Ultra-reliability and low-latency communications on the internet of things based on 5G network: Literature review, classification, and future research view

Article *in* Transactions on Emerging Telecommunications Technologies · April 2023

DOI: 10.1002/ett.4770

2 authors:

Seyed Salar Sefati
Polytechnic University of Bucharest

Simona Halunga
Polytechnic University of Bucharest

Received: 8 September 2022 | Revised: 14 March 2023 | Accepted: 14 March 2023
DOI: 10.1002/ett.4770

SURVEY ARTICLE

WILEY

Ultra-reliability and low-latency communications on the internet of things based on 5G network: Literature review, classification, and future research view

Seyed Salar Sefati | **Simona Halunga**

Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest, 060042, Bucuresti, Romania

Correspondence
Seyed Salar Sefati, Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest, 060042 Bucuresti, Romania.
Email: sefati.seyedsalar@upb.ro

Funding information
This study has been partially conducted under the project 'Mobility and Training for Beyond 5G Ecosystems (MOTOR5G)'. The project has received funding from the European Union's Horizon 2020 program under the Marie Skłodowska Curie Actions (MSCA) Innovative Training Network (ITN), having grant agreement No. 861219.

Abstract
A new technology known as the Internet of Things (IoT) uses several sensor devices and communication protocols. By implementing cutting-edge and modern equipment, people use IoT to make their lives easier. Home automation is one of them, and it works with actuators and sensors. However, increasing the number of devices in the IoT network could degrade the Quality of Service (QoS). Therefore, an appropriate framework in software and hardware can improve the Quality of Experience (QoE) and QoS for all users. One of the critical QoS measures in IoT is called Ultra Reliability and Low Latency Communication (URLLC). URLLC is essential in the IoT network released from the third Generation Partnership Project (3GPP) cellular. However, a systematic and comprehensive investigation of the practical procedures for URLLC in IoT needs to be done. This paper comprehensively investigates the existing methodologies in this subject. All the chosen techniques are separated into four categories to obtain a complete picture of the topic: structure-based, diversity-based, metaheuristic algorithm-based, and channel state information. In this paper, we also investigate more benefits and drawbacks of other QoS when URLLC is applied in the IoT network. This paper highlights the challenges of URLLC in IoT networks and describes future open issues in detail to provide an efficient way for researchers in this field.

1 | INTRODUCTION

The rapid growth of computer and communication technology has made society and industries more intelligent.[1] The Internet of Things (IoT) represents a critical technology for connecting numerous heterogeneous devices in an intelligent environment.[2] Unlike most mobile networks built for human-to-human communication, IoT aims to link many objects without human interaction.[3] Some typical applications of IoT networks include control systems, intelligent recognition, positioning, and monitoring.[4] IoT devices use diverse heterogeneous applications and become more complex than other technologies. Many IoT scenarios, such as automation in industrial Vehicle-to-Anything (V2X) networks, grid computing, and remote surgery, may demand Ultra-Reliability and Low Latency Communications (URLLC).[5] Also extending the life of IoT sensors is an essential issue for IoT networks.[6] Furthermore, IoT systems usually use fog/cloud computing, sensor networks, Internet protocols, and high-quality communication systems to provide intelligent gadgets.[7] Quality of Service (QoS) evaluates the system performance based on user necessities.[8] Two approaches are allowed in URLLC. Firstly, increasing the public mobile cellular networks to satisfy the needs of various IoT applications. Secondly, developing options in existing systems or even dedicated networks for essential applications.[9] Cellular networks have met the URLLC criteria in some favorable scenarios initially developed for human-to-human conversations.[10] Three significant use cases for 5G mobile networks are URLLC, Enhanced Mobile Broadband (eMBB), and Massive Machine Type Communications (mMTCs).[11] In 5G networks, the transmission time is less than 1 millisecond (ms), which might serve many IoT applications.[12] Furthermore, ultra-reliability and low latency in IoT networks have become increasingly prevalent due to the development of new applications.[13] New technology advancements are on the horizon in IoT devices,[14] and processing techniques such as cloud computing and fog computing have improved URLLC.[15] However, some significant obstacles still need to be overcome. The second method is to develop options in existing systems or even dedicated networks for essential applications.[16]

The 5G network is the latest cellular network technology that offers significant improvements over its predecessor, 4G.[17] 5G promises faster data transfer speeds, lower latency, greater network capacity, and improved reliability.[18] The technology also features MIMO (Multiple-Input Multiple-Output), which allows for the use of multiple antennas for data transmission and reception.[19] This technology significantly improves the efficiency and capacity of the network, enabling faster and more reliable data transfer.[20] MIMO technology enhances the reliability and efficiency of URLLC, enabling faster and more reliable data transfer.[21] IoT devices generate massive amounts of data that can be analyzed using big data analytics to provide valuable insights for businesses and organizations.[22] Big data analytics allows for the processing and analysis of large datasets, enabling businesses to make better decisions and improve their operations.[23] The technology can also enhance public safety by enabling real-time communication and collaboration among emergency responders, providing faster response times, and improving situational awareness.[24] Standardization is also crucial to enable the widespread adoption[25] of the technology and ensure its security and privacy.[26] This paper provides a concise overview of IoT systems, such as structure-based, diversity-based, metaheuristic algorithm-based, and channel state information. Furthermore, this paper emphasizes the standardization of URLLC and provides outstanding customer service. The study questions are designed to identify the most significant challenges and concerns in URLLC and their effects on other parameters, such as energy consumption, availability, reliability, QoS, cost, and latency. Another significant purpose of this work is to examine the challenging problem of providing URLLC without sacrificing other QoS parameters. This survey aims to answer the questions in Table 1 and help us understand the importance of URLLC systems in different IoT scenarios.

Figure 1 illustrates that each IoT layer has a communication system responsibility. The Industrial Internet of Things (IIoT) is expected to continue experiencing exponential growth over the next decade.[27] According to a Cisco report, the IIoT market was valued at 157 billion dollars in 2016,[28] and it is projected to reach 771 billion dollars by 2026. By 2030,

TABLE 1 Questions and goal of the paper related to URLLC.

List of questions	Where can find the answer
RQ1: What is the importance of URLLC in IoT that contributes to its increased popularity? This question aims to determine the significance of URLLC in IoT by examining the number of published studies related to it.	Section 3 will provide an answer to this question, while Section 7 will address the unresolved issue.
RQ2: How do the existing URLLC techniques satisfy the key performance indicators in IoT? The primary objective of this question is to evaluate the effectiveness of the current URLLC methodologies in the IoT for specific applications.	Section 5 deals with this issue.
RQ3: What concerns and issues related to URLLC in IoT have been identified as future trends? This question aims to highlight the significance of low latency in IoT for ensuring QoS in the environment and to identify potential concerns and issues related to URLLC that may arise in the future.	Section 7 contains the answer to this question.
RQ 4: Which methods researchers use to conduct their research?	This question is addressed in Sections 3.
RQ5: What are the less frequently raised issues in URLLC systems, and what is their significance?	Section 6 has the solution to this question.

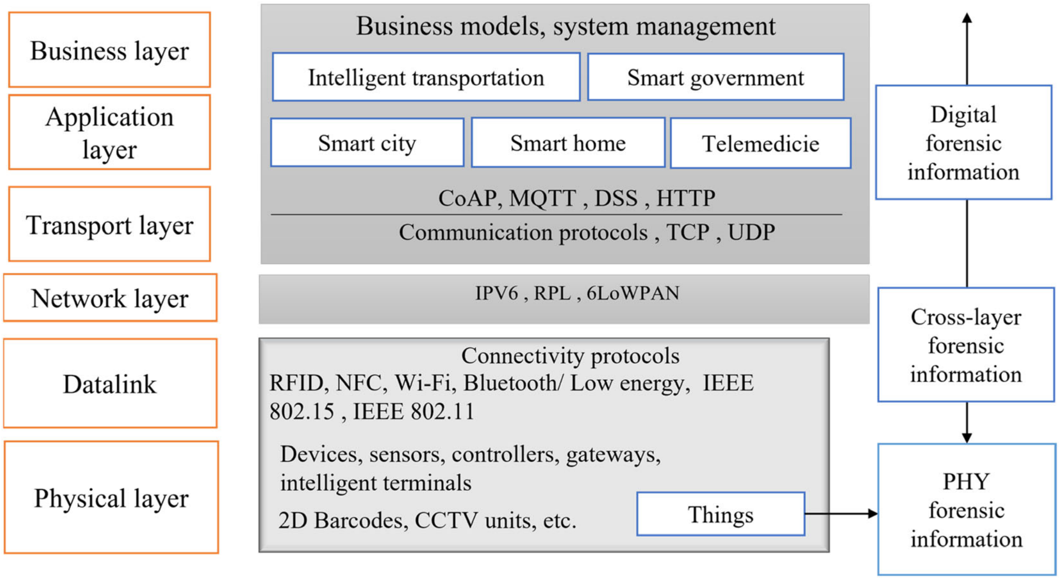

FIGURE 1 IoT-layered architecture.

Cisco predicts that there will be 500 billion things connected to the Internet.[29] This report may encourage IoT companies to develop new businesses. Unlike traditional computer devices that rely on wired connections, new IoT sensors operate using wireless communication. Therefore, a set of validity limits for QoS parameters is necessary for all IoT networks that depend on sensitive data for real-time decisions.[30]

In the coming years, the importance of URLLC will continue to increase, and more researchers in Information Technology (IT) will attempt to investigate this subject.[31] This paper provides a comprehensive and systematic overview of current research on structure-based, diversity-based, metaheuristic algorithm-based, and channel-state information approaches to handle the URLLC challenge in a broad range of IoT applications. The taxonomy of the paper is shown in Figure 2, and the main contributions of this study can be summarized as follows:

- Comparing IoT systems based on structure-based, diversity-based, metaheuristic algorithms and channel state information.
- Providing summaries and classifications of the URLLC algorithms.
- Covering various URLLC applications, including Intelligent Transportation Systems (ITS), IIoT, and Mobile Crowdsensing (MCS).
- Highlighting upcoming obstacles and suggesting future research activities to propel URLLC success in IoT applications.
- Demonstrating the importance of URLLC in IoT systems.

The rest of this paper is organized as follows: Section 2 provides an in-depth insight into a large number of related research. Section 3 describes the research methodology. Section 4 highlights the challenges in URLLC. Section 5 discusses the URLLC methods and classification. Section 6 compares all the previously discussed techniques, examines and categorizes the URLLC approaches in IoT applications, and compares the strategies. Finally, the outstanding concerns, future directions, and conclusions are examined.

2 | RELATED WORKS

Although more research is needed to survey URLLC approaches for IoT, most papers exist based on IIoT and V2X networks. Therefore, the topics covered by those articles are limited. This paper provides a quick overview of these

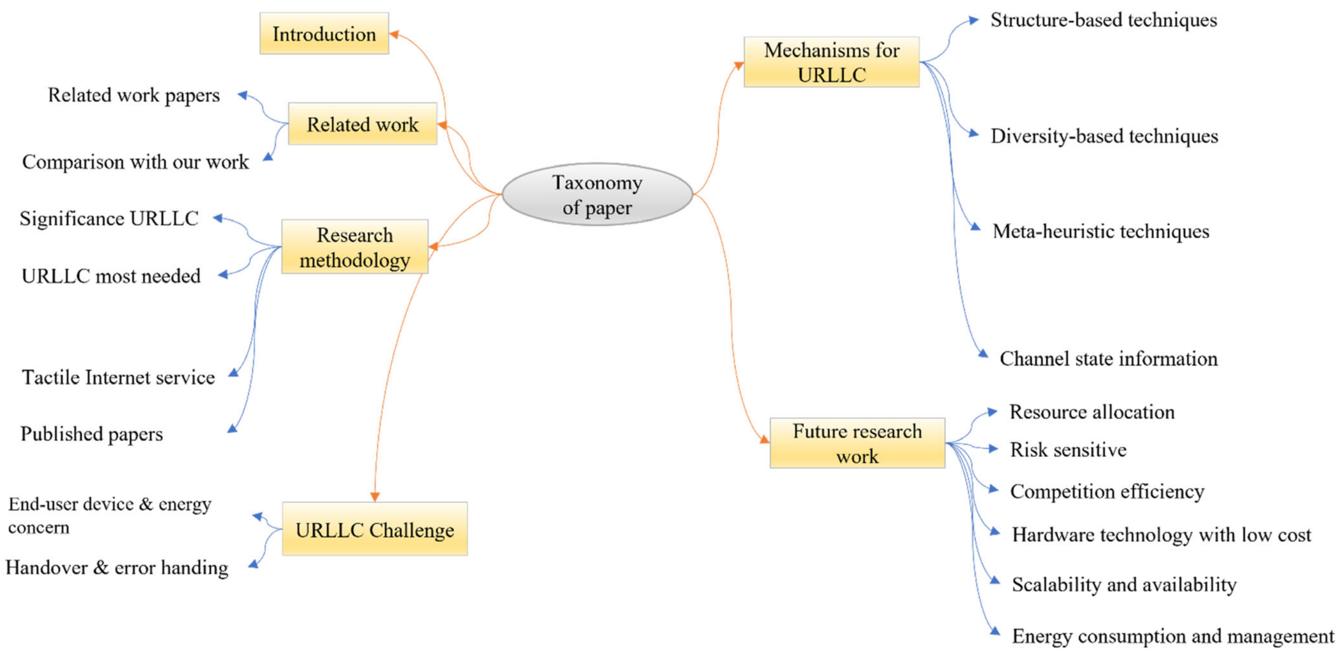

FIGURE 2 Taxonomy of survey paper.

publications and examines how they differ from ours. URLLC aims to deliver essential communication services using mobile networks as one of the 5G requirements.[32] However, the use of URLLC is significantly dependent on the 5G standardization process.[33] Therefore, the first field of study in the 5G network aims to decrease latency and enhance the reliability of these new cellular networks. Regarding transport protocols and solutions, we see an intense concentration on low-layer solutions and a need for more analysis of URLLC.[34] The URLLC for V2X networks has been presented in a few surveys that include examples of application development use-case scenarios. However, most of these are limited to specific contexts (such as industrial control applications or V2X networks) and are not generic URLLC IoT approaches. This section discusses several studies on URLLC survey papers and highlights their primary areas of interest and roadblocks.

Sutton et al[9] reviewed several papers from the conceptual perspective of IoT services to integrate the physical (PHY) layers and Medium Access Control (MAC). This article provided an overview based on the communication protocol between the 5G network and URLLC. The survey paper categorized the services from two points of view. Firstly, they analyzed communication patterns in Vehicle-to-Network (V2N), Vehicle-to-Pedestrian (V2P), Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and V2X. Secondly, they discussed the specific issues related to PHY and MAC layers on those communication patterns. This article tried to identify the best-proposed method for URLLC applications for solving these problems in PHY and MAC layer approaches. However, one drawback of this survey paper is that it would benefit from a systematic approach to article selection rather than focusing on metaheuristic algorithms in URLLC techniques.

Rico and Merino[35] surveyed several papers focusing on end-to-end protocols to reduce communication delays and improve system reliability. They first surveyed the network's functionality and technologies that fit current demands and then developed application interface programming to ensure protocols and technologies were used appropriately. This survey paper categorized the existing literature on end-to-end solutions for URLLC into three categories, outlining the merits and cons of each proposed protocol, as well as the URLLC's efficiency. The importance of improving QoS and higher layers was also emphasized in this paper. The research focused on network support, Application Programming Interfaces (APIs), new protocol developments, multipath protocols, multicast protocols, edge computing, Software-Defined Networking (SDN), Network Function Virtualization (NFV), and Information-Centric Networking (ICN). They also extracted the common approaches employed by the submissions to analyze current trends. However, a drawback of this survey paper is the need to pay attention to metaheuristic algorithms and cover cloud computing and fog computing in the URLLC domain.

Kelechi, et al[36] reviewed more than 50 research papers on URLLC in 5G networks to address IoT concerns. The study aimed to provide a flexible and scalable architecture for URLLC in 5G communications, comparing framework architectures in terms of cost, complexity, cross-layer capabilities, and computation. Upcoming technologies, such as SDN, NFV, and fog computing, were discussed. The review paper presented various methodologies, approaches, models, features, and applications to illustrate URLLC principles. However, the survey paper did not cover machine learning (ML) and artificial intelligence algorithms in URLLC, and more attention could be paid to data processing and management approaches that are fundamental to IoT.

Shukla and colleagues[37] conducted a systematic review of over 60 strategies and solutions aimed at reducing latency in the IoT and cloud environment, using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology for paper selection. The review highlighted 23 techniques and 32 technologies that could be used to address latency issues in IoT and cloud systems, while also identifying research gaps and limitations that require further investigation.

Khan, et al,[38] reviewed several papers following Beyond 5G (B5G) based on the Industrial Internet of Things, with an investigation on eMBB and URLLC. In this paper, they tried to discuss the B5G structure based on IoT layers. Moreover, this paper reviewed papers based on the bandwidth and data rate, with most articles providing the architectures of IIoT (hardware). At the end of the paper, the authors discuss the future of this topic, including ML /AI, network slicing optimization, dense IIoT applications, security, and privacy. As a drawback, this survey paper is not well organized and does not discuss the reviewing papers in depth.

Vaezi, et al,[39] reviewed several papers based on the context of cellular and non-terrestrial networks in IIoT networks. This paper examines the current and future standards of the 5G network that affect energy consumption, URLLC, and scalability. They divided all IoT networks into three categories: Cellular, WAN/MAN, LAN/PAN, and explained the differences between these methods. Also, they discussed the 6G vision and covered the 3D coverage, AI-Integrated, All frequencies, and IoT dominance. In other sections, they talked about massive connectivity, IoT security, and deep learning in IoT systems. This paper is well organized and discusses deeply the 5G network and the future of the 6G network. As a drawback, this survey paper needs to cover metaheuristics and heuristic algorithms in 5G and 6G networks.

Ma et al[2] reviewed various constraints and technological solutions for URLLC in IoT networks. They first examined how the physical, MAC, and network layers can impact latency and reliability. Next, they explained the underlying information-theoretic limitations for the physical layer of URLLC and offered insights into channel codes with limited block lengths suitable for these applications. Finally, the paper suggested the best network topology, traffic allocation techniques, and network coding strategies to reduce latency at the network layer. However, this study did not provide a systematic literature review of the existing URLLC approaches, classifications, or the crucial role of URLLC techniques in the IoT.

2.1 | Comparison with our work

In this paper, we tried to distinguish the many characteristics of URLLC from the state-of-the-art presented above, so we attempted to identify all areas covered by the pre-existing surveys. For example, in the last survey papers on URLLC, the authors did not concentrate on categorization; however, in this paper, we tried to identify some techniques used in URLLC and categorize them. In other survey papers, the framework and metaheuristic technique were unavailable; however, these techniques significantly participate in URLLC. Many papers use these techniques to solve URLLC issues; therefore, we attempted to expand the framework and metaheuristic approach. Since URLLC is an NP-hard problem,[40] metaheuristic algorithms can find suitable solutions.[41] As previously stated, implementing these four methodologies together is a new survey technique that can be applied to this issue. Researchers utilized a systematic approach to locate, propose, and evaluate algorithms for any given scenario based on predetermined criteria. Our paper includes a systematic review of publications concentrating on different elements of a URLLC system, such as communication, security concerns, and application instances. On the other hand, this study focuses on the technologies under development that will underpin future 5G networks.

3 | RESEARCH METHODOLOGIES

This section discusses different technological components described in URLLC or those that will be considered in the future. These technical components represent the "URLLC toolkit" used to increase system performance. The URLLC techniques are divided into two groups: the first aims to reduce latency, and the second aims to improve reliability.[42]

TABLE 2 Overview of requirements for URLLC in IoT.

