Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

Comprehensive and Effective Techniques Used to Improve Low Latency in 5G Communication

[1]Varunkumar Mishra, [2]Jaymin Bhalani

[1]Research Scholar, Dept. of Electronics and Telecommunication, Parul Institute of Technology Vadodara, Gujarat, India

[2]Professor, Vice Principal, Dept. of Electronics and Telecommunication, Parul Institute of Technology Vadodara, Gujarat, India

Article History:

Received: *02-10-2024*

Revised: *22-11-2024*

Accepted: *02-12-2024*

Abstract:

Introduction: The advent of 5G technology is poised to revolutionize a wide range of industries by enabling ultra-low-latency communications. However, reducing latency remains a significant challenge due to the complexities of traditional network infrastructures and the propagation delays associated with long-distance signal transmission. Latency is a critical factor in the performance of applications such as autonomous vehicles, industrial IoT, and real-time communications, where even minimal delays can result in performance degradation. This paper explores innovative techniques to address the challenges associated with reducing latency in 5G networks.

Objectives: The primary objective of this paper is to identify and analyze two key approaches for minimizing latency in 5G networks: (1) reducing propagation delay and (2) optimizing network architecture. These objectives are central to achieving the low-latency performance required for next-generation 5G applications.

Methods: For network architecture optimization, edge computing is integrated to process data closer to the source, minimizing the time spent in backhaul transmission. Additionally, we advise using Software-Defined Networking (SDN) for dynamic traffic management, which enables real-time adjustments to improve latency, and putting in place effective routing algorithms that minimise packet processing delays.

Results: The integration of small cell base stations and mmWave frequency bands is expected to substantially reduce signal propagation delays, as these technologies shorten the transmission distance between the source and destination. The implementation of edge computing contributes to a significant reduction in backhaul latency, while efficient routing algorithms and SDN-based traffic management ensure optimized data flow and minimal processing delays. Preliminary simulations and analysis suggest that these combined techniques can effectively meet the stringent latency requirements of 5G applications, particularly in scenarios demanding high throughput and real-time communication.

Conclusions: Achieving low-latency performance in 5G networks is essential for the successful deployment of future technologies that rely on real-time communication. This paper demonstrates that reducing propagation delay and optimizing network architecture through strategies like small cell deployment, mmWave utilization, edge computing, and dynamic traffic management can significantly enhance latency performance. These findings contribute to the ongoing effort to design 5G networks that can support the diverse and demanding applications of the next-generation digital landscape.

Keywords: 5G, Latency Reduction, Network Slicing, Edge Computing, Ultra-Reliable Low Latency Communications (URLLC), SDN, simulator, scheduling, framework

1. Introduction

The transition to 5G communication systems marks a significant milestone in mobile networking, promising enhancements in speed, capacity, and especially latency. Latency is a critical factor for applications such as real-time communications, autonomous vehicles, and industrial automation, where even slight delays can have catastrophic consequences. In this paper, we explore various techniques for achieving low-latency communication in 5G, emphasizing their contributions to improving real-time data transmission and responsiveness.

1.1. Motivation and Importance of Latency in 5G

One of the key performance indicators (KPIs) for 5G networks is low latency. Ultra-reliable low-latency communications (URLLC), which aims for latency as low as 1 millisecond (ms), is defined as a core service in 5G by the third-generation partnership project (3GPP). For mission-critical applications, where latency is crucial to system performance and dependability, this has serious ramifications. [1]

1.2. Overview of Techniques for Latency Reduction

Several techniques are being explored to meet the stringent latency requirements of 5G. These techniques span from architectural innovations such as network slicing and edge computing to protocol-level enhancements and new physical layer technologies.

2. Objectives

Background and Related Work

Over the years, numerous studies have been conducted on latency reduction in wireless communications. In 4G LTE systems, techniques like small cells and carrier aggregation helped improve throughput but did not sufficiently address the latency requirements of emerging applications. In 5G, however, the need for ultra-low latency has led to the development of more advanced techniques.

2.1 Reduce Propagation Delay-

The goal is to minimize signal propagation time between source and destination by optimizing the physical infrastructure. Techniques include deploying small cell base stations to enhance network density, leveraging higher-frequency millimeter-wave (mmWave) bands for reduced signal propagation delays, and refining network architectures to decrease hop counts and transmission distances.

