1. Introduction

With the everyday increase in data traffic requirements ranging from mission-critical to massive machine connectivity, the anticipation of the fifth-generation (5G) is growing at an exponential rate. Although such requirements open new doors to exciting features and business models, the provision of these requirements for such intense traffic and diverse services (Figure 1) remains a challenge for the telecom industry.

Figure 1. Road to 5G networks and their diverse services.
These challenges are further enhanced by elements such as expectations of node-to-node or machine-to-machine communications requirements, the uncertainty of topology, diverse application requirements, backward compatibility, user equipment resource limitations, and the rapidly increasing number of devices. These elements exacerbate the technical complications of the implementation of future 5G networks.

One of the crucial supporting technologies for the implementation of 5G is new radio (NR) [1], a new radio access technology. NR is the new radio interface and access method that is developed to facilitate the growing requirements of 5G in the coming future [2]. As shown in Figure 2, NR provides a flexible frame structure to support all 5G service requirements defined by the International Telecommunication Union (ITU).

![Figure 2. The flexible frame structure of 5G new radio (NR) [2]. URLLC, ultra-reliable low latency communications; eMBB, enhanced mobile broadband; mMTC, massive machine-type communications; IoT, Internet of Things; NB, Narrowband.](image)

The new NR facilitates communications between the base station and the user/mobile device. With such rapid changes in technology and user demands, NR is designed not only to facilitate the three defined categories of ITU, but also to provide space for future growth in technology, as shown in Figure 3. The three basic categories formulated by ITU for 5G [3] are as follows:

1. enhanced mobile broadband (eMBB),
2. ultra-reliable low-latency communications (URLLC), and
3. massive machine-type communications (mMTC).

![Figure 3. Features and improvement plan for future 5G [4].](image)
2. Importance of URLLC

Although all three categories defined by ITU have their significance as shown in Table 1. URLLC has recently been very intriguing for researchers worldwide. A field experiment carried out by the NTT DOCOMO Inc. and Huawei on URLLC showed quite positive results [5].

Table 1. Intended features for the categories defined by the International Telecommunication Union (ITU) [3,4]. URLLC, ultra-reliable low latency communications; eMBB, enhanced mobile broadband; mMTC, massive machine-type communications.

| Category | Basic Features |
|----------|----------------|
| eMBB     | eMBB focuses on a higher data rate, with a large payload and prolonged internet connectivity based applications. Potential applications could include cloud office/gaming, virtual/augmented reality (VR/AR) and three-dimension/ultra-high-definition (3D/UHD) video. |
| URLLC    | URLLC focuses on an ultra-responsive connection with ultra-low latency. The data rate is not expected to be very high in URLLC, but offers high mobility. Potential applications of URLLC include industrial automation, autonomous driving, mission-critical applications, and remote medical assistance. |
| mMTC     | mMTC focus on providing connectivity to a large number of devices (IoTs), but with low reliability. It can provide long-range communication with energy efficiency and asynchronous access. Such features are very suitable for low power devices in a massive quantity. |

Achieving URLLC requirements is quite a challenge for 5G networks and will require massive modifications to the system design of the current telecom infrastructure. Owing to the encouraging results achieved with URLLC, it can play an integral role in the 5G era. Although current user requirements are initially based on high bandwidth, latency and reliability are also expected to play a vital role in real-time applications and mission-critical networks. Table 2 clearly highlights the importance of reliability and latency in future applications.

Table 2. Importance of reliability and low-latency as per industrial/user requirements [6,7].

| Industry                  | Application                                | Importance of Reliability and Low Latency |
|---------------------------|--------------------------------------------|------------------------------------------|
| Medical and Health Care   | Remote surgery/patient diagnosis.          | Remote surgery or remote patient’s diagnosis might be carried out with the help of a robot. In such cases, the reliability of data transmitted as instruction for robot needs to be ultra-reliable because even a slight latency or delay could be very harmful to the patient. |
| Media/Entertainment/Business | Live reporting of an event, live sports events, online gaming, cloud-based entertainment (VR/AR). | With the help of technology, the entire world is shrinking in terms of communications. Users desire to be up to date on world events and entertainment in real-time. Even in terms of business, the delay could make a huge impact on trades carried out in the world. In online gaming, the lag could be very frustrating for gamers. |
| Transport                 | Drone-based delivery, remote driving, self-driven cars, traffic management, sub-station management (system synchronization, traffic management) | Through new features and attractions for users such as Amazon Prime Air [8] to deliver orders, it is very important for drones to respond in real-time. Similar to Amazon Prime Air, Google’s self-driven car (WAYMO) [9] is quite important for the future automobile industry. The importance of reliability and latency is self-explanatory in such projects. |
| Industrial Automation     | Control systems, automated assembly lines with robots, machine status reports, process surveillance, power grid management. | In order to maximize productivity, industries have moved toward automation. Higher reliability and productivity can be obtained by replacing humans with robots in the manufacturing process. Apart from the manufacturing industry, the agriculture, journalism, and education sectors have also moved towards automation [10]. In the mentioned industrial areas, reliability will be a key factor. Such as that the automated car assembly line must have minimum latency to keep up with the moving tray and high reliability to avoid any damage to the car parts during assembly. |
3. Issues in Implementing URLLC

3.1. Quality of Service (QoS) for URLLC

URLLC-focused applications require an end-to-end (E2E) delivery of data with reliability, security, and minimum latency. Such requirements have driven the 3rd Generation Partnership Project (3GPP) to set desired quality of service (QoS) requirements such as an air interface latency of 1 ms and 99.999% system reliability for URLLC [11]. These QoS requirements for URLLC, depending on its various applications, are shown in Table 3.

| Industry                      | Error Rate/Reliability | Latency (ms) |
|-------------------------------|------------------------|--------------|
| Augmented/Virtual Reality     | $10^{-3}$–$10^{-5}$    | 5–10         |
| Autonomies/guided vehicle     | $\geq 10^{-3}$         | 5–10         |
| Automated Industry            | $10^{-5}$–$10^{-9}$    | 1            |
| IoT (Internet of things/Tactile Internet) | $10^{-5}$    | 1            |

