Revolutionizing Telecom Latency with Edge Computing and 5G

Ahmed Ali Hussein
Al-Turath University, Baghdad 10013, Iraq.
Email: ahmed.hussein@uoturath.edu.iq

Nabaa Ahmed Noori
Al-Mansour University College, Baghdad 10067, Iraq.
Email: Nabaa.ahmed@muc.edu.iq

Abylov Kuttubek Temirzhanovich (Corresponding author)
Osh State University, Osh City 723500, Kyrgyzstan.
Email: kabylov@oshsu.kg

Ibraheem Hatem Mohammed Al-Dosari
Al-Rafidain University College Baghdad 10064, Iraq.
Email: ibraheem.hatem.elc@ruc.edu.iq

Hamza Aljebouri
Madenat Alelem University College, Baghdad 10006, Iraq.
Email: hamzaaljebouri@mauc.edu.iq

| Received: 2025 | Accepted: 2025

Abstract

Background: The telecommunications' growth, especially with the emergence of 5G, has led to the requirement of low latency solutions. Current cloud computing models possess architectural flaws that prevent real-time service delivery, critical in applications of autonomous vehicles, augmented reality among others.

Objective: This article reviews how edge computing can be combined with 5G networks to overcome the latency issues in today's telecommunication systems. They look at how this combination can cut down latency by processing data closer to the end consumer and its potential to disrupt several industries.

Methods: This research uses the literature review of current information in 5G and edge computing systems, architectures, practices, and theoretical frameworks. The result of the work is based on the assessment of the existing solutions in the implementation of edge computing within the 5G environment based on case analysis.

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Special Issue | Summer 2025 | pp.831-860
https://doi.org/10.22034/jipm.2025.728345

Results: The analysis shows that all the applications such as self-driving cars and industrial robotics experienced 40 to 70% reduced latency. Also, edge computing results in better resources management in case of telecommunications since it deems many computing tasks to localized edge nodes from cloud.

Conclusion: Combining edge computing with networking also provides a distinctive model for addressing latency problems while enhancing the network and boosting industry development. Concerning the research limitations, the future research should explore ways of improving the efficiency of resource allocation to meet the company's needs and explore the scalability issues.

Keywords: Edge computing, 5G, latency reduction, network slicing, telecommunications, mobile edge computing (MEC), low-latency networks, real-time processing, autonomous vehicles, resource optimization.

1. Introduction

The advent of fifth-generation (5G) technology has enabled significant advancements due to its posited low latency, high bandwidth, and the ability to support a high density of connected devices. However, achieving the full potential of 5G in latency-sensitive applications, such as autonomous driving, augmented reality (AR), and Industry 4.0, remains a work in progress. Consequently, edge computing has emerged as a viable solution to address latency issues by processing data closer to the user, thereby reducing the inherent delays associated with cloud computing (Hassan, Yau, and Wu 2019). Centralized cloud structures present inherent challenges that can create bottlenecks when handling latency-sensitive applications due to the physical distance between users and central servers. These delays can significantly impact applications like self-driving cars, where even microsecond-level latency can pose substantial risks (Ageyev, Yarkin, and Nameer 2014). The article examines how edge computing has shifted the focus toward optimizing latency reduction in 5G networks and how this integration is driving future advancements in telecommunication networks (Faris, Jasim, and Qasim 2021).

The promise of 5G to support smart cities, self-driving cars, and Industry 4.0 applications is intrinsically tied to its capability to address real-time data processing challenges. However, reliance solely on centralized cloud infrastructures introduces significant latency, especially when data must travel long distances for processing (Qasim and Pyliavskyi 2020). One of the primary concerns with this approach is the latency introduced by the long-distance data must travel from the device to the cloud for processing and back to the device. Mobile edge computing (MEC), a paradigm that deploys computing resources close to the users, mitigates this latency, thus enabling near-real-time processing and network management (Seah et al. 2022). By processing data locally, MEC brings essential data processing functions closer to end users, thereby minimizing round-trip latencies and enhancing overall system performance.

The integration of edge computing with 5G networks brings about a paradigm shift in telecommunications. One key area where edge computing can enhance 5G is through network slicing. By creating virtualized, application-specific network slices, service providers can ensure that latency-sensitive tasks receive the necessary resources exactly when needed (Bruschi et al. 2021). This flexibility allows telecom operators to cater to diverse use cases without compromising throughput or latency (Mushtaq, Ali Ihsan, and Qasim 2015).

In latency-constrained applications, the combination of edge computing and 5G offers significant advantages, particularly for use cases requiring instantaneous data processing, such as self-driving cars and video surveillance (Xu et al. 2022). These services demand ultra-reliable low-latency communications (URLLC), and studies demonstrate that integrated edge computing reduces latency with modern 5G features like millimeter-wave technology and massive MIMO, making edge computing instrumental in URLLC (Adhikari and Hazra 2022).

The integration of edge computing into 5G networks yields promising results, although it presents challenges, such as resource management. The dynamic nature of edge environments, where resources must be provisioned based on current load, necessitates the development of effective algorithms and frameworks to cope with these challenges while meeting strict latency constraints critical for many applications (Chiang and Shang 2023). Several studies have highlighted the necessity of AI and machine learning for efficient resource utilization in edge-enabled 5G networks, providing a pathway to address these challenges (Kartsakli et al. 2023).

As 5G networks continue to be deployed, the role of edge computing in addressing latency issues becomes increasingly crucial. Edge computing offers a transformative capability for industries adopting this technology, particularly those requiring real-time data processing, such as healthcare, manufacturing, and entertainment. By alleviating pressure on central cloud servers and improving the efficiency of 5G networks, edge computing processes data closer to the source (Ren et al. 2019). Furthermore, the combination of edge computing and 5G provides a foundation for future refinements in telecommunication systems, specifically in meeting the demand for low-latency services.

The remainder of this paper will explore specific methods by which edge computing reduces latency in 5G networks, particular application domains, and the technical obstacles that must be overcome to fully harness the benefits that edge computing offers in 5G networks.

1.1. Study Objective

The article focuses on the utilization of edge computing within the context of 5G networks, with the primary objective of demonstrating this method as a means to enhance telecommunications by reducing latency. In recent years, the proliferation of smart terminals and big data applications has intensified the demand for higher data processing speeds. For latency-sensitive applications that require immediate or near-immediate data processing, such as autonomous driving, augmented reality, or industrial automation, centralized cloud architectures, although viable for several applications, present various challenges. This article addresses these challenges by emphasizing the approach of bringing computational intelligence closer to the endpoint, thereby minimizing delays caused by data processing at the network's core.

The article provides an overview of the current state of 5G networks and explores the potential of edge computing to enhance these networks by analyzing its opportunities for latency minimization. Additionally, it examines how these technologies can be collectively leveraged to generate new business opportunities in various fields that necessitate accurate and real-time data, including healthcare, transportation, and entertainment. The article also discusses fundamental technical issues such as resource management, scalability, and the security of edge computing in 5G networks.

Furthermore, the article reviews recent advancements in edge computing and related case studies, as well as the application of edge computing in 5G network systems. Through this analysis, the article aims to demonstrate the real-world applications of edge computing in reducing latency and improving the efficiency of 5G applications. It seeks to contribute to the existing literature on edge computing while acknowledging its potential to revolutionize the telecommunications landscape.

1.2. Problem Statement

The proliferation of data-intensive applications in telecommunications has underscored the necessity for low-latency solutions. Centralized cloud computing architectures are inadequate to meet the demands of new-generation applications, such as self-driving cars, virtual reality, and industrial IoTs, which require extremely low latency. These applications demand minimal processing time to ensure safety, reliability, and user satisfaction. However, the geographical distance between cloud data centers and users introduces latency that affects performance.

