

Research Article

AI-Driven 5G Network Optimization: A Comprehensive Review of Resource Allocation, Traffic Management, and Dynamic Network Slicing

Dileesh Chandra Bikkasani^{1,*} , Malleswar Reddy Yerabolu² 

¹Department of Technology Management, University of Bridgeport, Bridgeport, USA

²Independent Researcher, North Carolina, USA

Abstract

The rapid advancement of 5G networks, coupled with the increasing complexity of resource management, traffic handling, and dynamic service demands, underscores the necessity for more intelligent network optimization techniques. This paper comprehensively reviews AI-driven methods applied to 5G network optimization, focusing on resource allocation, traffic management, and network slicing. Traditional models face limitations in adapting to the dynamic nature of modern telecommunications, while AI techniques—particularly machine learning (ML) and deep reinforcement learning (DRL)—offer scalable and adaptive solutions. These approaches facilitate real-time optimization by learning from network conditions, predicting traffic patterns, and managing resources intelligently across virtual network slices. The integration of AI into 5G networks enhances performance, reduces latency, and ensures efficient bandwidth utilization, which is essential for supporting emerging applications such as the Internet of Things (IoT), autonomous systems, and augmented reality. Furthermore, this paper highlights key AI techniques and their applications to 5G challenges, illustrating their potential to drive future innovations in network management. By laying the groundwork for autonomous network operations in 6G and beyond, this research emphasizes the transformative impact of AI on telecommunications infrastructure and its role in shaping the future of connectivity.

Keywords

5G, Telecommunication, Wireless Communication, Artificial Intelligence, Network Performance

1. Introduction

With the evolution of wireless communication came significant advancements in the telecommunications space, with data demand increasing 1000-fold from 4G to 5G [1]. Each new generation has addressed the shortcomings of its predecessors, and the advent of the 5th generation of wireless network (5G) technology, in particular, promises unprece-

dented data speeds, ultra-low latency, and multi-device connectivity. The new 5G-NR (New Radio) standard is categorized into three distinct service classes: Ultra-Reliable Low-Latency Communications (URLLC), massive Machine-Type Communications (mMTC), and enhanced Mobile Broadband (eMBB). URLLC aims to provide highly reliable

*Corresponding author: dbikkasa@my.bridgeport.edu (Dileesh Chandra Bikkasani)

Received: 24 October 2024; Accepted: 9 November 2024; Published: 28 November 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

and low-latency connectivity; eMBB focuses on increasing bandwidth for high-speed internet access; and mMTC supports many connected devices, enabling IoT on a massive scale [2]. Optimizing 5G performance is crucial for emerging applications such as autonomous vehicles, multimedia, augmented and virtual realities (AR/VR), IoT, Ma-

chine-to-Machine (M2M) communication, and smart cities. Built on technologies like millimeter-wave (mmWave) spectrum, massive multiple-input multiple-output (MIMO) systems, and network function virtualization (NFV) [3], 5G promises to revolutionize many industries.