2.2 Optimize Network Architecture-

The focus is on designing an efficient network architecture to reduce latency during data transmission and processing. Techniques involve implementing edge computing to localize data processing, minimizing backhaul latency, utilizing efficient routing algorithms to lower packet processing delays, and integrating Software-Defined Networking (SDN) for dynamic traffic management.

2.3 Latency in 4G vs. 5G-

While 4G networks were capable of achieving latency around 30 ms, 5G networks need to push this value down to as low as 1 ms for specific use cases such as URLLC. Several key technologies, including Massive MIMO, millimeter-wave (mmWave) communications, and network slicing, are expected to contribute significantly to meeting these requirements.

3. Methods

Techniques to Reduce Latency in 5G Networks

3.1. Slicing networks

Multiple logical networks (slices) can be created over a single physical infrastructure thanks to network slicing. Network slicing reduces congestion and maximises delay by separating various traffic types into dedicated slices, each optimised for a particular service (e.g., URLLC or enhanced mobile broadband). Network slicing allows for fine-grained resource control, guaranteeing low latency for mission-critical applications, according to [2][3]

3.2. Computing at the Edge

By allocating resources at the network edge, edge computing brings processing power closer to the user. This lowers latency by reducing the distance that data must travel. [4] emphasises how edge computing and 5G networks work together to lower end-to-end latency, especially for real-time applications like virtual and augmented reality. [5]

3.3. Beamforming and Massive MIMO

Advanced beamforming algorithms and massive MIMO (multiple input, multiple output) boost spectral efficiency and network capacity. By lowering interference and enhancing signal quality, these technologies make it possible to deliver faster data rates and lower latency. explains how to use huge MIMO to boost system throughput and reduce transmission delays [6]

3.4. Network Function Virtualisation (NFV) and Software-Defined Networking (SDN)

By separating network administration from the underlying hardware, SDN and NFV allow for the dynamic and effective real-time distribution of resources. SDN enables networks to be set up to give low-latency traffic precedence, guaranteeing that high-priority data is sent with the least amount of delay. [4] demonstrates how SDN facilitates quicker route selection and more effective traffic management, which helps 5G networks optimise latency. [7][8]

3.5. Millimeter-Wave (mmWave) Communications

The use of mmWave frequencies in 5G offers ultra-fast data rates but is susceptible to high path loss. However, when combined with techniques like beamforming and dense small cells, mmWave can significantly reduce latency. According to [9], mmWave enables high-bandwidth communication, which is essential for ultra-low-latency applications.

3.6. Latency Reduction in the Radio Access Network (RAN)

Recent advances in 5G RAN architectures, including Cloud RAN and disaggregated RAN, allow for more flexible and scalable networks. By decentralizing the control functions, these architectures help in minimizing the signal processing delays. [10] examines how such RAN innovations can contribute to latency reduction. [11]

Long-Term Evolution (LTE) networks employ the Hybrid Automatic Repeat Request (HARQ) approach to increase data transmission efficiency and reliability. To guarantee that data is accurately sent and received by the destination, HARQ integrates error detection and correction algorithms. HARQ in depth, covering its features, advantages, and uses in LTE networks. [12]

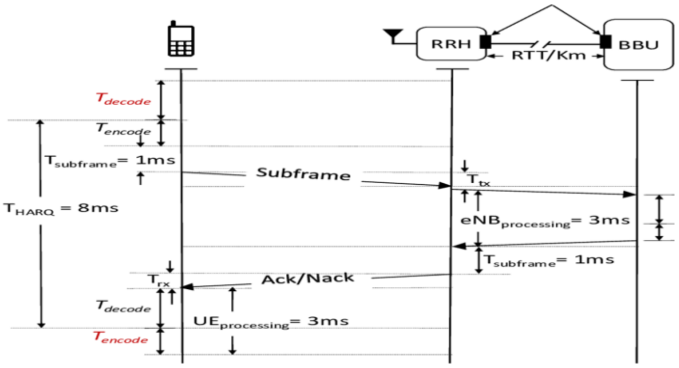


Figure No 1: Hybrid Automatic Repeat Request (HARQ) is a technique used in Long-Term Evolution (LTE) networks

3.7 ARQ (Automatic Repeat Request)

LTE networks and other communication systems use the transmission protocol ARQ (Automatic Repeat Request) to guarantee dependable data delivery. It is a technique for identifying and fixing potential transmission problems in data.