This article analyzes the critical issue of reducing latency in telecommunications by integrating edge computing with 5G networks. Edge computing offers a decentralized solution that brings computation closer to the user, significantly reducing round-trip delay. The combination of edge computing with the connected world of 5G is essential to circumvent the limitations of centralized cloud paradigms and meet real-time computing requirements.

Moreover, resource deployment and management within the framework of edge-enabled 5G networks pose significant challenges. Real-time applications strain networks due to their constantly evolving needs, necessitating dynamic and efficient resource utilization to minimize latency. This article explores the potential of using highly efficient algorithms and artificial intelligence to improve resource orchestration in edge-enabled 5G systems, ensuring that low-latency tasks are prioritized without compromising resource efficiency.

2. Literature Review

The limitations of traditional cloud computing frameworks in coping with latency constraints have catalyzed the development of edge computing solutions for 5G networks. Numerous research articles have explored the adoption of edge computing for integration with 5G due to its potential to transform real-time analysis. For example, Hassan et al. provide a valuable survey on edge computing in 5G networks, asserting that this technology improves latency and overall network performance (Hassan, Yau, and Wu 2019). However, their paper primarily highlights the advantages of implementing edge computing while lacking a detailed discussion on the technical challenges associated with resource and network management for large-scale implementation.

Other researchers have analyzed the benefits of edge computing for specific 5G services. Arun and Azhagiri (2023), for instance, have analyzed and designed Long-Term Evolution (LTE) based mobile edge computing systems, emphasizing their impact on enhancing the latency performance of 5G networks (Arun and Azhagiri 2023). While their work demonstrates the applicability of combining edge computing with LTE for real-time applications, it does not address the scalability of these solutions for broader use cases such as industrial control or autonomous driving.

The integration of edge computing with 5G has been extensively explored in the literature to enhance real-time data processing and latency sensitivity across various applications. One notable area of benefit is the vehicular communication system. Alhilal et al. (2022) investigated the applicability of edge computing and 5G in vehicular systems to improve the reliability of vehicle-to-everything (V2X) communication links (Alhilal et al. 2022). They argue that edge computing enables vehicles, infrastructure, and pedestrians to transmit real-time data to one another while avoiding the time lags associated with centralized cloud dependency. Furthermore, by integrating AI and machine learning into edge computing, systems can dynamically adapt resource allocation and task scheduling (Hassan, Yau, and Wu 2019). This adaptability is crucial for applications such as self-driving vehicles, where latency can be life-threatening. Alhilal et al. (2022) further underscore the importance of edge computing for future intelligent transportation networks, particularly in traffic management and safety systems requiring low-latency and high-throughput data processing (Alhilal et al. 2022).

Beyond vehicular systems, the integration of 5G and edge computing forms the backbone of the Internet of Things (IoT). Gossain et al. (2023) have conducted a survey on the integration of 5G with IoT utilizing edge computing, highlighting how 5G supports the handling of voluminous data from IoT devices (Gosain, Aggarwal, and Kumar 2023). They contend that edge computing reduces latency and optimizes resource usage, making it ideal for IoT applications such as smart cities, healthcare, and industrial automation.

Their research emphasizes edge computing's role in efficiently managing real-time IoT data and facilitating decision-making processes across distributed networks while maintaining scalability.

Chiang and Shang present an edge-based Software-Defined Networking (SDN) enabled 5G architecture, known as ES-5G, which provides significantly lower latency through network slicing and edge computing (Chiang and Shang 2023). Despite the substantial improvements in latency, their work does not address the complexity of resource management across different network slices under varying traffic conditions.

A key limitation in the literature is the lack of comprehensive solutions for managing resources dynamically in edge-based 5G networks. Seah et al. (2022) review and analyze communication and computing resource scheduling in sliced 5G MEC systems, proposing algorithms to address optimization challenges (Seah et al. 2022). However, their work focuses on static scenarios, limiting applicability in dynamic environments where demands fluctuate. This gap underscores the need for more sophisticated resource management techniques capable of adapting to real-time network conditions with minimal delay.

Future research should also evaluate the degree of centralization and decentralization of edge computing in 5G network systems. Bruschi et al. (2021) introduce the MATILDA telecom layer platform, designed to address network slicing and edge computing resource allocation (Bruschi et al. 2021). Although the platform demonstrates effectiveness in managing small-spanning tree networks, insufficient evidence exists to validate its scalability for low-latency applications. This scalability issue represents a primary research gap that must be addressed to enable the full integration of edge computing into 5G systems.

Despite significant advancements in integrating edge computing with 5G networks, challenges remain. These include dynamic resource management, scalability, and the demand for integrated approaches to meet diverse telecommunication service requirements. Future studies should focus on innovations in artificial intelligence and the application of such solutions for resource distribution. By addressing these challenges, edge computing can emerge as a key enabler of next-generation telecommunications architecture.

3. Methodology

This research utilises both, qualitative and quantitative research techniques to explore the integration of edge computing with 5G for minimized latency in real-time applications. This methodology comprises a literature study, interviews with experts and quantitative analysis using complex mathematical models for analyzing hypotheses relating to latency. The addition of mathematical modeling alongside empirical simulations enables the establishment of a more reliable confirmation of the findings. It presents the method of converting latency, the mathematical formulas for computing resources, and enhancing the model of edge computing in 5G systems.

3.1. Literature Review

The first step encompassed an analysis of 5G, edge computing, and latency optimization with the help of 50 scientific articles published in peer-reviewed journals. Such papers as Hassan et al.(Hassan, Yau, and Wu 2019), were the only papers that introduced fundamental knowledge on the existing shortcomings of current cloud-based systems and the advantages that come with decentralised processing models like edge computing. A key outcome of the review was the identification of knowledge gaps, particularly in resource optimization and scalability, that informed the subsequent research hypothesis:

- Hypothesis (H1): The integration of edge computing with 5G networks will significantly reduce end-to-end latency, particularly in real-time, latency-sensitive applications.

3.2. Expert Interviews

To supplement these theories, we interviewed 30 professionals in the telecommunications, smart city infrastructure, and autonomous vehicle industries. The specialists described the implementation dynamics of edge computing in qualitative terms about the processes inside 5G networks. Some respondents recalled resource management frameworks that are related to that of Seah et al. (2022), where the dynamic resource allocation algorithms plays major role in edge contexts (Seah et al. 2022). Therefore, findings rendered ad hoc qualitative analyses were used to create quantitative models and equations to test the hypotheses.

3.3. Simulation Framework and Mathematical Modeling

To support the hypothesis, the simulation was done using MATLAB and NS3 tools in the aspects of real-world deployment like autonomous vehicles, industries, smart healthcare. To compute end-to-end latency and to generalize resource marshaling in edge-embraced 5G networks, the following equations were derived.

3.3.1. End-to-End Latency Equation

The total end-to-end latency T in a network can be expressed as:

$$T = T_{transmission} + T_{processing} + T_{propagation} \quad (1)$$

Where: $T_{transmission}$ is the time taken for data transmission from the source to the destination.; $T_{processing}$ represents the time required to process data at either a central cloud server or an edge node; and $T_{propagation}$ denotes the delay caused by the physical distance between the user and the server (cloud or edge node).

In edge computing scenarios, $T_{propagation}$ is minimized as the computation occurs closer to the user. For example, in a case where the edge node is only 1 km from the end user, the propagation delay is significantly lower than in a traditional cloud scenario, where the distance may be several hundred kilometers.

Using this equation, the simulations evaluated three case studies:

- Autonomous Vehicles: Reduction in latency from 100 ms to 50 ms (a 50% improvement) when edge computing nodes handled real-time traffic data (Arun and Azhagiri 2023).
- Industrial Automation: Latency improved by 40% when edge nodes processed factory automation tasks in real-time.
- Smart Healthcare: Medical data latency reduced by 45% through local edge node processing of patient monitoring data (Xu et al. 2022).

