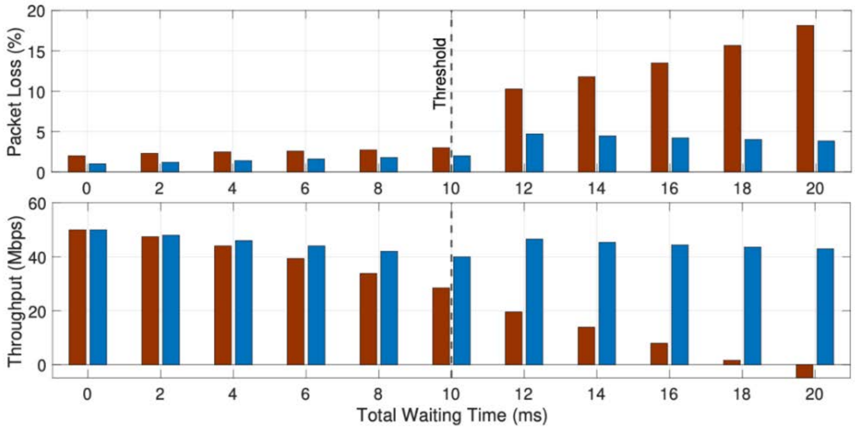
This paper uses the NS-3 network simulator to assess the performance of a dynamic resource allocation framework in a 6G environment. The simulation takes place in a 1000-meter by 1000-meter area, representing a dense urban deployment scenario. Within this space, macro cells are strategically placed to provide consistent, wide-area coverage and serve as the main connection backbone. To increase network capacity and alleviate the load on macro cells, small cells are deployed in a semi-random grid pattern. These small cells help densify the network and address coverage gaps, particularly in areas where macro connectivity may weaken due to obstacles or high user density. All base stations, including macro and small cells, remain stationary during the simulation to illustrate fixed network infrastructure. Users are randomly positioned within the simulation area to reflect the unpredictable nature of user distribution in real-world urban settings. Their movement follows the Random Waypoint Model, which simulates realistic user behavior. In this model, each user chooses a random destination within the area and moves toward it at a randomly selected speed, pausing briefly upon arrival before picking a new destination and repeating the process. This variability in speed and direction captures the unpredictable dynamics of human movement, making the evaluation more representative of actual 6G network usage. During the simulation, user devices connect to the nearest available cell, and handovers are managed based on signal strength and quality. The dynamic resource allocation mechanism continuously optimizes the assignment of network resources in response to changes in user density, traffic demand, and cell load conditions. A comprehensive list of the simulation parameters, including channel configurations, user behavior settings, and radio propagation models, is detailed in Table 3.

Table 3. Simulation Parameters

Parameter	Values
Frequency Band	28 GHz
Reception Power	0.6 Watts
Coverage Range	80 meters
Modulation Scheme	256-QAM
Antenna Gain	6 dBi
Transmission Power	1.8 Watts
MIMO Configuration	8x8
Carrier Frequency	28 GHz
Backhaul Capacity	2 Gbps
Scheduler Algorithm	Proportional Fair
Energy Efficiency	10 mW/bit
Network Density	50 users/cell



(a)



(b)

Figure 2. Expected System Length- and Total Waiting Time-Based Evaluation

In these simulations, we compare the proposed dynamic resource allocation approach with the static resource allocation methodology. Figure 2a illustrates the behavior of the proposed approach according to increasing system lengths ($E[L]$) to show its efficiency compared to the static resource allocation strategy. Here, we evaluate the efficiency in terms of latency, packet loss, and throughput. As shown in Figure 2a, latency and packet loss grow steadily with increasing system length due to prolonged queuing and saturation. Also, throughput declines with the degraded performance due to congestion blocks. Our proposed dynamic resource allocation mitigates these issues effectively after reaching the threshold. At that point, packet loss decreases due to better handling of incoming traffic. As shown in Figure 2a, latency becomes steady when more resources are added to lower waiting times. Throughput improves when network resources are adjusted for the workload.

We also analyze the performance of the proposed model according to the total waiting time, as shown in Figure 2b. We measure this metric with Eq. 5, and high waiting times indicate queuing delays due to congestion or under provisioned resources. In static resource allocation, packet loss continuously increases due to excessive buffering or the dropping of requests during prolonged waiting periods. In our proposed approach, we can improve performance after the threshold. More specifically, our proposed approach reduces packet loss by mitigating the negative impacts of traffic congestion. Similarly, throughput declines since the system cannot efficiently process incoming traffic within an acceptable timeframe. On the other hand, our proposed approach effectively handles queued traffic and reduces system delays to improve throughput. We evaluate the proposed approach's latency, packet loss, and throughput according to the busy period in Figure 3. As explained in Eq. 1, analyzing busy periods is essential to