Applications	Domain	Tolerable delay	Update frequency	Data rate
Health monitoring	Smart city	1 min	10 min	Low
Waste management	Smart city	1 min	1 h	Low
Virtual reality	Smart city	Milliseconds	Real-time	High
City air quality	Smart city	1 min	10 min	Low
Patients healthcare	Healthcare	Low (second)	One report per hour	High
Interlocking control	Industrial	Milliseconds	Milliseconds	Low
Monitoring and supervision	Industrial	Seconds or ms	Seconds	Low

TABLE 3 KPI analysis for modern IoT connectivity solutions.

KPI	ZigBee	BLE	Wi-Fi	SigFox	LoRa	eMTC	NB-IoT
Reliability	×	√	√	×	×	√	√
Low cost	√	√	√	√	√	√	√
Long duration of operation	√	√	√	√	√	√	√
Low latency	×	√	√	×	×	√	√
Scalability	×	×	√	×	×	√	√
Flexibility	×	×	×	×	×	√	√

For example, some technological components mainly designed to increase reliability may provide more retransmission opportunities. However, most issues originate from requirements related to latency, reliability, and coexistence with other services, all of which impact the physical layer architecture. One of the main issues in IoT applications is data rate and update frequency, which can be different in each domain. Table 2 shows some critical applications with tolerable delay and update rates. For example, patients' healthcare delays and data rates differ completely from city air quality, so URLLC and QoS criteria vary for each application. Finally, Table 3 provides a brief assessment of the 5G Key Performance Indicators (KPIs) of several existing and emerging technologies, which have been explained in detail in the following tables. Each KPI has a benefit in comparison with other methods. For example, Zigbee and Sigfox have the weakest performance,[43] while emMTC and NB-IoT have the highest execution.[44]

3.1 | What is the significance of URLLC in the 5G network?

This section focuses on 5G topology in IoT to achieve low latency and high-reliability requests. We thoroughly examined and assessed many research papers covering many IoT problems concerning low latency and high-reliability issues. Cloud computing, supported by a 5G network, can transfer data faster than LTE networks.[45] This is one of the critical benefits of using 5G technology. Cloud computing based on 5G will enable real-time data streaming, providing a significant boost to virtual business practices by supporting uninterrupted storage and improving productivity. In most 5G cloud methods, fog-computing works like the base station and initial data is entered into fog computing and connected to cloud computing in the necessary time.[46] This type of computing will require data centers located closer and adding micro data centers into or near 5G towers to create this access. This system usually connects sensors to fog computing for communications, and this connection uses 5G gNB techniques.[47] Industrial automation,[48] intelligent transportation,[49] smart electrical grid,[50] and surgery[8] are several examples of these application cases. One of the central issues in cloud computing is service selection and composition. Service selection and composition are critical processes to answer user requirements. Due to the complexity of IoT systems and large scalability, one service cannot respond to the user's needs, so the service composition technique will answer this complexity. The 5G network in services selection and composition can work with the highest rate and reliability. Figure 3 shows the 5G network based on cloud computing and its effect on URLLC in IoT networks.

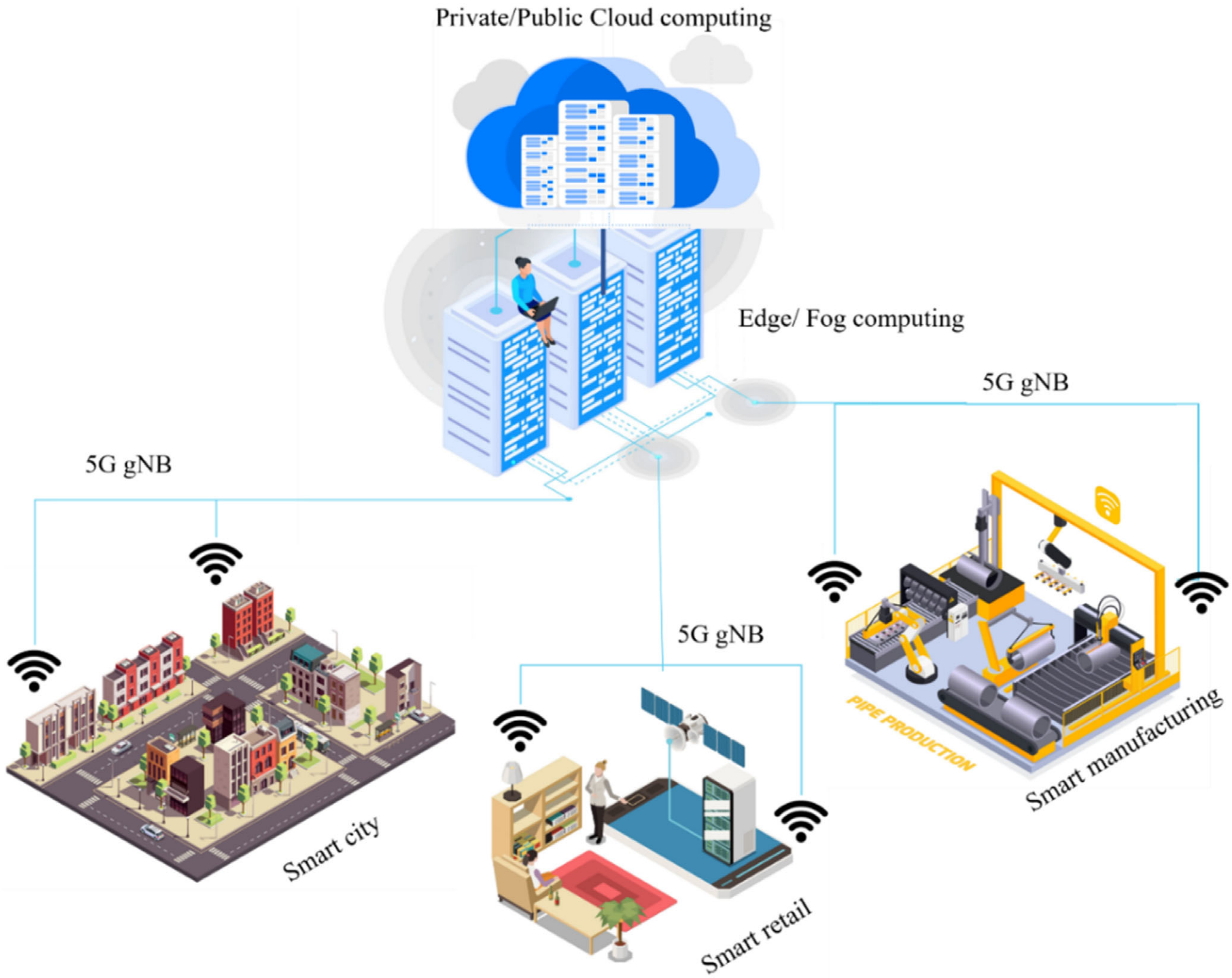

FIGURE 3 5G network based on cloud computing.

In 2019, cellular systems began installing fifth-generation technological standards.[51] URLLC has more than 99.999% reliability, and the packet transmission delay is less than 1 ms.[52] These two characteristics make the URLLC use case suitable for a 5G network.[53] Drone-based delivery uses URLLC to assess traffic flow in real-time.[54] Some applications require network connectivity, such as broadcast V2X servers, including smart cars and all other participants in the traffic system.[55] URLLC can also deliver data in wireless connections for predictive vehicle maintenance.[56] Multiple sensors linked to a cellular network records metrics such as vibration and temperature. The system analyzes sensor information to avoid possible vehicle maintenance faults, lower maintenance costs, and reduce idle time.[57] URLLC offers a smooth and economical communication platform for implementing new technologies to control power distribution networks.[58] One of the real-life use cases of 5G, linked to the energy grid and port automation, was provided by the Wireless for Verticals (WIVE) research project.[59] In this application, URLLC aims to guarantee the security of the energy grid infrastructure and automation used in protection applications.[60]

3.2 | Where is the URLLC most needed?

There has been much discussion and enthusiasm around URLLC development in the run-up to commercial 5G. The release of Third Generation Partnership Project (3GPP)[61] standards include many specifications to enable sub-millisecond latency. On the consumer side, the use cases revolve around Augmented Reality (AR),[62] Virtual Reality (VR),[63] and

TABLE 4 The specifications for the new radio non-standalone applications specified by ITU classifications.

Category	Basic features
Enhanced mobile broadband	eMBB is an extension of the services enabled by 5G for those applications that need significant data rates for many users clustered together, who can be fixed and mobile, an increased payload, and a full-time internet connection. Specific use cases include hotspots, enhanced multimedia, Three-Dimensional/Ultra-High-Definition (3D/UHD) video applications, smart offices, cloud office/gaming, and virtual/augmented reality (VR/AR).
URLLC	URLLC emphasizes a fast, responsive connection with minimal latency. URLLC is not developed to offer high data rates, but it needs to provide flexibility. URLLC can be used in industrial automation, autonomous driving, mission-critical applications, and remote medical care, among other things.
(Massive machine type communications)	mMTC focuses on delivering low-reliability communication to a very large number of devices that transfer usually a small quantity of data, like IoT applications. It can provide long-range communication while still being energy efficient and having asynchronous access. mMTC is ideal for low-power gadgets in Wireless Sensor Networks (WSNs), smart cities, energy management, etc.

TABLE 5 Reliability and low latency importance for different type of applications.

Industry	Application	Importance of reliability and low latency
Healthcare and medical services	Patient diagnosis/remote surgery	A robot might be used to do remote surgery or diagnose a patient from a far.
Entertainment/ Business/ Media/	Live event handling, live sporting events support, online gaming, and (VR/AR), entertainment based on cloud support	Users desire to participate online in various current events, like sports, concerts, or other forms of entertainment.
Transport	Drone-based deliveries, remote driving, self-driving automobiles, traffic control, and sub-station management.	Drones must adapt in real-time when new services and incentives for consumers emerge, such as Amazon Prime Air for order delivery. Google's self-driving car (WAYMO)
Industrial automation	Automated control systems including robot-assisted industries, machine status reporting, monitoring, and power monitoring.	Industries have turned toward automation to increase productivity. By substituting people with robots in the production process, more reliability and productivity may be achieved.

cloud gaming.[64] On the enterprise side, use cases have been developed related to sectors like public safety[65] (for image recognition applications like facial recognition[66]), transportation (vehicle-to-everything communications and intelligent transport systems[67]), and manufacturing (AR support and digital twins). Several research results have been published on these use cases and their requirements associated with low latency. Unfortunately, operators still need to develop strong business cases.[68] For example, a few safety contracts can bring high value and Service Level Agreements (SLAs).[69] On the other hand, various operators and companies are experimenting with low-latency offerings. For instance, Verizon is leveraging its 5G multi-access edge-computing platform through a partnership with Amazon Web Services (AWS).[70] As a result, QoS depends on the layer from the IoT architecture in which the users identify themselves.[71] Table 4 shows the new radio non-standalone applications specified by the International Telecommunication Union (ITU).[72] NTT DOCOMO Inc[73] and Huawei conducted[74] a field trial on URLLC that produced several promising results. For 5G networks, meeting URLLC criteria is a significant problem because it requires changes in the present telecom infrastructure's architecture. Nevertheless, the promising results obtained with URLLC can play a critical role in 5G future communications and user demands. The importance of latency and reliability in open applications is shown in Table 5.

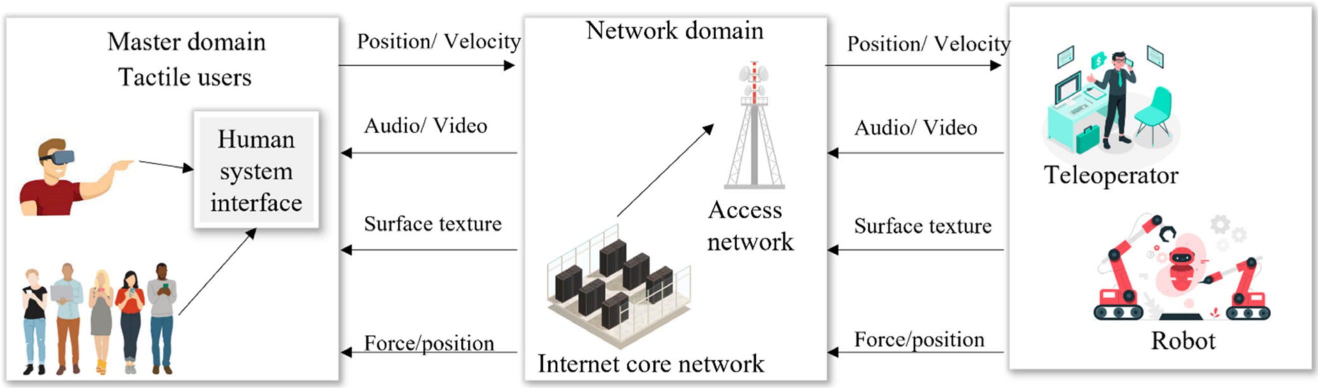

FIGURE 4 Tactile internet.

3.3 | URLLC technique in tactile internet services

URLLC is appropriate in TI services that significantly need low latency and payload.[75] TI's core requirements include ultra-reactive and reliable connectivity, additional intelligence at the border of the network, and efficient data transmission.[76] As opposed to the traditional Internet, which is used to transmit audio and video data, TI sends touch information and actuation data with audiovisual data.[77] In TI applications, the haptic experience is also bilateral[78]; for example, in a teleconferencing environment, the teleoperator receives motion while the environment's communication is transmitted back to the Human System Interface (HSI).[79] The tactile user and the HSI determine the master domain and change human sensorial inputs into tactile information using a suitable tactile conversion technique. The HSI produces tactile data that is transferred through the network domain. A remotely controlled robot or tele operator is necessary for the controlled field or environment. The master domain directly incorporates the controlled region based on specific command signals (eg, velocity, position).[80] The master domain receives response signals from the controlled domain (eg, force/position, surface texture). The master domain gets audio/visual feedback signals from the controlled environment and haptic feedback signals. The TI links the users (master domain) remotely to manage the environment (controlled territory).[81] Figure 4 shows the tactile internet using the HIS.

3.4 | Method of searching and finding the published papers

Two groups of reviewers scanned abstracts separately to seek out candidate theories and assess their applicability. We considered all types of articles published in English, except for opinion-driven materials (such as editorials, commentaries, and letters). We used the following three critical steps for the article selection strategy:

- Search the existing scientific databases using an automated search based on the keywords.
- Choose the papers based on their titles and abstracts.
- Evaluate the entire text of the selected papers.

Figure 5 depicts using Google Scholar and other electronic resources, such as ScienceDirect, SpringerLink, Web of Science, and IEEE Explore, in Stage 1. These databases were searched using three keywords: "information shared in project teams" and "team groups." First, we conducted an automated search using the Google Scholar search engine and other electronic databases to find primary studies. Next, we transferred all the publications' citation data, abstracts, and keywords to an Excel datasheet for further analysis. We identified six well-known publishers, resulting in 34 journal publications and 10 conference papers. The topic of URLLC became relevant in 2017 when the first papers were published and grew in importance between 2019 and 2021. In 2019, we identified 11 papers; in 2020, we found seven articles; in 2021, we discovered 10 papers; and in 2022, we came across five papers on URLLC topics. We identified several challenges based on our analysis of these papers, as presented in the section below.

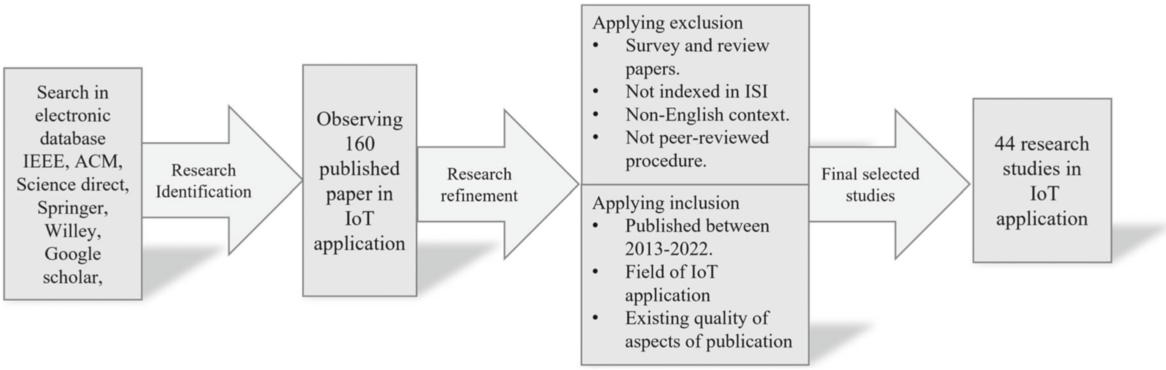

FIGURE 5 Electronic databases used in URLLC.

TABLE 6 Expected QoS requirements for URLLC.

Industry	Error rate / error probability	Latency (ms)
Augmented/virtual reality	10^{-3}–10^{-5}	5–10
Guided vehicles	$10^{-5} \geq 10^{-3}$	5–10
Automated industry	10^{-5}–10^{-9}	1
IoT	10^{-5}	1

4 | URLLC CHALLENGE IN THE IOT NETWORKS

URLLC needs to transfer data from exchange to exchange (E2E) with high reliability, low latency, and high security. Table 6 shows the various applications that require URLLC with different error rates and latency.

All URLLC applications need low latency and good reliability, while eMBB also demands effective data rates.[82] Therefore, systems must share physical resources to maintain the required QoS when URLLC and eMBB coexist. Figure 6 shows a practical coexistence approach. However, this collaboration between eMBB and URLLC on the same radio network can bring new challenges. The New Radio (NR)[83] describes two scheduling protocols: immediate and reservation scheduling. The immediate scheduling strategy recommends that packets be sent to the base station as soon as data is created.[84] As a result, this scheduling scheme may cause interruptions in data transmission. The reservation schedule is separated into two categories for active packet processing: semi-static and dynamic reservation. Both techniques employ an additional reservation frame for URLLC, which leads to control signaling overhead. Moreover, the reserved space may be better spent with URLLC data.