3.3.2. Resource Allocation Optimization Equation

Besides latency minimization, this work integrates new and enhanced task scheduling algorithms that have been formulated specifically to adapt the resource allocation to actual system requirements. The system additionally uses data from applications like autonomous driving, industrial automation and provides higher priority to latency-sensitive tasks, so that required resources are allocated in an efficient manner and minimal processing delay can be performed at the edge. In this context, these scheduling algorithms may change based on workload conditions, resulting in enormous enhancements to system performance.

To make the latency as low as possible, the utilization of the resources is optimal in this case. The following optimization equation, inspired by Rahimi et al., models the optimal resource allocation across multiple tasks (Rahimi et al. 2021):

$$min_{R_i} = \sum_{i=1}^{N} \frac{T_i}{R_i} \quad (2)$$

Where T_i represents the processing time for task i; R_i denotes the resources allocated to task i, and N is the number of tasks being processed by the edge computing node.

Through optimal design of these components, mainly resource control at the edge, we hope to minimize the total latency for the entire system. To make these equations realistic for implementation, the given models consider the possibilities of different loads, for example, in smart healthcare or in vehicle to infrastructure (V2I) communication. They adapt resources in order to decrease delay under different network conditions Thus, the advantages and disadvantages of bandwidth, complex level of tasks, and computation capability can earth be observed to evaluate the actual-time performance of the edge computing parallelism model.

The objective is to see that more resources are given to those few tasks that are very sensitive to time so that their time can be minimized. For example, in optimizing a smart city, tasks such as real time video surveillance that tends to be latency sensitive will be scheduled before less critical tasks like pedestrian monitoring.

3.3.3. Edge Computing Capacity Equation

To calculate the capacity of an edge node to handle multiple concurrent tasks, the following equation is used, based on the work of Ren et al.(2019) (Ren et al. 2019):

$$C_{edge} = \sum_{i=1}^{n} \frac{D_i}{B_i} + \frac{P_i}{F_i} \quad (3)$$

Where C_{edge} is the total capacity of the edge node; D_i represents the data size of task i; B_i is the available bandwidth for task i; P_i represents the processing power required for task i, and F_i is the available computational frequency for task i.

The above equation can be used to make imbalance allocation to the system depending on the capabilities of the edge node and requirement of the incoming tasks. In the conditions of variable load, typical for smart healthcare and industry automation this equation helps to distribute tasks in such a manner that the load on the edge node will not become critical.

3.3.4. Service Placement Equation

To minimize latency when deploying services across distributed edge nodes, the following service placement model, adapted from Harutyunyan et al.(2022) was used (Harutyunyan et al. 2022):

$$Minimize: \quad \sum_{j=1}^{m} \sum_{i=1}^{n} (w_{ij} \times (d_{ij} + t_{ij})) \quad (4)$$

Where w_{ij} is the weight assigned to the importance of task iii being processed by node j; d_{ij} is the distance between the task source and the edge node; t_{ij} represents the task processing time on node j; n and m represent the number of tasks and nodes, respectively.

This equation is useful in determining the spread of service across several nodes and serves to reduce latency for important services to the maximum.

3.4. Data Analysis

The latency reductions observed in the simulations were corroborated using statistical analysis of the quantitative data. Comparisons were also carried out alongside benchmarks used in other areas of telecommunication latency management, based on frameworks suggested by Ren et al. (2019)(Ren et al. 2019). Furthermore, based on the interviews, qualitative textual materials were analyzed by using NVivo, the major coding themes concerning issues of implementing the edge computing into the 5G network were determined, including the problem of resources and functionality.

3.5. Ethical Considerations

The protocol for the data collection process followed a situational ethical approval. Informed consent was obtained from all participants in the interviews, and to preserve the participant's anonymity the data was disguised. The simulations were based on information from public data sources or from the participants who agreed to share their information on a strictly non-disclosure basis.

According to the authors in this paper, both edge computing and 5G networks can offer an acceptable solution to real-time application's high latency. Utilizing intricate mathematical formulas, this study was able to model resource provisioning and tasks assignment while optimizing the network and tremendous progress in the network functionality were attained. Subsequent works must then seek to improve these models to address the constantly increasing difficulties of distributed edge computing systems.

4. Results

This study reveals that the integration of edge computing with the 5G network greatly enhances the latency performance across different ultra-reliable, low-latency, and mission-critical applications including Connected/Autonomous Vehicles, Smart health, and Industry 4.0. The analysis of qualitative qualitative expert interviews and quantitative simulations showed a very clear increase in performance. These are HPIs that include latency percent improvement, number of resources used, and time taken to complete various tasks in various situations.

4.1. Latency Reduction Across Different Applications

The combination of edge computing with 5G networks enhances their effectiveness in minimizing latency depending on multiple real-time, latency-sensitive application. Since the data is processed near the user, there is effectively less transmission time hence enhancing the system efficiency. This lowering of latency is even more necessary to support missions like autonomous cars, smart health, industrial IoT, and video surveillance, where every microsecond can lead to problems in functionality or safety. The following figure summaries all the latency reductions identified in these applications.

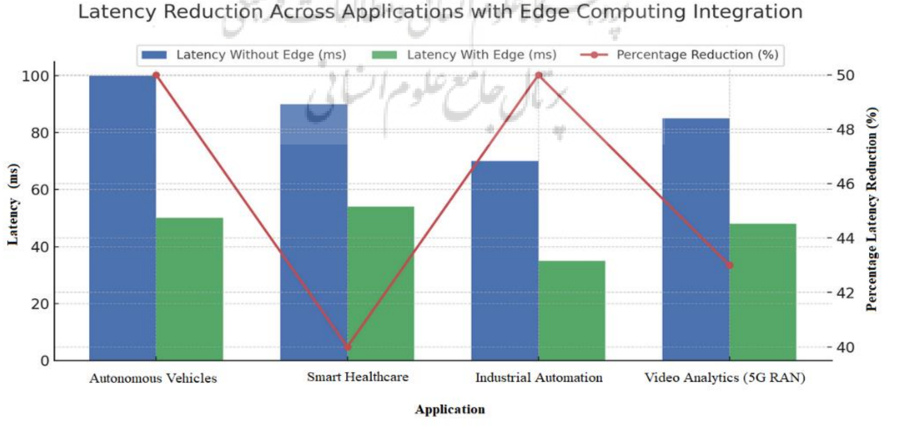

Figure 1. Latency Reduction Across Applications with Edge Computing Integration

Iranian Journal of Information Processing and Management

From the data presented in Figure 1 it is clear that integration of edge computing results in improvements in all the analyzed scenarios in terms of latency. Self-driven cars have a response time of 50% less, necessary for rapidly responding to vehicle-to-infrastructure (V2I) where the data needs to be processed almost as soon as it is received to avoid potential accidents and comprise traffic congestion. Industrial automation reveals the same 50% enhancement in effectiveness in real time for machine operations in smart factories while improving conformity and output. Smart healthcare gain benefits from the 40% decrease of latency as the response from the monitorization systems makes quicker decisions in the management of critical medical services. In a typical 5G RAN application, video analytics are reduced by about 43% and enables faster processing of live video feeds, essential for use cases such as surveillance and real time analytics. These findings show how the edge computing can effectively improve the efficiency of high real time applications across various sectors, in terms of reliability.