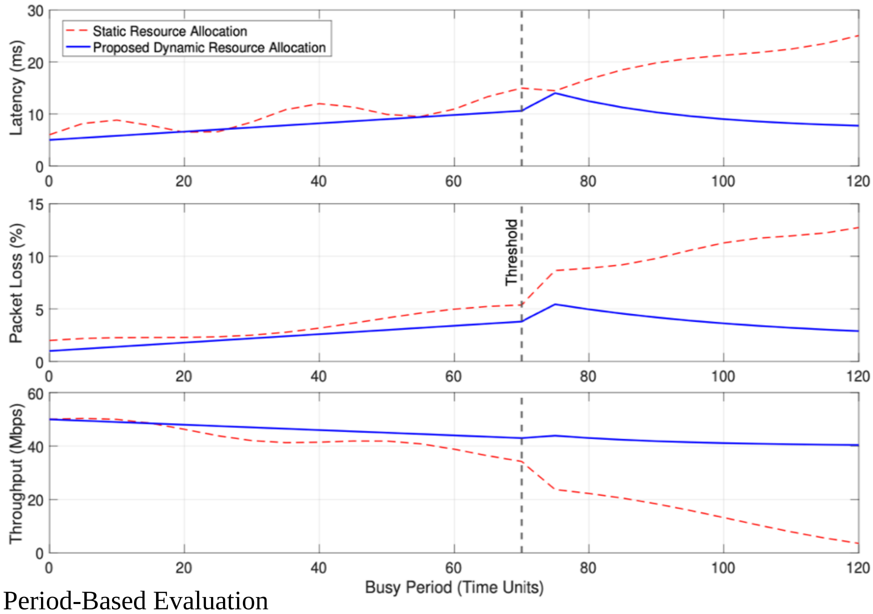


Figure 3. Busy Period-Based Evaluation

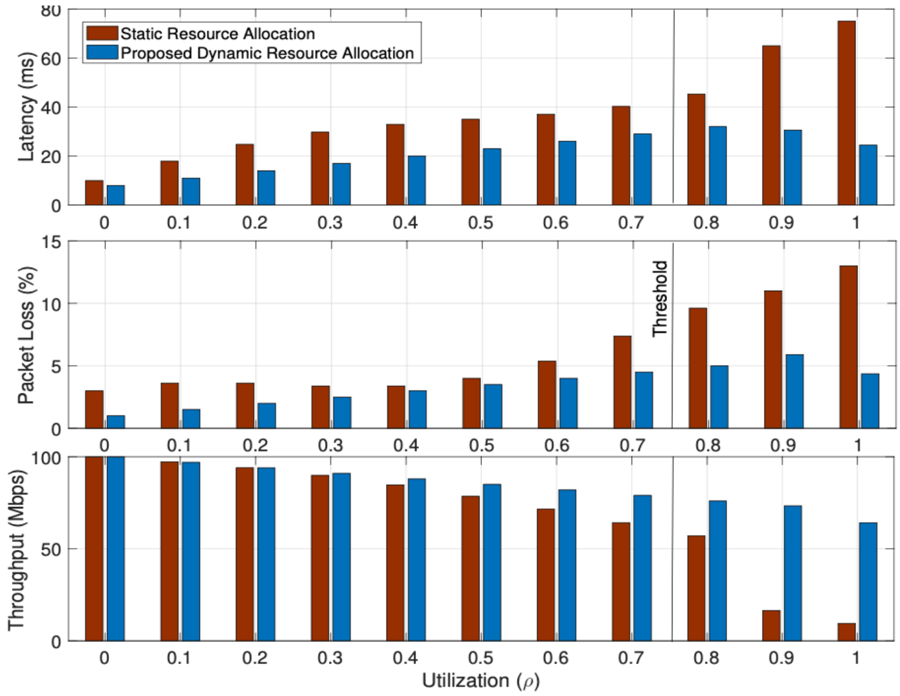


Figure 4. Utilization-Based Evaluation

understanding system overload and applying dynamic resource strategies. As shown in Figure 3, latency increases during long busy times because resources can't handle the continuous high demand. However, the proposed dynamic resource allocation approach shows improvements post-threshold. Here, dynamic resource allocation reduces prolonged queuing delays, leading to a rapid latency decrease. As shown in Figure 3, packet loss increases considerably during extended busy periods with throughput reduction. This indicates that the system's capacity to handle all incoming traffic has decreased. However, the proposed approach stabilizes it due to enhanced congestion handling as given with Eq. 5. Similarly, throughput increases as more resources optimize data processing during peak times, as summarized in Table 2.

