INVITED PAPER

5G-Based Systems Design for Tactile Internet

This article discusses the systems design of ultrareliable and low-latency communications in NR and LTE technologies from the physical and medium access control layer perspective.

By Chong Li, *Member IEEE*, Chih-Ping Li, Kianoush Hosseini, Soo Bum Lee, Jing Jiang, Wanshi Chen, Gavin Horn, Tingfang Ji, John E. Smee, *Senior Member IEEE*, and Junyi Li, *Fellow IEEE*

ABSTRACT | Tactile internet is defined as a network that offers a response to a physical process or an object in perceived real time. The development of new radio (NR) and long-term evolution (LTE) technologies with tailored high-reliability and low-latency design is expected to reliably transmit data in milliseconds, leading to the promising realization of tactile internet with applications in areas such as factory automation, education, gaming, and healthcare. In this paper, from a physical layer and medium-access control (PHY/MAC) perspective, we discuss the systems design of ultrareliable and low-latency communications (URLLC) in NR and LTE technologies, both belonging to the fifth-generation (5G) wireless technologies. Motivated by the safety and privacy requirements of tactile internet, we also outline the 5G security landscape, major categories of attackers, and potential countermeasures. Finally, we provide a case study of factory automation as an example of the proposed 5G-based tactile internet.

KEYWORDS | Fifth generation (5G); latency; long-term evolution (LTE); new radio (NR); reliability; tactile internet

I. INTRODUCTION

Tactile internet, a novel technology that shifts the network from content delivery to skill-set delivery, will revolutionize most aspects of our society in the near future. Recently, significant research effort from industrial and academic associations has been devoted into the study of tactile internet. These research activities and investments are mostly motivated by the numerous applications of tactile internet ranging from healthcare, robotics to autonomous driving and virtual/augmented reality. Concretely, the real-time interaction of human beings requires the response time to the environment on the order of milliseconds (ms). For instance, tactile steering and/or control of an object is carried out in less than 10 ms. The hearing–reacting latency without an observable delay requires a latency of up to 100 ms [1]. As a result, we see that the tactile internet will define a new human–machine interaction where the network provides a physiological latency of human beings to build real-time interactive systems. Numerous applications of tactile internet can be found in [2]–[4].

In the following, we list some key requirements rising from the nature of tactile internet [3].

1) Ultralow latency: As introduced above, the end-to-end latency in the network should be on the order of millisecond to avoid cyber-sickness, where the conflicts between visual, vestibular, and proprioceptive sensory systems occur constantly [5]. This end-to-end latency includes the time spent on the embedded computing on the sensing data, data transmission to/from the control/steering server possibly on the (mobile) edge cloud, and the data processing in the actuator. As a consequence, the latency over the air (or wired) interface needs to be less than a few milliseconds.
2) Ultrahigh reliability: Motivated by the attractive use cases of tactile internet, e.g., factory automation, process control, and remote surgery, demand for the ultrahigh-reliable communications is emerging along with the low-latency needs. For instance, in process/industrial automation and the control of smart grid/substations, the reliability requirement in terms of packet error rates (PERs) is as low as 10^{-5}. In factory automation, this requirement always turns out to be less than 10^{-7} [6].

Manuscript received December 31, 2017; revised June 18, 2018; accepted July 31, 2018. Date of publication August 31, 2018; date of current version January 22, 2019. *(Corresponding author: Chong Li.)*

C. Li and **J. Li** are with Qualcomm Technologies, Inc., Bridgewater, NJ 08807 USA (e-mail: chongl@ntlabs.io).

C.-P. Li, **K. Hosseini**, **S. B. Lee**, **J. Jiang**, **W. Chen**, **G. Horn**, **T. Ji**, and **J. E. Smee** are with Qualcomm Technologies, Inc., San Diego, CA 92121 USA.