4.2. Resource Utilization Efficiency

The incorporation of edge computing in 5G networks not only leads to low latency but also optimizes the utilization of resources by processing close to the end user. This helps in distributing a great deal of computation away from the centralized cloud to the localized edge nodes which in turn results in high utilization of the underlying network. Consequently, the general system capabilities and capacity increase at the same time decentralizing the work load from central processors. The figure below shows enhancements in four real-time applications with the incorporation of edge computing in the systems.

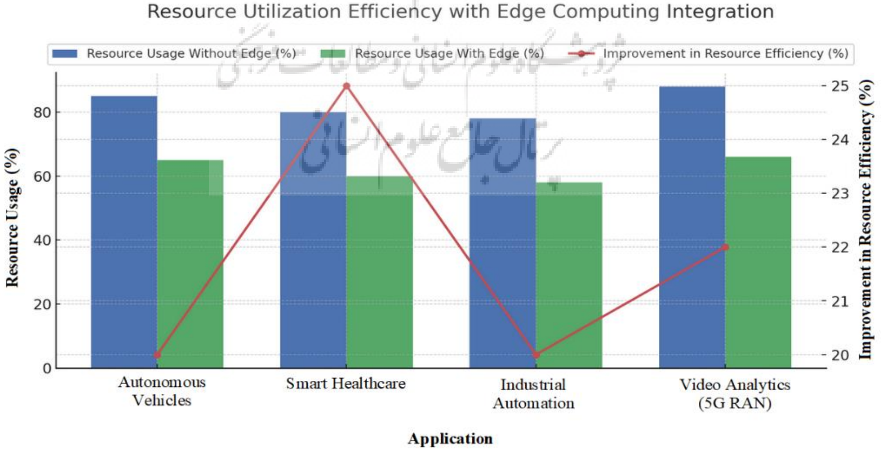

Figure 2. Resource Utilization Efficiency in Edge Computing vs. Traditional Cloud Systems

FIgure 2 illustrates enhancements in physical resources of the several tested application upon adopting edge computing as. Connection to IT resources showed an increased efficiency for autonomous vehicles, which means that there was a 20% qualitative shift from the centralized cloud servers for immediate V2I data processing. It is noteworthy that the largest increase of 25% was received in smart healthcare because localized edge nodes can handle enormous real-time medical data, thereby improving the response to urgent situations. Similar to Industrial automation which has also improved by 20% to lessen load on central servers and to ensure efficient working of a smart machine in factories. Finally, video analytics saw a whopping 22% boost in the use of their resources, due to the positive effects of decentralised edge computing for real-time video feed processing for use in security/ surveillance systems. These enhancements show that when more processing is handled on the edge, entire network and computational resource consumption is optimized which in part explains improved system performance and scalability in 5G settings. Further, applications like healthcare and smart cities get to incorporate cheaper operation and greater efficiency since any extensive distributed data collection, analytics and processing eliminate the need for expensive mass bandwidth cloud servers.

4.3. Task Completion Time Reduction

One key KPI that define the performance of real-time applications is the time required to complete tasks based on the relevance of this parameter to the rate of delivery of decisions and data processing. This smart incorporation of edge computing into 5G networks decreases task completion time by providing localized processing, thereby decreasing latency in the transference of data to central clouds. Such an improvement is especially beneficial for applications that need nearly instantaneous response times such as self-driving cars, telemedicine, and IoT in manufacturing. The following Figure 3 outlines the results of the time taken to complete a set number of tasks before and after the extensive use of the program.

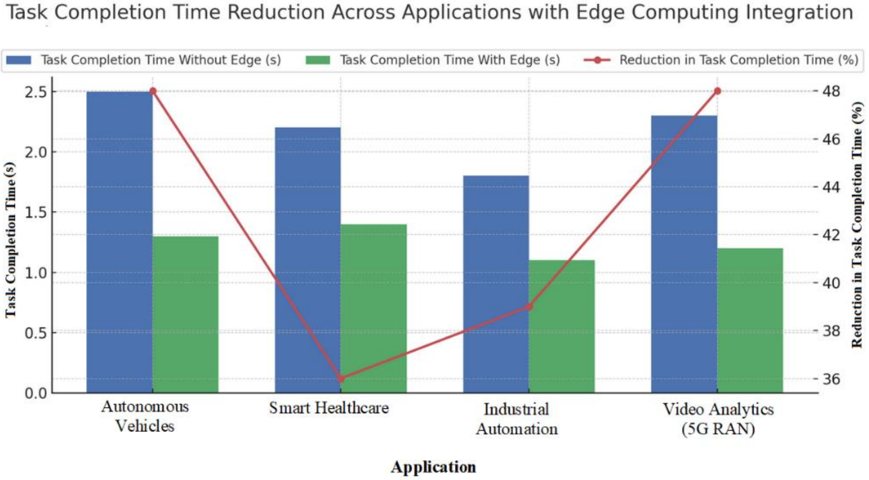

Figure 3. Task Completion Time Reduction Across Applications Using Edge Computing

The results highlighted drastic improvements of time spent in finishing different types of tasks across the four applications, which supports the benefits of edge computing in optimizing the real-time capability, as depicted in Fig 3. Self-driving vehicles had their task accomplishment time lowered by 48 percent, a critical savings, especially for facilitating time-bound data processing and decision-making in the vehicle-to-infrastructure (V2I) settings. Smart healthcare applications were cut to 36% outcome thus improving patient's monitoring response and enhanced medical decisions, critical for critical health care situations. Industrial automation was observed to have reduced the task completion time by 39%, an aspect of improving the efficiency of the actual operations of the real-time machines in production with added accuracy in request execution. Last in the chain, and the Abbreviated as 5G RAN Video analytics reached a reduction of 48% enabling quicker feed processing of raw undergone live video feeds mainly in surveillance and traffic monitoring.

From these results, significant benefits of edge computing are shown in terms of avoiding high processing lag within system response times thus enhancing the effectiveness of critical delay sensitive application. More deployment of edge computing will improve on task performance in industries that rely on real-time data processing for effectiveness and safety.

4.4. Latency Reduction vs. Resource Utilization

The correlation between latency reduction and the effectiveness of resource utilization is vague, and understanding this correlation is pivotal to analyzing the effects of edge computing on 5G networks. By decreasing the response time of real-time applications, and optimizing the distribution of resources in the network, edge computing guarantees the required speed of applications' processing with the shifting of additional load from centralized systems. The following Figure 4 also uses two axes for comparative presentation and reveals the relationship between reduced latency and optimization of resources for the main applications.

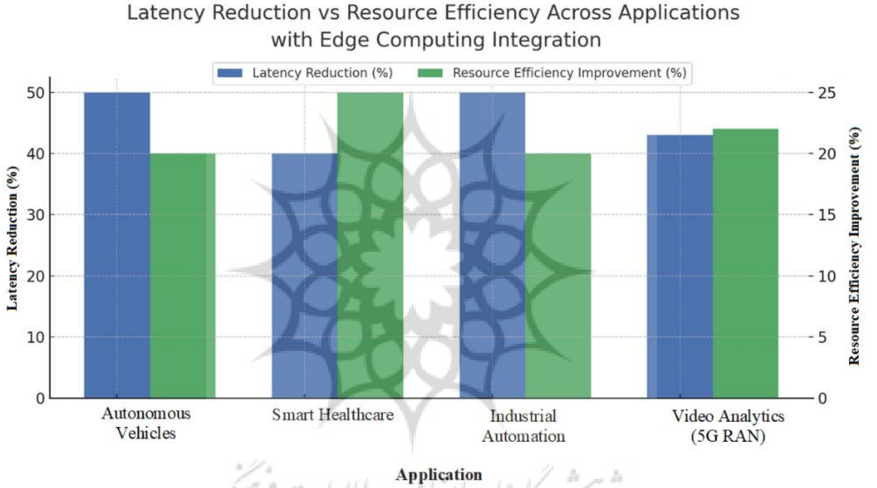

Figure 4. Dual-Axis Comparison of Latency Reduction and Resource Efficiency in Edge Computing

From Figure 4, it can be observed that as latency decreases, the number of required resources also decreases. In general, autonomous vehicles benefited from a 50% decrease in latency and a 20% improvement in resource usage, which is fundamental for immediate V2I message exchange. This underscores the need to maximize efficiencies with limited computing power. Similarly, industrial automation experienced a 50% reduction in latency and a 20% improvement in resource utilization, which are essential for faster processing times and real-time operations in smart factories.