4.1 | End user device & energy efficiency concern

New technologies, features, services, and applications have emerged in mobile communication since the introduction of the 5G network.[84] However, in some countries, the previous telecom network operation and administration systems are still in use (LTE and 3G). These networks must meet users' demands and data rates and may require increased network efficiency and better control of operational expenses. The industry has recognized the need for a knowledgeable and automated network for the 5G era.[85] Developing an intelligent autonomous system is critical for creating mobile communications, and it will become a crucial component of communications networks in the 5G future.[86] The 5G and beyond 5G eras introduce AI and ML techniques, which may be considered an unavoidable necessity for network design.[86] For example, reducing unnecessary power usage is essential in IoT networks, and AI and ML can perform these tasks very well. During peak and off-peak hours, the network traffic flow changes dramatically.[87] However, the equipment continues to operate, and power consumption is not constantly adjusted in response to traffic. Therefore, building the capability of "zero bits, zero watts" is necessary. In standard networks, the attributes in various situations can vary substantially from one application to another. Therefore, recognizing multiple conditions and designing suitable energy-saving measures is

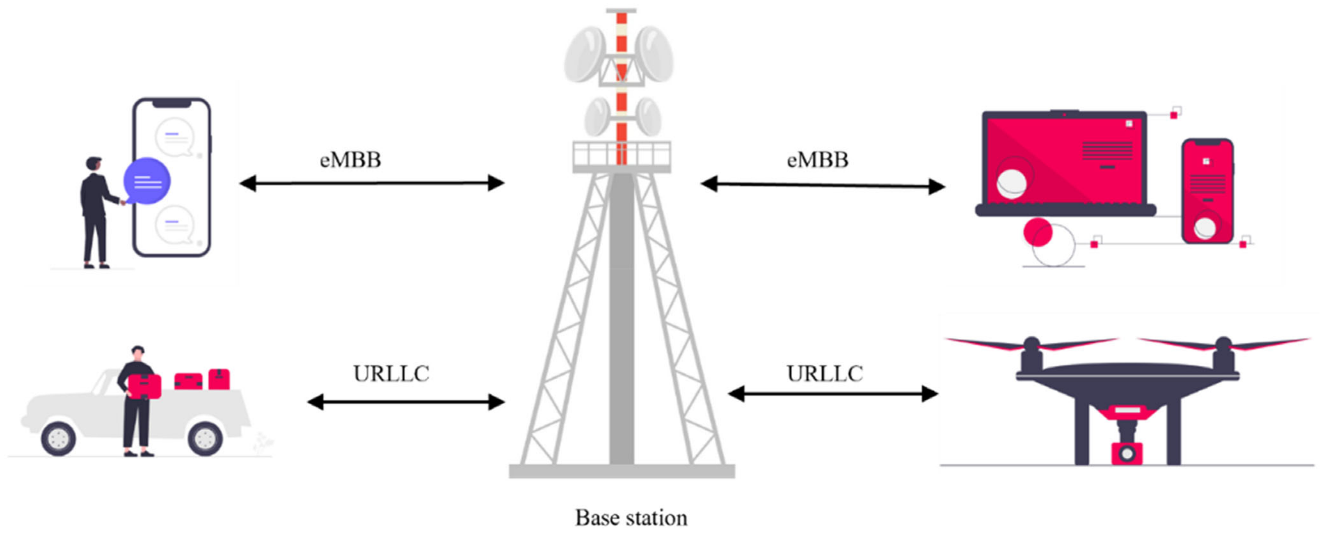

FIGURE 6 Benefits of the URLLC network.

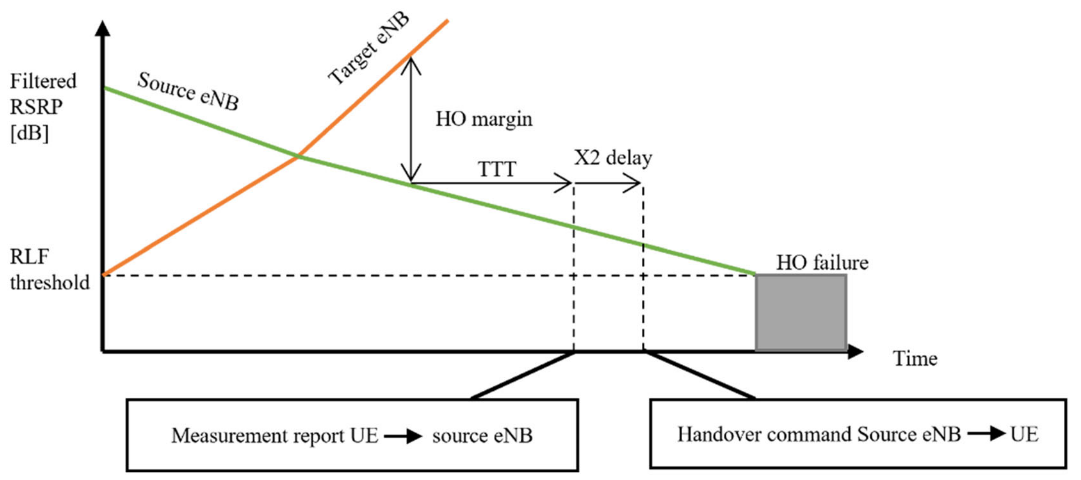

FIGURE 7 Handover decision principle.

crucial in IoT and 5G networks. To save energy, most wireless devices operate in sleep mode.[88] Furthermore, the devices regularly check for pending packets on the network to prevent delays.

4.2 | Handover & error handing issues for URLLC

In 5G network defines numerous handover criteria and thresholds for initiating the handover mechanism.[89] These parameters should be carefully established in the network design. For example, as illustrated in Figure 7, two critical parameters determine the handover, Hand Over margin (HO margin) and Time-To-Trigger (TTT) value. The UE measurements are made regularly or on-demand, assuming that the source eNB's Reference Signal Received Power (RSRP) decreases while the target eNB's RSRP increases. The source eNB starts the TTT timer when the target eNB suits the HO margin better than the source eNB. If the entry condition persists during the TTT, the source eNB determines the UE's handover decision. At this point, the target eNB asks the source eNB to perform the admission control procedure. After completing the admission control, the source eNB launches the transfer. One of the most important aspects of any telecom infrastructure is handover (handoff). NR must support the mobility criteria shown in Table 7 for 5G.

TABLE 7 Mobility requirements for URLLC.

User	Speed
Normal vehicle	120 km/h
Drones	160 km/h
High-speed vehicle	250 km/h
Trains	500 km/h

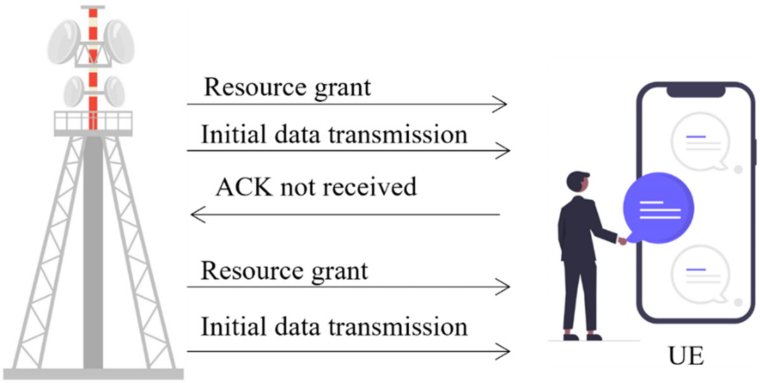

FIGURE 8 Downlink data transfer signaling mechanism. ACK/NACK stands for acknowledgment/non-acknowledgment.

The fundamental handover mechanism LTE uses smooth handover, and NR provides two levels of handover to improve this process further. LTE technology is regulated using a Radio Resource Control (RRC) layer.[90] In addition, beam-level mobility is managed via physical and MAC layers to achieve low latency rather than relying on RRC. However, mobility in NR inherits two unsolved issues: strength and Mobility Interruption Time (MIT). To achieve very low Handover Interruption Time (HIT) and Handover Failure (HOF), NR is being explored and suggested. When data arrives at the base station, a request for a Resource Grant (RG) is delivered to the target UE, as shown in Figure 8. The UE decodes the data and replies with a positive or negative acknowledgment (ACK/NACK). The BS must retransmit the data if the UE does not respond within the allotted time interval. The 5G network has a shorter TTI than LTE and demands a faster reaction from the user to prevent retransmission.

5 | URLLC TECHNIQUE IN IOT

The URLLC techniques are divided into structure-based, diversity-based, metaheuristic algorithm-based, and channel state information. These four categories came up from most of the papers on IoT that have been studied. We evaluated the approach and techniques used in each article and then categorized the documents based on the methods used. Additionally, we included the articles for each category to offer a more accurate comparison and result. An objective examination of the paper under the abovementioned consideration reveals the presence of eight factors for assessing the acquired findings. The parameters are specified as follows:

Low latency in IoT indicates the amount of time needs for a device to receive and respond to a command or request. It is an essential factor in the performance of IoT systems, as it affects the responsiveness and reliability of the devices and the overall system. The τ_1 shows the physical layer, where τ_t represents the time to transmit and τ_{prop} is a signal propagation time, τ_{proc} represents a time for preceding and decoding, τ_{ret} represents time to re-transmit, and τ_{sig} is the pre-processing time.[91]

$$\tau_1 = \tau_t + \tau_{prop} + \tau_{proc} + \tau_{ret} + \tau_{sig} \quad (1)$$

The processing time using the URLLC systems should not exceed 0.5 ms.[92] It represents the necessary time to deliver a packet using the new physical layer frame structure required to enable URLLC.[93] Furthermore, the latency criterion is not fulfilled if the time-to-transmit is higher than 1 ms. The physical layer must consider various traffic characteristics[94] and the different QoS for various services.[95]

The reliability in the 5G network implies an error probability of 10^{-5},[96] so transferring in the channel must be performed in short packets. The goal of improving automated industrial and control is achieve high reliability in terms of successful packet rate delivery.[76] In a one packet communication, the error probability is:

$$p_e = 1 - (1 - p_c)(1 - p_d) \tag{2}$$

Where P_e shows the transmission error probability, P_c the error probability of the Physical Downlink Control Channel (PDCCH), and P_d the probability of the Physical Downlink Sharing Channel (PDSCH).

The goal of third formula is created a new URLLC application can fulfill reliability criteria such as packet loss probability between $10^{-5}\sim 10^{-7}$ in PDSCH and PDCCH less than 10^{-6}:

$$P = P_c^* P_{d_1} + (1 - P_c) * P_c^* P_{d_1}^* P_{DTX} + P_c^* \left(1 - P_{d_1}\right) * P_N^* P_C^* P_{d_2} \tag{3}$$

Where P_{d1} shows the probability of correct detection of a single PDSCH. P_{d2} the success probability of retransmitted PDSCH, and PDTX the successful probability of Discontinuous Transmission (DTX). Specific algorithms can train AI in reliability to send the data more precisely. The AI algorithm can increase the reliability of network, however thus algorithms can use a high energy in IoT network.

The response time is defined as a service capacity to fulfill several responsibilities under specific circumstances for a certain amount of time, as assessed by the following criteria[97]:

$$RT_{R_K} = \frac{RES_K}{REC_K}, K = 1, \ldots, M \tag{4}$$

where RES_K shows the amount of jobs submitted to each service R_K in a specified time period, for each $K = 1, 2 \ldots, M$, where M is the amount of services, and REC_K is the total number of demands.

The availability shows the ability of IoT service to be functioning when requested, and it may be computed using the following formula:

$$AV_{R_K} = \frac{A_K}{J_K}, K = 1, \ldots, M \tag{5}$$

where $R_1, R_2, \ldots, R_M$ **represents** the assets, $J_K = 1, 2, \ldots, M$ are the number of activities submitted to R_K, and A_K represents the number of jobs accepted by R_K.

The cost is the amount of funds spent to satisfy an IoT node demands, based on the amount of memory, number of operations, and bandwidth required. We can determine the costs based on:

$$Cos\ t = \sum_{i=1}^{k} \left(C_i^* T_i\right) \tag{6}$$

where K is the size of the concrete service requested by an IoT node, C_i is the number of nodes necessary to fulfill the users' requests, and T_i is the time interval in which the user has access to the nodes.

Devices with IoT potential are heavily impacted by energy usage. Therefore, the next step is for each possible service to supply a parameter indicating how much energy it consumes.

$$E = \frac{E_{init} - E_c}{E_{init}} \tag{7}$$

where E_{init}, indicate the initial energy of sensor, and E_c the current energy of sensor.

Packet Delivery Ratio (PDR) is based on the sum of the total received data packets into the BS:

$$PDR = \frac{\sum No.\text{ of packets recieved}}{\sum No.\text{ of packets sent}} * 100\% \tag{8}$$

Throughput indicates the packets delivered to the BS per unit of time.

$$Throughput = \frac{\sum No.\text{of packets sent} * \text{Packet size}}{Time\text{ taken}} \tag{9}$$

5.1 | Structure based techniques

Structure-based strategies rely on structured assumptions, principles, ideas, and practices. This technique chooses the most efficient IoT slots, symbol duration, sub-carrier spacing mapping, and modulation methods. Structural techniques present frameworks such as frame structure, waveform design, and finite block length information theory. There have been extensive discussions on frame structure both in academia and industry. The frame structure is a fundamental approach to designing the 5G network. Several papers have researched this method, which is vital in URLLC. Meanwhile, the frame structure ensures high reliability in 5G networks and achieves high bandwidth. Orthogonal frequency-division multiplexing (OFDM) is the primary waveform used in 4G and has an essential role in 5G networks. OFDM can support the 3GPP and is also mandatory for UE. Most authors use filtering, time-domain, guard band insertion, and spectral precoding in this design. Finite block-length information is based on argument and random coding, which causes radio resource allocations. This strategy has a relationship between the desired reliability and bandwidth in applications. In the following section, we will discuss the structure techniques in-depth.

Zhao et al[98] proposed a shared Finite-Block Length Coding (FBC) for the multi-user downlinks. In this method, only a few symbol lengths are accessible to users. Expanding the FBC algorithm can cause a more effective coding rate and can be performed by jointly encoding users. The authors employ a matrix-based strategy to develop the multi-user collaborative encoding design. Then, using a nonlinear bipartite pairing problem, they achieved the best possible power constrained within extremely low latency. They created a two-step approach to implementing the unified FBC policy across numerous users. Firstly, they identified the user occupancy status for each accessible Resource Block (RB) by detecting the users who would send data. Then, a shared encoding design scheme was developed among the RBs with similar occupations. In this paper, the authors also established an IP issue to optimize a weighted sum of users' throughput under a power consumption constraint. They obtained the best energy usage based on a unified approach by transforming the IP issue into a nonlinear bipartite matching problem. However, this approach has a long execution time.

Park and Saad[99] developed the finite memory multi-state architecture to enable several IoT devices to share constrained connection resources. This proposed method uses the suggested learning architecture to understand critical signals and allocate the proper communication resources. Also, it can deliver delay-tolerant, periodic, and urgent packets. Furthermore, the proposed learning architecture compensates for memory usage and computational constraints for IoT devices. The authors have shown that this learning framework can achieve the shortest predicted latency in IoT devices. Moreover, the presented learning algorithm's effectiveness in IoT systems with varying delay targets, detection probabilities, memory sizes, and network densities. However, the method is time-consuming and energy-intensive.

Avranas et al[100] analyzed URLLC by using Incremental Redundancy (IR) and Hybrid Automatic Repeat Request (HARR). In addition, this study investigated energy and latency using a finite block length domain. A dynamic programming technique solved the non-convex median energy reduction problem with URLLC limitations. The primary outcome of this paper is the latency compared to the packet size, and a well-optimized IR-HARR method may be energy-efficient. The benefits of this mechanism are low complexity and low energy usage. However, it has limited scalability and is relatively expensive.

Yu et al[101] proposed a connection between consensus and communication link transmission. However, they discovered that consensus latency and consensus reliability are incompatible. Therefore, they aim to develop a new approach using the Raft architecture in their study. The leader node in a Raft network must pack the instructions into log entries and continually duplicate these entries to all followers through downlink communications. Then, the followers confirm and transmit the log to the leader via uplink communications based on the successful receipt of the request. Consensus nodes in the IIoT may be actuators or can function as a group to offer consensus to the actuators. The actuators can

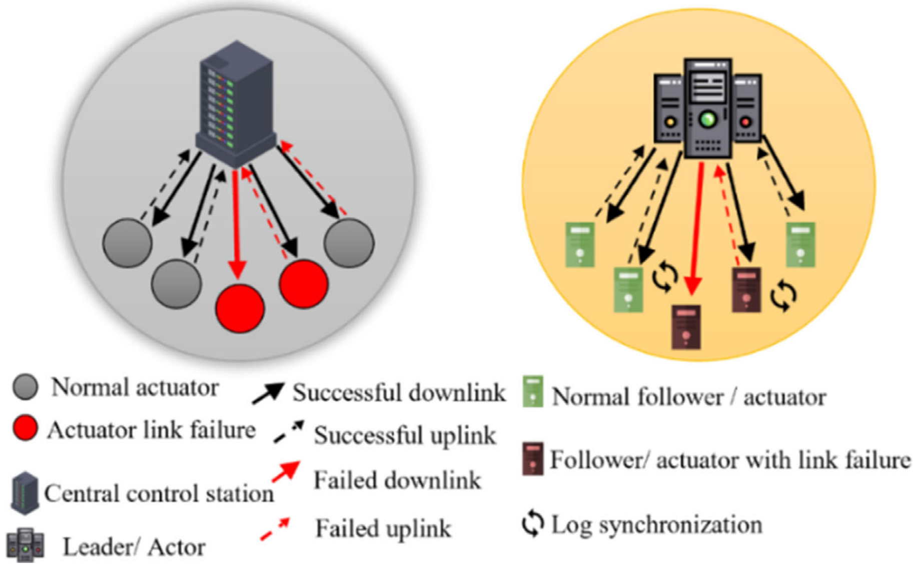

FIGURE 9 The centralized and distributed consensus systems in Raft framework.

operate only if the Consensus Mechanism (CM) network agrees on the critical choice. This framework lowers the rate of consensus failure and boosts user satisfaction. It also provides several essential benefits; the most important is high availability. However, this method is limited in scalability, and the concept has not been tested in a real-world IoT logistics system. Raft's centralized and distributed consensus system is shown in Figure 9.

Din et al[102] developed a system that combines green IoT with a 5G network for healthcare systems. They attempted to connect heterogeneous networks using the least amount of energy possible. Furthermore, the proposed protocol supports the 5G network architecture by mapping Internet Protocol (IP), MAC, and Location Identifiers (LOC). In a clustering strategy, many mobile devices were grouped based on Received Signal Strength (RSS) data. They created a mobility supervision system for the nearby cluster. Furthermore, they sought to design a 5G network with green IoTs as the sensing layer. The sensing layer was proposed to improve efficiency and minimize energy by collecting data from the medical care system. To verify the performance and viability of the suggested technique, they used the C programming language to simulate it. They created a mobility supervision system where each mobile node searched for a nearby cluster and joined it with a smaller amount of energy. The suggested 5G network architecture for medical applications is shown in Figure 10.

Zeng et al[103] proposed an IoT system based on Massive Multiuser (MU-MIMO) and Pilot-Aided Channel Estimation (PACE). The PACE technique assumes that all users are equally distributed and the radio channel is only affected by log-normal shadowing. The error probability of connecting users with a given delay is determined using Finite Block Length (FBL) data theory, allowing large MU-MIMO to transmit short packets. In addition, the Golden Section Search Method (GSSM) has been used to calculate the pilot's length and limit the possibility of failure. The numerical results demonstrate that massive MU-MIMO can accommodate many URLLC users, even when users are placed randomly under shadow fading. According to the analytical findings, MU-MIMO may achieve high reliability under shadow fading with several Rx antennas, even when many users are supplied simultaneously. They also used GSSM to reduce the pilot's length to achieve fast convergence. Even under severe shadow fading and random user placement, large MU-MIMO enhances reliability.

Mahyoub et al[104] proposed a Routing Protocol for Low-Power (RPL) based on a low-latency and Lossy Networks (LNs). To provide compatibility with RPL requirements, the network has to use the original control features and transfer a data packet. First, the source node should retrieve a route to the destination from the local cache. Next, the source node should send a Route Request (RREQ) note to the original path's root. Finally, if this path does not exist, the source node should request a path to the intended destination by sending an RREQ message to the root. It is worth mentioning that the RPL sends RREQ and Non-Storing RPL mode (NSRPL) using uplink and downlink routes. It also proved to have an average latency reduced by 74%, energy usage reduced by 23% for the investigated traffic intensities, and high PDR. However, this article achieved poor load balancing, and the system was not tested in a real environment. Therefore, the suggested solution was probed based on simulation in Cooja.