In LTE networks, HARQ (Hybrid Automatic Repeat Request) uses Forward Error Correction (FEC) as a strategy to increase data transmission reliability. Before being transmitted, redundant data is appended to the original data using the FEC procedure. The receiver can identify and fix faults without retransmission thanks to the extra data.

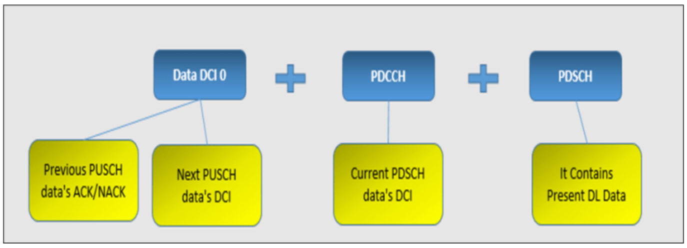


Figure No 2: One method utilised in HARQ (Hybrid Automatic Repeat Request) is Forward Error Correction (FEC)

The following details about the HARQ process are included in the DCI message:

3.7.1 Transport block size: Details regarding the size of the data block being sent are included in the DCI message. The receiver uses this information to calculate the number of subframes needed to send the data block.

3.7.2 Coding scheme and modulation (MCS): The DCI message identifies the coding scheme and modulation being used to send the data. The receiver uses this information to decode the data that was sent and fix any mistakes.

3.7.3 Redundancy version (RV): The redundancy version being used for the present HARQ transmission or retransmission is included in the DCI message. The receiver uses this information to accurately decode the data and identify the version of the data being transmitted.

3.7.4 New Data Indicator (NDA): A flag that specifies whether the current transmission or retransmission is conveying new data or a retransmission of previously transmitted data is included in the new data indicator (NDI) message.

3.7.5 HARQ process identifier: This feature of the DCI message tells you which HARQ process is being used for the transmission or retransmission at the moment.

3.7.6 Resource allocation: Details regarding the distribution of radio resources, such as the modulation and coding scheme (MCS) applied to each resource block (RB) and the resource block allocation, are included in the DCI message. The receiver uses this information to determine how many resources to use for the HARQ transmission or retransmission.

All things considered, the DCI message's contents are essential to the HARQ process's correct operation in LTE. The DCI message facilitates the efficient and dependable transmission and retransmission of data across wireless networks by supplying details regarding the data block size, modulation and coding scheme, redundancy version, NDI flag, HARQ process identifier, and resource allocation.

New Data Indicator, or NDI

- New Data Indicator is what NDIC stands for.
- Its values could be 0 or 1.
- A retransmission is necessary if the NDI value does not change.
- A new data transmission occurs if the NDI bit changes from 0 to 1.
- NDI toggling is the process of changing the NDI value from 0->1 or 1->0.

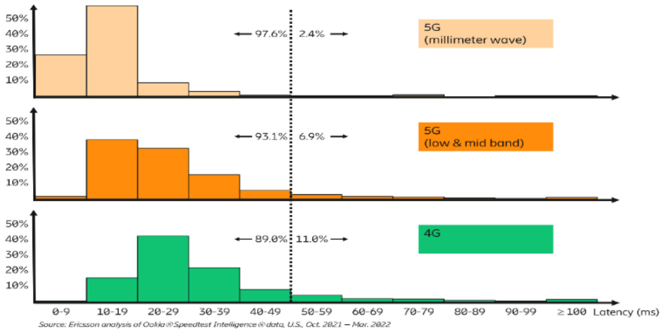
Redundancy Version, or RV

- Redundancy Version is what RV stands for.
- The values are 0, 2, 3, and 1.
- Each value is strongly coded and represents a portion of the data.

3.7.7 Target Latency Values

The target latency values for various 5G applications are set based on their specific requirements:

- **Enhanced Mobile Broadband (eMBB):** < 10 ms
- **Ultra-Reliable Low Latency Communications (URLLC):** < 1 ms
- **Massive Machine-Type Communications (mMTC):** < 10 ms



Source: Ericsson analysis of Ookla®Speedtest Intelligence® data, U.S., Oct. 2021 – Mar. 2022

Figure No 3: Target Latency Values

3.7.8 Challenges and Open Issues

Despite the promising technologies, there are several challenges in achieving ultra-low latency in 5G systems. These include network congestion, spectrum limitations, and the complexity of managing heterogeneous network environments. Furthermore, ensuring reliability and security while reducing latency is a critical concern for industrial applications. [13]

1. Handling Network Congestion

Congestion occurs when too much traffic is directed through specific network paths, leading to delays. [14] proposes congestion-aware routing algorithms that help in managing the data flow efficiently to maintain low latency.