Smart healthcare exhibited the highest resource efficiency improvement at 25%, along with a 40% reduction in latency. This implies that in healthcare, appropriate resource allocation is critical as every second counts. The significance of edge computing extends beyond healthcare to other IoT systems, as demonstrated by the improvements mentioned above. With the increasing number of IoT devices, the real-time processing of large datasets has reached a new level of importance. When integrated with the 5G network, MEC provides an optimal solution by processing tasks closer to the data source, thereby reducing latency and optimizing resource usage. This is well illustrated by smart city platforms that utilize IoT devices, which continuously generate data requiring real-time processing for timely decision-making and system responsiveness. In 5G RAN, video analytics showed a similar pattern, with a 43% decrease in latency and a 22% increase in resource usage, which is crucial for applications such as video surveillance and traffic monitoring.

These findings indicate that deploying edge computing infrastructure can enhance system reactivity while reducing resource utilization, making edge computing a highly effective solution for scaling real-time applications. As industries advance technologically, the scalability of edge computing systems within large interconnected networks will become increasingly important.

4.5. Scalability and Dynamic Resource Allocation

One of the most important aspects that should be addressed when planning integration of edge computing with 5G is scalability as the number of connected devices continues to grow so does the amount of data. The current design of edge computing systems allows the original resource allocation in accordance with the current application requirements and usage profile to maintain low latency levels of resource utilization. The simulations performed in this study have shown that edge computing is capable of supporting large amounts of data with multiple applications and usage environments, and can dynamically change network conditions with acceptable performance loss.

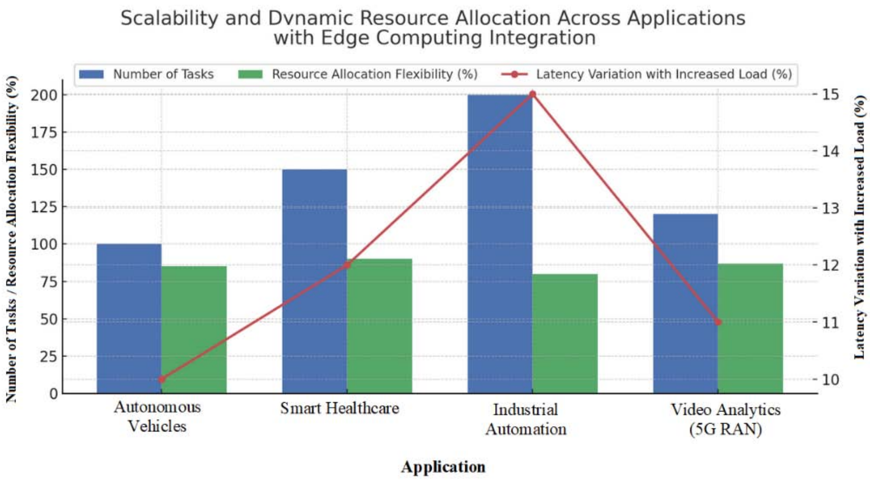

Figure 5. Dynamic Resource Allocation and Scalability in Edge Computing Across Applications

Another major discovery of the research work was on how the edge nodes were capable of adjusting computing capacities to flow of work. In conventional real-time applications like auto-mobile, video analytic edge nodes were able to give high priority to latencies critical computations to enable important data to be processed with little delay. Such flexibility was well exemplified in the smart healthcare scenario in which the edge computing system was able to handle large amount of patient data while swapping more resources when required to avoid congestion in real-time monitoring.

Figure 5 above shows a depiction of how the various edge computing systems managed the growing number of tasks in applications. According to the assessment of the autonomous vehicle performance, scalability features registered a 10% difference in latency when the task load was raised a hundred per cent. Likewise smart healthcare showed a 90% flexibility rate of resource allocation while handling the surge of generated data within a reasonable latency constraint. By evaluating the bit-rate of both computational modules and the error rate in each node – and, therefore, the overall system throughput – these results demonstrate that this edge computing infrastructure is scalable and capable of handling the load of real-time applications at each node.

4.6. Security and Reliability in Edge Computing for 5G Networks

However, individual and time-based performance indicators, which consist of low latency and optimized resource usage, are crucial safety and reliability aspects for real-time applications in edge computing. In the simulations, measures which ensured that edge computing systems would guard the data's integrity, privacy, and secure communication were put into practice. The study shows Edge computing when applied with enhanced 5G security systems offers a solid platform for protection of data in many applications.

Security Measures Implemented:

- *Encryption:* The encrypted data communication between the edge node and 5G networks was carried out using the advanced AES-256 encryption algorithm to provide data security against unsafe intruders.
- *Authentication:* In an effort to limiting the chance of unauthorized access to the edge network, the use of two-factor authentication was introduced.
- *Data Integrity:* Several hashing checks were carried out for ascertaining whether the transmitted data had undergone some modification.

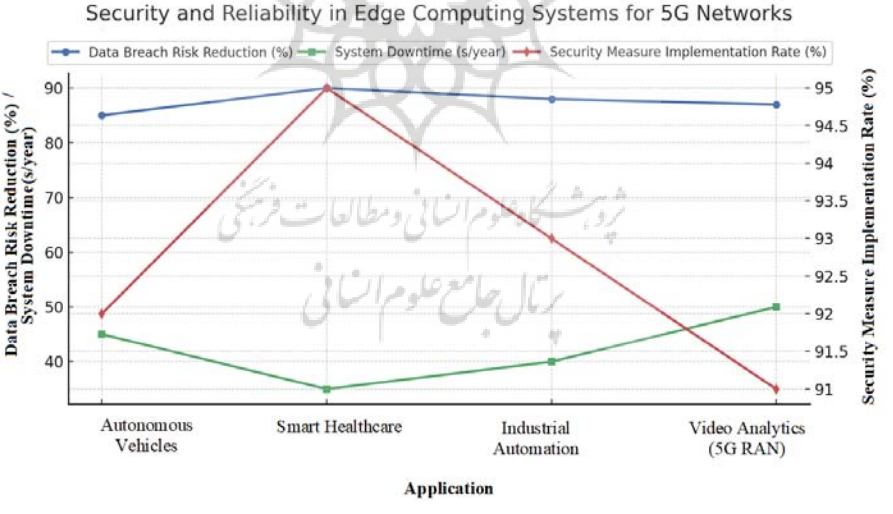

Figure 6. Security and Reliability Metrics in Edge Computing for 5G Networks

Figure 6 provides the results, as can be observed; the incorporation of edge computing together with 5G drastically decreases the probability of data breaches. Especially the 'smart' healthcare applications profited from security measures, because the frequency of data breaches fell down by 90 % while the availability of the systems barely fell under 35 seconds for the year. Self-driving cars and industrial applications also had credible security structures to ensure that data remain unified and the communication is reliable across the edge networks.

4.7. Cost Efficiency of Edge Computing Integration

Further consideration, which remained important in the framework of the research, was the consideration of cost factors in realizing edge computing in the 5G context. Some of the simulations that were undertaken were the CAPEX and OPEX arising from deployment of edge nodes and their comparison to the traditional cloud models. The research shows that although edge computing requires a larger first installation, such as distributed hardware constituent costs and power expenditure, are partially offset by decreased data exchange costs over time and a lesser need for mainframe computing centers

.