The channel quality and lack of dedicated bandwidth can be an obstacle to meeting the desired latency requirement for URLLC [13]. To achieve the desired reliability in URLLC is also a challenge. As several mobile applications rely on different methods of retransmission, the retransmission of data in URLLC can degrade latency [11], unless the retransmission methods are designed as per URLLC requirements. The current 4G long-term evolution (LTE) and NR hybrid automatic repeat request (HARQ) are not quite appropriate to handle URLLC requirements. These methods depend on the complete or partial retransmission (RTX) of the error packet with additional time for HARQ processing, which is not suitable for time-critical applications [14]. An alternative method to achieve a low block error rate (BLER) and avoid RTX is to allocate high resources to the system; however, this might result in poor system capacity and low spectral efficiency. Therefore, there exists a trade-off between reliability and latency, which can be based on the application requirements. The physical layer plays a major role to achieve such a low latency and reliability; however, three major concerns exist [15]. First, system overhead in term of channel access, user schedule, and allocation of resources should be minimized. Second, the packet error probability should be minimized to achieve lower latency because the retransmission of packets can affect the latency, as mentioned earlier. Third, the transmission of URLLC packets should be prioritized, and they should be transmitted as soon as they are generated. Although the requirements mentioned by 3GPP and ITU for URLLC are based on one-way communication, latency should be defined on the basis of E2E communication [16].

3.2. Coexistence with eMBB

The emerging 5G network must provide services to diversified applications with different requirements. Applications relying on URLLC require low latency with high reliability, whereas eMBB requires high data rates. For the existence of URLLC and eMBB in the same physical resource, as shown in Figure 4, an efficient coexistence method is needed to maintain the required QoS. Such coexistence on the same radio spectrum will open doors to new concerns in the scheduling optimization [17].

![Figure 4. Coexistence of URLLC and eMBB. BS, base station.](image_url)
The proposed agile 5G frame structure [18] shows promising results for URLLC latency requirements by utilizing different transmission time intervals (TTIs) for URLLC and eMMB to meet their desired spectral efficiencies (SE). For example, URLLC traffic can be scheduled on a smaller TTI duration to achieve its low latency goal, and eMBB traffic can be scheduled with a long TTI duration to maintain its extreme SE requirements. However, such a case will bring an additional overhead to the control signaling, which can result in the degradation of the control channel (CCH) capacity [17].

3.3. URLLC Packet Design

Packet design is one of the key issues in URLLC. With an effective packet structure, the latency can be minimized in terms of packet processing time and packet transmission time [15]. Packet processing involves the time of acquiring a packet, accessing channel information, extracting scheduling (control) information, decoding the packet, and checking errors. As per URLLC requirements, the 5G NR system employs a non-square-shaped packet in the frequency domain with polar code for the control channel and low-density parity-check (LDPC) for the data channel to minimize the transmit latency. However, in LTE, a square-shaped packet is generally utilized for effective spectrum utilization [19].

3.4. URLLC Scheduling

The scheduling of an unexpected packet generation by URLLC is one of the most significant issues. When user data arrives, it is stored in a user-specific transmission buffer, as illustrated in Figure 5. The transmission of each packet takes no less than one TTI. However, radio channel conditions, payload size, and availability of resources may force scheduling to increase the TTI of a packet [20].

![Figure 5. User-specific transmission buffer [20]. UE, user equipment.](image)

The NR defined by 3GPP for 5G has proposed two scheduling schemes, instant scheduling and reservation-based scheduling, to handle URLLC packets [20]. The instant scheduling approach proposes to facilitate URLLC packets whenever they are generated. Therefore, this scheduling can interrupt the ongoing data transmission. Consequently, this approach can result in a drastic degradation of other services. The reservation-based scheduling is further divided into two types, semi-static and dynamic reservation, for effective handling of packets. Both approaches use a reservation-based frame for URLLC, which results in overheads in the control signaling. In the case of no URLLC data, the reserved slot may be wasted.

3.5. Energy Efficiency Concern for End-User Device

Most of the wireless devices employ a sleep mode operation to save energy. The devices need to act immediately upon receiving a packet from a network to avoid any delay. Similarly, the devices periodically check awaiting packets on the network to avoid latency [21]. The current energy-saving states defined for user equipment (UE) are not suitable for URLLC-based service. However, the UE can lose battery drastically as a result of the high frequency of data checks over the network.

3.6. Handover Issues for URLLC

Handover (handoff) is one of the most integral parts of any telecom infrastructure. NR for 5G must be able to support the mobility requirements illustrated in Table 4.
### Table 4. Mobility requirements for URLLC [22].

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

In 5G NR, the basic handover process is quite similar to that of the LTE handover [23]. NR supports handover at two different levels to manage seamless handover. Cell level mobility is managed using a radio resource control (RRC) layer in the same way as in the LTE handover. In addition, the beam level mobility is handled using physical and medium access control (MAC) layers without involving RRC for low latency [24]. As NR adopts the same handover signaling procedures as the LTE [25], it inherits two unresolved issues of mobility robustness and mobility interruption time (MIT). Because of the mentioned concerns, enhancements in NR are considered and proposed to achieve zero handover interruption time (HIT) and handover failure (HOF) for URLLC [24]. Further studies are still needed to satisfy the requirements of URLLC QoS.

#### 3.7. Error Handling

Owing to the faulty nature of the data channel [26], the handling of packet errors is another issue to meet along with the defined latency requirements for URLLC; for example, the 1 ms latency deadline [17]. The current LTE provides a very low error rate at the cost of higher latency, which is not suitable for URLLC [27].