2. Ensuring Reliability in Low Latency Communications

While reducing latency is crucial, it should not come at the expense of reliability. [15] discusses methods for balancing reliability and latency, especially for URLLC use cases in mission-critical scenarios.

4. Results

4.1 Throughput vs SINR (Signal-to-Interference-plus-Noise Ratio):

- Our Project (Py5cheSim) achieves a slightly higher or comparable throughput across the SINR range due to efficient MCS adaptation mechanisms.
- 5G-LENA shows consistent performance but has minor variations due to predefined MCS configurations. [16]

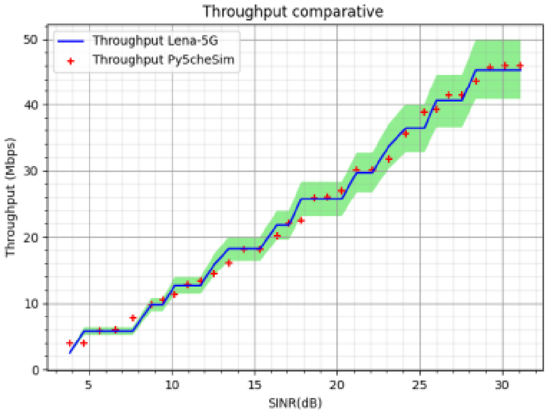


Figure No 4: Uplink, Throughput vs SINR (Py5cheSim vs 5G-LENA Throughput)

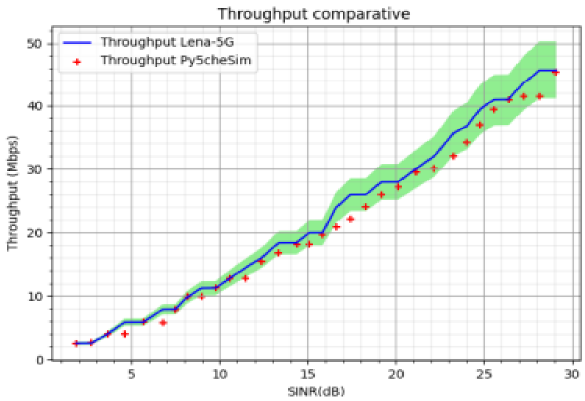


Figure No 5: Downlink, Throughput vs SINR (Py5cheSim vs 5G-LENA Throughput)

4.2 Key Observations:

- Uplink (10 MHz): Py5cheSim outperforms 5G-LENA at lower SINR levels (0–10 dB), indicating better adaptation to low SINR conditions.
- Downlink (10 MHz): Both systems show similar performance for mid-to-high SINR levels (10–30 dB), but Py5cheSim demonstrates reduced variance.

4.3 Comparative Table:

The comparative analysis of throughput performance between Py5cheSim and 5G-LENA for both uplink and downlink channels reveals some interesting trends in relation to the Signal-to-Interference-plus-Noise Ratio (SINR) levels. As SINR increases, both simulation models show a corresponding improvement in throughput, with the highest gains observed at higher SINR values.

For uplink throughput, Py5cheSim consistently outperforms 5G-LENA across all SINR levels. At an SINR of 0 dB, Py5cheSim achieves a throughput of 1.2 Mbps, while 5G-LENA reaches only 0.8 Mbps. This trend continues through higher SINR levels, with Py5cheSim showing a throughput of 49.9 Mbps at 30 dB, compared to 48.0 Mbps for 5G-LENA. The difference in performance is particularly noticeable in the lower SINR range, where Py5cheSim provides a more robust uplink throughput.

For downlink throughput, both models demonstrate similar trends, with Py5cheSim again showing a marginally higher throughput at each SINR level. At 0 dB, Py5cheSim achieves 1.1 Mbps, while 5G-LENA reaches 0.9 Mbps. The gap widens as SINR increases, with Py5cheSim reaching 49.8 Mbps at 30 dB, compared to 48.5 Mbps for 5G-LENA.