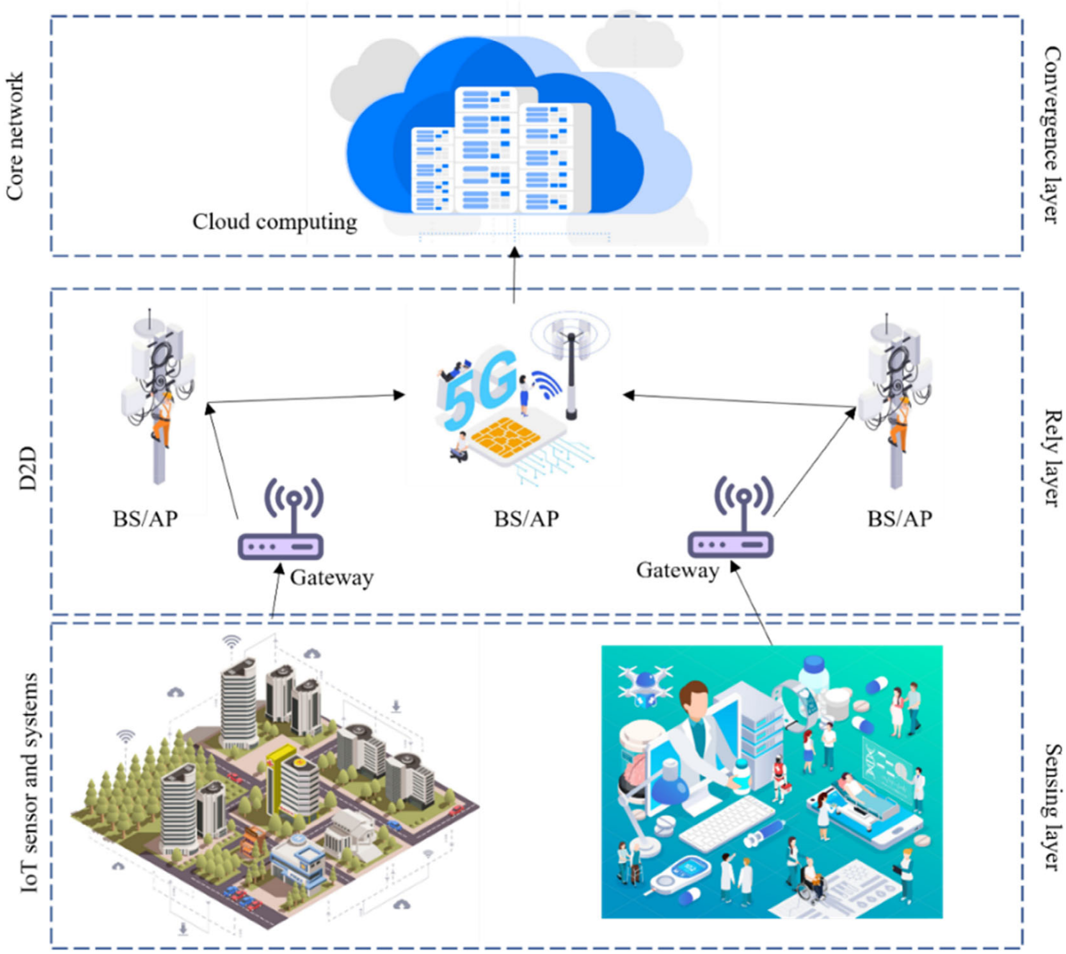

FIGURE 10 5G network architecture in LOC architecture.

El Haber et al[105] proposed a unique method to solve the URLLC problem in UAV networks, which helps future IoT services. They attempted to maximize the process rate by optimizing the positions of the UAVs. The issue was split into two phases: firstly, a planning problem that optimized UAV deployment, and secondly, an operational problem that determined the best offloading and resource allocation options with limited UAV energy. The authors represented both phases' concerns as non-convex mixed-integer programs due to their non-convexity. Figure 11 shows the system model of the UAV-aided mechanism in the URLLC. This method achieved high performance and low energy consumption and could quickly discover a solution under strict QoS limitations. However, it had a long convergence time, and the simulation results had not been verified yet in a real-world setting.

Pang et al[106] proposed a Fog-Radio Access Network (F-RAN) that includes small cells and a macro base station to address URLLC requirements. They considered the low-latency architecture an optimization challenge, where F-RAN balances communication vs computational effort over numerous nodes. The authors described suitable task computing techniques for the simultaneous selection of F-RAN nodes in a multi-user environment with heterogeneous resource allocation. Simulation results revealed that the one-for-all idea might drastically reduce the overall latency of collaborative task computing and provide a win-win scenario for all users. In addition, they discovered a dynamic programming method that may limit the overall running time as the number of users grows, proving the scheme's viability and scalability. However, as a limitation, this method achieved low availability, and the concept had not been tested yet in a real-world IoT system.

Zhang et al[107] used an adaptive routing technique to enable nodes to choose receivers based on the current state. In this paper, they proposed a new Data Collection Framework (DCF) for low latency. Furthermore, the DCF technique allowed many nodes to communicate in each time slot. They offered two methods for generating local schedules in DCF, enabling nodes to set their schedules based solely on information from their neighbors. To overcome this challenge,

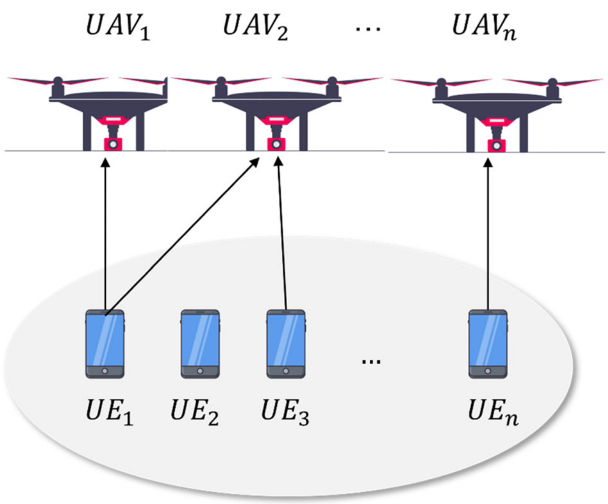

FIGURE 11 System model of UAV-aided mechanism.

they proposed a distributed data-gathering architecture based on DCF, utilizing spatial parallelism to prevent collisions. Additionally, nodes in the DCF could alter their routing algorithms. This technique offered several benefits, including high performance, scalability, and low energy consumption. However, the cost had not been determined.

Zhang et al[108] proposed a multi-cell grant-free uplink in an IoT network where services with hard deadlines coexist. In this paper, the packet loss rate is calculated for each service, and sensor nodes are communicated with the nearest base station. All sensor nodes are fixed in position and do not have mobility options. Furthermore, a lower bound of the successful decoding probability (SDP) is derived. Finally, the uplink communication is settled, and a pre-specified average received power is needed for transmitting. This technique offered various benefits, including successful decoding probability, high resource consumption, and low packet loss; however, the cost has not been determined. Table 8 shows the advantages and disadvantages of the structure methods.

5.2 | Diversity technique

The diversity technique is based on the computational effort to simulate actions. It describes components based on relationships, information structures, user needs, and business problems. Structured diversity gives abstractions of attribute features and explains other ideas. For example, error probability and noise problems cannot cause packet loss. Furthermore, most researchers prefer to use these techniques in a noisy environment. The modulation-coding scheme (MCS) can be considered as obtaining diversity from redundancy over time. Diversity techniques present frameworks such as Frequency/Time/Space Diversity and Modulation and Coding Schemes. Frequency/time/space diversity can achieve Ultra reliability communication without affecting latency. Spatial diversity consists of two approaches: micro-scale and macro-scale. Additionally, this technique significantly affects system capacity in both noise-limited and interface-limited scenarios. This section will analyze several diversity-based strategies used in IoT to solve the URLLC problem.

Elgabli et al[109] proposed using the Age of Information (AoI) to share reliable data remotely. AoI indicates the amount of time since a transmitter generates a packet. This work investigated time-sensitive remote monitoring issues, and URLLC was used to minimize the chance that each sensor's AoI exceeds a predetermined threshold. Furthermore, the authors anticipated that sensors would accept various thresholds and different-sized output packets. Finally, they introduced a low-complexity reinforcement algorithm to solve the suggested formulation inspired by the success of ML. They used the state-of-the-art actor-critic method to train their system using a collection of public bandwidth traces. The authors did not study energy consumption or test the system in a natural environment. Simulation results demonstrated that the proposed method achieved robust scalability in a vast region with minimal packet loss.

Lee, et al[94] proposed a model for data transmission through the relevant data available. The proposed technique used the most reliable data as a virtual pilot for increasing the estimated channel accuracy. This technique prompted a short training time with a small packet shape. However, small packets have significantly degraded the target channel and

TABLE 8 Structure based techniques.

Authors	Methodology	Pros	Cons	Assessment
Zhao, et al	Finite-Block length Coding (FBC)	(+) Optimal power consumption (+) High throughput	(−) Long execution time	Real environment
Park and Saad	Finite memory multistate sequential learning	(+) Discover critical messages (+) Reduced memory sizes	(−) High energy & time consumption (−) High complexity	Simulation N/A
Avranas, et al	Incremental redundancy (IR)	(+) Low complexity (+) Low energy consumption	(−) Low scalability (−) High cost	Real environment
Yu, et al	Raft framework	(+) High availability (+) User satisfaction rate	(−) Low scalability (−) Not tested in real environment	Simulation Matlab
Din, et al	Internet protocol (IP), Medium access protocol (MAC)	(+) High data rate (+) Inefficient bandwidth usage	(−) Low scalability (−) Not using big data	Simulation using C programming
Zeng, et al	Finite block length (FBL)	(+) Low error probability (+) High throughput (+) High reliability (+) High convergence	(−) Low scalability (−) High energy consumption	Simulation 2D area
Mahyoub, et al	Routing protocol low-power (RPL)	(+) High packet delivery rate (+) Low energy consumption	(−) Low load balancing (−) Not tested in a real environment	Cooja emulator Contiki OS
El Haber, et al	Optimize the placement of UAVs	(+) High performance (+) Low energy consumption	(−) High convergence time	Simulation N/A
Pang, et al	Fog-radio access network (F-RAN)	(+) Reduced total running time	(−) Medium reliability and availability	ARToolkit Based on LTE
Zhang, et al	Distributed framework (DCF)	(+) High performance (+) High scalability (+) Low energy consumption	(−) The cost and processing during data gathering has not been evaluated	NS-3 simulator

interfered with covariance matrix estimation. This research emphasized low-latency communication, but the primary principle could be expanded to mMTC and high-throughput MIMO scenarios. In both instances, the accuracy of the channel was critical for achieving the intended result. Also, for URLL communication, a noncoherent technique does not depend on the pilot signal. The benefits of this mechanism were considered to be the short time training, high throughput, and high availability; however, other characteristics, such as the number of users and service in overall results, have not been examined in this paper.

He et al[110] proposed a multi-device IoT network that utilizes shared radio resources. They described the network's resilience and throughput in the FBL domain. Furthermore, they investigated a multi-UE IoT network's reliability and throughput performance while running retransmission. They initially described the FBL performance model and then proposed two design architectures: the first minimized error probability and the second maximized throughput in the multi-UE network. The authors discovered a considerable difference in the FBL network and Infinite Block Length (IBL) performance. Finally, they conducted simulations using the Monte Carlo technique to verify the analytical model and assess the system's performance. Although the suggested approach achieved high throughput and a low error probability, this algorithm has not been tested in a real-world scenario.

Seo, et al[111] proposed a Compressed Sensing-based random-access protocol (CS-RACH) to manage machine-type communication in an IoT network. They found more advantages in CS-RACH than LTE. The compressed sensing approach allowed them to identify users concurrently with reasonable accuracy. Furthermore, the user detection scheme could eliminate preamble impacts and minimize collision probability compared to standard LTE. Normalized throughput, access probability, and average access delay have been evaluated using the minor absolute reduction and selection operator technique. Their simulations showed that the suggested method was effective in throughput and latency. However, this study did not include scalability and has not been tested in the real world.

van Rensburg et al[112] proposed a low-cost wireless network for reliable data transfers in South Africa. The IoT network was developed inside a building, and the nodes were linked to send millions of packets. The researchers used "big data" to assess the network's effectiveness and reliability. Statistical analysis showed that the quality of service improved between the network's multiple asynchronous and transmitting nodes. Additionally, the authors concluded that the network's point-to-point and mesh connections delivered high reliability, but this approach had a significant level of complexity. However, scalability and costs were not evaluated.

Nakao et al[113] described a random graph optimization to solve the order issue. Using graph symmetry, the suggested technique improved the Average Shortest Path Length (ASPL). The proposed approach is applied to both generic and grid graphs. Finally, the authors evaluated various graph issues with up to 1 million vertices. Therefore, graphs with better symmetry properties have a lower diameter and reduced ASPL. However, symmetric graphs have less complexity than random graphs and networks without balance, and this algorithm did not evaluate energy usage, and the reliability has not been proven in a real system.

Ye, et al[114] proposed a Non-orthogonal Multiple Access (NOMA) system to solve the difficulties of delivering fast, responsive, and highly reliable connections for significant devices in the IoT. The neural network and the associated multi-class function produced symbol-spreading signatures. The technique eliminated the time-consuming human-crafted effort and allowed the automated construction of spreading signatures. According to simulation results, the suggested method outperforms traditional grant-free NOMA schemes regarding reliability. Furthermore, a multitask learning method was employed since many IoT sensors require low power utilization and high reliability, becoming more critical when designing radio-frequency access schemes. Based on simulation results, the suggested technique improved reliability and had a lower symbol error rate, but it does not enable conflict analysis.

Liu, et al[115] proposed a Half-Duplex Relay-Assisted NOMA (HDR-NOMA) to transmit data using relay-assisted NOMA for 5G V2X systems. Although none of the specified problems can be formulated as concave or convex functions, the issues studied were demonstrated to be modeled as a quasi-concave function. Therefore, a bisection-based power allocation method was developed to find the best solutions to the challenges. As a result, they converted it into a series of convex feasibility functions and solved them using a bisection-based power allocation technique. According to simulation results, the suggested approach provided a considerable performance gain over the Fractional Transmit Power Allocation (FTPA) method. Furthermore, when self-interference is adequately suppressed, the suggested full-duplex relay-assisted NOMA (FDR-NOMA) technique achieved a higher max-min attainable rate than the proposed HDR-NOMA.

Sun and Yang[116] proposed an unsupervised deep-learning algorithm to tackle resource allocation issues in URLL communications. They used a deep learning algorithm to reduce the necessary bandwidth and enhance the QoS of the network. Simulation results revealed that the learning-based approach performed and the optimal solution in the symmetric situation can save roughly 40% of the available bandwidth with low convergence time and computational cost.

Liu, et al[117] proposed an open-loop system and multi-cell grouping in a Heterogeneous Network (HetNet). Before analyzing communication reliability and latency, they discussed how mobile users in a HetNet employed the suggested Proactive Multi-cell Association (PMCA) strategy to build their virtual cells. They demonstrated that the PMCA system could significantly increase communication reliability and be optimized by adjusting the number of users. The delays encountered on ascending and descending links were also evaluated. The PMCA system indicated very low latency in a single user's cell. The suggested open-loop communication and PMCA strategy could meet the target URLLC users' requirements in a HetNet. The proposed technique could achieve a low error probability rate and good reliability values; however, this approach might lead to a long convergence time.

Zhang, et al[118] proposed semi-supervised learning for improving URLLC in IoT networks' radio frequency identification (RFI) framework. This paper analyzes the error probability and availability of MIMO for sending data to the base station. Received signals and processing by the base station can cause an increase in latency. Furthermore, essential and

TABLE 9 Diversity techniques.

Authors	Proposed method	Advantage	Disadvantage	Assessment
Elgabli, et al	Age of information (AoI)	(+) Good scalability in large area (+) Low packets drop	(−) Energy consumption has not been evaluated	Simulation Python based on TensorFlow
Lee, et al	Mini-mental state examination	(+) Low execution time (+) High throughput	(-) Low number of users and service in simulation	Simulation MATLAB
He, et al	Finite block lengths (FBL) regime for a multi-UE IoT network	(+) High optimal throughput (+) Low error probability	(−) High convergence time (−) Not tested in real environment	Simulation N/A
Seo, et al	Compressed sensing-based random-access protocol (CS-RACH)	(+) High throughput	(−) Not pay attention to a scalability	Simulation N/A
van Rensburg, et al	The packets were transmitted and logged between interconnected nodes	(+) Low packet losses (+) Tested in real environment	(−) High complexity (−) Scalability and	Simulation N/A
Nakao, et al	Average shortest Path length	(+) Workability (+) Lower complexity	(−) High Energy consumption	Simulation Framework SimGrid-3.25
Ye, et al	Nonorthogonal multiple access (NOMA)	(+) High reliability	(−) The approach does not provide conflict analysis for the workflow	Simulation Tensorflow
Liu, et al	Full-duplex relay non-orthogonal multiple access (FDR-NOMA)	(+) Fixed power allocation (+) High performance	(−) The simulation environment is not mentioned (−) Low availability	Simulation N/A
Sun and Yang	Unsupervised deep learning	(+) Rapid convergence (+) Low computational complexity	(−) High packet loss probability (−) High energy consumption	Simulation MATLAB
Liu, et al	Proactive multi-cell association (PMCA)	(+) Low error probability rate (+) High dynamics	(−) High programming complexity	Simulation N/A

valuable data are forwarded to the cloud for computing, analysis, and decision-making. The proposed method divides the vital information into short packages and sends them to the base station. The proposed technique has acceptable results in data transmission, low cost, and computational complexity. Nevertheless, using the ML algorithm has high-energy consumption. Table 9 shows the advantages and disadvantages of the diversity-based technique.

5.3 | Metaheuristic-based techniques

In engineering challenges, detecting the highest and lowest function values is critical.[119] To address specific issues, efficient analytical-based methods are available in the existing literature. However, due to computational demands, correct algorithms are limited, and heuristics or meta-heuristic approaches are required.[120] Using a heuristic methodology to solve an optimization issue does not guarantee the best solution. These methodologies, like humans, employ a heuristic function to aid in the search. Meta-heuristic algorithms also handle complex global optimization problems by relying on natural phenomena. As a repetitive creation process, meta-heuristic algorithms combine numerous notions to identify and use search regions. Learning procedures are used to organize the data and find near-optimal solutions.

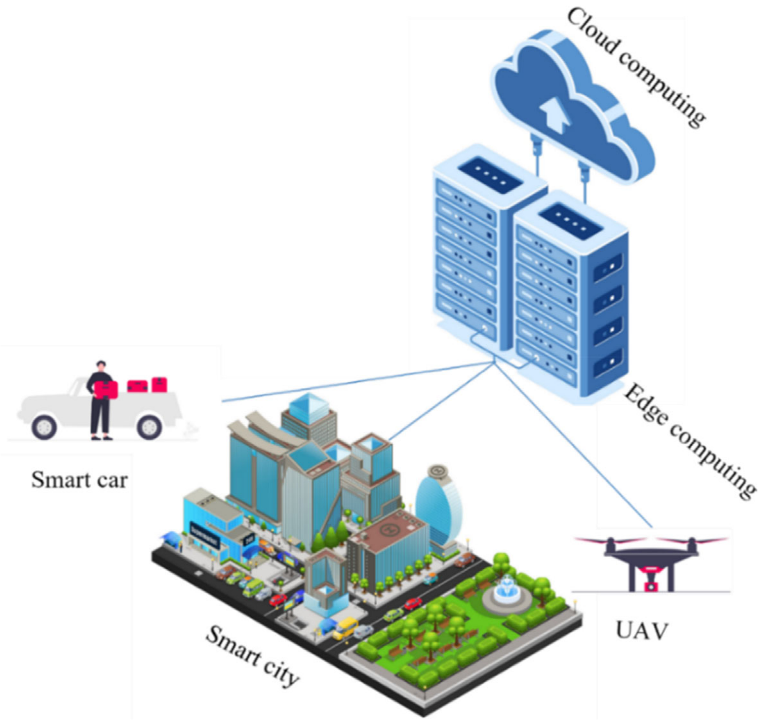

FIGURE 12 Edge-computing-based system model.