As shown in Figure 6, when data arrives at the base station (BS) buffer, a request for resource grant (RG) is transmitted to the target UE. As data is received, the UE decodes the data and responds with either a positive or negative acknowledgment (ACK/NACK) based on the success of the data decoding. In a case where the UE fails to respond within the allocated time, the BS retransmits the data. Compared with the LTE, URLLC operates in a shorter TTI and requires a faster response from the UE to avoid retransmission. Wireless channel impairments can be another concern. If the BS does not receive an ACK/NACK within the mentioned deadline as a result of channel fading, the BS retransmits the data. This can increase latency, and hence the wastage of resources [27]. To overcome this issue, the stronger channel coding and multiple antenna technologies can be considered. However, the stronger channel coding may require longer decoding latency, for example, more iterations in an iterative channel decoding scheme. As a result of high power consumption and space limitation, multiple antennas cannot be equipped on the Internet of things (IoT) devices.

**Figure 6.** Signaling procedure for downlink data transmission. ACK/NACK, acknowledgment/ negative acknowledgment.

#### 3.8. Beamforming and mmWave Frequency Communications

The next-generation mobile networks will operate with mmWave frequencies in order to increase bandwidth. In mmWave communications, beamforming between the UE and the BS becomes an important aspect, as shown in Figure 7.
The process of beam selection can affect the E2E performance and the QoS [28]. In order to counter the key issues of achieving precise beamforming for performance enhancement [29], the 3GPP NR standard included new MAC and physical layer (PHY) features. The new MAC and PHY features support directional communications [30], inter-network, and multi-network mechanism for LTE [31]. Despite the additional standards, some of the issues related with the directional communications and multi-connectivity are still unsolved. In the directional link, the requirement for precise beamforming can affect E2E performance. On the other hand, the provided solution [31] for multi-connectivity improves the mmWave network’s E2E performance by merging a reliable sub-6 GHz link using LTE [32]. In 3GPP, the deployment of NR networks can be in standalone (SA) and non-standalone (NSA) modes. In the SA mode, the NR core and radio access network (RAN) are included, while LTE evolved packet core (EPC) and the LTE RAN are used in the NSA mode. However, the practical implementation of such systems to support beam management with SA and NSA is still an open issue [30].

4. Role of URLLC in Operating IoT

Although mMTC is specifically categorized and designed to meet the IoT requirements, URLLC holds the key ingredients for effective IoT operations. When multiple operators control time-critical devices remotely, as shown in Figure 8, the latency and reliability play a vital role in the smooth operation of the IoT devices. It is quite challenging to operate mission-critical and real-time IoT devices over a wireless connection [33]. A massive multiple-input multiple-output (MIMO) technology has recently become quite applicable to manage a massive number of devices. However, the struggle to meet latency and reliability requirements remains problematic.

Many potential real-time IoT operational issues can be overcome with the integration of the tactile Internet [34], URLLC, and MIMO radio access technologies. In the following subsections, some of the basic URLLC disputes are explained when operating IoT devices.

4.1. URLLC and Massive Device Connectivity

Present mobile services and specifications are not completely equipped to deliver URLLC cost-effectively at scale [35]. Furthermore, they lack the capacity to deliver a reliable low latency communication to multiple users at the same time. It is particularly difficult to ensure link-level reliability and latency over a wide area and in a remote scenario, as shown in Figure 4. As wide-area
cases involve many elements such as transitional nodes, backhaul, core/cloud, and fronthaul, they can play a vital role in degrading latency.

However, the resources such as energy and computing power of IoT devices also play a vital role when operating over URLLC. To meet the latency requirements for URLLC, the IoT devices are forced to utilize excessive power and processing ability that is not appropriate for the life span of IoT devices. However, most of the IoT devices have limited resources [36].

4.2. On-Device Artificial Intelligence and URLLC

Traditionally, communication networks are designed with the concept of achieving high data rates with centralized management of resources. To accomplish the upcoming extreme latency and reliability requirements, the communication network architecture is now being pushed to be more non-centric and proactive. Most of the IoT devices are designed to be remotely controlled or to operate in a limited non-complex environment. However, some of the machines/applications require machine learning (ML) or artificial intelligence (AI) in order to be more effective and efficient to achieve the goals of the applications.

Clearly, the customary machine learning approach based on the centralized architecture, as shown in Figure 9, is not very suitable for delicate latency applications [21]. However, most of the IoT devices have limited resources [36], and such devices may not be able to carry out ML or AI-based algorithms effectively while meeting the latency requirements. Consequently, researchers are investigating decentralized approaches such as distributed ML or AI on edge that involve collective problem solving [37]. Even with on-device machine learning, devices require a significant amount of storage and computational ability, which most of the IoT devices lack.

![Figure 9. Traditional machine learning (ML) concept (centralized). AI, artificial intelligence.](image)

Most of the AI algorithms usually require a large data set to provide effective results. In URLLC, however, it is a challenge to provide such a big data set for the mission-critical IoT devices with reliability and low latency [38].

4.3. URLLC and Vehicle-to-Vehicle (V2V)

One of the most promising and important applications of the future 5G network is V2V communication. V2V communication is one of the technologies that can lead to an intelligent transport system [39]. Naturally, for V2V, road safety (distance awareness to avoid any collision, speed limits, location-based traveling, environment information, road condition) plays a vital role and is extremely time-critical, as shown in Figure 10.

Because of the safety concern, European Telecommunications Standard Institute (ETSI) has standardized safety protocols based on two awareness-based messages: decentralized environmental notification message (DENM) and cooperative awareness message (CAM) [40]. To reflect vehicles based on the mentioned safety standards, V2V communication should have the low latency characteristic of URLLC. As discussed in Section 3.6, handover is still an issue in URLLC implementation owing to the mobility of vehicles (as shown in Table 4) delivering safety standard messages.
with sensor-type equipments with limited resources, for example, limited battery and computing power [41]. A number of URLLC applications require a lot of computation, which is not handled by some IoT devices [42]. From a PHY perspective, it is a challenge for URLLC to achieve low latency and high reliability in mission-critical IoT devices. The use of short packet in order to achieve low latency can degrade channel-coding gain, and it causes reliability issues in wireless channels. To mitigate reliability issue, re-transmission is required, but it involves additional resources and increases latency [43].