To provide a clearer comparison of our approach with existing solutions, we offer a qualitative assessment of various advanced resource allocation strategies in Table 4. Unlike static or purely heuristic methods, our model combines queuing theory and GBM-based prediction to support both traffic adaptation and user mobility. While many learning-based models, such as Q-learning or deep reinforcement learning, show strong performance, they often require extensive training and centralized control, which may not be suitable for scaling in 6G networks. In contrast, our approach is lightweight and scalable, ensuring real-time responsiveness with predictive capabilities. This balance makes our method a practical

Table 4. Qualitative Comparison of Resource Allocation Approaches

Study / Approach	Adaptability to Traffic	Mobility Awareness	Learning-Based	Real-Time Capability	Scalability	Notes
Static Resource Allocation	✗	✗	✗	✗	✓	Simplest; lacks responsiveness
Heuristic Rule-Based Models [4, 12]	~	✗	✗	✓	✓	Scenario-tuned; limited generalization
Q-Learning Models [8, 9]	✓	~	✓	~	~	Requires training phase; limited by mobility complexity
Deep RL Approaches [5, 6, 11]	✓	✓	✓	✓	~	High accuracy but computationally demanding
Federated Learning-Based Models [13]	✓	✓	✓	✓	✓	Scalable and privacy-preserving; model synchronization overhead
Our Proposed Method	✓	✓	✗	✓	✓	Intelligent resource allocation without complex training

and advanced alternative, particularly for latency-sensitive and dynamic environments like smart warehouses or digital twin platforms.

The simulation results reveal notable improvements in latency reduction, resource utilization efficiency, and throughput stability, especially under conditions of fluctuating user mobility and variable traffic loads. These enhancements are particularly relevant for real-world applications that require ultra-reliable low-latency communication, such as smart warehouses, Industry 4.0 settings, autonomous transportation systems, and real-time digital twin platforms in logistics and manufacturing. In these contexts, tasks like the precise coordination of autonomous guided vehicles, remote robotic control, and time-sensitive sensor-actuator interactions depend on reliable network performance, which our proposed model effectively supports. By facilitating predictive and adaptive resource allocation, the system can proactively address network congestion and user mobility patterns, minimizing service disruptions and ensuring stable quality of service. For example, in smart logistics hubs where numerous interconnected devices operate simultaneously, the dynamic nature of our model ensures continuous communication without over-provisioning or underutilizing resources. This approach not only enhances operational efficiency but also promotes energy savings by reducing idle resource allocation and unnecessary signaling overhead. Furthermore, the improved spectral efficiency and reduced packet delays highlighted in our evaluation indicate tangible economic advantages, including lower energy costs, better hardware utilization, and scalability for dense network deployments. These outcomes align with the objectives of sustainable and intelligent 6G infrastructure. Overall, the findings of this study suggest that intelligent, model-driven resource orchestration mechanisms are not only theoretically beneficial but also essential for meeting the rigorous performance demands of next-generation cyber-physical systems.

CONCLUSION

We have presented an integrated framework for dynamic resource allocation in 6G small cell networks. This approach combines the M/G/1-based queueing system with GBM. This approach obtains the small cell states with M/G/1-based queue modelling. Then, these states are used with the GBM to predict the resource allocation decisions. This combined approach allows for adaptable resource management by ensuring the dynamic allocation of bandwidth, computing power, and energy usage. Simulation results also show that the proposed dynamic resource allocation model significantly outperforms static allocation strategies. Specifically, the model achieves a reduction of up to 35% in latency, a decrease of up to 42% in packet loss, and an improvement of up to 28% in throughput under high traffic and congested conditions. These enhancements demonstrate the effectiveness of our predictive and adaptive framework in efficiently managing resources in dense 6G small cell deployments.

FUTURE WORK

While the proposed queuing theory and GBM-based dynamic resource allocation framework shows significant improvements in latency, packet loss, and throughput, several extensions could further enhance its applicability and performance in real-world 6G scenarios. The current implementation treats small cells as independent decision-making units. Future work may explore federated learning or cooperative reinforcement learning to facilitate resource sharing and joint optimization among neighboring cells. As energy harvesting small cells become more common, future work could also incorporate models of renewable energy availability to enable energy-aware resource allocation while ensuring QoS guarantees.