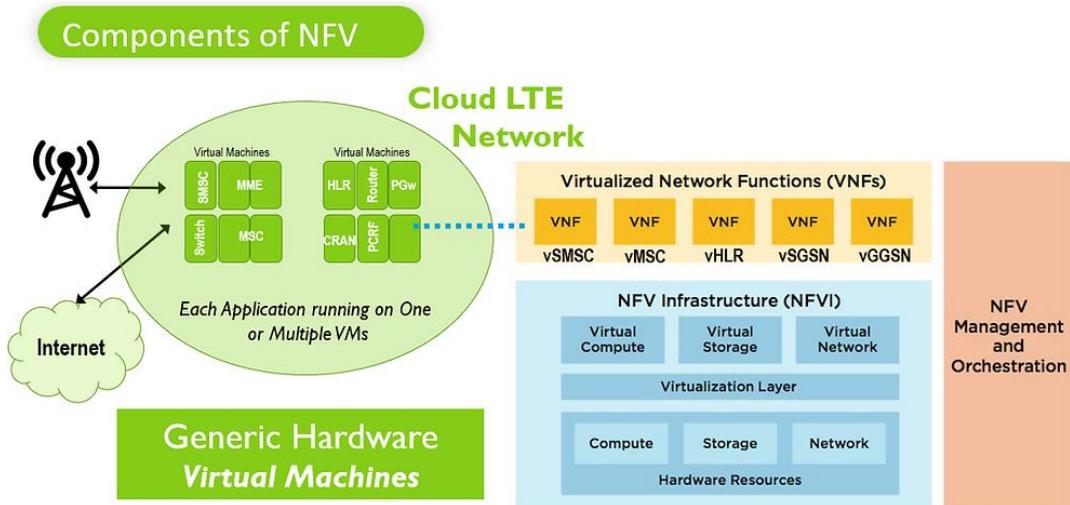


Figure 1. Components of NFV.

Figure 1 illustrates the components of Network Function Virtualization (NFV), a key enabler for 5G. NFV decouples network functions from proprietary hardware, allowing these functions to run as software on standardized hardware. By virtualizing network functions—such as firewalls, load balancers, and gateways—NFV supports dynamic and scalable network management, making it easier to allocate resources flexibly across different network slices and use cases. This flexibility is critical in managing the growing demands of 5G applications, where real-time adaptability and resource optimization are paramount. The significance of NFV lies in its ability to decouple hardware from software, allowing network operators to deploy new services and scale existing ones more efficiently. For example, in a 5G network, operators can allocate resources dynamically to different virtual network functions (VNFs), optimizing for the specific needs of applications such as autonomous vehicles or telemedicine, which demand high reliability and low latency. **Figure 1** showcases the architectural elements of NFV, including the virtualization layer, hardware resources, and the management and orchestration functions that control resource allocation and scaling. NFV plays a pivotal role in enabling network slicing, a critical feature of 5G, which allows operators to create virtual networks tailored to specific application requirements.

However, the complexity and heterogeneity of 5G networks present several challenges, including quality of service (QoS) provisioning, resource management, and network optimization. As 5G networks scale, traditional rule-based ap-

proaches to network management become inadequate. Efficient resource allocation, traffic management, and dynamic network slicing [4] are necessary to handle demanding use cases without compromising speed or reliability. Additionally, with the increase in mobile traffic flow, meeting customer demands on time requires addressing the allocation of bandwidth for heterogeneous cloud services [5].

Network resource allocation (RA) in 5G networks plays a critical role in optimizing the efficient utilization of spectrum, computing power, and energy to meet the demands of modern wireless communication. Resource allocation is pivotal for data-intensive applications, IoT devices, and emerging technologies like AV and AR. It ensures these technologies receive adequate network resources, enhancing overall performance and QoS in dynamic and heterogeneous environments. Traditional resource allocation relies on channel state information (CSI), which necessitates significant overhead costs, thereby increasing the overall expense of the process [6]. Section 2 focuses on various AI techniques applied to resource management, highlighting their impact on network slicing, energy efficiency, and overall quality of service (QoS). By leveraging reinforcement learning (RL), optimization methods, and machine learning (ML) models, these advanced strategies address the dynamic and complex requirements of 5G networks, providing adaptive and intelligent solutions for enhanced connectivity and sustainability.

Network slicing allows creating multiple virtual networks on top of a shared physical infrastructure, each optimized for specific service requirements. Network slices can be inde-

pendently configured to support diverse applications with varying performance needs, such as low-latency communication, massive device connectivity, or high-throughput data services. Using network slicing, 5G can provide tailored experiences for different user types while maximizing the use of network resources. This capability is crucial for emerging use cases like smart factories, telemedicine, and autonomous systems, where performance requirements can differ significantly across applications. The rest of the paper's sections are structured as follows: The first section covers the general understanding of the topic and defines why AI is necessary; the second contains the traditional AI-ML methods used in current spaces; and the third section consists of the DL-based techniques.