4.5. Base Station Densification and Device-to-Device (D2d) Communications

In typical automated industry, clusters of sensors and actuators are working in a fixed area. One of the crucial use cases for the 5G URLLC is to support the wireless industrial automation (e.g., Industry 4.0 [44]). With the emerging industrial automation, M2M and D2D communications require URLLC features to deliver short messages from a controller to a cluster of sensors or machines. A reasonable amount of traffic is expected to be handled by WiFi and small-cell-technology based on mmWave frequencies, as shown in Figure 11. The METIS project estimated that dense metropolitan areas might have up to 200 devices per km², with an expected data volume generated by each device could be 500 Gbyte/month [45]. Such an immense number of devices could force a drastic change in network infrastructure to avoid congestion and availability of service. With the limited frequency bands, improvement of the spectral efficiency (SE) could be an answer to support massive data.

Figure 10. Basic vehicle-to-vehicle (V2V) road safety.

Figure 11. Illustration of a single user data packet and multiple user data packets relayed by a server/cluster head.

The massive MIMO network topology can theatrically support high-density traffic [46]. The theoretical performance and limitations of massive MIMO communications are extensively studied by a number of researchers [46–48].

One of the main issues with massive MIMO is to manage data generated with unpredictable behavior [46]. It is essential to highlight that the building block of next-generation networks will be data packets, which show an unpredictable data generation behavior in non-streaming applications (such as social network applications and web browser).
5. 3GPP Standardization for URLLC

To guarantee the desired reliability, 3GPP and its allies are still working and planning for improvements at multiple aspects of 5G architecture. Some of the recent critical points highlighted by 3GPP Release 16 [49] include the following:

5.1. Handover

Handover, which is one of the most crucial prominent issues in supporting URLLC requirements, was also a part of the discussion in Release 16 of 3GPP. A way to support handover while keeping low latency and jitter remains a significant concern. In 3GPP TS 23.502 [50], the handover process requires a lossless handover. The source RAN node forwards data directly or indirectly to the target RAN node. The use of tunnel is the current approach, but it introduces additional jitter and latency. With this current approach, the issue remains open, and further studies were suggested by 3GPP on the matter.

5.2. User Mobility

The second issue on improving session stability while keeping UE application uninterrupted. The main focus of the discussion was to enhance runtime synchronization between UE and 5G systems to support application transfer without breaking application sessions and service continuity. In Release 15 of 3GPP, the mobility-related issue was partially solved by introducing a “notification” mechanism, but the purposed solution was not suitable for non-human devices. The proposed solution requires the UE to alter its behavior (e.g., reduction of speed in case of a vehicle) so that the RAN can restore the required QoS level with that cell. The purposed solution also involves UE to exchange numerous signal messages regardless of the potential link quality or congestion-related issues with the RAN to achieve the desired guaranteed bit rate (GBR). Then, it clearly degrades the E2E performance.

5.3. QoS Monitoring to Support URLLC

As mentioned in Table 2, URLLC services require high reliability and very low latency. Such requirements pose quite a challenge for 5G systems, because such requirements could affect E2E QoS performance. In the current 5G system, QoS notification control (NC) is maintained by a 5G access network (AN) to monitor GBR. Though the 5G-AN mechanism supports the guaranteed flow bit rate (GFBR) [49], which might not be sufficient for URLLC E2E services, features such as packet loss, jitter, and packet latency will also play a major part for URLLC services. To counter such QoS relates issues, 3GPP has branched out further aspects of studies, which include specification of the UE requirements for URLLC and improvement of QoS monitoring ability of the defined mechanisms.

In Release 16 [49], a number of solutions to handle the mentioned issues were purposed and evaluated. The idea that server supporting user equipment should be kept close to the user (i.e., topologically, geographically) was proposed. With this idea, the transmission latency between the server and the base station could be minimized. In order to achieve high consistency, backhaul reliability was suggested for further improvement.

5.4. Possible 5G Integration Plan by 3GPP

Previous generations of cellular networks required access and core networks, which belong to the same generation, to be installed. It means that 4G systems were composed of LTE and evolved packet core (EPC). The deployment of a 5G system comes with the flexibility of integrating elements of previous cellular generations in different configurations [51]:

1. SA using only one radio access technology
2. N-SA is combining multiple radio access technologies.
5.4.1. Standalone (SA)

In SA, the evolved LTE radio or the 5G NR cells and the core network are operated alone, so that the NR or evolved LTE radio cells are used for both user and control planes. The SA provides a simple solution for operators to provide services to both 4G and 5G customers using normal inter-generation handover as shown in Table 5. The three variations of SA, as defined by 3GPP, are as follows:

1. EPC and LTE Evolved Node B (eNB) access (i.e., based on current 4G LTE networks)
2. 5G core (5GC) and NR 5G Node B (gNB) access.
3. 5GC and LTE ng-eNB access

Table 5. Contrast of core network and 5G RAN [51]. SA, standalone; NSA, non-standalone; EPC, evolved packet core; 5GC, 5G core; LTE, long-term evolution; NR, new radio.

| Radio Access Network | Core Network |
|----------------------|-------------|
|                       | SA         | NSA | EPC | 5GC |
| **Advantages**        | Simple management | Supports existing LTE deployment | Supports current EPC deployment | Cloud-native multiple access is easy to support |
|                      | Support handover between 4G and 5G | Tight interworking of LTE and NR is necessary | Optional Cloud support | The new deployment is essential |
|                      | Will not be able to support existing LTE deployment if NR is used in SA | End-user experience may be degraded | |
| **Disadvantages**     | |

5.4.2. Non-Standalone (NSA)

In NSA, the LTE radio cells and NR radio cells are combined using dual-connectivity to provide radio access. The core network could be either 5GC or EPC based on the operator’s choice [52]. On the basis of the operators, they can provide 5GC for 5G customers or facilitate the existing 4G deployment combining NR radio and LTE resources with current EPC. NSA needs tight integration with the LTE RAN. Three variations of NSA, as defined by 3GPP, are as follows:

1. LTE eNB and EPC as master and NR en-gNB as secondary.
2. NR gNB and 5GC as master and LTE ng-eNB acting as secondary.
3. LTE ng-eNB and 5GC as master and NR gNB as secondary.