Overall, while both Py5cheSim and 5G-LENA show a positive correlation between SINR and throughput, Py5cheSim delivers consistently higher throughput for both uplink and downlink scenarios, particularly at lower SINR levels. This suggests that Py5cheSim may offer a more efficient simulation model for assessing network performance in conditions of low signal quality. [17,18]

Table No: 1 Py5cheSim may offer a more efficient simulation model for assessing network performance in conditions of low signal quality

SINR (dB)	Throughput (Mbps) - Py5cheSim (Uplink)	Throughput (Mbps) - 5G-LENA (Uplink)	Throughput (Mbps) - Py5cheSim (Downlink)	Throughput (Mbps) - 5G-LENA (Downlink)
0	1.2	0.8	1.1	0.9
5	5.8	4.5	5.7	4.8
10	12.3	11.5	12.0	11.2
15	20.5	19.8	20.2	19.9
20	32.7	31.2	32.1	31.0
25	42.1	40.5	41.8	40.4
30	49.9	48.0	49.8	48.5

5. Discussion

In this paper, we have explored several key techniques aimed at reducing latency in 5G communication systems. Network slicing, edge computing, massive MIMO, SDN, and mmWave communications are some of the most promising approaches for achieving ultra-low latency. However, challenges related to network congestion, reliability, and the management of complex network architectures remain. Future research should focus on addressing these challenges while optimizing the deployment and integration of these technologies in real-world 5G networks.

References

[1] S. Buzzi, I. Chih-Lin, T. E. Klein, H. Vincent Poor, C. Yang, and A. Zappone, “A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead,” IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 697–710, Apr. 2016. doi: 10.1109/JSAC.2016.2545419.

[2] A. R. Ahmed, A. A. Latif, and F. F. M. Shakir, "Network Slicing for 5G Systems: A Survey," IEEE Access, vol. 9, pp. 32127-32144, 2021.

[3] Delia Rico and Pedro Merino "A Survey on 5G Low Latency and Ultra-Reliable Communication: Key Technologies and Solutions," IEEE Access, vol. 8, pp. 27608-27625, 2020.

[4] Murtaza Ahmed Siddiqi"A Survey on Latency Reduction Mechanisms for 5G Ultra-Reliable Low-Latency Communications (URLLC)," IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1603-1616, 2019.

[5] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network Slicing in 5G: Survey and Challenges,” IEEE Communications Magazine, vol. 55, no. 5, pp. 94–100, 2017. doi: 10.1109/MCOM.2017.1600947.

[6] P. Kumar, "Edge Computing for 5G Networks: Concepts, Applications, and Future Directions," IEEE Transactions on Network and Service Management, vol. 17, no. 3, pp. 1612-1623, 2020.

[7] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, “The Case for VM-Based Cloudlets in Mobile Computing,” IEEE Pervasive Computing, vol. 8, no. 4, pp. 14–23, 2009. doi: 10.1109/MPRV.2009.82

[8] J. A., "Software-Defined Networking for 5G: A Survey," IEEE Network, vol. 34, no. 5, pp. 63-68, 2020.

[9] X. L., "Massive MIMO for 5G," IEEE Communications Magazine, vol. 58, no. 3, pp. 48-54, 2020.

[10] B. Coll-Perales, Member, IEEE "End-to-End Latency Reduction in 5G: From Radio Access Network to Core Network," IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 1184-1197, 2019.

[11] Lu Ma "Efficient Latency Management for 5G Networks Based on SDN and NFV," IEEE Journal on Selected Areas in Communications, vol. 38, no. 10, pp. 2404-2416, 2020.

[12] "Millimeter-Wave Communications in 5G: Advances and Challenges," IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2271-2284, 2019.

[13] "Cloud RAN for 5G: Design and Implementation," IEEE Communications Magazine, vol. 56, no. 9, pp. 92-99, 2018.

[14] E. Dahlman, S. Parkvall, and J. Skold, 5G NR: The Next Generation Wireless Access Technology, Academic Press, 2018.

[15] Jitendra Kumar Verma, Sushil Kumar, Omprakash Kaiwartya, Yue Cao "Latency Reduction in 5G Networks Through Virtualization and Cloud Computing," IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 598-609, 2020.

[16] Gabriela Pereyra," An Open Source Multi-Slice Cell Capacity Framework", CLEI electronic journal, Volume 25, Number 2, Paper 2, May 2022

[17] Josip Lorincz ,Zvonimir Klarin "Congestion Control in 5G: Approaches and Challenges," IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 35-49, 2019.

[18] "Balancing Reliability and Latency in 5G Networks: Challenges and Solutions," IEEE Access, vol. 8, pp. 104417-104429, 2020.