Digital Object Identifier 10.1109/JPROC.2018.2864984

3) Security and privacy: How to protect the transferred information in the tactile internet is always a key factor to determine the efficiency of the network. Today, the best effort Internet of Things (IoT) suffers from denial of service (DOS) attacks and targeted cyberattacks, a fact that results in only inefficient best effort communications for cloud services [7]. Given the massive connectivity applications in the tactile internet, the safety and privacy issue becomes more challenging.

Other requirements, including data integrity, edge intelligence, etc., are also imperative to establish and enhance a tactile internet. For instance, edge intelligence is referred to as intelligence-based predictive actuation in close proximity of the tactile edge. With predictive caching of human actions, the tactile internet can overcome the range limit of tactile services due to the finite speed of light (otherwise, for instance, the service is limited to 100 km if using fiber). In this paper, we will focus on the solutions to the aforementioned three key issues by leveraging the ongoing development of new radio/long-term evolution (NR/LTE) ultrareliable and low-latency communications (URLLC) technologies in fifth generation (5G). Discussions on other requirements and the corresponding potential solutions can be partly found in [3] and the references therein.

In what follows, we introduce the background and the development of the current NR/LTE URLLC technologies and their security considerations, upon which we will present the 5G-based physical layer/medium-access control (PHY/MAC) systems design for the tactile internet.

A. New Radio Ultrareliable and Low-Latency Communications

The emerging 5G wireless systems have key features including high throughput, high reliability, low latency, high mobility, and high connection density [8]–[11]. International Telecommunication Union (ITU) has divided 5G network services into three categories: evolved mobile broadband (eMBB), URLLC, and massive machine-type communications (MMTC). Detailed scenarios and requirements for 5G can be found in 3rd Generation Partnership Project (3GPP) TR38.913 [13]. As one of the three major service categories in 5G, URLLC is targeted at emerging applications in which data messages are time sensitive and must be securely delivered end-to-end subject to high reliability and hard latency requirements. The hard latency requirement means that a data transmission that cannot be decoded at the receiver before a deadline is of no use and can be dropped from the system, resulting in the loss of reliability. As an example, the current 3GPP requirement for URLLC includes the hard latency of 1 ms and the average latency of 500 μs over the air interface, and the system reliability of 99.999% [12]. That is, the quality of service (QoS) of URLLC is violated if more than one out of 10^5 packets fails to be delivered in 1 ms over the wireless medium. This is in contrast to the QoS of mobile broadband applications that optimize data throughput and average delay.

The QoS of URLLC introduces several new challenges in the wireless systems design. Mobile broadband applications can rely on a sufficient number of hybrid automatic repeat request (HARQ) retransmissions to achieve ultrahigh reliability, but this approach is limited for URLLC due to the hard latency requirement. Low latency can potentially be achieved by allocating enough resources to minimize the block error rate (BLER) of HARQ transmissions, but it may result in low spectral efficiency and system capacity. Furthermore, all URLLC users, once connected to the radio access network, are expected to satisfy their QoS requirements and receive equal grade of service (EGOS), for which the outage capacity is of interest, but not the ergodic capacity. According to information theory, the outage capacity is defined as the largest system throughput at certain outage probability, e.g., 0.001%, in which the outage here is defined as the event that at least one URLLC user's QoS is not satisfied. Therefore, one can see that the outage capacity is affected by the tail behavior of the wireless system, e.g., the tail of random traffic demand and intracell/intercell interference, and users that are at the cell edge, power limited, or in deep fade. Last but not least, when the services of URLLC and eMBB in 5G are deployed in the same radio spectrum, their coexistence needs to have a future-proof design to best accommodate the mixed demand that is unknown at the moment.

From the standardization perspective, 3GPP has been developing the core standards for 5G, which involve the development of an NR interface and a new core network architecture. To accommodate early commercial deployment, the first phase of 5G specification in Release 15, including URLLC, has been completed in 2018, and the second phase in Release 16, as an IMT-2020 technology candidate that will be submitted to the ITU, will be completed in 2020.