Options defined by 3GPP for 5G deployment use either existing EPC [53] or the 5GC [54]. Both architectures follow different design principles. EPC can be considered as an evolution of earlier generation packet-based core networks. 5GC is designed based on “cloud-native” approach, with virtualization and cloud computing as its core. 5GC provides improved QoS features and superior network slicing.

6. Future Research Areas

In this paper, many key issues related to the URLLC implementation are discussed. Once the standards for 5G NR are matured, the researchers can work towards improving the core network, backhaul, and transport delays [39]. Issues related to UE energy efficiency (discussed in Sections 3.5 and 4.1) can be further discussed in the light of the new NR defined energy-efficient state called INACTIVE [55]. This state is defined as a state existing between the CONNECTED and IDLE states of a UE. The performance evaluation of the INACTIVE state remains open for further examination.

One of the concerns related to the coexistence of URLLC with eMBB (discussed in Section 3.2) is not only limited to the existence of the mentioned categories of 5G. With the backward compatibility of 5G NR, the coexistence of URLLC with distributed system architecture and diverse application requirements will need further study to avoid latency concerns [55]. Likewise, handover related
issues (discussed in Sections 3.6 and 4.3) can be further discussed in terms of the reestablishment of connection in the case of the radio link failure, keeping in view that LTE and NR traffic will coexist in the same carrier.

Many areas need further investigation to operate IoT devices over URLLC. Although URLLC can provide low latency with high reliability, the characteristics of massive connectivity can degrade such requirements, as discussed in Section 4.1. As discussed in Section 4.2, on-device AI and ML will play a major role in an upcoming era, and hence researchers can further study the designing of AI/ML algorithms that do not require high resources and are better suited for devices with limited resources.

A comprehensive summary of the issues discussed in the paper is presented in Table 6.

### Table 6. Summary of issues discussed in the paper. AI, artificial intelligence; ML, machine learning.

| Issue                        | Reference       | Section Summary                                                                                                                                                                                                 |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| QoS                          | [11,13,15,16]   | In this Section 3.1, QoS requirements of URLLC (latency and reliability) and factors, which are a hindrance in achieving the desired QoS for URLLC, are discussed.                                                       |
| Coexistence with eMBB        | [17–19]         | In the 5G networks, many different applications with diverse requirements will exist in the same physical medium. Such a coexistence of services will raise many challenges for telecom companies. In Section 3.2, the problems with the coexistence of eMBB and URLLC with different service requirements are discussed. |
| URLLC Packet Design          | [15,20]         | Packet design plays a vital role in achieving low latency. Minimizing the packet processing time will be a key factor in enabling low latency for URLLC. Packet structure proposed by LTE and NR to achieve low latency is discussed in Section 3.3. |
| URLLC Scheduling             | [20]            | Because of the unpredictable packet generation of URLLC, scheduling is a challenging task. In Section 3.4, some of the proposed scheduling schemes for URLLC and issues with those schemes are discussed.     |
| Energy issues for UE         | [21,56]         | To keep up with the latency requirement of URLLC, UEs are forced to perform extra tasks, which can result in low battery life for the UEs. Such power consumption related issues are discussed in Section 3.5.  |
| Handover issues for URLLC    | [22–25]         | Providing uninterrupted services to a mobile user is the most significant facility of any telecom infrastructure. Providing such an uninterrupted service to a user using URLLC based services is quite difficult. Issues related to handover when it comes to strict latency are discussed in Section 3.6. |
| Error Handling               | [17,26,27]      | Wireless services are prone to many challenges, and providing highly reliable service in wireless communication is quite a tough task. The issues related to the handling of error packets and retransmission are covered in Section 3.7. |
| Role of URLLC in operating IoT | [33,34]       | IoT will play a major role in the coming era of technology. URLLC will play a vital role in supporting IoT services. In Section 4, the importance of URLLC to operate IoT is discussed.             |
| URLLC and Massive device connectivity | [35,36] | Although URLLC fulfills the basic requirement of reliability and latency for mission-critical IoT, it is a challenge for URLLC to provide simultaneous services to a vast number of devices. Section 4.1 covers the issues that URLLC brings in operating massive IoT devices. In earlier sections importance of URLLC for time-critical applications is highlighted. However, the provision of low latency service to massive devices is also a challenge, as cited in Section 4.1. It is provoking researchers to seek new solutions to achieve low latency with high reliability. Among such solutions developing intelligent machines is quite prominent. In Section 4.2, issues related to AI/ML-based machines and relying on URLLC services for such machines are discussed. |
| On-device AI and URLLC       | [21,36–38]      | An automated vehicle is one of the most anticipated services of the upcoming era. However, providing highly reliable and time-critical connectivity is still a challenge for URLLC. V2V connectivity opens a whole new level of disputes. Among them, some issues are discussed in Section 4.3. |
| URLLC and V2V                | [39,40]         |                                                                                                                                                                                                                  |
6.1. Possible Solutions for Reliability and Latency Requirements

Edge communications will play an essential role in future networks; researchers have provided a number of solutions to overcome the strict QoS requirements in URLLC. Current cellular networks follow a centralized approach, while edge communications bring resources close to the UE. Despite some issues, the edge communications based solutions that is, mobile-edge computing (MEC) [57,58], local area communication [59,60], and wide-area large-scale communications [61,62], are very promising, as shown in Table 7.

| Communication Type          | Current Issue                               | Possible Solution                                      |
|-----------------------------|---------------------------------------------|--------------------------------------------------------|
| Local-Area Communication    | Shadowing, channel estimation overhead      | Multi-connectivity, 5G NR, grand-free access            |
| Mobile Edge Computing       | E2E delay and reliability, optimizing communication | Optimizing scheduling methods in computing system and communication |
| Wide-Area Communication     | Reliable and precise communication between slave and master controller | Forecast mobility and communication methods to be co-design to improve QoS |

The mentioned communication methods in Figure 12 can probably be the key point in improving overall efficiency for URLLC. A number of researchers have further proposed methods to improve the communication methods in edge computing.