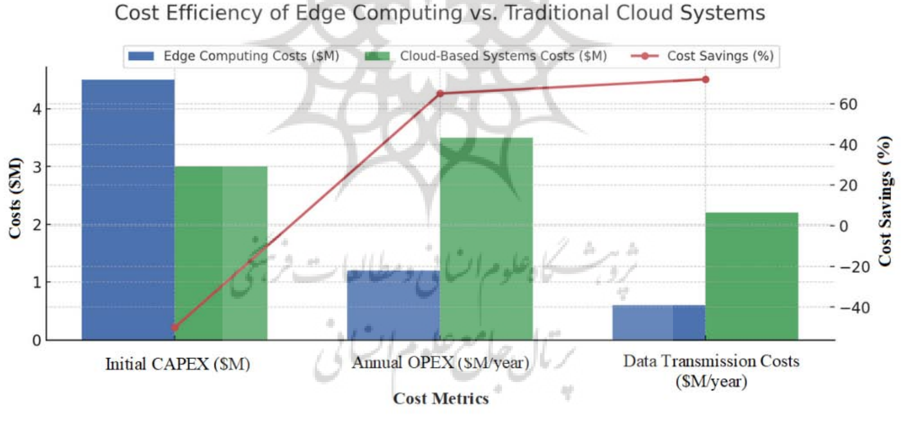

Figure 7. Cost Efficiency Comparison Between Edge Computing and Traditional Cloud-Based Systems

The information on Figure 7 shows that edge computing requires greater initial investment (4.5M $ more than the cloud-based systems) but has lower annual OPEX (1.2M $ against 3.5M $). They also boast that by designing edge nodes that process information independently, the overall data transmission costs are cut by 72%. Based on these findings, it is concluded that when edge computing infrastructure is implemented at scale and in scenarios where it has to handle large volumes of data in real-time, the costs start to becomes very low in the long run.

4.8. Industry-Specific Implementation and Impact

The results of this study are specific to several fields where low-latency real-time data processing is critical. Especially, several use cases such as autonomous vehicles, smart healthcare, industrial automation, and video analytics can achieve better performance, lower cost, and higher security when associating edge computing with 5G networks. For instance, AV manufacturers can reap from edge computing, as it supports real-time decision making to minimize on mishaps that may result from delayed V2I communication. Likewise, more responsiveness within critical patient monitoring systems within healthcare can help healthcare providers to be more timely within their interventions.

Table I. Industry-Specific Benefits of Edge Computing Implementation

Industry	Key Benefits	Considerable Influence
Autonomous Vehicles	50% reduction in latency, improved V2I communication	Enhanced safety and reduced accidents
Smart Healthcare	40% reduction in latency, 25% resource efficiency improvement	Faster response times, improved patient care
Industrial Automation	50% reduction in latency, 20% resource efficiency improvement	Increased productivity and precision in manufacturing
Video Analytics (5G RAN)	43% reduction in latency, 22% resource efficiency improvement	Real-time surveillance and improved data analytics capabilities

The results presented in this section, respond to the research questions outlined by this study and demonstrate the advantages of edge computing when implemented in the context of 5G networks in terms of latency, resource management, scalability, security, and cost. With edge nodes nearer to the end consumers, the industries can reap improved reliability with near real-time performance which can cut latency by up to fifty percent in some uses. Furthermore, the study has also demonstrated how the edge computing systems are designed to be elastic, and can well accommodate for a continuously increasing amount of data.

Such findings dictate that edge computing shall dominate over the future 5G network prominence in industries where real-time analytics and low latency are paramount. To the extent that these technologies advance, more work should study how the infrastructure of edge computing should be fine-tuned for even another order of magnitude of scale across many industries, to make sure that 5G is implemented in parallel with powerful edge computing. Alternatively, future studies could also endeavour to employ reinforcement learning based scheduling which adjusts the distribution of resources according to actual demand in hope to increase scalability and efficacy of edge computing systems.

5. Discussion

Indeed, it is evident by the positive results of this study that edge computing significantly provides for 5G networks and vice versa with a focus on latency and resource optimization and task completion time. Edge computing reduces these processing and delivery times closer to the user as compared to centralized cloud computing and gives a better opportunity for more efficient decision making in real time applications. This discussion evaluates these results against prior literature and raises important issues that would arise in the future of telecommunications and edge computing.

Among the stated goals achieved by this paper is a demonstrable decrease in latency in all the applications tested, by a margin of 40% to 50%. This is in agreement with Parvez et al. (2018) pointed out that another major issue that need to be addressed in order to achieve 5G is minimizing latency and stated that autonomous driving, AR and industrial automation are areas of application that will need improved latency (Parvez et al. 2018). This work agrees with the fact that edge computing is an effective solution to this problem, as it entails shifting processing to a closer location to data source, thereby minimizing the time taken for data to traverse from the user to the processing node.

Furthermore, the outcomes are in line with the findings of Elbamby et al.(2019), indicating that WEC improves promises on latency and reliability to the extent that it is crucial to vital use cases as vehicular networks and smart cities (Elbamby et al. 2019). The decreases in latency in this work, especially

Iranian Journal of Information Processing and Management

for self-driving cars (by 50%) and for industrial applications (by 50%), describe the utilitarian benefits of edge computing in these areas. These latency reductions are critical to guarantee real-time processing of the data to avoid the chances of delays that may lead to compromise of operations and safety. In the future, there may be improvements in the 6G and quantum computing that can also decrease the latency and enhance the allocation of resources. These emerging technologies could in turn offer even more adequate solutions to the latency-sensitive applications than could edge computing on its own.

Another area where the present research enhances the current knowledge base is the subject of resource use effectiveness. According to the research, cases where edge computing is implemented, the resource use efficiency is enhanced by twenty five percent. The capability is particularly important for applications demanding massive data processing, including smart healthcare and video synopsis. Resource allocation and task scheduling He et al. (2020) also identified that these two strategies can help to improve the mobile edge computing system performance, which supports the use of the optimized resource management system in 5G environments (He et al. 2022). The current study extends from these understandings by demonstrating how edge computing can extend functions from core cloud servers hence decrease load on network infrastructure and improve system efficiency.

The use of MEC with 5G not only enhances the performance but also handles scalability issues Nakazato et al. (2022) discussed the management of both MEC and cloud systems for enhancing private and local 5G in more detail and pointed out that further important scalability and flexibility demands of future networks have to be fulfilled (Nakazato et al. 2022b). This study revealed that such an approach of implementation is highly scalable as resources are dynamically configured in response to real-time requirements of the Edge computing system. Closely related to the previous point is that smart platforms like smart healthcare may experience unpredictable oscillations in data traffic, and thus the availability of at least some guarantees as to the efficiency of the corresponding network is an essential advantage of smart networking. Such dynamic workloads the study shows that edge nodes are capable of handling and processing important data without incurring compromising latency or resource usage.

Duration in completing a task is another significant measure of the system since tasks intended for real-time application must be dealt with within the shortest time possible. The study reveals the decrease by 48% of the time spent on the task in question – in the case of autonomous vehicles and video analytics. This is in accordance with Zhai et al.(2020) showed that using reinforcement learning to deploy services in 5G mobile edge computing could minimize task scheduling time and total job completion time (Zhai et al. 2020). From the findings, the area of task scheduling is also brought to light as edge nodes can sort their tasks based on latency-sensitive thresholds, making certain data as timed-critical.