B. Long-Term Evolution Ultrareliable and Low-Latency Communications

Since the first release of LTE technology, the core focus of the standardization efforts has been predominantly placed on enabling higher data rates by aggregating many component carriers, adopting higher order modulation and coding schemes, and the introduction of higher order multiple-input–multiple-output (MIMO) schemes. However, until recently, there has not been an attempt to reduce the LTE air interface latency. Although the data rates supported by the LTE technology have been enhanced considerably in both downlink (DL) and uplink (UL) over the years, all PHY and MAC layer procedure timelines remain the same as originally specified.

Motivated by the need for supporting applications, e.g., tactile internet, with stringent latency and reliability requirements, certain enhancements to LTE air interface

have gained significant traction. In fact, one of the main challenges faced by the 3GPP standardization body in Release 15 was to reduce the LTE PHY and MAC layer latencies by shortening the transmission time interval (TTI) from 1 ms to only a few symbols, while maintaining the LTE waveform and its overall channel structures unchanged [14]. Reducing the TTI length enables faster user scheduling in both downlink and uplink, reduces the HARQ timeline, and allows for access to more accurate channel state information [15]. Consequently, not only the network capacity increases (i.e., a larger number of users can be served in each cell), but also a larger number of retransmission opportunities can be accommodated within the latency bound, which in addition to the possibility for more accurate rate control, lays the groundwork for providing LTE URLLC services. As will be discussed later, other enhancements on the current LTE control and data channels, which are being contributed to 3GPP, are also necessary to fulfill the URLLC requirements.

C. Security in 5G

3GPP systems including 4G LTE and prior generations have had security integrated as part of the fundamental framework since their initial deployments. Though some privacy issues [25] and SIM database breaches [26] have occurred, overall the system has functioned security in terms of confidentiality and integrity, which are two of the three pillars of information security, the other being availability.

Fifth-generation systems are intended to address a larger diversity of services beyond the existing ones in 4G and will introduce new security procedures to address these use cases. Examples of such services include eMBB services with much higher throughput and lower latency, various IoT services for different domains (e.g., smart home/city, healthcare, environmental monitoring), vehicular services, and other critical services (e.g., tactile internet for industrial systems). Some of those services may be established within the 5G systems, whereas others be established via 5G systems. To support different security requirements of various services, 5G system is designed to support various authentication methods, enhanced key management, and subscription identifier privacy.

This paper is intended to provide the fundamental PYH/MAC systems design for 5G URLLC with particular applications to the tactile internet. The paper is organized as follows. Sections II and III present the PHY/MAC design for NR URLLC and LTE URLLC, respectively. Section IV discusses the security architecture and development in 5G URLLC. Section V presents a case study of tactile internet where NR URLLC is applied to an application of factory automation.

II. NR URLLC SYSTEMS DESIGN

In this section, we present the basic PHY/MAC design for NR URLLC in a systematic manner, which is supported by both theoretical queueing analysis and system-level simulation results. It is agreed that 5G system design is not necessary to be backward compatible, thus resulting in novel proposals that could be quite independent from 4G LTE and all previous generations. In what follows, we focus on the downlink of URLLC transmissions in a frequency-division-duplexing (FDD) system, and the same design principles are applicable to the uplink and also time-division-duplexing (TDD) systems [27]. Specifically, we have the following:

- a queueing model is introduced which captures the PHY/MAC layer behavior of a 5G base station (also known as gNB) for URLLC service;
- an efficient HARQ design with shortened TTI and round trip time (RTT) is shown to improve the outage capacity of URLLC and the minimally achievable low latency;
- theoretical queueing analysis and system-level simulations are provided to demonstrate the fundamental tradeoff between the outage capacity, system bandwidth, and the latency requirement for URLLC; these tradeoffs lead to the optimized systems design;
- we show that multiplexing the URLLC and eMBB traffic by frequency-division multiplexing (FDM) is highly inefficient, and wideband resource allocation is needed for URLLC to maximize the outage capacity. Dynamic multiplexing between URLLC and eMBB in both time and frequency domains is also desired to optimize the system spectral efficiency and the capacity of both services.