2. Data Science and AI Techniques for Resource Allocation

AI has shown promising results in resource allocation through continuous learning and adaptation to network changes. Unlike traditional mathematical model-based paradigms, Reinforcement Learning (RL) employs a data-driven model to efficiently allocate resources by learning from the network environment, thereby improving throughput and latency [7]. However, achieving fully distributed resource allocation is not feasible due to virtualization isolation constraints in each network slice [8].

In 4G LTE and LTE Advanced (LTE/A) networks, IP-based packet switching primarily focuses on managing varying user numbers in a given area. By employing ML-based MIMO for channel prediction and resource allocation, these technologies enhance CSI accuracy through data compression and reduced processing delays while adapting to channel rates despite imperfect CSI [9]. However, this technique proves inefficient due to the complexity and traffic load of 5G networks. Traditional resource allocation using CSI struggles with system overhead, which can reach up to 25% of total system capacity, making it suboptimal for 5G Cloud Radio Access Network (CRAN) applications. The conventional method also falters with an increasing number of users [10].

Traditional optimization techniques include using an approximation algorithm to connect end-users with Remote Radio Heads (RRH). This algorithm estimates the number of end-users linked to each RRH and establishes connections between end-users, RRHs, and Baseband Units (BBU) [11]. Challenges such as millimeter-wave (mmWave) beamforming and optimized time-delay pools using hybrid beamforming for single and multiple-user scenarios in 5G CRAN networks have also been explored [12]. In another instance, a random forest algorithm was proposed for resource allocation, where a system scheduler validates outputs from a binary classifier; although robust, further research and development are necessary [13].

Many Deep Reinforcement Learning (DRL) techniques have been applied for network slicing in 5G networks, allowing dynamic resource allocation that enhances throughput and latency by learning from the network environment. DRL-based approaches can handle the complexity and overhead issues of traditional centralized resource allocation methods [14]. Balevi and Gitlin (2018) proposed a clustering algorithm that maximizes throughput in 5G heterogeneous networks by utilizing machine learning techniques to improve network efficiency and adjust resource allocation based on real-time network conditions [15]. A Graph Convolution Neural Network (GCN) resource allocation scheme was explored, addressing the problem of managing resource allocations effectively. Optimization techniques such as heuristic methods and Genetic Algorithms (GAs) are being explored to solve resource allocation problems efficiently by minimizing interference and maximizing spectral efficiency. Genetic algorithms, for instance, utilize evolutionary principles like selection, crossover, and mutation to evolve solutions toward optimal resource allocation configurations. Heuristic methods like simulated annealing and particle swarm optimization (PSO) are employed to further enhance resource management. The integration of AI-driven algorithms, such as RL and DRL, into 5G networks enables real-time, adaptive resource allocation based on changing network conditions and user demands, significantly improving network performance and efficiency [16]. By utilizing AI-driven optimization, 5G networks can achieve higher efficiency and better manage the interplay between different network elements, ensuring seamless connectivity and high performance.

AI and machine learning techniques are revolutionizing resource allocation in 5G networks. By shifting from traditional models to adaptive, data-driven approaches like RL and DRL, these technologies can significantly enhance network throughput, reduce latency, and efficiently manage system overhead. As traditional methods struggle with rising complexity and user demand, AI-driven optimization provides dynamic solutions that adapt to real-time network conditions, enabling more efficient and effective resource management in the 5G era and beyond.

3. Data Science and AI in Traffic Management

Integrating AI technologies in 5G network traffic management aims to achieve traffic volume prediction, enhance real-time computational efficiency, and ensure network robustness and adaptability to fluctuating traffic patterns. Mobile data traffic is anticipated to grow significantly, with 5G's share increasing from 25 percent in 2023 to around 75 percent by 2029 [17]. This growth and the complexity of 5G networks necessitate advanced techniques for efficient traffic prediction, crucial for optimizing resource allocation and ensuring net-

work reliability. Machine learning (ML) offers diverse applications in network traffic management, from predicting traffic volumes and identifying security threats to optimizing traffic engineering (TE). These capabilities enable proactive network monitoring, enhanced security measures, and improved traffic flow management, leading to efficient and resilient network operations [18]. By combining time, location, and frequency information, researchers have identified five basic time-domain patterns in mobile traffic, corresponding to different types of urban areas such as residential, business, and transportation hubs [19].