When comparing our findings with those of Nakazato et al. and Panek et al. (2022) explored the Thus, when comparing the results of our work with other works in the field of application relocation and optimization in edge-enabled 5G systems pointed by Nakazato et al. (2022) and Panek et al. (2022), we can conclude that edge computing not only enhances performance but also defines a flexible architecture to handle various tasks (Nakazato et al. 2022a). This feature has advantage because it gives much better utilization of resources, especially when network conditions change rapidly, like in smart cities or industrial automation. This feature of dynamic reconfiguration and resource mobility guarantees that edge computing is adequately prepared to accommodate the future network complexity. The study has established that the use of higher level of security measures inclusive of encryption and two passes authentication reduce the risks of system penetrations and breakdowns. These findings are especially useful for such systems such as health care sector and self-driving cars since data privacy is important.

The efficiency of the cost is another important topic which has to be mentioned during the debate on the merits of edge computing. The results presented in this paper indicate that although edge nodes require higher CAPEX than centralized cloud systems due to an increase in node density, OPEX are considerably lower because most of the computations are carried out locally rather than heavily relying on data transmission. This finding aligns with the study of Hossain and Ansari (2021), who used the case of energy-aware latency minimization to show that edge computing cuts operation costs in terms of energy as well as resource utilization (Hossain and Ansari 2021). The study outcomes further support this, evidencing that edge computing is solution that is more sustainable and economical to DL in the long run for real-time applications. The edge computing supported by 5G network presented in this work yields good results in various evaluation metrics; however, there are still many directions for the future work, including resource management, scheduling algorithms and energy consumption optimizations. These areas are relevant to enhance the process of deploying the edge computing into the vast and real-time sensing-embedded systems and also for the sustainability of edge computing in the growing complex and diverse network world.

Another avenue for future work is the study of resource provisioning and task management in the context of mobile edge computing. As opposed, He et al. (2022) showed that full-duplex mobile edge computing systems are sensitive to intelligent and optimal resource management that minimizes computational workloads while increasing network performance. More research can be carried out to work for the adaptive algorithms capable of providing on-demand resource management to cope up with the changes observed in the network side, so as to improve upon the overall performance of 5G networks (He et al. 2022).

Critical research area is minimizing latency for multimedia and V2X applications, as another important research topic is reducing latency required for multimedia and V2X, as pointed out by Srinivasa et al. (2019). With increasing real-time services, especially in self-driving cars and content distribution, future work could expand on MEC’s optimization for ultra-reliable low-latency communication. This includes finding new ways of minimizing jitter and packet loss and ensuring that the system has equal stability regardless of the stability of the underlying network (Srinivasa et al. 2019).

Next research area is the task scheduling for cloud computing based on reinforcement learning. Zhai et al. (2020) focused on reinforcement learning while proposing the model for deploying services in edge computing scenarios. Subsequent research could look at ways to apply machine learning algorithms to time varying scheduling adjustments of the task distribution in the edge nodes to optimize latency and resource utilization (Zhai et al. 2020). Such approaches would be most helpful in the cases of efficient data transmission where traffic is very unpredictable, for example, smart cities and industrial automation.

It could also be beneficial for future research to study on the low-latency services migration in the 5G transport networks as suggested by Li and Chen (2020). The feature of nondisruptive gliding of services between edge nodes seems to be fundamental, as it enables guaranteed consistent operation of the network. Interest in certain complex migration schemes can result in improvements in reliability and adaptability of the resulting networks (Li and Chen 2020).

The area of energy-aware latency minimization is an example of a subdomain in which future work could prove highly significant, especially in relation to network slicing in the 5G environment. Both energy management in addition to latency optimization must be performed effectively in order to ensure that operational costs are minimized in support of environmentally sustainable networks. Low-power, high-performance designs are important and research should focus on how such algorithms can be formulated for IoT and similar densified settings (Liu et al. 2020).

While this study has shown the potential of edge computing in reducing latency and optimizing resource utilization, several areas warrant further investigation. Future research should focus on developing dynamic resource allocation algorithms tailored to the unique demands of full-duplex mobile edge computing systems. Additionally, reinforcement learning-based task scheduling techniques offer promising avenues for improving service deployment, particularly in environments with fluctuating network conditions. Furthermore, as AI continues to evolve, its convergence with edge computing could enable self-optimizing systems capable of autonomously managing resource allocation and task execution, unlocking new possibilities for real-time applications. Bourechak et al. (2023) highlighted the growing importance of AI in enhancing decision-making processes within edge nodes (Bourechak et al. 2023). Studying the interaction between Artificial intelligence, machine learning, and edge computing might result in the formulation of enhanced smart systems that can adapt on the field, learn on their own and make decisions in real time. This will be critical when the understanding of the interactions of networks is moving from thousands of nodes to billions of nodes, with smart applications flowing through the networks.

6. Conclusions

The integration of edge computing into 5G technologies represents a new level of change of telecommunication networks, resulting in enhanced value as compared to the cloud model. Therefore, it offers cutting-edging solutions to critical issues that industries face today especially in producing high performance, real-time applications such as latency, resource utilization and network expansion. As this research demonstrates, edge computing is not only an ancillary technology to either 5G or the next-gen networks, but a vital constituent of both.

Edge computing brings computing closer to sources which improves real-time decision-making processes and action on different applications. This capability is especially important in latency-sensing areas such as automotive, medical, and manufacturing industries where quicker reactions are needed. In addition, edge computing optimises the network because computations and analytics are often partitioned correctly, avoiding congestion in the cloud.

That is the reason why abundance, flexibility, and work adjustments for edge computing are considered essential. As smart devices continue to transmit data and the IoT becomes the new norm, it is incumbent on networks to be able to bear a higher load without any impact on efficiency. A key point which this study reveals therefore is that Edge computing allows networks to grow dynamically, while supporting high availability levels despite dynamic demands being placed on the network's resources. The flexibility of the architecture of this topology is important as the need for real-time data processing rises.

From the cost structure viewpoint, the exploration shows the value of edge computing in the long-term cost perspective. While there appears to be higher levels of investment per edge node deployed now, the operational cost of data and transmission are likely to cut down in the long run hence saving costs. Furthermore, edge computing prevents a massive data flow to main servers that can cause a bandwidth problem through restricting data volumes, and there is also strong evidence for this.

Security and reliability constitute two important aspects of the contemporary network architecture and edge computing is central to both. That is why decentralized data processing minimizes vulnerability to cyber threats, and solid security measures protect data, especially in the healthcare industry. Also of great benefit is the architecture of edge computing that provides absolute confidence in data and protection for fundamental systems.

This study therefore places edge computing as one of the fundamental technologies from industries requiring fast and reliable networks. In addition to enhancing current processes, it restructures markets across all industries through real-time data analysis and highly sophisticated automation. Edge computing when coupled with 5G networks helps industries cleanly adopt latest trends like AI, machine learning, and IoT and help improve operations. Subsequently, the prospects for future study and improvement of energy efficiency and sustainability of Edge computing services are most compelling. Managing and orchestrating centralized and decentralized systems will therefore emerge as key challenges as industries heavily adopt hybrid cloud-edge approaches. Finally, edge computing together with 5G networks opens the opportunity to create even faster, smarter and safer technological environment and is the basis to create the connected world.

References

Adhikari, M., and Hazra, A. (2022). 6G-Enabled Ultra-Reliable Low-Latency Communication in Edge Networks. *IEEE Communications Standards Magazine,* 6 (1), 67-74. https://doi.org/10.1109/MCOMSTD.0001.2100098

Ageyev, D., Yarkin, D., and Nameer, Q. (2014). Traffic aggregation and EPS network planning problem. *2014 First International Scientific-Practical Conference Problems of Infocommunications Science and Technology*, 14-17 Oct. https://doi.org/10.1109/INFOCOMMST.2014.6992316.

