Introduction to Ultra-Reliable Low-Latency Communication (URLLC)

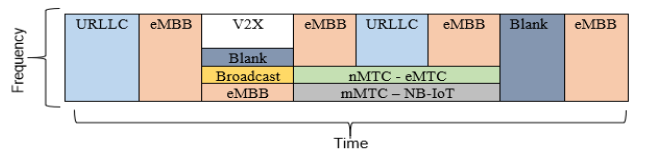
Survey Paper by Arju Reza

COMP 4320: Intro to Computer Networks

**I. Introduction**

**A. 5G and new radio (NR)**

In our world, technology is constantly and rapidly evolving which changes how we live our daily lives. A heavily anticipated grouping of technology is the fifth generation (5G). Within 5G, there are many technologies that are needed to support the upcoming development including new radio (NR). NR is a new air interface presented by the Third Generation Partnership Program that is going to bring technologies that will be better than the current technologies in the fourth generation (4G) long term evolution (LTE) [3], and NR is the new access method and radio interface that is going to help meet requirement for 5G [10]. To support NR, the International Telecommunication Union has introduced a flexible frame structure (Figure 1) using three key categories that will help with communication between the base station and device users [10]. The three categories presented are enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable low-latency communication (URLLC).



**Figure 1.** Flexible frame structure for NR [10] URLLC = ultra-reliable low-latency communications, eMBB = enhanced mobile broadband, V2X = vehicle-to-everything, mMTC = massive machine-type communications, NB = narrowband, IoT = Internet of Things

**1. Enhanced mobile broadband (eMBB)**

As explained in [3], eMBB is a category that entails services that require high bandwidth, like high-resolution video streaming, virtual reality, and augmented reality. In 4G systems, physical layer technologies were developed that increased throughput which included multiple-input multiple-output (MIMO) transmission, carrier aggregation, and cell densification. To fulfill 5G requirements, the current 4G system technologies need to be heavily improved upon. To improve current technology, improving spectral efficiency, which is the rate at which information can be transmitted over a bandwidth, and using the unexplored spectrum are needed. Technologies that can possibly achieve this include full-dimension and massive MIMO, millimeter-wave communication, and spectrally localized waveforms.

**2. Massive machine-type communications (mMTC)**

Also explained in [3], mMTC is a category of services that support the access of many machine-type devices. These services include ones that need energy-efficient, connectively-dense implementations like sensing and monitoring. Recently, there have been attempts to improve upon the requirements for mMTC, but approaches became lacking when presented with more devices than current resources could handle. A couple of ways thought to make mMTC 5G achievable are using a non-orthogonal spreading sequence or using user-specific interleaving.

**3. Ultra-reliable low-latency communication (URLLC) [3]**

Explained in [3], URLLC is a service category that is aimed at implementing two seeming conflicting issues: ultra-reliability and low latency. Services include remote control, autonomous driving, and tactile internet. The aim for 5G URLLC is to make packet transmission time faster than human perceptron time which would require packet transmission time to be around tens to hundred of microseconds. For context, 4G technologies operate are around 30 to 100 milliseconds (ms) for end-to-end latency. To reduce the latency of end-to-end communications, changes in wireless linking and backbone networking are needed. For backbone networking, solutions like software defined networking have been proposed which creates a fast, private connection for the URLLC service. For wireless linking, solutions that reduce overhead and create a streamlined transmission mechanism are needed. As seen from these needed improvements, many parts of the physical layer need to be improved upon, or even redesigned.

**B. Requirements for URLLC**

In [1], three main requirements for URLLC are outlined: low latency, ultra-high reliability, and co-existence.



**Figure 2.** Equation for physical layer latency [1]

**1. Low Latency**

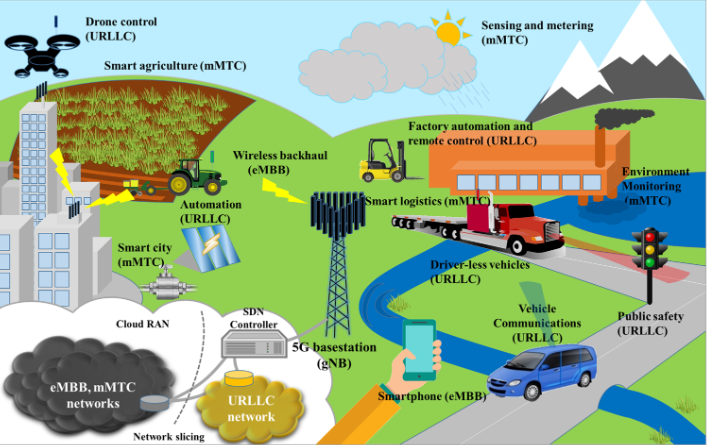
In [1], physical-layer latency (TL) is defined as the sum of time-to-transmit latency (Tttt), propagation (travel time for signal from sender to receiver) latency (Tprop), the time to process for a channel estimation (Tproc), and the time for retransmission (TreTx) as seen in Figure 2. In accordance with the International Telecommunication Union, The Third Generation Partnership Program has determined that the physical-layer latency for URLLC should not be more than 0.5 ms.