Traditional models like ARIMA have been widely used for seasonality because they can model temporal dependencies in time series data. The ARIMA model combines autoregression (AR), integration (I), and moving average (MA) components to predict future values based on past observations [20]. Variations such as seasonal ARIMA (SARIMA) use spectrum analysis to describe traffic patterns and predict parameter estimation using maximum likelihood methods [21]. The seasonal ARIMA (SARIMA) model algorithm describes a procedure for fitting seasonality to traffic data, determining seasonality periods through spectrum analysis, estimating differencing parameters, and identifying model orders using information criteria that can predict the parameter estimation using maximum likelihood methods. Despite their effectiveness, ARIMA and SARIMA often struggle with the non-linear and complex traffic patterns characteristic of 5G networks. Machine learning models, including Support Vector Machines (SVM) and Random Forests, have been implemented to overcome these limitations. SVMs capture non-linear relationships, particularly in their regression form (SVR) [22]. The Random Forest algorithm constructs decision trees by training each one on randomly selected data points and features, improving prediction accuracy and robustness [23]. This approach builds multiple decision trees and merges them to improve prediction accuracy and robustness, effectively handling the heterogeneity of 5G network traffic. Support Vector Machines focus on maximizing the distance from the separating plane to the nearest data points, known as support vectors, using dot products and kernel functions. This approach enables faster training than methods like Bagging and Random Forest, which require using the entire dataset.

Deep learning (DL) has revolutionized many facets of network traffic management, including traffic prediction, estimation, and smart traffic routing. Models like Long Short-Term Memory (LSTM) networks are particularly effective because they can capture and learn long-term dependencies in sequential data, making them highly suitable for predicting network traffic. DL also presents promising alternatives for interference management, spectrum management, multi-path usage, link adaptation, multi-channel access, and traffic congestion [24]. For instance, an AI scheduler using a neural network with two fully connected hidden layers can reduce collisions by 50% in a wireless

sensor network of five nodes [25]. Advanced techniques such as transformer-based architectures leverage self-attention mechanisms to efficiently process vast amounts of data [26]. DRL techniques have been proposed to schedule high-volume flexible traffic (HVFT) in mobile networks. This model uses deep deterministic policy gradient (DDPG) reinforcement learning to learn a control policy for scheduling IoT and other delay-tolerant traffic, aiming to maximize the amount of HVFT traffic served while minimizing degradation to conventional delay-sensitive traffic [27].

Several studies highlight innovative applications of DRL and AI frameworks in this context. One study introduced a DRL approach for decentralized cooperative localization scheduling in vehicular networks [28]. An AI framework using CNN and RNN enhanced throughput by approximately 36%, though it incurred high training time and memory usage costs [29]. Another DRL model based on LSTM enabled small base stations to dynamically access unlicensed spectrum and optimize wireless channel selection [30]. A DRL approach for SDN routing optimization achieved configurations comparable to traditional methods with minimal delays [31]. Work on routing and interference management, often reliant on costly algorithms like WMMSE, has advanced by approximating these algorithms with finite-size neural networks, demonstrating significant potential for improving Massive MIMO systems.

In summary, integrating AI technologies into 5G network traffic management offers significant advancements in multiple facets, such as traffic prediction, resource allocation, and network management. Techniques such as ML and DL using models like LSTM and advanced frameworks utilizing CNN, RNN, and DRL address the complex and dynamic nature of 5G networks. AI-driven solutions improve network efficiency and reliability by enhancing interference management, spectrum access, and routing capabilities and adapting to varying traffic patterns and demands. These innovations highlight the transformative potential of AI in achieving robust, adaptive, and efficient 5G network operations, paving the way for future research and development in this critical field.