REFERENCES

[1] M. Shafi, R. K. Jha, and S. Jain, 6g: Technology evolution in future wireless networks, IEEE Access, vol. 12, pp. 57 548–57 573, 2024.
[2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, Toward 6g networks: Use cases and technologies, IEEE Communications Magazine, vol. 58, no. 3, pp. 55–61, 2020.
[3] S. Gopi, S. Kalyani, and L. Hanzo, Cooperative 3d beamforming for small-cell and cell-free 6g systems, IEEE Transactions on Vehicular Technology, vol. 71, no. 5, pp. 5023–5036, 2022.
[4] Y. Sadi, S. Erkucuk, and E. Panayirci, Flexible physical layer-based resource allocation for machine type communications towards 6g, in 2020 2nd 6G Wireless Summit (6G SUMMIT), 2020, pp. 1–5.
[5] S. Mhatre, F. Adelantado, K. Ramantas, and C. Verikoukis, Aiaas for oran-based 6g networks: Multi-time scale slice resource management with drl, in ICC 2024 - IEEE International Conference on Communications, 2024, pp. 5407–5412.
[6] F. Zhou, R. Ding, Q. Wu, D. W. K. Ng, K.-K. Wong, and N. Al-Dhahir,A partially observable deep multi-agent active inference framework for resource allocation in 6g and beyond wireless communications networks, in GLOBECOM 2023 - 2023 IEEE Global Communications Conference, 2023, pp. 2662–2667.
[7] A. Nouruzi, A. Rezaei, A. Khalili, N. Mokari, M. R. Javan, E. A. Jorswieck, and H. Yanikomeroglu, Smart resource allocation model via artificial intelligence in software defined 6g networks, in ICC 2023 - IEEE International Conference on Communications, 2023, pp. 5141–5146.
[8] W. Sun, H. Xu, H. Wang, S. Chang, S. Sun, and D. Miao, Research on user-centric wireless resource allocation of is tn based on reinforcement learning, in 2023 IEEE Globecom Workshops (GC Wkshps), 2023, pp. 141–146.
[9] D. Li, S. R. Khosravirad, T. Tao, P. Baracca, and P. Wen, Power allocation for 6g sub-networks in industrial wireless control, in 2024 IEEE Wireless Communications and Networking Conference (WCNC), 2024, pp. 1–6.
[10] M. Wu, Y. Xiao, Y. Gao, and X. Lei, Design of quality-of-experience criteria for resource allocation toward 6g wireless networks: A review and new directions, in 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), 2022, pp. 1–7.
[11] X. Shen, W. Liao, and Q. Yin, “A novel wireless resource management for the 6g-enabled high-density internet of things,” IEEE Wireless Communications, vol. 29, no. 1, pp. 32–39, 2022.
[12] H. Zhang, Y. Zhang, X. Liu, K. Sun, and Y. Zhang, “Resource allocation and mobility management for perceptive mobile networks in 6g,” IEEE Wireless Communications, vol. 31, no. 4, pp. 223–229, 2024.
[13] Q. Guo, F. Tang, and N. Kato, “Federated reinforcement learning-based resource allocation in d2d-enabled 6g,” IEEE Network, vol. 37, no. 5, pp. 89–95, 2023.
[14] F. Rezazadeh, H. Chergui, S. Siddiqui, J. Mangues, H. Song, W. Saad, and M. Bennis, “Intelligible protocol learning for resource allocation in 6g o-ran slicing,” IEEE Wireless Communications, vol. 31, no. 5, pp. 192–199, 2024.
[15] J. Wang, Y. Li, J. Liu, and N. Kato, “Intelligent network slicing for b5g and 6g: Resource allocation, service provisioning, and security,” IEEE Wireless Communications, vol. 31, no. 3, pp. 271–277, 2024.
[16] J. Zhao, Y. Chen, and Y. Huang, “Qoe-driven wireless communication resource allocation based on digital twin edge network,” IEEE Journal of Radio Frequency Identification, vol. 8, pp. 277–281, 2024.
[17] T. Bilen, K. Ayvaz, B. Canberk, “Qos-based distributed flow management in software defined ultra-dense networks,” Ad Hoc Networks, vol. 78, pp. 24-31, 2018.
[18] T. Bilen and B. Canberk, "Handover-Aware Content Replication for Mobile-CDN," in IEEE Networking Letters, vol. 1, no. 1, pp. 10-13, March 2019, doi: 10.1109/LNET.2018.2873982.
[19] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, Fundamentals of Queueing Theory, 4th ed. USA: Wiley-Interscience, 2008.
[20] L. V. Cakir, T. Bilen, M. Özdem and B. Canberk, "Digital Twin Middleware for Smart Farm IoT Networks," 2023 International Balkan Conference on Communications and Networking (BalkanCom), İstanbul, Turkiye, 2023, pp. 1-5, doi: 10.1109/BalkanCom58402.2023.10167962.
[21] T. Bilen, E. Ak, B. Bal and B. Canberk, "A Proof of Concept on Digital Twin-Controlled WiFi Core Network Selection for In-Flight Connectivity," in IEEE Communications Standards Magazine, vol. 6, no. 3, pp. 60-68, September 2022, doi: 10.1109/MCOMSTD.0001.2100103.