**2. Ultra-High Reliability**

As explained in [1], one way to measure ultra high reliability is through block error rate which is 10-2­ in 4G. This rate is achieved through retransmission mechanisms like hybrid automatic repeat request and channel coding. To reach the 5G requirement for URLLC, the block error rate within a 1 ms period should be at least 10-5.

**C. Conflicting Issues in URLLC**

Explained in [1], probably the most challenging requirement for URLLC is achieving low latency and ultra-high reliability at the same time. To achieve low latency, packet size needs to be short, which hurts the quality of channel coding gain which creates more error, therefore decreasing reliability. On the other hand, increasing reliability requires more resource usage which increases the latency because more information is being sent. This paper addresses the balance needed to achieve both ultra-high reliability and low latency.



**Figure 3.** Depiction of URLLC and other services [3]

**D. Applications of URLLC**

As seen in the table in [10] and in Figure 3, many industries can make great use of URLLC technology like the medical field, entertainment, industrial automation, and transportation. In the medical field, new possibilities like remote surgery through a robot can be made possible with immediate responses and reliability. In entertainment, online gaming and devices that utilize virtual reality and augmented reality can use URLLC. For industrial automation, as robot presence in manufacturing increases, they need more reliability and quicker response time through low latency. Finally, in transportation, technologies like drone-based delivery and remote/self-driving cars need URLLC to ensure safety and quality of service. To better understand an URLLC, we will be overviewing problems and solutions in connected vehicles as presented in [4].

**II. URLLC in Connected Vehicles**

**A. Introduction**

This section will be going in depth through the problems and solution proposed in [4]. With the emergence and preparation toward 5G, the idea of connected vehicles is becoming more and more feasible. If implemented correctly, connected vehicles can rid of driver-included errors and improve traffic and travel efficiency. Autonomous vehicles, being the ultimate form of connected vehicles, need very strong communication networks to exchange information instantaneously being the ultimate form of connected vehicles. In response to this, the solution of vehicular networks (VNETs) was introduced to ensure low latency and high reliability due to their essentiality. With current 4G LTE technology, it is infeasible for URLLC VNETs to be realized, so 5G technologies are being focused on. This section focuses on addressing the challenges of implementing URLLC VNETs for connected vehicles and offers solutions to those challenges.

**B. Vehicle-to-everything (V2X) communications with scenarios**

As explained in [4] and seen in Figure 4, vehicle-to-everything (V2X) communications are split into four categories: vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-network (V2N), and vehicle-to-pedestrian (V2P).

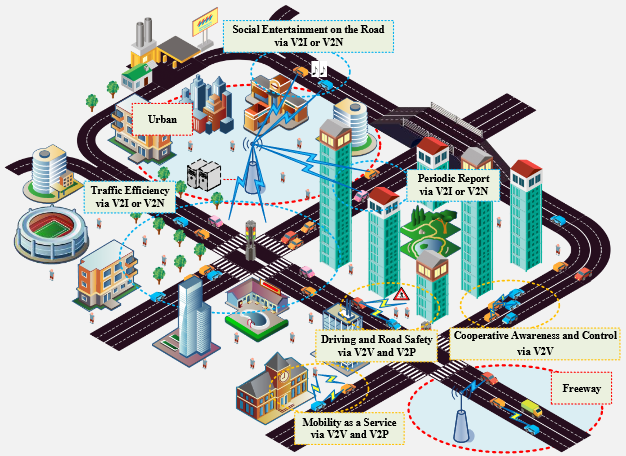
**1. V2V and V2P communications**

Due to their similarity, V2V and V2P communications will be discussed together. In [4], the purpose of V2V and V2P communications is the improve driving safety and experience. Using URLLC VNETs, safety on the road can be heavily reduced. Instead of relying on vision, URLLC VNETs will rely on vehicles and pedestrians using communication to know where each other is located. This heavily depends on low latency for nearly instant communication and ultra-reliability for safety due to the risk of these communications failing. A key component of V2V and V2P communications is the idea of mobility as a service. Mobility as a service offers a user the most efficient means of transporting and communicating. With established URLLC VNETs, pedestrians can send out their requirements for transportation and vehicles can meet those

requirements quickly. Another application is vehicles acting as an intermediate relay for vehicles in areas with poor connectivity.