4. Network Slicing in 5G: Data Science and AI Approaches

Network slicing is one of 5G's most transformative features. It enables the partitioning of a single physical network into multiple virtual networks, each modified and adjusted to meet specific service requirements. These network slices can be dynamically created, modified, and terminated to optimize resources for various applications, ranging from massive IoT deployments to ultra-reliable low-latency communications (URLLC). The challenge lies in managing the complexity of creating and maintaining these slices in real time, a task where Artificial Intelligence (AI) plays a crucial role.

AI technologies are increasingly being adopted in 5G to

automate dynamic network slicing. The traditional manual approach to network management is insufficient for handling the large-scale, highly heterogeneous environments enabled by 5G. AI, particularly machine learning (ML), offers advanced capabilities in real-time decision-making, predictive analytics, and adaptive control, which are critical for the efficient deployment and management of network slices.

AI models predict traffic patterns, analyze network conditions, and dynamically adjust resource allocation to meet the specific needs of each slice. This ensures that slices maintain optimal performance, even under fluctuating traffic and varying service demands. Reinforcement learning (RL) and deep learning (DL) algorithms are frequently used to handle the complex decision-making processes required for slice orchestration. These algorithms can autonomously learn from network data, optimize resources, and balance loads between slices without human intervention.

AI-driven resource allocation plays a critical role in the success of network slicing. Each network slice may have distinct bandwidth, latency, and reliability requirements, making it necessary to allocate resources dynamically. AI can help predict and pre-allocate resources based on historical data and real-time network traffic patterns. For instance, ML

algorithms like neural networks can predict peak traffic times for specific services, enabling proactive resource allocation to avoid congestion.

Reinforcement learning, particularly in a multi-agent environment, is also becoming popular for resource allocation in network slicing. Multi-agent reinforcement learning (MARL) allows different network entities, such as base stations and user equipment, to collaborate as independent agents to maximize overall network performance. The result is more efficient resource utilization, minimizing waste and ensuring that each slice receives the appropriate resources to maintain its service-level agreements (SLAs).

Traffic management in network slicing is another area where AI excels. The diversity of services in a 5G network, such as enhanced mobile broadband (eMBB), URLLC, and massive IoT, demands intelligent traffic prioritization. AI algorithms analyze traffic patterns in real time, enabling the system to prioritize slices that require lower latency or higher reliability automatically. This dynamic traffic management helps ensure that critical services, like autonomous vehicles or remote surgeries, get priority over less critical applications like video streaming see [Figure 1](#).

5G Network Slicing

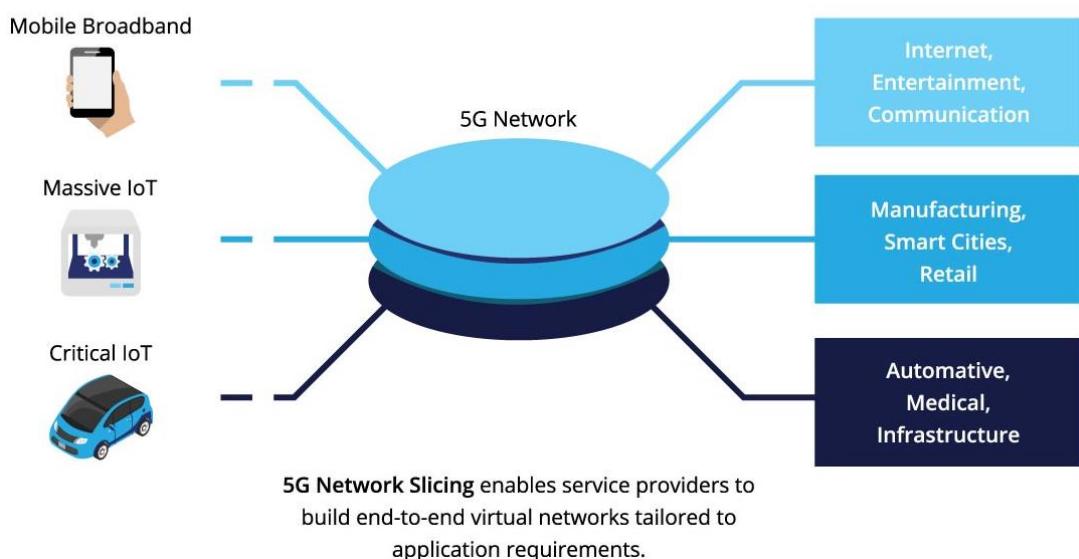


Figure 2. Network Slicing in 5G.