![Figure 12. Three communication solutions for URLLC quality of service (QoS) requirements.](image)

In the papers explained in Table 8 and numerous other papers on edge computing, it is highlighted that the next generation networks should have a cloud-native approach in order to achieve high reliability and low latency. Although some challenges still exist in implementing the edge-computing systems (ECS) based approach for URLLC, ECS can further be improved and can provide better QoS for 5G URLLC networks.

| Reference | Proposed Solutions Using MEC to Support URLLC |
|-----------|-----------------------------------------------|
| [64]      | Minimizing E2E communication delay            |
| [65]      | Highlighting the MEC role to support URLLC in mission-critical applications with further optimization parameters for significant use cases |
| [66]      | Minimizing E2E communication delay            |
| [42]      | Proposing an algorithm for energy efficiency (EE) in mobile devices by optimizing queue complexity of the communication process |
| [67]      | Reducing computation and latency for IoT devices using MEC |

AI and 5G Networks Traffic Management

AI can be integrated with 5G networks to improve the efficiency of resource and network management. The network architecture and user requirements for 5G networks, the traffic management
AI including ML and deep learning can assist 5G networks in predicting and managing the unpredictable network traffic. AI can analyze and cope with the unpredictable requirements of 5G networks traffic [69]. Additionally, ML can play an integral part to support the MEC architecture [70]. However, the current AI-based research has some limitations [71]. Most of the existing studies focuses on utilizing AI/ML to solve core network and routing related issues. Only a few researchers have worked on applying AI to the semantic and application layer to propose traffic management solutions.

6.2. 5G and Beyond

With the expected deployment of global 5G networks in the 2020s, it is time to raise the imperative question on the future of mobile network, that is, beyond 5G or 6G. One of the prominent areas of beyond 5G will be software-defined network (SDN) and network function virtualization (NFV) [71]. The SDN and NFV will play a vital role in enabling management and control systems for E2E structures. Software-based transformation (i.e., softwarization) [72] cannot be implemented within the defined 5G implementation time frame. Other areas of beyond 5G include security [73,74], spectral and energy efficiency, resiliency (i.e., tolerance to interference and maintaining QoS), and MEC [58]. MEC and softwarization are among the promising candidates to provide a cost-effective, secure, manageable, and flexible architecture for 5G and beyond.

7. Conclusions

The main contribution of the paper is to provide researchers a fast and brief reference to some of the core issues in the implementation of URLLC. Keeping in view the importance of IoT in the coming era, this paper also covers a few most critical aspects of IoT and V2V communication over URLLC. On the basis of issues being covered in this paper, some of the areas that are still open for further investigation in URLLC improvements are also provided to readers. At the end of the article, a possible solution using edge computing is proposed for URLLC implementation. This paper can provide a comprehensive platform for researchers who are looking to study URLLC and its issues with diverse services and applications.

Author Contributions: M.A.S. and H.Y. have written this paper and have done the research which supports it. J.J. has reviewed the work.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) (2019R1A2C1083988) and by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2016-0-00313) supervised by the IITP (Institute for Information & communications Technology Promotion) and in part by the NRF grant funded by the Korea government (MSIT) (2018R1A4A1023826).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moakes, P. Embedded Signal Processing and RF Modules. Available online: https://www.commagility.com/images/pdfs/white_papers/CommAgility_5G_New_Radio_white_paper.pdf (accessed on 20 August 2019).
2. Ghosh, A. 5G New Radio (NR): Physical Layer Overview and Performance. IEEE Communication Theory Workshop. Available online: http://ctw2018.ieee-ctw.org/files/2018/05/5G-NR-CTW-final.pdf (accessed on 20 August 2019).
3. Popovski, P.; Trillingsgaard, K.F.; Simeone, O.; Durisi, G. 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View. IEEE Access 2018, 6, 55765–55779. [CrossRef]
4. Mallinson, K. 3GPP-The Path to 5G: As much Evolution as Revolution. The Mobile Broadband Standard. Available online: http://www.3gpp.org/news-events/3gpp-news/1774-5g_wiseharbour (accessed on 20 August 2019).
5. Iwabuchi, M.; Benjebbour, A.; Kishiyama, Y.; Ren, G.; Tang, C.; Tian, T.; Gu, L.; Takada, T.; Kashima, T. 5G Field Experimental Trials on URLLC Using New Frame Structure. In Proceedings of the 2017 IEEE Globecom Workshops (GC work shop), Singapore, 4–8 December 2017.

6. Lema, M.A.; Layla, A.; Mahmoodi, T.; Cuevas, M.; Sachs, J.; Markendahl, J.; Dohler, M. Business Case and Technology Analysis for 5G Low Latency Applications. *IEEE Access* **2017**, *5*, 5917–5935. [CrossRef]

7. Ericsson. NR URLLC Rel-16 Use Cases and Requirements. Available online: https://list.etsi.org (accessed on 20 August 2019).

8. Amazon-Prime-Air. Amazon Inc Technology. Available online: https://www.amazon.com/Amazon-Prime-Air/b?ie=UTF8&node=8037720011 (accessed on 20 August 2019).

9. Google. Waymo LLC Technology. Available online: https://waymo.com (accessed on 20 August 2019).

10. Karaa. Setpoint USA Custom Automation. Available online: https://www.setpointusa.com/blog/industrial-automation-examples/ (accessed on 20 August 2019).

11. Li, C.P.; Jiang, J.; Chen, W.; Ji, T.; Smee, J. 5G Ultra-Reliable and Low-Latency Systems Design. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017.