Bhardwaj and Kim proposed[121] the Dragonfly Node Identification Algorithm (DNIA) to optimize criteria for URLLC systems. They evaluated the suggested approach using several benchmark functions and found that DNIA produced an optimal global solution with improved convergence behavior compared to other optimization algorithms. Although DNIA could not find globally optimal solutions for increasingly complex Congress on Evolutionary Computation (CEC) benchmark functions, its performances were classified as accurate, consistent, and efficient. This technique also studied the conflicting tradeoff for the best solutions. In addition, DNIA benefited from reduced packet loss, an acceptable Cumulative Distributive Function (CDF), and a short convergence time. However, this approach has limited scalability and is only appropriate for a few users.

Wang and Chen[122] proposed the M/M/1 technique to solve the queuing problem in wireless channels. Firstly, they calculated the capacity of local computation and processing resource distribution. Secondly, they separated the original optimization issue into two independent sub-issues. In this paper, they offered a hybrid genetic algorithm to improve offloading choice based on low latency. Eventually, the numerical results have shown that their suggested scheme outperforms other alternative strategies regarding achievement time and energy usage. However, the proposed technique had limited sensor and service coverage in simulations.

Singh and Nagaraju[123] proposed using nature-inspired computational approaches to enhance the operation of a sensor network using three distinct procedures: sink node position, route creation, and optimization. Furthermore, opportunistic coding was used at prospective relays to reduce the number of transmissions, significantly improving data transfer. In this paper, two algorithms were used. First, the Particle Swarm Optimization (PSO) algorithm was used for sink placement. Second, the Artificial Bee Colony (ABC) algorithm optimized the routing from the source node to the base station. These techniques offer high throughput, a high packet delivery rate, and low packet delay. However, the approach required a long convergence time and high-energy consumption.

Babar et al[124] proposed an ABC algorithm to control edge computing while decreasing latency and response time. Additionally, they suggested an ABC algorithm to distribute the workload to allocate appropriate resources between IoT devices and servers. The proposed algorithm can be efficiently used for IoT devices under stringent energy constraints with acceptable latency. The ABC algorithm achieved good results compared to PSO, Ant Colony Optimization (ACO), and Round-Robin (RR) scheduling algorithms regarding response time and management effort. On the other hand, this strategy ignores power efficiency and cost issues. Figure 12 shows the system model of the ABC algorithm for URLLC in edge computing.

Chang, et al,[125] proposed a Machine Learning and Genetic Algorithms (MLPGA) approach to enhance the performance of ultra-reliable and low-latency wireless sensor networks (WSNs). They utilized MLPGA to select the best chromosome and generate a near-optimal clustering by ML methods. The proposed two-layer network architecture, based on K-means clustering, ensured effective communication in WSNs, satisfying multiple network objectives simultaneously. The proposed technique built a multi-objective optimization based on important network metrics. Simulation results showed that this technique improved network performance, increased the average network lifetime, and reduced execution time. However, real-world evaluations and scalability issues still need to be considered.

Aburukba, et al,[126] proposed a GA algorithm to schedule IoT queries and reduce total latency. The GA evaluated the dynamic environment, determined the problem size, created a population of viable solutions (chromosomes), and determined each chromosome's fitness function. Finally, they selected chromosomes for crossover and possible alteration to create a new population based on the existing one. The GA's performance was compared to Waited-Fair Queuing (WFQ), Priority-Strict Queuing (PSQ), and RR methodologies. The proposed technique considerably reduced total latency, from 21.9% up to 46.6%. However, scalability and availability were not analyzed, and the approach's performance has not been validated in a real-world scenario.

Cui, et al[127] proposed a method to identify the balance between energy usage and latency using a restricted multi-objective optimization and a selective nondominated genetic algorithm to determine the best solutions. They also proposed a unique encoding method and genetic algorithm to increase performance. This research focused on computing offloading issues to assess energy and delay, with offloading associated with servers and local computing in the decision space. Offloading may save time and money by reducing computational effort, while local computing might minimize transmission delay. This method achieved low energy consumption and low latency in mobile edge computing. However, the results were obtained based on MATLAB simulation and were not tested in a real environment.

Javanmardi, et al[128] proposed the Fuzzy PSO Fog Work Scheduler (FPFTS) for delay sensitivity and scheduling in a fog-computing scenario. The suggested method aimed to use most of the fog resources in order to decrease the network and application loop time. In this proposed paper, fog-computing compares with similar approaches previously proven suitable for task scheduling. Furthermore, the suggested method may be used in delay-sensitive and delay-tolerant applications. They design and assess their process using an IoT architecture based on three layers. According to the results presented in terms of latency and network usage, their solution outperformed First-Come, First-Served (FCFS) mechanisms. On average, FPFTS reduced the application latency by 86% while increasing networking utilization by 81%. However, as a limitation, the overall QoS was not considered in object-connected settings, and the small number of nodes cannot provide scalability.

Seyfollahi and Ghaffari[129] introduced the concept of reliable data distribution in IoT networks. In this paper, the Reliable Data Dissemination (RDDI) approach uses the behavior of node information to identify and report potential attacks. The Harris Hawks Optimization (HHO) algorithm also integrated routing facilities, energy awareness, and geographical data transmission to provide reliability and nature-inspired optimum routing. In the RDDI approach, all nodes send information to the Cluster Head (CH). The CH stores the behavior of the member nodes, and the system will detect and block the CH if it exceeds a certain threshold. This strategy improves user satisfaction while reducing resource monitoring. It also provides several benefits, such as high reliability, PDR, and low energy usage. However, this method is limited in scalability. Furthermore, the concept has not been tested in an IoT logistics system. Figure 13 shows the proposed RDDI architecture plan.

Kanagaraj, et al[130] proposed the Cuckoo Search (CS) optimization algorithm enhanced by a Genetic algorithm. This proposed method is named CS-GA to handle the reliability and redundancy allocation issue. In addition, integrating genetic operators in regular CS can enhance the balance between exploration and exploitation. A comparison of the results of this algorithm confirmed that the suggested algorithm was a better solution for reliability-redundancy allocation issues. The CS-GA algorithm was a population-based algorithm that employed several solutions to get the global answer. The best solution is obtained for the current population by executing a Lévy flight before the end of each generation. This strategy has a high customer satisfaction rate, a short reaction time, and a high level of reliability. However, concerning limitations, this algorithm has not been tested in a real-world IoT scenario, and the developed system reached high complexity.

Sefati and Halunga[120] proposed service selection and composition for improving reliability and availability using the adaptive penalty function in genetic and artificial bee colony algorithms. In this paper, the authors improved the URLLC of cloud computing to serve the highest QoS for IoT networks. The first genetic algorithm selects the nearest services according to the user's needs, then, using the ABC algorithm, tries to combine the user's requests. In complex and large scalability, one service cannot respond to the user's needs, so those systems need service composition. This paper has

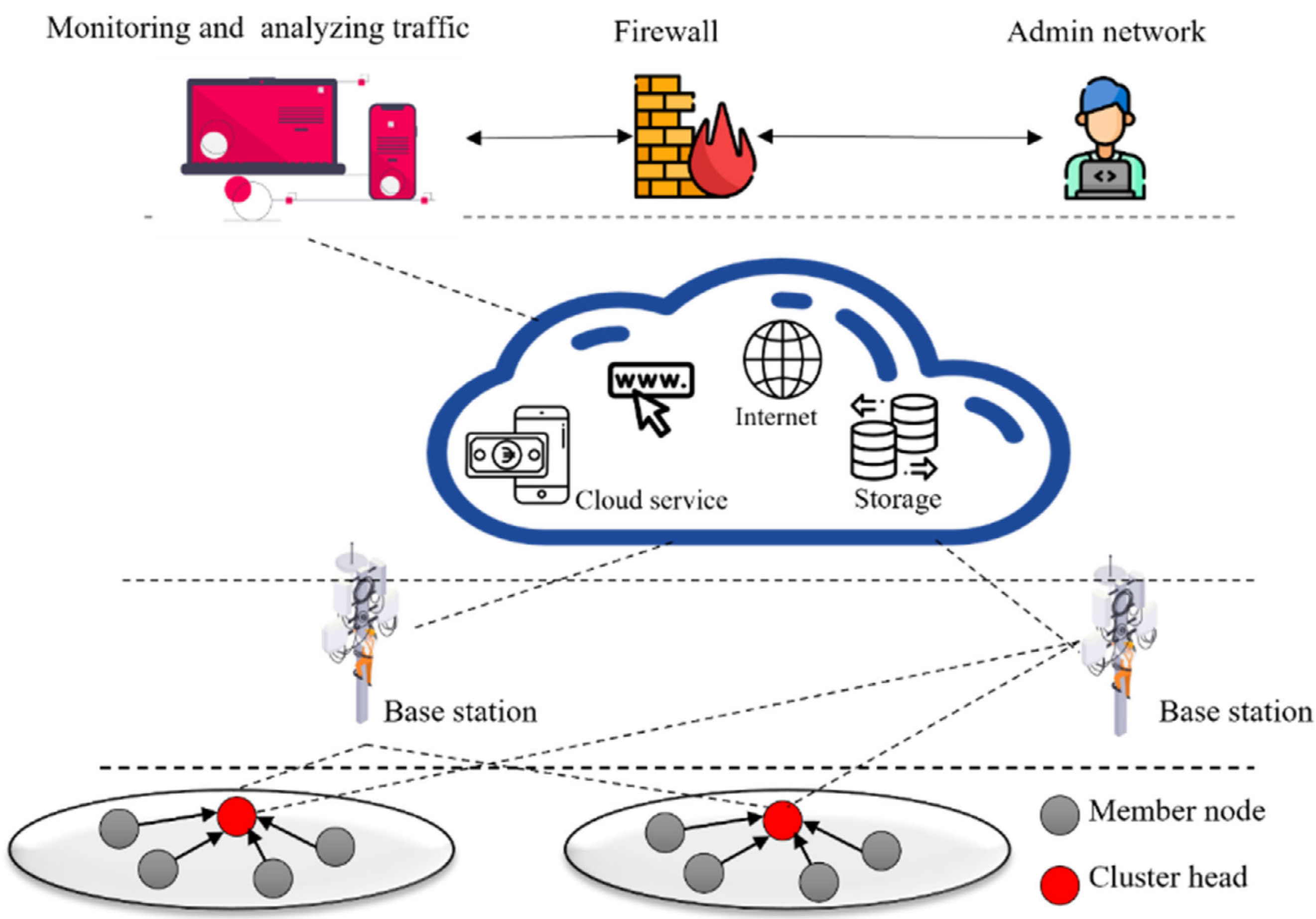

FIGURE 13 The communication architecture of the RDDI history system.

good reliability, availability, and low latency performance but has not been tested in real environments, and QoE is not considered. Table 10 shows the advantages and disadvantages of the metaheuristic algorithm.

5.4 | Channel state information technique

In IoT-based 5G networks, channel state information is critical in ensuring reliable and efficient communication between devices. Channel state information refers to the information about the wireless channel between a transmitter and receiver, including signal strength, phase, and delay. One technique for acquiring channel state information in 5G IoT networks is channel estimation. This involves transmitting a known signal, such as a pilot signal, and then using the received signal to estimate the channel parameters. The estimated channel state information can then be used to adapt the transmission parameters, such as the power and modulation scheme, to optimize communication performance. Another technique is feedback-based channel state information acquisition, where the receiver sends feedback to the transmitter about the quality of the received signal. This feedback can include information about the channel conditions, such as the signal-to-noise ratio (SNR) and the number of detected errors. The transmitter can then adjust the transmission parameters based on this feedback to improve performance. Channel state information is a well-known communication technique in IoT and WSN communications.

Atutxa, et al[131] proposed a method for an IIoT network to minimize response time by utilizing in-network computing. In the data plane, their solution analyzed the Message Queuing Telemetry Transport (MQTT) of the data packets provided by a sensor. In their study, the authors proposed a solution based on In-Network Computing (INC) to improve the network architecture in industrial operations. Their solution implements INC in industrial applications utilizing Data Plane Programming (DPP) to improve reaction time and throughput. The article's actual part tests and analyzes how a

TABLE 10 Metaheuristic algorithm techniques.

Authors	Proposed method	Advantage	Disadvantage	Assessment
Bhardwaj and Kim	Dragonfly node identification algorithm (DNIA)	(+) Low packets lose (+) Low convergence time	(−) Low scalability and not suitable for big environment.	Simulation MATLAB
Wang and Chen	Hybrid genetic simulated annealing (HGSA) algorithm	(+) Low energy consumption (+) High convergence speed and quality	(−) Low number of devices and services used in simulation (−) High cost	Simulation N/A
Singh and Nagaraju	Minimum Weiner spanning tree (MWST)	(+) High throughput (+) High packet delivery rate	(−) High convergence time (−) High complexity	Simulation MATLAB 2017a version 9.2.0
Babar, et al	Artificial bee colony (ABC)	(+) Scales the edge server to meet the demand of high QoS for IoT applications	(−) The proposed method does not pay attention to low power efficiency and low cost.	Simulation MATLAB
Chang, et al	Machine-learning-based parallel genetic algorithms	(+) Network performance (+) Average network lifetime	(−) Performances have not been tested in real environment (−) Low scalability	N/A
Aburukba, et al	Genetic algorithm	(+) Improvement in the overall latency (+) Succeeded in meeting the requests deadlines	(−) Comparison by another heuristic algorithm energy consumption is less. (−) Approach has not been tested in the real-world IoT application	Simulation MATLAB R2017a
Cui, et al	Nondominated sorting genetic algorithm (NSGA-II)	(+) Improved Latency in mobile edge computing latency (+) Low energy consumption	(−) Simulated only in Matlab; needs to be checked also in Cooja simulation or NS3 (−) High convergence time	Simulation MATLAB R2017a
Javanmardi, et al	Fuzzy pso fog task scheduler (FPFTS)	(+) Improves application loop delay (+) Improves network utilization	(−) QoS trust as an important factor does not consider in object-connected environments (−) Limited set of nodes cannot really ensure scalability	iFogSim testbed
Seyfollahi and Ghaffari	Harris Hawks optimization (HHO) algorithm	(+) High reliability (+) High packet delivery rate (+) Low energy consumption	(−) Low scalability (−) Performances have not been evaluated in real-world logistics system in IoT	MATLAB R2016b
Kanagaraj, et al	Cuckoo search (CS) & genetic algorithm (GA) CS–GA	(+) High user satisfaction rate (+) Low response time (+) High reliability	(−) Performances have not been evaluated in real-world logistics system in IoT (−) High complexity	Simulation C++
Sefati & Halunga	Service selection and composition based on GA and ABC	(+) High reliability (+) High availability (+) Low latency	(−) Performances have not been tested in real environment, and QoE is not considered	Simulation cloud sim

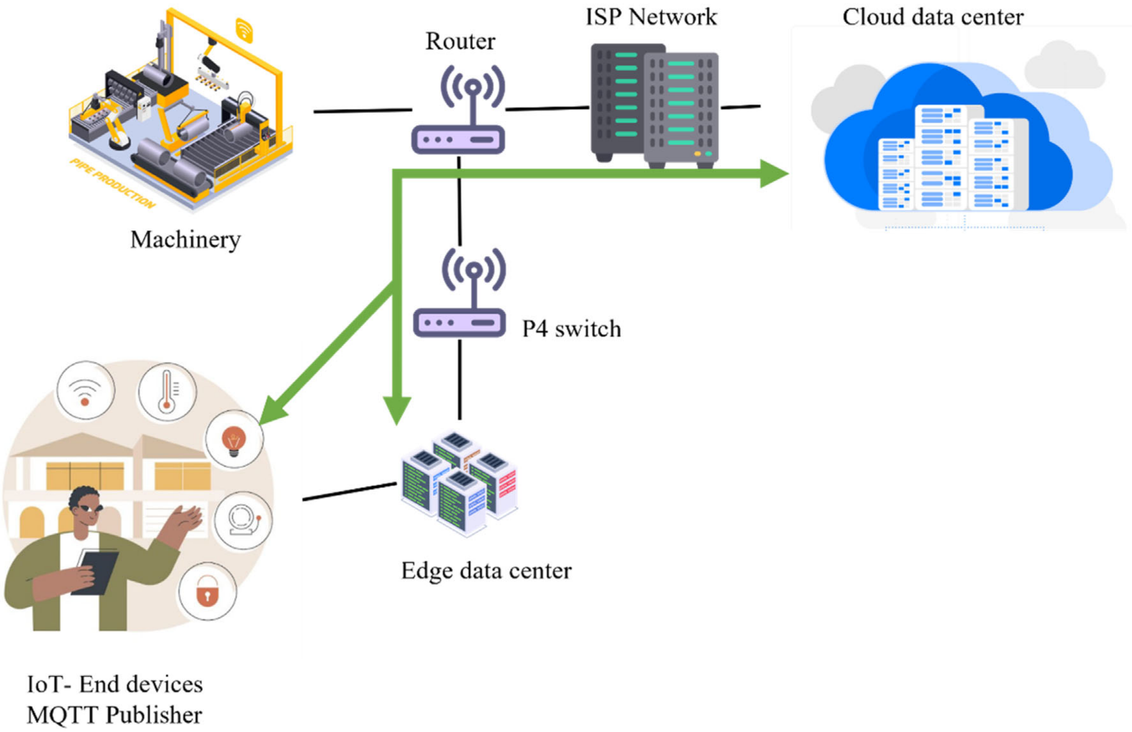

FIGURE 14 depicts the suggested DPP architectural technique.

concept works in real-world situations. DPP proved an effective technique for dealing with communication issues specific to industrial applications. Figure 14 depicts the suggested DPP architectural technique, including a programmable switch and combining edge and cloud paradigms.

Jurdi, et al[132] proposed using a channel assisted with the variable rate in the URLLC system. The proposed method includes two phases: First, a training phase in which the controller estimates the channel state. In the second stage, the controller delivers a message to all devices at an appropriate rate following the current channel state. Compared to other alternatives, the recommended URLLC system achieved better performance. The proposed system delivered ultra-reliable communication across various payload sizes by modifying the rate in each connection. The mechanism's benefits included multiuser variety and great scalability, but the small number of nodes cannot guarantee energy usage.

Chen, et al[133] proposed a method to use a high-order complex power system to improve the system's reliability. Instead of developing a new time-delay controller, they examined how frequency controllers monitored the settings. In addition, the search space for two-dimensional active and reactive power tuning shifted from a "line" to a "plane." The proposed method indicated that the tuned system's stability region increased after adjusting the parameter to the ideal value. Furthermore, dynamic system responses to latency assaults with a more extensive temporal range. This method's format proved complicated, resulting in a high transmission cost.

Lin, et al[134] proposed the True Random Number Generator (TRNG) for a single cell in an IoT network. They suggested the changes in the pulse number as the random generator. They were producing random bits by using the parity of the pulse number. TRNG throughput achieved more than 1 Mbit/sec for a single cell. Implementing chip-level parallel processing on several cells has been experimentally tested. All the random bitstreams produced to pass the National Institute of Standards and Technology (NIST) from −40 to +125°C. The endurance issue was considerably improved using an optimized small analog switching window since TRNG capability is kept after 1011 incremental switching cycles. One of the significant measures of TRNG was single-cell throughput; thus, they progressively increased the amplitude of the impulses to get the Resistive Random-Access Memory (RRAM). Experimental validation at the chip level of the TRNG is shown that 16 RRAM cells operated in parallel with the same pulse configuration. This technique offered several benefits, including high performance, throughput, and low power consumption. Nevertheless, the main limitation was that the convergence time is quite long.