**2. V2I and V2N communications**

In [4], V2I and V2N communications are grouped together due to both focusing on large-scale cooperation. One area that URLLC VNETs can improve for V2I and V2N communications is traffic efficiency. V2I and V2N communications deal with the macroscopic aspects of transportation which can allow for controlling of traffic and even pollution in areas with a large amount of vehicles. This can also be applied to emergency situations where the macroscopic view of traffic is very essential knowledge. Another aspect of V2I and V2N communications is the distribution of entertainment on the road. With URLLC VNETs, the driving experience can be made not just safer and faster but more enjoyable.



**Figure 4.** Scenarios in V2X communications [4]

**C. Challenges for VNETs**

**1. Frameworks for Different Scenarios**

The current cellular networks are designed to be energy efficient and spectrally efficient (able to transmit a high information rate at a given bandwidth) which does not prioritize URLLC. As explained before, implementing both ultra-reliability and low latency is difficult. [4]

**2. Mobility Models**

Implementing the microscopic and macroscopic aspects of VNETs is difficult since each rely upon different types of latencies being low. The macroscopic aspect depends on the PHY layer transmission latency being low while the microscopic aspect depends on the MAC layer scheduling latency being low. Created a URLLC VNET system that accounts for both is a challenge. [4]

**3. Deployment**

In [4], the deployment of URLLC VNETs is a challenge. In the case of an urban area, regulating vehicles with stoplights causes difficulty in achieve highly efficient travel. Also, organizing network topology in urban area is a big challenge that the implementation of URLLC VNETs face.

**D. Solutions for VNETs**

**1. PHY Layer**

A solution presented in [4] to improve the PHY layer is to use short-packet transmissions. Short-packet transmission allows for transmission time to remain low to achieve low latency while keeping a high reliability, addressing the problem of implementing both reliability and latency. Another solution is to use a multiple inputs multiple outputs (MIMO) framework to make the allocation of resources for multiple users easier. This helps to reduce the transmission latency due to better organization.

**2. MAC Layer**

In [4], MAC layer solutions are presented. One solution is making it so that cellular users and vehicular users share available uplink radio resources with each other. This allows for cellular users to rid of unnecessary which makes transmission more reliable, and it allows for vehicular users to meet their URLLC requirement.

**E. Future for VNETs**

**1. Scheduling**

One problem with current scheduling is the problem of request-grant procedure between access points. To fix this, research into grant-free access should be investigated, which eliminates the overhead gained through requesting and granting access [4].

**2. Slicing**

In [4] another solution that could prove useful for URLLC VNETs in the future is network slicing which splits physical networks into multiple virtual networks. This would allow for different networks to account for difference priorities present in URLLC VNETs.

**III. Third Generation Partnership Program (3GPP) URLLC Standardization**

As presented in [8], the Third Generation Partnership Program (3GPP) identifies the low latency requirement as less than 1 ms for transmission of small payloads and the high reliability requirement as 99.999% probability of no error. As outlined in [2] and in Section I.B, the 3GPP explained the three service categories for NR: eMBB, mMTC, and URLLC. One of the main issues addressed by the 3GPP in Release 16 was the URLLC physical design model by improving the performance with changes like reducing the number of bits in the downlink control information format to lower the aggregation level which reduces resource shortage. More critical points were addressed by the 3GPP as seen in [10].

**A. Handover and User Mobility**

In [10], handover is identified as a very important issue in support URLLC by the 3GPP. Handover entails transferring transmission from one base station to another without losing information, increasing latency, or introducing jittering. Approaches like using a tunnel were introduced but the performance is not up to par, so the issue is still being researched. User mobility entails maintaining reliability and stability while user equipment utilizing URLLC is on the move. Currently solutions, which include changing behavior of the user equipment to allow the radio area network to restore the quality of service of the cell, create more latency, so more experimenting is needed in the area.

**B. Quality of Service Monitoring**

[10] also explains how the 3GPP is concerned with maintaining the quality of service required while reaching high reliability and low latency. The 3GPP introduced the idea that the supporting server for user equipment and the user should be near to keep transmission latency low. Along with this, 3GPP also saw backhaul reliability, which introduces intermediate connections, to achieve high consistency.

**IV. Frame and Packet Structure**

**A. Problem**

In [1], 4G physical layer structure cannot be used to achieve the low latency requirement set by the 3GPP, so a new frame structure is needed to support URLLC. The new frame structure would ideally minimize the processing latency and transmission latency of a URLLC packet [1].