12. Holfeld, B.; Wieruch, D.; Wirth, T.; Thiele, L.; Ashraf, S.A.; Huschke, J.; Aktas, I.; Ansari, J. Wireless Communication for Factory Automation: an opportunity for LTE and 5G systems. *IEEE Commun. Mag.* **2016**, *54*, 36–43. [CrossRef]

13. Maaz, D.; Galindo-Serrano, A.; Elayoubi, S.E. URLLC User Plane Latency Performance in New Radio. In Proceedings of the 2018 25th International Conference on Telecommunications (ICT), St. Malo, France, 2–28 June 2018.

14. Pocovi, G.; Soret, B.; Pedersen, K.I.; Mogensen, P. MAC layer enhancements for ultra-reliable low-latency communications in cellular networks. In Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, France, 21–25 May 2017.

15. 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. [CrossRef]

16. Sachs, J.; Wikström, G.; Dudda, T.; Baldemair, R.; Kittichokechai, K. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. *IEEE Netw.* **2018**, *32*, 24–31. [CrossRef]

17. Esswie, A.A.; Pedersen, K.I. Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks. *IEEE Access* **2018**, *6*, 38451–38463. [CrossRef]

18. Pedersen, K.; Pocovi, G.; Steinier, J.; Maeder, A. Agile 5G Scheduler for Improved E2E Performance and Flexibility for Different Network Implementations. *IEEE Commun. Mag.* **2018**, *56*, 210–217. [CrossRef]

19. Liao, Q.; Baracca, P.; Lopez-Perez, D.; Giordano, L.G. Resource Scheduling for Mixed Traffic Types with Scalable TTI in Dynamic TDD Systems. In Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016.

20. Pocovi, G.; Pedersen, K.I.; Mogensen, P. Joint Link Adaptation and Scheduling for 5G Ultra-Reliable Low-Latency Communications. *IEEE Access* **2018**, *6*, 28912–28922. [CrossRef]

21. Meredith, J.M. Technical Specification Group Radio Access Network, NR (Release 15); 3GPP: Sophia Antipolis, France, 2017.

22. Bennis, M.; Debbah, M.; Poor, H.V. UltraReliable and Low-Latency Wireless Communication: Tail, Risk and Scale. *Proc. IEEE* **2018**, *106*, 1834–1853. [CrossRef]

23. Norp, T. 3GPP (5G Service Requirements). Available online: http://www.3gpp.org/news-events/3gpp-news/1831-sa1_5g (accessed on 20 August 2019).

24. John, J.K.; Meredith, M. 3GPP (NR; Overall Description; Stage-2). Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3191 (accessed on 20 August 2019).

25. Park, H.S.; Lee, Y.; Kim, T.J.; Kim, B.C.; Lee, J.Y. Handover Mechanism in NR for Ultra-Reliable Low-Latency Communications. *IEEE Netw.* **2018**, *32*, 41–47. [CrossRef]

26. Korhonen, J. 3GPP (E-UTRA and E-UTRAN; Overall description; Stage 2). Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430 (accessed on 20 August 2019).

27. Shariatmadari, H.; Li, Z.; Iraji, S.; Uusitalo, M.A.; Jäntti, R. Control channel enhancements for ultra-reliable low-latency communications. In Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, France, 21–25 May 2017.
28. Giordani, M.; Polese, M.; Roy, A.; Castor, D.; Zorzi, M. Standalone and Non-Standalone Beam Management for 3GPP NR at mmWaves. *IEEE Commun. Mag.* 2019, 57, 123–129. [CrossRef]
29. Rangan, S.; Rappaport, T.S.; Erkip, E. Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges. *Proc. IEEE* 2014, 102, 366–385. [CrossRef]
30. 3GPP. NR—Physical Channels and Modulation—Release 15. TS 38.211, V15.4.0; 3GPP: Sophia Antipolis, France, 2019.
31. 3GPP. NR—Multi-Connectivity—Overall Description (Stage 2) TS 37.340, V15.4.0; 3GPP: Sophia Antipolis, France, 2019.
32. Polese, M.; Giordani, M.; Mezzavilla, M.; Rangan, S.; Zorzi, M. Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks. *IEEE JSAC* 2017, 35, 2069–2084. [CrossRef]
33. Pocovi, G.; Shariatmadari, H.; Berardinelli, G.; Pedersen, K.; Steiner, J.; Achieving, Z.L. Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements. *IEEE Netw.* 2018, 32, 8–15. [CrossRef]
34. Tarneberg, W.; Karaca, M.; Robertsson, A.; Tufvesson, F.; Kihl, M. Utilizing Massive MIMO for the Tactile Internet: Advantages and Trade-Offs. In Proceedings of the 2017 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops), San Diego, CA, USA, 12 June 2017.
35. ITU. T. Watch. The Tactile Internet. Available online: https://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000320001PDFE.pdf (accessed on 20 August 2019).
36. Hoymann, C.; Astely, D.; Stattin, M.; Wikström, A.; Frenne, M.; Blasco, R.; Huschke, J.; Gunnarsson, F. LTE release 14 outlook. *IEEE Commun. Mag.* 2016, 54, 44–49. [CrossRef]
37. Li, R.; Zhao, Z.; Zhou, X.; Ding, G.; Chen, Y.; Wang, Z.; Zhang, H. Intelligent 5G: When Cellular Networks Meet Artificial Intelligence. *IEEE Wirel. Commun.* 2017, 24, 175–183. [CrossRef]
38. Shah, S.A.A.; Ahmed, E.; Imran, M.; Zeadally, S. 5G for Vehicular Communications. *IEEE Commun. Mag.* 2018, 56, 111–117. [CrossRef]
39. Gordon, J.S.; Zeng, J.; Liu, R.P.; Ni, W.; Nguyen, N.D.; Jayawickrama, B.A.; Huang, X.J.; Abolhasan, M.; Zhang, Z.; Dutkiewicz, E.; 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.
40. Dong, R.; She, C.Y.; Hardjawana, W.; Li, Y.H.; Vucetic, B. Improving Energy Efficiency of Ultra-Reliable Low-Latency and Delay Tolerant Services in Mobile Edge Computing Systems. In Proceedings of the 2019 IEEE International Conference on Networking Workshops (ICCW Workshops), Shanghai, China, 20–24 May 2019.
51. Road to 5G: Introduction and Migration to 5g: Introduction and Migration by GSMA. Available online: https://www.gsma.com/futurenetworks/wp-content/uploads/2018/04/Road-to-5G-Introduction-and-Migration_FINAL.pdf (accessed on 11 August 2019).