Qin, et al[135] proposed energy-efficient task offloading for Latency-Sensitive Computing service activities (L2SC) in enhanced Mobile-Edge Computing (MEC) and Multiple Radio Access Technologies (multi-RAT) wireless networks. They

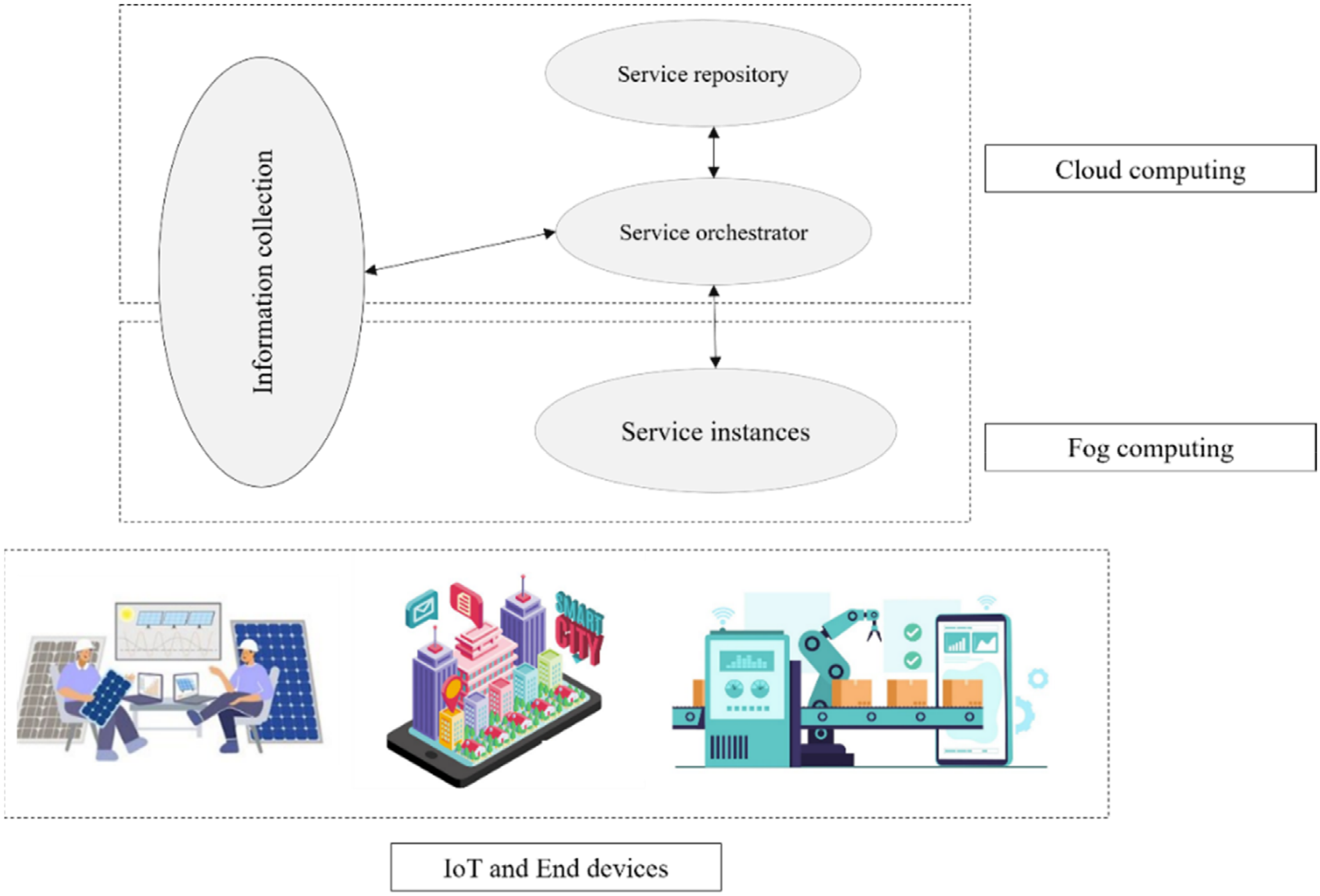

FIGURE 15 Proposed architecture of ILP.

developed an energy-latency challenge in a partial task-offloading environment as a weighted sum of latency expenses and energy use. However, solving this mathematical problem in non-convex and non-smooth form proved challenging. They transformed the tradeoff issue into a smooth biconvex problem and suggested an Alternate Convex Search (ACS) strategy to significantly reduce the computational effort and execution time of the algorithms. They developed a MEC framework to compute task offloading in IoT networks, including a multi-RAT mechanism for low-latency transmission. At the same time, the MEC server provided efficient task computing to meet the UEs' QoS. They also explained the relationship between the UE's battery energy and the overall network cost. The results obtained in this study showed poor availability and insufficient load balancing.

Velasquez, et al[136] suggested a service placement architecture to improve the system's overall performance. Their architecture aimed to place services on convenient fog servers and continuously relocate these services in response to changing network circumstances. Certain concepts of the service and some implementations used Integer Linear Programming (ILP). The strategy included a central module and an Information collection that enabled the users to track changes in the network status and user-system interactions. While exploiting the fog advantages, the architecture allowed intelligent service deployment to ensure latency decrease based on the present network condition as well as the location of the user and servers. Figure 15 shows the ILP method.

Tian, et al[137] proposed using Radio Frequency Fingerprint (RFF) to create a lightweight IIoT architecture. The RFF-based device access authentication technique was proposed to meet the requirements of IIoT. They divided 10 wireless devices into eight legitimate and two unlawful ones. The eight legal wireless devices were included in the training set. The simulation results revealed that at SNR = 10 dB, the recognition rate might approach 95%. Furthermore, the proposed solution demonstrated that it might identify particular IIoT devices and avoid Man-in-the-Middle (MitM) attacks in an IIoT context. As a limitation of this research, scalability has not been considered.

Wang, et al[138] proposed the use of Open Platform Communications (OPC) and Time-Sensitive Software-Defined Networking (TSSDN) for industrial systems. This paper presents efficient solutions and reliable communication links in a three-layer architecture for industrial designs to help transition from old automation systems to cyber-physical systems. The Undefined Architecture (UA-based) and TSSDN switch have been merged in the proposed design. The TSSDN enhanced data transmission reliability and latency, while the OPC gateway protected various communication protocols and data formats. Transmission Control Protocol (TCP) was used in the intelligent industry testbed to communicate between an OPC and a server. The proposed strategy obtained good scalability and reduced complexity.

Zhu, et al[139] proposed an IP wireless communication method using a Building Automation (BA) system. They conducted a practical assessment of IPv6 on the multi-standard wireless platform CC2650. The findings revealed that the implicit CoAP Retransmission Timeout (RTO) was inefficient and caused the “Stair Effect.” They tested various CoAP RTOs in multiple scenarios. The BA system is connected to the primary networks based on the Gateway and internal network. The BA system based on Gateway is used to develop an extensively scalable network.

Kurma, et al[140] proposed a new method based on the channel state protocol. The proposed paper considers the energy consumption mechanism in each sensor and sends the data with a direct cooperative mechanism. This article considers the downlink transmission occurring from the control node (CN) to the target device (TD). N sensors can collaborate with the CN and send the message to TD. The active sensors select the cooperative device based on maximizing the signal-to-noise ratio. The proposed method offers good performance in low energy consumption and low cost but suffers from high complexity and needs to be tested in a real environment.

Karem, et al[141] proposed non-orthogonal multiple access (NOMA) with a UAV system used for increasing the reliability and throughput of the IoT networks. NOMA resource allocation is used for many IoT devices in this method. This paper tries, at first to guarantee the delay of each device by time domain packet scheduler (TPS), and, after that tries to increase the spectral efficiency by maximizing the system rate by a frequency domain packet scheduler. However, the processing time strategy for up linking NOMA is complex. The proposed method performs well with respect to the delay but cost but suffers from high complexity and needs to be tested in a real environment. Table 11 shows the advantage and disadvantage of the channel state information techniques.

6 | DISCUSSION

This section compares different URLLC approaches used in IoT systems. First, the state-of-the-art URLLC is divided into four categories to identify the key concerns and obstacles. Then, based on several characteristics, we examine the similarities and differences among the presented URLLC approaches for IoT. Finally, all recommended methodologies are split into four categories and summarized in Tables 8, 9, 10, and 11 to provide a thorough picture of the presented issue. Next, our paper investigates the benefits and drawbacks of other QoS parameters when URLLC is applied in the IoT network. Numerous studies have shown that various QoS parameters decrease latency and increase reliability. These benefits were identified using information from the articles and comparisons to other approaches in the same area. According to the papers studied, scalability, execution time, and low energy usage have received the most attention from researchers for URLLC scenarios. This distinction may be attributed to the researchers’ focus on time complexity and user satisfaction.

In response to question RQ1 from Table 1, it has been demonstrated that URLLC is vital through increased URLLC papers in journals and conferences. However, this paper identifies specific devices, such as medical sensors and fire sensors, that require low latency and high reliability for communication during critical times. URLLC can enhance user satisfaction and other QoS metrics in most situations. Additionally, IoT devices are spread throughout an active ecosystem, and their communication, energy, and processing capabilities are frequently restricted. Figure 16 indicates that heterogeneity is one of the most apparent characteristics of the IoT, encompassing variations in device types, abilities, and message types sent over the network. Intelligent meters and environmental sensors are examples of IoT devices that regularly transmit short data packets but may also need to send urgent and critical notifications. As a result, we investigate the coexistence of periodic messages, which usually require minimal delay for data transmissions and critical alerts. IoT devices will select the most appropriate Resource Allocation Protocol (RAP)[142] based on the message type.

All categories have pros and cons; this paper cannot advise which techniques are better than others. Still, a solution that can minimize the latency and simultaneously increase reliability, scalability, and execution speed has not been found. Furthermore, only focusing on the URLLC techniques may lose the other QoS parameters. Ten articles on diversity techniques attempted to improve throughput and low packet delivery rate besides URLLC. However, many mathematical operations have been used in these techniques, so the energy consumption is higher than in other categories. Also, according to our investigation, focusing on URLLC can cause to increase the energy consumption. Although many metaheuristic algorithms can be used to solve this issue, most authors try to solve this problem by genetic algorithm. According

TABLE 11 Channel state information techniques.

Authors	Proposed method	Advantage	Disadvantage	Assessments
Atutxa, et al	Message queuing telemetry transport (MQTT)	(+) Low response time (+) High throughput (+) Low threshold	(−) Does not pay attention to energy consumption (−) Need programming ability to understand the concept of a framework	Real environment and simulation
Jurdi, et al	Pilot assisted variable rate by using downlink (DL)	(+) Exploits multiuser diversity (+) High scalability	(−) Set of nodes cannot really ensure energy consumption	Simulation N/A
Chen, et al	Secondary frequency Control (SFC)	(+) Stability region of the tuned system is enhanced (+) Good dynamic system responses under latency	(−) High complexity (−) High cost of communication	Simulation N/A
Lin, et al	True random number generator (TRNG)	(+) High performance (+) High throughput (+) Low power consumption	(−) High convergence time (−) No attention has been paid to latency	Real environment
Qin, et al	Latency sensitive computing service tasks (L2SC)	(+) Low total network cost (+) Low battery energy	(−) Low load balancing (−) Low availability	Simulation N/A
Velasquez, et al	Architecture for service placement	(+) Reducing the latency (+) High performance	(−) Some requirements do not have an exact true or false value (−) Conflict analysis for the workflow	Real environment
Tian, et al	Radio frequency fingerprint (RFF)	(+) Low signal noise ratio (+) Low energy consumption	(−) Scalability is not considered (−) Need programming ability to understand the concept of a framework	Simulation N/A
Zhu, et al	Constrained application protocol (CoAP), and RPL	(+) Large scalability with more than 500 nodes	(−) High cost (−) Does not pay attention to energy consumption	Simulation Matlab
Wang, et al	Open platform communications (OPC) gateways and time-sensitive software-defined networking (TSSDN)	(+) High scalability (+) Low complexity	(−) Many important parameters are ignored (−) Neither implementation nor evaluation is presented	Real environment
Kurma et al	Downlink transmission occurring based on the control node (CN)	(+) Low energy consumption (+) Low cost	(−) High complexity (−) Not tested in real environment	Simulation MATLAB
Karem, et al	Non-orthogonal multiple access (NOMA) with a UAV system	(+) Low energy consumption (+) Low cost	(−) High complexity (−) Not tested in real environment	Simulation MATLAB

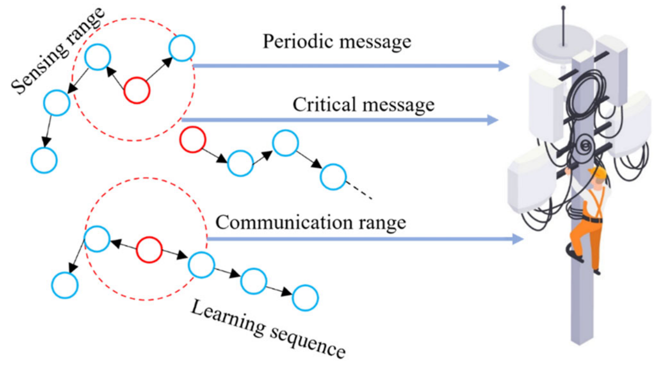

FIGURE 16 All red devices send crucial signals, some blue devices transmit periodic messages, and learning sequences propagate across devices within the communication range in the investigated issue setting diagram.

to the structure of the genetic algorithm, they are suitable for a small area network, but they prove non-counter limitations when the network size or the number of users increases.

On the other hand, URLLC is an Np-hard problem; so many metaheuristic algorithms can effectively solve the latency issue. Furthermore, over 36% of channel state information techniques attempted to enhance scalability. Finally, 33.3% of structure-based methods tried to increase scalability and affordability. Therefore, it is critical to determine the most significant and influential QoS factors. The number of events in which a specific parameter I occur (occur_no[i]) has been recorded individually and divided into all the number of appearances of all parameters to determine the importance percentage (Imp_percentage[i]), as given in (10).

$$imp\_percentage(i) = \frac{occur\_no(i)}{\sum_{j=1}^{param\_no} \text{occur\_no(j)}} * 100\% \tag{10}$$

To answer question RQ3, we discovered that new and recent metaheuristic algorithms have yet to be widely deployed for IoT to solve QoS issues in URLLC. More technology is also needed to enable QoS in the IoT. Section 7 delves into these topics in depth.

7 | FUTURE RESEARCH OVERVIEW & CONCLUSION

The purpose of this section is to examine and analyze the proposed methods. Figure 17 shows the percentage of URLLC metrics in each category. For example, nearly 40% of the studied papers are concerned with energy consumption and availability, while only 10% cover complexity, costs, and scalability. Also, only 10% of the papers presented results obtained after implementation in a real test bed environment; the rest are based on simulation results. Figure 17 shows the average QoS in using the URLLC techniques, while Figure 18 shows the percentage of the methods that attempted to improve a specific parameter concerning URLLC.

7.1 | URLLC in other areas & resource allocation issues

URLLC is a new feature of 5G and 6G networks that may become critical for some scenarios, especially for various applications. However, most research papers and industries ignore this issue. Reliability is the capacity to send a certain quantity of data in a predefined time with a high likelihood of success. Furthermore, reliability is the probability that a system or service will remain operational for a specific period. Availability refers to the probability that a system or service will operate when required. Both concepts are necessary to ensure the network's resilience. Resource allocation offers an acceptable level of service even when facing routine operations. Consequently, URLLC networks provide the required services in all situations and overcome obstacles such as network connection failure, user collisions caused by channel access coordination, virtual network function failure, and an overloaded edge cloud. The authors categorize various resource allocation

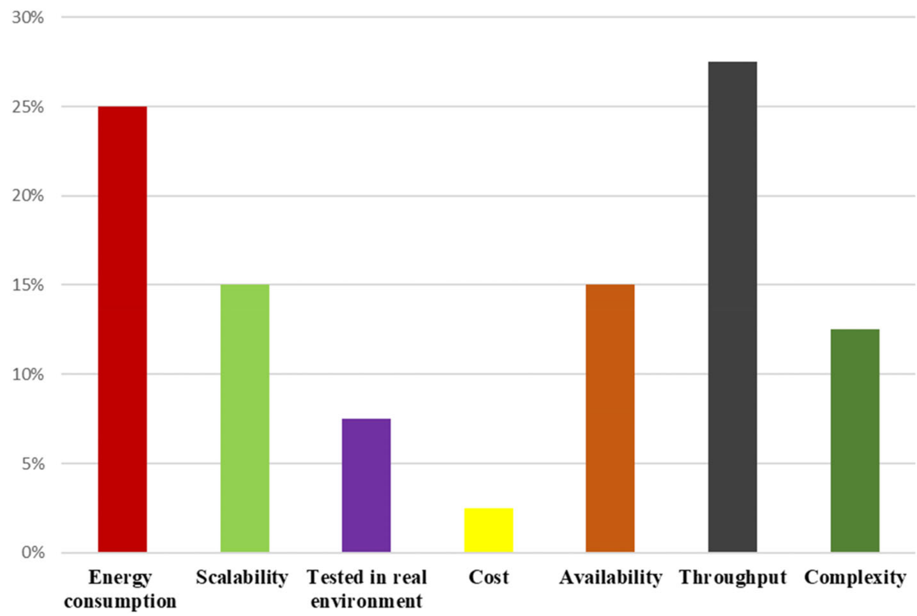

FIGURE 17 Average of URLLC effect on other parameters and improve the other QoS.

algorithms into categories depending on their goals. These models include several game models, such as (i) economic; (ii) predictive; and (iii) robust/failure and recovery. As previously stated, most of the studied approaches fall into the first two types; thus, URLLC slices need more backup resources in the event of failure. However, most consumers do not want or cannot afford the costs of a premium subscription for increased availability. Therefore, future URLLC techniques will be used for various situations, including social networks, SDN, V2V, mobile cloud computing, WSN, VANET, and peer-to-peer networks. Other upcoming research topics include assessing possible energy efficiency benefits in various application areas such as innovative governance, e-commerce, and disaster recovery scenarios, as well as examining the potential advantages of experimental optimization strategies for URLLC.

7.2 | RISK-sensitive approach in URLLC

The LTE-Advanced extensions eMBB, mMTC, and URLLC are designed to enhance peak data rates. mMTC is intended to work with many IoT devices that only send modest amounts of data during their active periods. According to 3GPP, the primary aim of URLLC is to reduce latency to 1 ms while ensuring packet error rates of less than 10. These characteristics are essential for applications such as IIoT, self-driving automobiles, and virtual reality. However, due to strict latency constraints, URLLC traffic has to be transferred promptly and could interfere with previously scheduled eMBB transfers. Therefore, the URLLC technique has recently received significant attention from academia and industry. For eMBB networks, different models such as linear, convex, and threshold, have been explored. Additionally, these authors offer a scheduling strategy for low-latency traffic transmitted using several multiplexing schemes over a combined channel with eMBB transmission. The Conditional Value at Risk (CVaR) is a risk measure defined as a risk-sensitive approach in URLLC. The RBs are distributed across the eMBB network using techniques that ensure proportional fairness. Since the problem is non-convex, researchers have divided it into user scheduling and URLLC device placement. The eMBB user scheduling is solved using integer programming.