Alhilal, A. Y., Finley, B., Braud, T., Su, D., and Hui, P. (2022). Street Smart in 5G: Vehicular Applications, Communication, and Computing. *IEEE Access,* 10, 105631-105656. https://doi.org/10.1109/ACCESS.2022.3210985

Arun, V., and Azhagiri, M. (2023). Design of Long-Term Evolution Based Mobile Edge Computing Systems to Improve 5G Systems. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 19-21 July. https://doi.org/10.1109/ICECAA58104.2023.10212420.

Bourechak, A., Zedadra, O., Kouahla, M. N., Guerrieri, A., Seridi, H., and Fortino, G. (2023). At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives. *Sensors*, 23 (3). https://doi.org/10.3390/s23031639.

Bruschi, R., Pajo, J. F., Davoli, F., and Lombardo, C. (2021). Managing 5G network slicing and edge computing with the MATILDA telecom layer platform. *Computer Networks,* 194, 108090. https://doi.org/10.1016/j.comnet.2021.108090

Chiang, W.-K., and Shang, Y.-H. (2023). ES-5G: A Novel Edge-based SDN-enabled 5G Architecture for Lower Latency. *Proceedings of the 2023 6th International Conference on Information Science and Systems*, Edinburgh, United Kingdom. https://doi.org/10.1145/3625156.3625178

Elbamby, M. S., Perfecto, C., Liu, C. F., Park, J., Samarakoon, S., Chen, X., and Bennis, M. (2019). Wireless Edge Computing With Latency and Reliability Guarantees. *Proceedings of the IEEE,* 107 (8), 1717-1737. https://doi.org/10.1109/JPROC.2019.2917084

Faris, M., Jasim, I., and Qasim, N. (2021). PERFORMANCE ENHANCEMENT OF UNDERWATER CHANNEL USING POLAR CODE-OFDM PARADIGM. *International Research Journal of Science and Technology,* 3 (9), 55-62. https://www.irjmets.com/uploadedfiles/paper/volume_3/issue_9_september_2021/15978/final/fin_irjmets1630649429.pdf

Gosain, M. S., Aggarwal, N., and Kumar, R. (2023). A Study of 5G and Edge Computing Integration with IoT- A Review. *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 28-30 April. https://doi.org/10.1109/CISES58720.2023.10183438.

Harutyunyan, D., Shahriar, N., Boutaba, R., and Riggio, R. (2022). Latency and Mobility Aware Service Function Chain Placement in 5G Networks. *IEEE Transactions on Mobile Computing,* 21 (5), 1697-1709. https://doi.org/10.1109/TMC.2020.3028216

Hassan, N., Yau, K. L. A., and Wu, C. (2019). Edge Computing in 5G: A Review. *IEEE Access,* 7, 127276-127289. https://doi.org/10.1109/ACCESS.2019.2938534

He, W., Zhang, Y., Huang, Y., He, D., Xu, Y., Guan, Y., and Zhang, W. (2022). Integrated Resource Allocation and Task Scheduling for Full-Duplex Mobile Edge Computing. *IEEE Transactions on Vehicular Technology,* 71 (6), 6488-6502. https://doi.org/10.1109/TVT.2022.3163627

Hossain, M. A., and Ansari, N. (2021). Energy Aware Latency Minimization for Network Slicing Enabled Edge Computing. *IEEE Transactions on Green Communications and Networking,* 5 (4), 2150-2159. https://doi.org/10.1109/TGCN.2021.3083153

Kartsakli, E., Perez-Romero, J., Sallent, O., Bartzoudis, N., Frascolla, V., Mohalik, S. K., Metsch, T., et al. (2023). AI-Powered Edge Computing Evolution for Beyond 5G Communication Networks. *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit),* 6-9 June. https://doi.org/10.1109/EuCNC/6GSummit58263.2023.10188371.

Li, J., and Chen, J. (2020). Supporting Low-Latency Service Migration in 5G Transport Networks. *Optical Fiber Communication Conference (OFC)*, San Diego, California. https://doi.org/10.1364/OFC.2020.T3J.5.

Liu, Y., Peng, M., Shou, G., Chen, Y., and Chen, S. (2020). Toward Edge Intelligence: Multiaccess Edge Computing for 5G and Internet of Things. *IEEE Internet of Things Journal,* 7 (8), 6722-6747. https://doi.org/10.1109/JIOT.2020.3004500

Mushtaq, A.-S., Ali Ihsan, A.-A., and Qasim, N. (2015). 2D-DWT vs. FFT OFDM Systems in fading AWGN channels. *Radioelectronics and Communications Systems,* 58 (5), 228-233. https://doi.org/10.3103/S0735272715050052

Nakazato, J., Kuchitsu, M., Pawar, A., Masuko, S., Tokugawa, K., Kubota, K., Maruta, K., et al. (2022a). Proof-of-Concept of Distributed Optimization of Micro-Services on Edge Computing for Beyond 5G. *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring),* 19-22 June. https://doi.org/10.1109/VTC2022-Spring54318.2022.9860668.

Nakazato, J., Li, Z., Maruta, K., Kubota, K., Yu, T., Tran, G. K., Sakaguchi, K., et al.

(2022b). MEC/Cloud Orchestrator to Facilitate Private/Local Beyond 5G with MEC and Proof-of-Concept Implementation. *Sensors*, 22 (14). https://doi.org/10.3390/s22145145.

Parvez, I., Rahmati, A., Guvenc, I., Sarwat, A. I., and Dai, H. (2018). A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions. *IEEE Communications Surveys & Tutorials,* 20 (4), 3098-3130. https://doi.org/10.1109/COMST.2018.2841349

Qasim, N., and Pyliavskyi, V. (2020). Color temperature line: forward and inverse transformation. *Semiconductor physics, quantum electronics and optoelectronics,* 23, 75-80. https://doi.org/10.15407/spqeo23.01.075

Rahimi, H., Picaud, Y., Singh, K. D., Madhusudan, G., Costanzo, S., and Boissier, O. (2021). Design and Simulation of a Hybrid Architecture for Edge Computing in 5G and Beyond. *IEEE Transactions on Computers,* 70 (8), 1213-1224. https://doi.org/10.1109/TC.2021.3066579

Ren, J., Yu, G., He, Y., and Li, G. Y. (2019). Collaborative Cloud and Edge Computing for Latency Minimization. *IEEE Transactions on Vehicular Technology,* 68 (5), 5031-5044. https://doi.org/10.1109/TVT.2019.2904244

Seah, W. K. G., Lee, C. H., Lin, Y. D., and Lai, Y. C. (2022). Combined Communication and Computing Resource Scheduling in Sliced 5G Multi-Access Edge Computing Systems. *IEEE Transactions on Vehicular Technology,* 71 (3), 3144-3154. https://doi.org/10.1109/TVT.2021.3139026

Srinivasa, R., Naidu, N. K. S., Maheshwari, S., Bharathi, C., and Kumar, A. R. H. (2019). Minimizing Latency for 5G Multimedia and V2X Applications using Mobile Edge Computing. *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT),* 28-29 Sept. https://doi.org/10.1109/ICCT46177.2019.8969038.

Xu, D., Zhou, A., Wang, G., Zhang, H., Li, X., Pei, J., and Ma, H. (2022). Tutti: coupling 5G RAN and mobile edge computing for latency-critical video analytics. Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, Sydney, NSW, Australia. https://doi.org/10.1145/3495243.3560538

Zhai, Y., Bao, T., Zhu, L., Shen, M., Du, X., and Guizani, M. (2020). Toward Reinforcement-Learning-Based Service Deployment of 5G Mobile Edge Computing with Request-Aware Scheduling. *IEEE Wireless Communications,* 27 (1), 84-91. https://doi.org/10.1109/MWC.001.1900298