AI-powered traffic management can also mitigate congestion and improve the overall quality of service (QoS) by re-routing traffic through less congested paths or adjusting bandwidth allocations. Predictive models, trained on historical traffic data, can forecast potential bottlenecks and allow the network to take preemptive measures, ensuring smooth operations even during peak usage periods.

One key advantage of integrating AI in network slicing is

self-optimization capability. AI can continuously monitor network performance metrics such as latency, throughput, and error rates across different slices. When deviations from expected performance are detected, AI systems can autonomously adjust configurations, redistribute resources, or even alter the slice architecture to restore optimal performance.

For instance, in cases where a slice serving IoT applications experiences a sudden increase in device connections, AI can

scale the slice's capacity by reallocating resources from less critical slices. Similarly, slices that require ultra-low latency can be dynamically reconfigured to prioritize routing through lower-latency paths.

AI-driven approaches are fundamental in overcoming the complexity of network slicing in 5G networks. By leveraging AI technologies like reinforcement learning, neural networks, and multi-agent systems, 5G networks can achieve greater efficiency, adaptability, and scalability. AI ensures that network slices are dynamically created, maintained, and optimized, providing tailored services to meet the varying demands of modern digital ecosystems.

5. Challenges and Future Directions

Integrating AI in 5G networks for resource allocation, traffic management, and network slicing presents significant potential and numerous challenges. One major challenge is the complexity of managing increasingly dense and heterogeneous networks. As 5G supports various applications with differing requirements, like eMBB, URLLC, and massive IoT, the need for real-time optimization of resources becomes critical. Traditional rule-based systems fail to manage dynamic traffic and user demands efficiently, necessitating AI-driven adaptive solutions. However, deploying AI models for real-time decision-making at scale requires significant computational power and efficient learning algorithms to avoid system delays and bottlenecks. A key issue is the overhead and latency of AI-based resource allocation, mainly when using deep reinforcement learning (DRL) models. DRL systems effectively learn from the network environment and make dynamic resource adjustments but often suffer from high training costs and memory consumption. This can lead to inefficiencies in real-time operations, especially when networks are large and involve many interconnected devices, such as in smart cities or autonomous vehicle networks.

Moreover, multi-agent reinforcement learning (MARL) methods used in network slicing require extensive coordination between network entities, which can result in system overhead and resource wastage if not correctly managed. Another challenge is the reliance on accurate channel state information (CSI) for resource allocation. This practice incurs considerable system overhead and is particularly inefficient in CRAN and mmWave-based 5G applications. Existing solutions like heuristic algorithms, genetic algorithms, or clustering techniques provide partial improvements but often fail to scale effectively as user demand increases. Future directions involve improving the efficiency and scalability of AI-based solutions in 5G. Research is needed to optimize learning algorithms to reduce training costs and memory usage, potentially through federated learning or edge computing, where processing is distributed closer to the network edge. Additionally, hybrid AI models combining multiple machine learning techniques like convolutional neural networks (CNNs) for traffic prediction and reinforcement

learning for resource allocation, could offer more adaptable solutions to 5G's heterogeneous environments.

Network slicing in 5G also requires more sophisticated AI-driven orchestration mechanisms. Real-time prediction and adaptation of network slices based on AI algorithms will become crucial, particularly in managing different services' varying latency, reliability, and bandwidth requirements. Integrating AI models with software-defined networking (SDN) and network function virtualization (NFV) can help optimize slice management dynamically.

6. Conclusion

Integrating AI-driven techniques into 5G networks presents a transformative approach to overcoming the inherent challenges of resource allocation, traffic management, and network slicing. As 5G networks scale in complexity, traditional methods struggle to provide the real-time adaptability required for dynamic, high-performance environments. AI models, particularly those based on machine learning (ML) and deep reinforcement learning (DRL), offer adaptive, data-driven solutions that can continuously learn from network conditions to optimize performance, reduce latency, and manage system overhead.