7.3 | Competition efficiency

Implementing URLLC faces the challenging task of improving computational efficiency and can progressively achieve encouraging results in IoT. Due to the increased computing complexity in IoT, heuristic techniques can achieve precise identification and explicit inductive inference models. Therefore, heuristic algorithms are critical in increasing efficiency

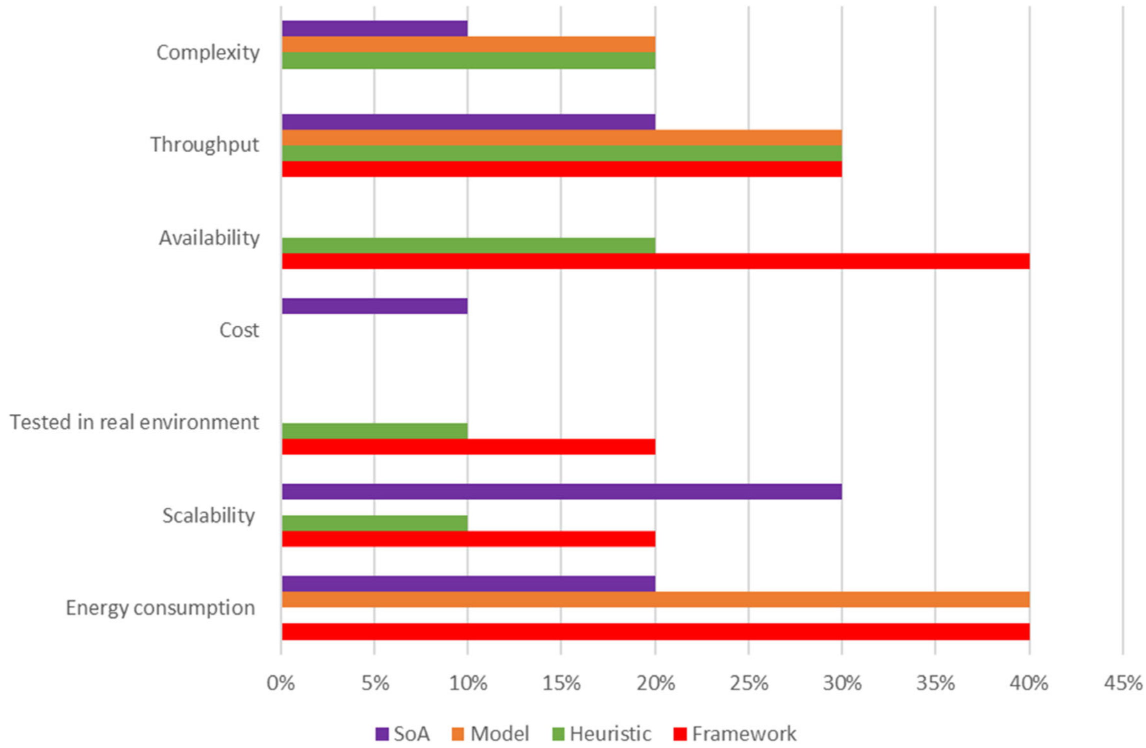

FIGURE 18 Percentage of the improved certain parameter with respect to other method in URLLC techniques.

by analyzing acquired data and assessing transmission connections with IoT device makers. In 5G IoT, the channel state information technique achieves the best computation efficiency. Furthermore, high-performance computing tools can provide excellent computational efficiency for designing better research methodologies using the structure-based technique. On the other hand, the diversity technique may create challenging distribution based on offloading systems to network structures. Therefore, the future trend for improving competition efficiency in IoT-based URLLC in 5G networks is expected to focus on developing advanced technologies and techniques. One of the critical areas of focus will be using ML algorithms to enable proactive resource allocation and improve network performance. For example, AI and ML algorithms can enable predictive maintenance, traffic prediction, and anomaly detection to optimize resource allocation and improve overall network efficiency. Another area of focus will be the development of advanced radio resource management techniques, such as beamforming, interference management, and dynamic time slot allocation. These techniques can help to manage the limited radio resources effectively and ensure reliable communication among IoT devices. Additionally, integrating edge computing and cloud computing technologies with 5G networks will enable processing and storing massive amounts of data closer to the edge, reducing latency and improving overall network efficiency. This will enable the developing of more sophisticated IoT applications, such as autonomous vehicles and smart factories.

7.4 | Hardware technology with low cost in URLLC

Low-cost hardware technology is crucial in implementing URLLC based on IoT, as it can facilitate the deployment of IoT devices on a large scale. One such technology is microcontrollers, which offer a low-cost and low-power solution for IoT devices. Microcontrollers are small, integrated circuits that combine a processor, memory, and input/output peripherals in a single chip, making them ideal for implementing IoT devices that require low-cost, low-power, and real-time processing capabilities. In addition, the use of System-on-Chip (SoC) technology can also contribute to the development of low-cost URLLC solutions in IoT. SoC integrates all the components required for a complete system onto a single chip, including processors, memory, storage, and communication interfaces, which can result in significant cost savings. SoC technology can also enable the development of compact and power efficient IoT devices with real-time processing capabilities. Another approach to achieving low-cost URLLC solutions in IoT is using software-defined radios (SDRs). SDRs are programmable radios that can be reconfigured using software, enabling them to adapt to different wireless standards and communication protocols. As a result, SDRs can provide a low-cost solution for implementing URLLC in IoT by enabling off-the-shelf hardware and reducing the need for specialized hardware components. As radio-frequency access devices and IoT become more critical in meeting future connectivity demands, one of the main challenges is hardware development in the 5G network. This includes the need for additional hardware and developing new software algorithms for non-line-of-sight communications that can function in the Terahertz (THz) range. Low-cost components are essential for hardware development. Furthermore, massive MIMO technologies will expand from 5G to 6G, requiring a new sophisticated architecture, including communication protocols and algorithm design. The progression of radio access towards THz bandwidths, through decreasing hardware costs and lowering interference, will significantly impact the transceiver. The growth in the price of IoT devices is caused by the increased storage capabilities of IoT devices through precise sensing.

7.5 | Scalability and availability

URLLC approaches can provide sufficient capacity to transfer massive data by boosting the efficiency of the infrastructure. Additionally, massive MIMO antennas in IoT can expand the scalability of IoT networks. ML can be used for enormous data sets, and those acquired can extend the availability and scalability of IoT networks. Furthermore, big data administration requires increased capacity for regulated and uncontrolled data at a large scale. Learning architectures must be developed to accommodate many interacting entities while maintaining a high quality of service to address these limitations. A viable design for a real-world implementation is necessary to enable a scalable epistemic uncertainty to estimate in deep learning. However, to build a robust variable for a ML technique, it is necessary to use accurate forecasts for extensive data sets in 5G. The existing trained models can be used to estimate the required scalability, improving overall communication efficiency and reducing processing latency. Nonetheless, the compensated learning phases suggested in the literature offer an accurate ML estimator that can assess availability without expensive testing and provide reliable networks. The future trend for scalability and availability in URLLC based on IoT is expected to focus on developing advanced technologies and techniques to handle the growing demands of IoT applications. One critical focus area will be developing distributed computing architectures that can scale horizontally by adding more nodes to the network. This will enable the system to handle many devices and users without significant degradation in performance. Another focus area will be developing more efficient and reliable communication protocols that can handle the increased data traffic in IoT networks. For example, using edge computing and fog computing can help reduce latency and improve overall system performance. Finally, regarding availability, the focus will be on developing redundancy and fault-tolerant mechanisms to ensure that the system is always available, even during failures or network outages. This can be achieved through the use of redundant hardware, backup power supplies, and distributed data storage.

7.6 | Energy consumption and management

Energy management for 5G networks will be one of the most challenging tasks in the future. Energy efficiency will become even more crucial when lowering the energy consumption per bit (J/bit) as more power is used due to intelligent connections for massive data processing and ultra-large antenna handling. Furthermore, energy management ensures that the received energy is used efficiently. Energy harvesting circuit development for 5G/6G networks is possible due to the improved circuit power consumption and transmission stacks based on design energy awareness. This allows different devices to be self-powered with high efficacy and reduced energy use. 5G/6G networks employ innovative energy management techniques that are vital. Moreover, diversity methodologies can assist these infrastructures and devices implement intelligent energy consumption control strategies. Diversity approaches are also used to enhance energy management. In THz communications systems, finding the optimal power in a vast antenna system is crucial for achieving the performance tradeoff between energy efficiency and overall system reliability. The use of energy harvesting has demonstrated that it could provide the most significant level of energy management reliability. To achieve high throughput using the least energy, efficient energy prediction systems must be used. Additionally, expanded Kalman Filtering methods use an energy control strategy to regulate and reduce energy consumption in IoT and 5G networks. This is accomplished by anticipating the harvesting power for model technology using adaptive security specifications. Energy consumption and management are critical considerations in URLLC-based IoT systems. Due to the limited battery life of IoT devices, energy-efficient communication protocols, and strategies must be implemented to ensure reliable and sustainable operation. One approach to minimizing energy consumption is to optimize the transmission power and modulation scheme based on the channel conditions and signal-to-noise ratio (SNR). Another technique is to utilize sleep modes and wake-up mechanisms to reduce energy consumption during inactivity. In addition, using renewable energy sources, such as solar or kinetic energy harvesting, can help extend the battery life of IoT devices.

7.7 | Conclusion

URLLC is one of the most critical issues in IoT based on available reasons and debates. According to our literature review, the event-driven nature of IoT and inherent constraints are the primary sources of unexpected network demand. Additionally, any growth in the number of IoT devices may result in increased energy consumption, decreased throughput, and packet loss. Therefore, this paper covers the structure, diversity, heuristic, and channel estimation information techniques. This study provides a comprehensive overview and analysis of prominent URLLC approaches in IoT. The review addresses current concerns and research priorities in URLLC. The benefits and drawbacks of various tactics and procedures are emphasized and briefly reviewed. Most research in this sector aims to improve IoT reliability and decrease average end-to-end delays. However, numerous factors are involved in the IoT network, so investigating other QoS factors when URLLC is applied is essential. As a result, employing URLLC may be beneficial for the network's applications.

ACKNOWLEDGMENTS

This study has been partially conducted under the project "Mobility and Training for Beyond 5G Ecosystems (MOTOR5G)". The project has received funding from the European Union's Horizon 2020 program under the Marie Skłodowska Curie Actions (MSCA) Innovative Training Network (ITN), having grant agreement no. 861219.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Seyed Salar Sefati https://orcid.org/0000-0002-7208-3576

REFERENCES

1. Franco J, Aris A, Canberk B, Uluagac AS. A survey of honeypots and honeynets for internet of things, industrial internet of things, and cyber-physical systems. *IEEE Commun Surv Tutor*. 2021;23:2351-2383.
2. Ma Z, Xiao M, Xiao Y, Pang Z, Poor HV, Vucetic B. High-reliability and low-latency wireless communication for internet of things: challenges, fundamentals, and enabling technologies. *IEEE Internet Things J*. 2019;6:7946-7970.
3. Montori F, Bedogni L, Di Felice M, Bononi L. Machine-to-machine wireless communication technologies for the internet of things: taxonomy, comparison and open issues. *Pervasive Mob Comput*. 2018;50:56-81.
4. Bello O, Zeadally S. Toward efficient smartification of the internet of things (IoT) services. *Future Gener Comput Syst*. 2019;92:663-673.
5. Arasteh B, Sefati SS, Halunga S, Fratu O, Allahviranloo T. A hybrid heuristic algorithm using artificial agents for data replication problem in distributed systems. *Symmetry*. 2023;15:487.
6. Amjad Z, Sikora A, Hilt B, Lauffenburger J-P. Low latency V2X applications and network requirements: performance evaluation. Paper presented at: 2018 IEEE Intelligent Vehicles Symposium (IV). 2018 220–225.
7. Touqeer H, Zaman S, Amin R, Hussain M, Al-Turjman F, Bilal M. Smart home security: challenges, issues and solutions at different IoT layers. *J Supercomput*. 2021;77:14053-14089.
8. Shabana N, Velmathi G. Advanced tele-surgery with IoT approach. Paper presented at: Intelligent Embedded Systems: Select Proceedings of ICNETS2, Volume II. 2018 17–24.
9. Sutton GJ, Zeng J, Liu RP, et al. Enabling technologies for ultra-reliable and low latency communications: from PHY and MAC layer perspectives. *IEEE Commun Surv Tutor*. 2019;21:2488-2524.
10. Callebaut G, Leenders G, Van Mulders J, Ottoy G, De Strycker L, Van der Perre L. The art of designing remote IoT devices—technologies and strategies for a long battery life. *Sensors*. 2021;21:913.
11. Sefati S, Navimipour NJ. A qos-aware service composition mechanism in the internet of things using a hidden-markov-model-based optimization algorithm. *IEEE Internet Things J*. 2021;8:15620-15627.

12. Dhanvijay MM, Patil SC. Internet of things: a survey of enabling technologies in healthcare and its applications. *Comput Netw*. 2019;153:113-131.
13. Zhu M, Anwar AH, Wan Z, Cho J-H, Kamhoua CA, Singh MP. A survey of defensive deception: approaches using game theory and machine learning. *IEEE Commun Surv Tutor*. 2021;23:2460-2493.
14. Goulart AE, Sahu A. Cellular IoT for mobile autonomous reporting in the smart grid. *Int J Interdiscip Telecommun Netw*. 2016;8:50-65.
15. Shahraki A, Abbasi M, Piran M, Taherkordi A. "A comprehensive survey on 6G networks: applications, core services, enabling technologies, and future challenges," arXiv preprint arXiv:2101.12475. 2021.
16. Ren H, Pan C, Deng Y, Elkashlan M, Nallanathan A. Resource allocation for URLLC in 5G mission-critical IoT networks. Paper presented at: ICC 2019–2019 IEEE International Conference on Communications (ICC). 2019 1–6.
17. He J, Yang K, Chen H-H. 6G cellular networks and connected autonomous vehicles. *IEEE Netw*. 2020;35:255-261.
18. Gamage H, Rajatheva N, Latva-Aho M. Channel coding for enhanced mobile broadband communication in 5G systems. Paper presented at: 2017 European Conference on Networks and Communications (EuCNC). 2017 1–6.
19. Sharma SK, Wang X. Toward massive machine type communications in ultra-dense cellular IoT networks: current issues and machine learning-assisted solutions. *IEEE Commun Surv Tutor*. 2019;22:426-471.
20. Slalmi A, Saadane R, Chehri A, Kharraz H. How will 5G transform industrial IoT: latency and reliability analysis. Paper presented at: Human Centred Intelligent Systems: Proceedings of KES-HCIS 2020 Conference. 2021 335–345.
21. Sefati SS, Tabrizi SG. Cluster head selection and routing protocol for wireless sensor networks (WSNs) based on software-defined network (SDN) via game of theory. *J Electr Electron Eng*. 2021;9:100-115.
22. Schulz P, Matthe M, Klessig H, et al. Latency critical IoT applications in 5G: perspective on the design of radio interface and network architecture. *IEEE Commun Mag*. 2017;55:70-78.
23. Lee I, Lee K. The internet of things (IoT): applications, investments, and challenges for enterprises. *Bus Horiz*. 2015;58:431-440.
24. Mountaser G, Condoluci M, Mahmoodi T, Dohler M, Mings I. Cloud-RAN in support of URLLC. Paper presented at: 2017 IEEE Globecom Workshops (GC Wkshps). 2017 1–6.
25. Liu T, Li J, Shu F, Han Z. Optimal task allocation in vehicular fog networks requiring URLLC: an energy-aware perspective. *IEEE Trans Netw Sci Eng*. 2019;7:1879-1890.
26. Puliafito C, Mingozzi E, Longo F, Puliafito A, Rana O. Fog computing for the internet of things: a survey. *ACM Trans Internet Technol*. 2019;19:1-41.
27. Husnoo MA, Anwar A, Chakrabortty RK, Doss R, Ryan MJ. Differential privacy for IoT-enabled critical infrastructure: a comprehensive survey. *IEEE Access*. 2021;9:153276-153304.
28. Mumtaz S, Alsohaily A, Pang Z, Rayes A, Tsang KF, Rodriguez J. Massive internet of things for industrial applications: addressing wireless IIoT connectivity challenges and ecosystem fragmentation. *IEEE Ind Electron Mag*. 2017;11:28-33.
29. Zikria YB, Ali R, Afzal MK, Kim SW. Next-generation internet of things (IoT): opportunities, challenges, and solutions. *Sensors*. 2021;21:1174.
30. Baroudi U, Bin-Yahya M, Alshammari M, Yaqoub U. Ticket-based QoS routing optimization using genetic algorithm for WSN applications in smart grid. *J Ambient Intell Humaniz Comput*. 2019;10:1325-1338.
31. Qureshi AR, Nasir AA, Masood M, Abdallah S. Resource allocation for a multi-device URLLC network. *Trans Emerg Telecommun Technol*. 2022;33:e4424.
32. Williams I. 11th CMI International Conference, 2018: Prospects and Challenges Towards Developing a Digital Economy within the EU: IEEE. 2018.
33. Pilz J, Holfeld B, Schmidt A, Septinus K. Professional live audio production: a highly synchronized use case for 5G URLLC systems. *IEEE Netw*. 2018;32:85-91.
34. Zemouri S, Djahel S, Murphy J. A fast, reliable and lightweight distributed dissemination protocol for safety messages in urban vehicular networks. *Ad Hoc Netw*. 2015;27:26-43.
35. Rico D, Merino P. A survey of end-to-end solutions for reliable low-latency communications in 5g networks. *IEEE Access*. 2020;8:192808-192834.
36. Kelechi AH, Alsharif MH, Ramly AM, Abdullah NF, Nordin R. The four-C framework for high capacity ultra-low latency in 5G networks: a review. *Energies*. 2019;12:3449.
37. Shukla S, Hassan MF, Tran DC, Akbar R, Paputungan IV, Khan MK. Improving latency in internet-of-things and cloud computing for real-time data transmission: a systematic literature review (SLR). *Clust Comput*. 2021;1-24.
38. Khan BS, Jangsher S, Ahmed A, Al-Dweik A. URLLC and eMBB in 5G industrial IoT: a survey. *IEEE Open J Commun Soc*. 2022;3:1134-1163.
39. Vaezi M, Azari A, Khosravirad SR, et al. Cellular, wide-area, and non-terrestrial IoT: a survey on 5G advances and the road toward 6G. *IEEE Commun Surv Tutor*. 2022;24:1117-1174.
40. Wang L, Yuan J, Jiang X, Cui J, Zheng B. Dynamic resource scheduling strategy with QoE-aware for the coexistence of eMBB and URLLC traffic. Paper presented at: 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). 2021 1–5.
41. Sefati SS, Halunga S. Data forwarding to fog with guaranteed fault tolerance in internet of things (IoT). Paper presented at: 2022 14th International Conference on Communications (COMM). 2022 1–5.
42. Wu L, Gan C, Xu Z, Hui J. (Network value)-based adaptive dynamic bandwidth allocation algorithm for 5G network slicing. *Trans Emerg Telecommun Technol*. 2022;34:e4722.
43. Zhang L, Liang Y-C, Xiao M. Spectrum sharing for internet of things: a survey. *IEEE Wirel Commun*. 2018;26:132-139.