**B. Solutions**

**1. Non-rectangular packet structure**

In [1], a solution to the frame and packet structure problem is to a non-rectangular-shaped packet as opposed to a rectangular-shaped one as seen in Figure 5. In 4G, a rectangular-shaped packet is used to be efficient is using the spectrum of the fading channel. In 5G URLLC however, a structure that stretches the frequency axis is good since it will reduce the time it takes to receive a packet. In this packet structure, the pilot, control, and data components are grouped to be transmitted to allow for channel estimation, data detection, and channel decoding to become possible.

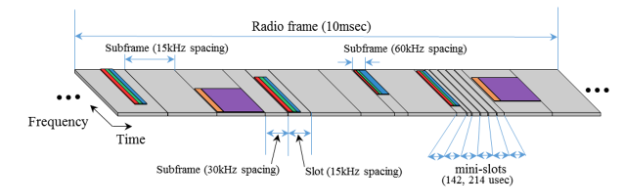
**2. Reducing the OFDM Symbols in the frame**

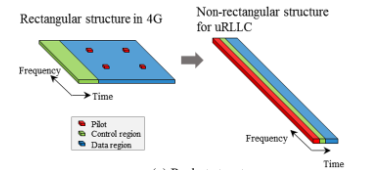
Another solution is presented in [6] to reduce the number of orthogonal frequency-division multiplexing (OFDM) symbols in a frame as seen in Figure 5. By reducing the number of OFDM symbols, the frame size can be reduced which reduces the time it takes to transmit a packet which results in lower latency. This solution is also presented in [1].

**V. Co-existence with Other Services**

**A. Problem**

Detailed in [1], In 5G NR, whenever a service request for URLLC is generated, the URLLC packets need to be transmitted immediately. This transmission should be immediate even during a data transmission period. This means that resources being saved for other services like eMBB and mMTC need to be used for URLLC whenever requested. However, this method does not notify the services like eMBB and mMTC which greatly increases their packet error rate. To create good environment for URLLC, the co-existence problems between URLLC other services, notably eMBB and mMTC, need to be considered and fixed.





**Figure 5.** frame structure (top) and packet structure (bottom)

**B. Solutions**

**1. Retransmission of selected code blocks**

Re-transmission of selected code blocks is a type of reactive strategy [1]. In this strategy, instead of retransmission of everything in error cases, we retransmit the corrupted code block instead. Doing this gives better code gain to URLLC as well as other services that were interrupted by the transmission of the URLLC packet. This allows for URLLC to interrupt other services while maintaining reliability for URLLC and the interrupted services.

**2. Proactive strategies**

In [3], the proactive strategies utilize improved robustness and more resource sharing. A method to improve robustness like adding parity bits to non-URLLC packets, so when URLLC packets interfere with other services’ packets, the error rate of the other services’ packets will be reduced.

**VI. Multiple Input Multiple Output (MIMO) in URLLC**

In 5G, services such as industrial automation and URLLC in areas that are dense in population have trouble avoiding congestion in the infrastructure as well keeping a service available with a massive number of users accessing URLLC [10]. To solve this, research has been looking into massive input massive output (MIMO) solutions.

**A. Base station densification**

Base station densification is a type of MIMO topology that is detailed in [7]. Base station densification is used to ensure reliability to many users by making the presence of base stations denser. Doing this allows for each user to have access to a base station that is a short distance away, and it allows for resources to be used better for each user since network traffic is less congested and more organized.

**B. Multi-antenna diversity**

In [7] multi-antenna diversity is presented as a MIMO solution. The benefits of MIMO are that the signal-to-noise ratio links are very high which means that there is less noise to cause errors which creates more reliability. Having multiple antennas opens more paths for users to use in a base station which helps establish more stable links in URLLC.

**VII. Other Solutions to Various Problems**

**A. Error rate handling**

**1. Problem**

According to [1], block error rate in 4G systems is around 10-2 which is achieved using methods that cannot be used to achieve block error rate of 10-5 for 5G URLLC [1]. To fix this problem, we need a new method that reduced the latency of error handling while maintaining the same high reliability.

**2. Solution**

**a. Asymmetric signal detection (ACK/NACK)**

One way of reducing the latency of error handling is to augment the ACK/NACK error handling system that causes large amounts of latency when an acknowledgement (ACK) or negative acknowledgement (NACK) is received due to having to retransmit which takes time. One solution in [11] is to use asymmetric signal detection which allows for better utilization of resources that maintains reliability as well as lower latency.