52. 3GPP. TS 23.214 Architecture Enhancements for Control and User Plane Separation of EPC Nodes; 3GPP: Sophia Antipolis, France, 2017.

53. 3GPP. TS 23.401 General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access; 3GPP: Sophia Antipolis, France, 2011.

54. 3GPP. TS 23.501 System Architecture for the 5G System; Stage 2; 3GPP: Sophia Antipolis, France, 2019.

55. Lyamin, N.; Vinel, A.; Jonsson, M.; Bellalta, B. Cooperative Awareness in VANETs: On ETSI EN 302 637-2 Performance. IEEE Trans. Veh. Technol. 2018, 67, 17–28. [CrossRef]

56. Mukherjee, A. Energy Efficiency and Delay in 5G Ultra-Reliable Low-Latency Communications System Architectures. IEEE Netw. 2018, 32, 55–61. [CrossRef]

57. Liu, J.H.; Zhang, Q. Offloading Schemes in Mobile Edge Computing for Ultra-Reliable Low Latency Communications. IEEE Access 2018, 6, 12825–12837. [CrossRef]

58. SK Telecom Unveils 5G Mobile Edge Computing Open Platform. Available online: https://www.mobileeurope.co.uk/press-wire/sk-telecom-unveils-5g-mobile-edge-computing-open-platform (accessed on 12 August 2019).

59. Vivek, K.; Singh, P.; Samundiswary, M. Sivasindhu. Cluster based Reliable Communication for 5G Network. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 4–6 April 2019.

60. Franco, D.; Aguado, M.; Toledo, N. An Adaptable Train-to-Ground Communication Architecture Based on the 5G Technological Enabler SDN. MDPI Electron. 2019, 8, 660. [CrossRef]

61. Antonakoglou, K.; Xu, X.; Steinbach, E.; Mahmoodi, T.; Dohler, M. Towards haptic communications over the 5G tactile internet. IEEE Commun. Surv. Tutor. 2018, 20, 3034–3059. [CrossRef]

62. Sofana, R.S.; Tomislav, D.; Pierluigi, S.; Prabaharan, S.R.S. Future Generation 5G Wireless Networks for Smart Grid: A Comprehensive Review. MDPI Energ. 2019, 12, 2140.

63. Feng, D.Q.; She, C.Y.; Ying, K.; Lai, L.F.; Hou, Z.W.; Quek, T.Q.S.; Li, Y.H.; Vucetic, B. Toward Ultra-reliable Low-Latency Communications: Typical Scenarios, Possible Solutions and Open Issues. IEEE Veh. Technol. Mag. 2019, 14, 94–102. [CrossRef]

64. Hu, Y.L.; Schmeink, A. Delay-Constrained Communication in Edge Computing Networks. In Proceedings of the 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 25–28 June 2018.

65. Elbamby, M.S.; Perfecto, C.; Liu, C.F.; Park, J.; Samarakoon, S.D.; Chen, X.F.; Bennis, M. Wireless Edge Computing with Latency and Reliability Guarantees. Proc. IEEE 2019, 107, 1717–1737. [CrossRef]

66. Duan, Y.F.; She, C.Y.; Zhao, G.D.; Quek, T.Q.S. Delay Analysis and Computing Offloading of URLLC in Mobile Edge Computing Systems. In Proceedings of the 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), Hangzhou, China, 18–20 October 2018.

67. Elbamby, M.S.; Bennis, M.; Saad, W.; Latva-aho, M.; Hong, C.S. Proactive edge computing in fog networks with latency and reliability guarantees. J. Wirel. Commun. Netw. 2018, 209. [CrossRef]

68. Eugenio, M.; Cayamcela, M.; Lim, W. Artificial Intelligence in 5G Technology: A Survey. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 19–17 October 2018.

69. Lin, I.C.; Sun, Q.; Liu, Z.M.; Zhang, S.M.; Han, S.F. The Big-Data-Driven Intelligent Wireless Network: Architecture, Use Cases, Solutions, and Future Trends. IEEE Veh. Technol. Mag. 2017, 12, 20–29.

70. Jiang, C.X.; Zhang, H.J.; Ren, Y.; Han, Z.; Chen, K.C.; Hanzo, L. Machine Learning Paradigms for Next-Generation Wireless Networks. IEEE Wirel. Commun. 2016, 24, 98–105. [CrossRef]

71. Fu, Y.; Wang, S.; Wang, C.X.; Hong, X.M. Stephen McLaughlin. Artificial Intelligence to Manage Network Traffic of 5G Wireless Networks. IEEE Netw. 2018, 32, 58–64. [CrossRef]

72. IEEE. IEEE 5g and Beyond Technology Roadmap White Paper. Available online: https://futurenetworks.ieee.org/images/files/pdf/ieee-5g-roadmap-white-paper.pdf (accessed on 27 August 2019).
73. Condoluci, M.; Mahmoodi, T. Softwarization and virtualization in 5G mobile networks: Benefits, trends and challenges. Elsevier Comput. Netw. 2018, 146, 65–84. [CrossRef]

74. Ahmad, I.; Shahabuddin, S.; Kumar, T.; Okwuibe, J.; Gurtov, A.; Ylianttila, M. Security for 5G and Beyond. IEEE Commun. Surv. Tutor. 2019. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).