44. Qian LP, Feng A, Huang Y, Wu Y, Ji B, Shi Z. Optimal SIC ordering and computation resource allocation in MEC-aware NOMA NB-IoT networks. *IEEE Internet Things J*. 2018;6:2806-2816.
45. Zheng G, Zhang H, Li Y, Xi L. 5G network-oriented hierarchical distributed cloud computing system resource optimization scheduling and allocation. *Comput Commun*. 2020;164:88-99.
46. Mahmud R, Kotagiri R, Buyya R. Fog computing: a taxonomy, survey and future directions. *Internet of Everything: Algorithms, Methodologies, Technologies and Perspectives*. Springer, Singapore; 2018:103-130.
47. Na W, Bae B, Cho S, Kim N. Deep-learning based adaptive beam management technique for mobile high-speed 5G mm Wave networks. Paper presented at: 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin). 2019 149–151.
48. Deshpande A, Pitale P, Sanap S. Industrial automation using internet of things (IOT). *Int J Adv Res Comput Eng Technol*. 2016;5:266-269.
49. Muthuramalingam S, Bharathi A, Rakesh Kumar S, Gayathri N, Sathiyaraj R, Balamurugan B. IoT based intelligent transportation system (IoT-ITS) for global perspective: a case study. *Internet Things Big Data Anal Smart Gener*. 2019;154:279-300.
50. Rabie AH, Saleh AI, Ali HA. Smart electrical grids based on cloud, IoT, and big data technologies: state of the art. *J Ambient Intell Humaniz Comput*. 2021;12:9449-9480.
51. Gallagher JC, DeVine ME. Fifth-generation (5G) telecommunications technologies: issues for congress. *Congress Res Serv*. 2019;1:1-39.
52. Vora LJ. Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G. *Int J Modern Trends Eng Res*. 2015;2:281-290.
53. She C, Yang C, Quek TQ. Radio resource management for ultra-reliable and low-latency communications. *IEEE Commun Mag*. 2017;55:72-78.
54. Hampel G, Li C, Li J. 5G ultra-reliable low-latency communications in factory automation leveraging licensed and unlicensed bands. *IEEE Commun Mag*. 2019;57:117-123.
55. Haque KF, Abdelgawad A, Yanambaka VP, Yelamarthi K. Lora architecture for v2x communication: an experimental evaluation with vehicles on the move. *Sensors*. 2020;20:6876.
56. Moubayed A, Shami A, Heidari P, Larabi A, Brunner R. Edge-enabled V2X service placement for intelligent transportation systems. *IEEE Trans Mob Comput*. 2020;20:1380-1392.
57. Isermann R, Schwarz R, Stolzl S. Fault-tolerant drive-by-wire systems. *IEEE Control Syst Mag*. 2002;22:64-81.
58. Abdukhakimov A, Bhardwaj S, Gashema G, Kim D-S. Reliability analysis in smart grid networks considering distributed energy resources and storage devices. *Int J Electr Electron Eng Telecommun*. 2019;8:233-237.
59. Wen M, Li Q, Kim KJ, et al. Private 5G networks: concepts, architectures, and research landscape. *IEEE J Sel Topics Signal Process*. 2021;16:7-25.
60. Sefati SS, Halunga S, Farkhady RZ. Cluster selection for load balancing in flying ad hoc networks using an optimal low-energy adaptive clustering hierarchy based on optimization approach. *Aircr. Eng. Aerosp*. 2022;94:1344-1356.
61. Rezaei A, Khalili A, Jalali J, Shafiei H, Wu Q. Energy-efficient resource allocation and antenna selection for IRS-assisted multicell downlink networks. *IEEE Wirel Commun Lett*. 2022;11:1229-1233.
62. Erol-Kantarci M, Sukhmani S. Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G. Paper presented at: Ad Hoc Networks: 9th International Conference, AdHocNets 2017, Niagara Falls, ON, Canada, September 28–29, 2017, Proceedings. 2018 169–177.
63. Bastug E, Bennis M, Médard M, Debbah M. Toward interconnected virtual reality: opportunities, challenges, and enablers. *IEEE Commun Mag*. 2017;55:110-117.
64. Sefati S, Mousavinasab M, Zareh Farkhady R. Load balancing in cloud computing environment using the Grey wolf optimization algorithm based on the reliability: performance evaluation. *J Supercomput*. 2022;78:18-42.
65. Fraga-Lamas P, Fernández-Caramés TM, Suárez-Albela M, Castedo L, González-López M. A review on internet of things for defense and public safety. *Sensors*. 2016;16:1644.
66. Mehmood I, Ullah A, Muhammad K, et al. Efficient image recognition and retrieval on IoT-assisted energy-constrained platforms from big data repositories. *IEEE Internet Things J*. 2019;6:9246-9255.
67. Lusvarghi L, Merani ML. On the coexistence of aperiodic and periodic traffic in cellular vehicle-to-everything. *IEEE Access*. 2020;8:207076-207088.
68. Palattella MR, Dohler M, Grieco A, et al. Internet of things in the 5G era: enablers, architecture, and business models. *IEEE J Sel Areas Commun*. 2016;34:510-527.
69. Girs S, Sentilles S, Asadollah SA, Ashjaei M, Mubeen S. A systematic literature study on definition and modeling of service-level agreements for cloud services in IoT. *IEEE Access*. 2020;8:134498-134513.
70. Mohammadzadeh A, Masdari M. Scientific workflow scheduling in multi-cloud computing using a hybrid multi-objective optimization algorithm. *J Ambient Intell Humaniz Comput*. 2021;1-21.
71. Zhang Y, Yu R, Xie S, Yao W, Xiao Y, Guizani M. Home M2M networks: architectures, standards, and QoS improvement. *IEEE Commun Mag*. 2011;49:44-52.
72. Liu G, Huang Y, Chen Z, Liu L, Wang Q, Li N. 5G deployment: standalone vs. non-standalone from the operator perspective. *IEEE Commun Mag*. 2020;58:83-89.
73. Wang J, Jin A, Shi D, et al. Spectral efficiency improvement with 5G technologies: results from field tests. *IEEE J Sel Areas Commun*. 2017;35:1867-1875.
74. Tekir G. Huawei, 5G network and digital geopolitics. *Int J Politics Secur*. 2020;2:113-135.

75. Siddiqi MA, Yu H, Joung J. 5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices. *Electronics*. 2019;8:981.
76. Simsek M, Aijaz A, Dohler M, Sachs J, Fettweis G. 5G-enabled tactile internet. *IEEE J Sel Areas Commun*. 2016;34:460-473.
77. Steinbach E, Strese M, Eid M, et al. Haptic codecs for the tactile internet. *Proc IEEE*. 2018;107:447-470.
78. Ge X, Zhou R, Li Q. 5G NFV-based tactile internet for mission-critical IoT services. *IEEE Internet Things J*. 2019;7:6150-6163.
79. Ahmadi K, Ghassemian M, Condoluci M, Dohler M. X2-Based signalling mechanisms for downlink uplink decoupling in next generation communication systems. *IEEE Access*; 2022;10:88941-88955.
80. Liu Y, Habibnezhad M, Jebelli H. Brain-computer interface for hands-free teleoperation of construction robots. *Automat Construct*. 2021;123:103523.
81. Gupta R, Tanwar S, Tyagi S, Kumar N. Tactile internet and its applications in 5G era: a comprehensive review. *Int J Commun Syst*. 2019;32:e3981.
82. Li C-P, Jiang J, Chen W, Ji T, Smee J. 5G ultra-reliable and low-latency systems design. Paper presented at: 2017 European Conference on Networks and Communications (EuCNC). 2017 1–5.
83. Kim Y, Kim Y, Oh J, et al. New radio (NR) and its evolution toward 5G-advanced. *IEEE Wirel Commun*. 2019;26:2-7.
84. Sichitiu ML. Cross-layer scheduling for power efficiency in wireless sensor networks. Paper presented at: 2004 IEEE INFOCOM. 2004 1740–1750.
85. Liu Y, Tong K, Mao F, Yang J. Research on digital production technology for traditional manufacturing enterprises based on industrial internet of things in 5G era. *Int J Adv Manuf Technol*. 2020;107:1101-1114.
86. Chochliouros IP, Kourtis M-A, Spiliopoulou AS, et al. Energy efficiency concerns and trends in future 5G network infrastructures. *Energies*. 2021;14:5392.
87. Fisher W, Suchara M, Rexford J. Greening backbone networks: reducing energy consumption by shutting off cables in bundled links. Proceedings of the First ACM SIGCOMM Workshop on Green Networking. 2010 29–34.
88. Shih E, Bahl P, Sinclair MJ. Wake on wireless: an event driven energy saving strategy for battery operated devices. Proceedings of the 8th Annual International Conference on Mobile Computing and Networking. 2002 160–171.
89. Duan X, Wang X. Authentication handover and privacy protection in 5G hetnets using software-defined networking. *IEEE Commun Mag*. 2015;53:28-35.
90. Pedersen KI, Kolding TE, Frederiksen F, Kovács IZ, Laselva D, Mogensen PE. An overview of downlink radio resource management for UTRAN long-term evolution. *IEEE Commun Mag*. 2009;47:86-93.
91. Salh A, Audah L, Shah NSM, et al. A survey on deep learning for ultra-reliable and low-latency communications challenges on 6G wireless systems. *IEEE Access*. 2021;9:55098-55131.
92. Ji H, Park S, Yeo J, Kim Y, Lee J, Shim B. Ultra-reliable and low-latency communications in 5G downlink: physical layer aspects. *IEEE Wirel Commun*. 2018;25:124-130.
93. Le T-K, Salim U, Kaltenberger F. An overview of physical layer design for ultra-reliable low-latency communications in 3GPP releases 15, 16, and 17. *IEEE Access*. 2020;9:433-444.
94. Lee B, Park S, Love DJ, Ji H, Shim B. Packet structure and receiver design for low latency wireless communications with ultra-short packets. *IEEE Trans Commun*. 2017;66:796-807.
95. Jin J, Gubbi J, Luo T, Palaniswami M. Network architecture and QoS issues in the internet of things for a smart city. Paper presented at: 2012 International Symposium on Communications and Information Technologies (ISCIT). 2012 956–961.
96. Pocovi G, Esswie AA, Pedersen KI. Channel quality feedback enhancements for accurate URLLC link adaptation in 5G systems. Paper presented at: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). 2020 1–6.
97. Ghobaei-Arani M, Rahmanian AA, Souri A, Rahmani AM. A moth-flame optimization algorithm for web service composition in cloud computing: simulation and verification. *Softw: Pract Exp*. 2018;48:1865-1892.
98. Zhao X, Chen W, Poor HV. Achieving extremely low-latency in industrial internet of things: joint finite blocklength coding, resource block matching, and performance analysis. *IEEE Trans Commun*. 2021;69:6529-6544.
99. Park T, Saad W. Distributed learning for low latency machine type communication in a massive internet of things. *IEEE Internet Things J*. 2019;6:5562-5576.
100. Avranas A, Kountouris M, Ciblat P. Energy-latency tradeoff in ultra-reliable low-latency communication with retransmissions. *IEEE J Sel Areas Commun*. 2018;36:2475-2485.
101. Yu D, Li W, Xu H, Zhang L. Low reliable and low latency communications for mission critical distributed industrial internet of things. *IEEE Commun Lett*. 2020;25:313-317.
102. Din S, Ahmad A, Paul A, Rho S. MGR: multi-parameter green reliable communication for internet of things in 5G network. *J Parallel Distrib Comput*. 2018;118:34-45.
103. Zeng J, Lv T, Liu RP, Su X, Guo YJ, Beaulieu NC. Enabling ultrareliable and low-latency communications under shadow fading by massive MU-MIMO. *IEEE Internet Things J*. 2019;7:234-246.
104. Mahyoub M, Mahmoud A, Abu-Amara M, Sheltami T. Reliable and low-latency routing mechanism for IoT-based networks. Paper presented at: 2019 IEEE 14th Malaysia International Conference on Communication (MICC). 2019 81–85.
105. El Haber E, Alameddine HA, Assi C, Sharafeddine S. UAV-aided ultra-reliable low-latency computation offloading in future IoT networks. *IEEE Trans Commun*. 2021;69:6838-6851.
106. Pang A-C, Chung W-H, Chiu T-C, Zhang J. Latency-driven cooperative task computing in multi-user fog-radio access networks. Paper presented at: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). 2017 615–624.

107. Zhang J, Gao H, Zhang K, Chen Q, Li J. A distributed framework for low-latency data collection in battery-free wireless sensor networks. *IEEE Internet Things J*. 2021;9:8438-8453.
108. Zhang Y, Tang W, Liu Y. Multicell Grant-free uplink IoT networks with hard deadline services in URLLC. *IEEE Wirel Commun Lett*. 2022;11:1448-1452.
109. Elgabli A, Khan H, Krouka M, Bennis M. Reinforcement learning based scheduling algorithm for optimizing age of information in ultra reliable low latency networks. Paper presented at: 2019 IEEE Symposium on Computers and Communications (ISCC). 2019 1–6.
110. He Q, Zheng P, Zhu Y, Hu Y, Schmeink A. Multi-device low-latency internet of things networks with blind retransmissions in the finite blocklength regime. Paper presented at: 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. 2020 1–6.
111. Seo H, Hong J-P, Choi W. Low latency random access for sporadic MTC devices in internet of things. *IEEE Internet Things J*. 2019;6:5108-5118.
112. van Rensburg R, Mellado B, Tobon CAM. Wireless mesh network data communications and reliability analysis for anti-theft application deployment in educational institutions. *J Eng Des Technol*. 2019;17:474-489.
113. Nakao M, Sakai M, Hanada Y, Murai H, Sato M. Graph optimization algorithm for low-latency interconnection networks. *Parallel Comput*. 2021;106:102805.
114. Ye N, Li X, Yu H, Wang A, Liu W, Hou X. Deep learning aided grant-free NOMA toward reliable low-latency access in tactile internet of things. *IEEE Trans Industr Inform*. 2019;15:2995-3005.
115. Liu G, Wang Z, Hu J, Ding Z, Fan P. Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications. *IEEE Internet Things J*. 2019;6:7828-7838.
116. Sun C, Yang C. Unsupervised deep learning for ultra-reliable and low-latency communications. Paper presented at: 2019 IEEE Global Communications Conference (GLOBECOM). 2019 1–6.
117. Liu C-H, Liang D-C, Chen K-C, Gau R-H. Ultra-reliable and low-latency communications using proactive multi-cell association. *IEEE Trans Commun*. 2021;69:3879-3897.
118. Zhang T, Ren P, Xu D, Ren Z. Intelligent radio frequency identification for URLLC in industrial IoT networks. *Symmetry*. 2022;14:801.
119. Sefati SS, Halunga S. Mobile sink assisted data gathering for URLLC in IoT using a fuzzy logic system. Paper presented at: 2022 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom). 2022 379–384.
120. Sefati SS, Halunga S. A hybrid service selection and composition for cloud computing using the adaptive penalty function in genetic and artificial bee Colony algorithm. *Sensors*. 2022;22:4873.
121. Bhardwaj S, Kim D-S. Dragonfly-based swarm system model for node identification in ultra-reliable low-latency communication. *Neural Comput Appl*. 2021;33:1837-1880.
122. Wang Q, Chen S. Latency-minimum offloading decision and resource allocation for fog-enabled internet of things networks. *Trans Emerg Telecommun Technol*. 2020;31:e3880.
123. Singh A, Nagaraju A. Low latency and energy efficient routing-aware network coding-based data transmission in multi-hop and multi-sink WSN. *Ad Hoc Netw*. 2020;107:102182.
124. Babar M, Khan MS, Din A, Ali F, Habib U, Kwak KS. Intelligent computation offloading for IoT applications in scalable edge computing using artificial bee colony optimization. *Complex*. 2021;2021:1-12.
125. Chang Y, Yuan X, Li B, Niyato D, Al-Dhahir N. Machine-learning-based parallel genetic algorithms for multi-objective optimization in ultra-reliable low-latency WSNs. *IEEE Access*. 2018;7:4913-4926.
126. Aburukba RO, AliKarrar M, Landolsi T, El-Fakih K. Scheduling internet of things requests to minimize latency in hybrid fog–cloud computing. *Future Gener Comput Syst*. 2020;111:539-551.
127. Cui L, Xu C, Yang S, et al. Joint optimization of energy consumption and latency in mobile edge computing for internet of things. *IEEE Internet Things J*. 2018;6:4791-4803.
128. Javanmardi S, Shojafar M, Persico V, Pescapè A. FPFTS: a joint fuzzy particle swarm optimization mobility-aware approach to fog task scheduling algorithm for internet of things devices. *Softw: Pract Exp*. 2021;51:2519-2539.
129. Seyfollahi A, Ghaffari A. Reliable data dissemination for the internet of things using Harris hawks optimization. *Peer-to-Peer Netw Appl*. 2020;13:1886-1902.
130. Kanagaraj G, Ponnambalam S, Jawahar N. A hybrid cuckoo search and genetic algorithm for reliability–redundancy allocation problems. *Comput Ind Eng*. 2013;66:1115-1124.
131. Atutxa A, Franco D, Sasiain J, Astorga J, Jacob E. Achieving low latency communications in smart industrial networks with programmable data planes. *Sensors*. 2021;21:5199.
132. Jurdi R, Khosravirad SR, Viswanathan H. Variable-rate ultra-reliable and low-latency communication for industrial automation. Paper presented at: 2018 52nd Annual Conference on Information Sciences and Systems (CISS). 2018 1–6.
133. Chen C, Chen Y, Zhang K, Ni M, Wang S, Liang R. System redundancy enhancement of secondary frequency control under latency attacks. *IEEE Trans Smart Grid*. 2020;12:647-658.
134. Lin B, Gao B, Pang Y, et al. A high-speed and high-reliability TRNG based on analog RRAM for IoT security application. Paper presented at: 2019 IEEE International Electron Devices Meeting (IEDM). 2019 14.8.1–14.8.4.
135. Qin M, Cheng N, Jing Z, et al. Service-oriented energy-latency tradeoff for IoT task partial offloading in MEC-enhanced multi-RAT networks. *IEEE Internet Things J*. 2020;8:1896-1907.
136. Velasquez K, Abreu DP, Curado M, Monteiro E. Service placement for latency reduction in the internet of things. *Ann Telecommun*. 2017;72:105-115.

137. Tian Q, Lin Y, Guo X, et al. New security mechanisms of high-reliability IoT communication based on radio frequency fingerprint. *IEEE Internet Things J*. 2019;6:7980-7987.
138. Wang R, Ji L, Ren T, He S, Shi Z. A low-latency and interoperable industrial internet of things architecture for manufacturing systems. Paper presented at: 2020 IEEE 18th International Conference on Industrial Informatics (INDIN). 2020 859–864.
139. Zhu H, Pang Z, Xie B, Bag G. IETF IoT based wireless communication for latency-sensitive use cases in building automation. Paper presented at: 2016 IEEE 25th International Symposium on Industrial Electronics (ISIE). 2016 1168–1173.
140. Kurma S, Sharma PK, Singh K, Mumtaz S, Li C-P. URLLC-based cooperative industrial IoT networks with nonlinear energy harvesting. *IEEE Trans Industr Inform*. 2022;19:2078-2088.
141. Karem R, Ahmed M, Newagy F. Resource allocation in uplink Noma-iot based UAV for URLLC applications. *Sensors*. 2022;22:1566.
142. Sefati SS, Tabrizi SG. Detecting Sybil attack in vehicular ad-hoc networks (vanets) by using fitness function, signal strength index and throughput. *Wirel Personal Commun*. 2022;123:1-21.

How to cite this article: Sefati SS, Halunga S. Ultra-reliability and low-latency communications on the internet of things based on 5G network: Literature review, classification, and future research view. *Trans Emerging Tel Tech*. 2023;e4770. doi: 10.1002/ett.4770