**B. Relaying**

In [5], relaying is presented as a solution to improving the reliability of URLLC. Relaying involves the idea of having an intermediate access point for URLLC to use instead of relying solely on end-to-end transmissions to communication. Using the relay point gives more reliability because it reduces the chance of packet loss during the transmission by creates intermediate points that can transmit the data as well.

**C. Access protocols**

**1. Problem**

The use of access protocols revolves around the idea of reducing the latency portion of URLLC [1]. When a URLLC packet is generated, the packet needs to be transmitted without delay to reduce the latency as much as possible. To satisfy the latency problem while not ridding of reliability, difference access protocols have been presented.

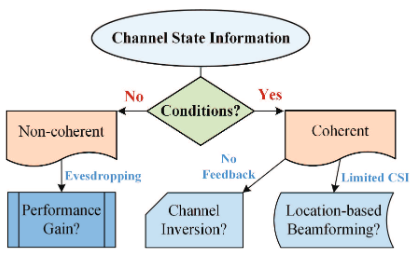
**2. Solutions**

**a. Grant-free**

In [7], grant free is presented as an access protocol. The idea of grant free access is to skip the reservation stage where the user equipment schedules a request to the access point, and the access point sends a transmission grant back to the user equipment. By not using this process, the latency of transmitting a packet is lower.

**b. Contention-based**

In [9] A contention-based access protocol is introduced. Contention-based means that many users can access the access point at the same time without preemption. Although doing this improves the latency, it increases the chance of collision which reduces the reliability of the system.



**Figure 6.** Problems and solutions for security involving channel state information (CSI)

**D. Security**

In [12], the security concerns of URLLC are outlined. In communication methods, the channel state information (CSI) is exchanged which presents the opportunity for an eavesdropping attack if the CSI intercepted by a malicious user. To avoid these problems, this solution suggests modifying the physical layer and not relying on CSI in transmission. Doing so does not impact the reliability or latency requirement of URLLC but also prevents the security concerns that CSI presents. Physical layer security modifications include the use of beamforming to induce location and channel inversion. The problems and solution process can be seen in Figure 6.

**VII. Conclusion**

Considering all the information presented about URLLC, the future for URLLC and 5G NR is a hopeful one. Although there are concerns about the financial feasibility of URLLC and whether the current technology to implement URLLC exist, with more time and research that is being done or will be done, finding a balance between ultra-reliability and low latency will become achievable and technology that will greatly benefit our communication space and our lives will be on the rise.

Bibliography

[1] Ji, Hyoungju, et al. "Introduction to ultra reliable and low latency communications in 5G." *arXiv preprint arXiv:1704.05565* (2017).

[2] Le, Trung-Kien, Umer Salim, and Florian Kaltenberger. "An overview of physical layer design for Ultra-Reliable Low-Latency Communications in 3GPP Release 15 and Release 16." *arXiv preprint arXiv:2002.03713* (2020).

[3] Ji, Hyoungju, et al. "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects." *IEEE Wireless Communications* 25.3 (2018): 124-130.

[4] Yang, Haojun, et al. "Ultra-reliable and low-latency communications for connected vehicles: Challenges and solutions." *IEEE Network* 34.3 (2020): 92-100.

[5] Hu, Yulin, M. Cenk Gursoy, and Anke Schmeink. "Relaying-enabled ultra-reliable low-latency communications in 5G." *IEEE Network* 32.2 (2018): 62-68.

[6] Feng, Daquan, et al. "Toward ultrareliable low-latency communications: Typical scenarios, possible solutions, and open issues." *IEEE Vehicular Technology Magazine* 14.2 (2019): 94-102.

[7] Popovski, Petar, et al. "Wireless access for ultra-reliable low-latency communication: Principles and building blocks." *Ieee Network* 32.2 (2018): 16-23.

[8] Pocovi, Guillermo, et al. "MAC layer enhancements for ultra-reliable low-latency communications in cellular networks." *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017.

[9] Singh, Bikramjit, et al. "Contention-based access for ultra-reliable low latency uplink transmissions." *IEEE Wireless Communications Letters* 7.2 (2017): 182-185.

[10] Siddiqi, Murtaza Ahmed, Heejung Yu, and Jingon Joung. "5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices." *Electronics* 8.9 (2019): 981.

[11] Shariatmadari, Hamidreza, et al. "Asymmetric ACK/NACK detection for ultra-reliable low-latency communications." *2018 European Conference on Networks and Communications (EuCNC)*. IEEE, 2018.

[12] Chen, Riqing, et al. "Physical layer security for ultra-reliable and low-latency communications." *IEEE Wireless Communications* 26.5 (2019): 6-11.