Resource allocation in 5G is especially critical given the rise of data-intensive applications like autonomous vehicles, augmented reality, and massive IoT deployments. AI-based methods, such as DRL and genetic algorithms, provide scalable approaches to efficiently manage spectrum, compute power, and energy resources. These intelligent methods address the shortcomings of conventional models, such as channel state information (CSI)-based allocation, by offering lower overhead and better adaptability to fluctuating conditions. By leveraging AI, 5G networks can dynamically allocate resources to meet the needs of different applications, from low-latency services to high-throughput data demands. Traffic management is another area where AI significantly enhances the operation of 5G networks. Through advanced traffic prediction and real-time analysis, AI models such as LSTM and transformer-based architectures offer sophisticated tools to predict traffic patterns and optimize network load distribution. These capabilities are crucial in managing the expected exponential increase in mobile data traffic, ensuring efficient bandwidth utilization, and maintaining network robustness even under high demand. Furthermore, network slicing, a cornerstone of 5G's architecture, benefits immensely from AI's ability to orchestrate and optimize virtual network slices in real time. AI techniques such as multi-agent reinforcement learning (MARL) enable more granular control over resource allocation across slices, ensuring each slice meets its specific service-level agreements (SLAs) while optimizing overall network efficiency.

AI's integration into 5G is not just a complementary technology but a necessity to fully realize the potential of next-generation networks. The shift from static, rule-based

systems to intelligent, adaptive algorithms marks a paradigm shift that will define future telecommunications, enabling more resilient, efficient, and scalable network operations that support a wide array of emerging technologies. This convergence of AI and 5G lays the foundation for autonomous networks and opens new research directions to further enhance performance, efficiency, and scalability in the era of 6G and beyond.

Abbreviations

ML	Machine Learning
DL	Deep Learning
DRL	Deep Reinforcement learning
NR	New Radio
URLLC	Ultra-Reliable Low-Latency Communications
MARL	Multi-agent Reinforcement Learning
HVFT	High-volume Flexible Traffic
DDPG	Deep Deterministic Policy Gradient
mMTC	Machine-Type Communications
MIMO	Massive Multiple-input Multiple-output Systems
CSI	Channel state information

Author Contributions

Dileesh Chandra Bikkasani is the lead author, and Malleswar Reddy Yerabolu is the co-author. The authors read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] An, J., et al., *Achieving sustainable ultra-dense heterogeneous networks for 5G*. IEEE Communications Magazine, 2017. 55(12): p. 84-90.
- [2] ITU. *Setting the Scene for 5G: Opportunities & Challenges*. 2020 [cited 2024 07/13]; Available from: <https://www.itu.int/hub/2020/03/setting-the-scene-for-5g-opportunities-challenges/>
- [3] Sakaguchi, K., et al., *Where, when, and how mmWave is used in 5G and beyond*. IEICE Transactions on Electronics, 2017. 100(10): p. 790-808.
- [4] Foukas, X., et al., *Network slicing in 5G: Survey and challenges*. IEEE communications magazine, 2017. 55(5): p. 94-100.
- [5] Abadi, A., T. Rajabioun, and P. A. Ioannou, *Traffic flow prediction for road transportation networks with limited traffic data*. IEEE transactions on intelligent transportation systems, 2014. 16(2): p. 653-662.
- [6] Imtiaz, S., et al. *Random forests resource allocation for 5G systems: Performance and robustness study*. in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. 2018. IEEE.
- [7] Wang, T., S. Wang, and Z.-H. Zhou, *Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks*. China Communications, 2019. 16(1): p. 165-175.
- [8] Baghani, M., S. Parsaeefard, and T. Le-Ngoc, *Multi-objective resource allocation in density-aware design of C-RAN in 5G*. IEEE Access, 2018. 6: p. 45177-45190.
- [9] Shehzad, M. K., et al., *ML-based massive MIMO channel prediction: Does it work on real-world data?* IEEE Wireless Communications Letters, 2022. 11(4): p. 811-815.
- [10] Chughtai, N. A., et al., *Energy efficient resource allocation for energy harvesting aided H-CRAN*. IEEE Access, 2018. 6: p. 43990-44001.
- [11] Zarin, N. and A. Agarwal, *Hybrid radio resource management for time-varying 5G heterogeneous wireless access network*. IEEE Transactions on Cognitive Communications and Networking, 2021. 7(2): p. 594-608.
- [12] Huang, H., et al., *Optical true time delay pool based hybrid beamformer enabling centralized beamforming control in millimeter-wave C-RAN systems*. Science China Information Sciences, 2021. 64(9): p. 192304.
- [13] Lin, X. and S. Wang. *Efficient remote radio head switching scheme in cloud radio access network: A load balancing perspective*. in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. 2017. IEEE.
- [14] Gowri, S. and S. Vimalanand, *QoS-Aware Resource Allocation Scheme for Improved Transmission in 5G Networks with IOT*. SN Computer Science, 2024. 5(2): p. 234.
- [15] Bouras, C. J., E. Michos, and I. Prokopiou. *Applying Machine Learning and Dynamic Resource Allocation Techniques in Fifth Generation Networks*. 2022. Cham: Springer International Publishing.
- [16] Li, R., et al., *Intelligent 5G: When cellular networks meet artificial intelligence*. IEEE Wireless communications, 2017. 24(5): p. 175-183.
- [17] Ericsson. *5G to account for around 75 percent of mobile data traffic in 2029*. [cited 2024 07/13]; Available from: <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast>
- [18] Amaral, P., et al. *Machine learning in software defined networks: Data collection and traffic classification*. in *2016 IEEE 24th International conference on network protocols (ICNP)*. 2016. IEEE.
- [19] Wang, H., et al. *Understanding mobile traffic patterns of large scale cellular towers in urban environment*. in *Proceedings of the 2015 Internet Measurement Conference*. 2015.
- [20] Box, G. E., et al., *Time series analysis: forecasting and control*. 2015: John Wiley & Sons.

- [21] Shu, Y., et al., *Wireless traffic modeling and prediction using seasonal ARIMA models.* IEICE transactions on communications, 2005. 88(10): p. 3992-3999.
- [22] Kumari, A., J. Chandra, and A. S. Sairam. *Predictive flow modeling in software defined network.* in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON).* 2019. IEEE.
- [23] Moore, J. S., *A fast majority vote algorithm.* Automated Reasoning: Essays in Honor of Woody Bledsoe, 1981: p. 105-108.
- [24] Arjoune, Y. and S. Faruque. *Artificial intelligence for 5g wireless systems: Opportunities, challenges, and future research direction.* in *2020 10th annual computing and communication workshop and conference (CCWC).* 2020. IEEE.
- [25] Mennes, R., et al. *A neural-network-based MF-TDMA MAC scheduler for collaborative wireless networks.* in *2018 IEEE Wireless Communications and Networking Conference (WCNC).* 2018. IEEE.
- [26] Vaswani, A., et al., *Attention is all you need.* Advances in neural information processing systems, 2017. 30.
- [27] Chinchali, S., et al. *Cellular network traffic scheduling with deep reinforcement learning.* in *Proceedings of the AAAI Conference on Artificial Intelligence.* 2018.
- [28] Peng, B., et al., *Decentralized scheduling for cooperative localization with deep reinforcement learning.* IEEE Transactions on Vehicular Technology, 2019. 68(5): p. 4295-4305.
- [29] Cao, G., et al., *AIF: An artificial intelligence framework for smart wireless network management.* IEEE Communications Letters, 2017. 22(2): p. 400-403.
- [30] Challita, U., L. Dong, and W. Saad, *Proactive resource management for LTE in unlicensed spectrum: A deep learning perspective.* IEEE transactions on wireless communications, 2018. 17(7): p. 4674-4689.
- [31] Stampa, G., et al., *A deep-reinforcement learning approach for software-defined networking routing optimization.* arXiv preprint arXiv:1709.07080, 2017.