A. Queueing Model

In the downlink of a radio access network (RAN), the end-to-end delay of a successful data transmission between the MAC entities comprises scheduling delay (the time between packet arrival and the next scheduling instant), queueing delay, transmission delay, receiver-side processing and decoding delay, and multiple HARQ RTTs, where the queueing delay results from the statistical multiplexing of data flows destined for multiple URLLC users. Note that the data flows may also be sporadic and bursty because the traffic models of various URLLC use cases may not be standardized to have a periodic arrival pattern as in the voice-over-IP (VoIP) services in 4G LTE. Therefore, the queueing effect needs to be considered in the systems design of URLLC because of the hard latency requirement. In general, sufficient HARQ retransmissions are needed to achieve high reliability while leaving enough delay margin from the hard latency bound to mitigate the queueing effect. One should understand that the queueing effect could be worsened as the admitted traffic load increases so as to maximize the spectral efficiency for the URLLC services.

To capture the queueing effect and the QoS requirements of URLLC, Fig. 1 provides a simplified queueing model that describes the PHY/MAC layer behavior when

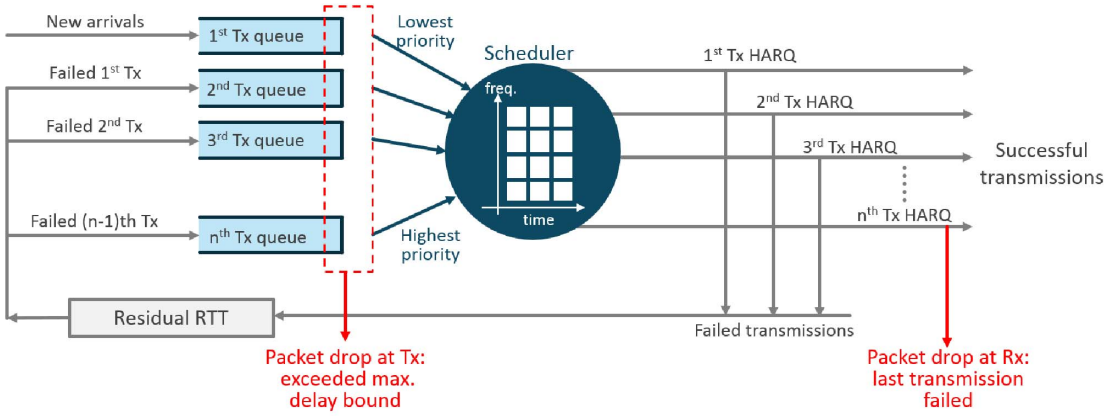

Fig. 1. ***Single-user queueing model that describes the PHY/MAC layer behavior at the gNB for URLLC applications. Given the hard delay bound, a packet can be dropped at the transmitter (Tx) and also at the receiver (Rx).***

a gNB schedules a user on the downlink. In particular, data packets of the user arriving at the gNB, are buffered at the first-transmission queue of the user, and await to be scheduled for their first HARQ transmissions. If the first HARQ transmission of a data packet fails, the packet is available at the second-transmission queue for future retransmission opportunities after an RTT. Whenever a packet that is buffered at the gNB misses its deadline, the packet is of no use and dropped by the system, resulting in the loss of reliability for this user. Additionally, a HARQ failure for a data packet may be declared by the gNB if the packet cannot be decoded after n HARQ transmissions, resulting in the loss of reliability as well. At every scheduling instant, the gNB allocates time and frequency resources to new transmissions and retransmissions of the buffered packets, in order to satisfy the QoS requirements of all users.

In what follows, we use queueing models to get qualitative insight on the performance scaling between reliability, delay, and outage capacity, and justify the observations from the system-level simulation results. In particular, the multiserver aspect of the queueing model can capture the FDM packets in an orthogonal frequency-division multiplexing (OFDM) system. The finite-buffer aspect of the queueing model can capture the fact that packets are dropped from the transmitter side if queueing delay is longer than the latency requirement.

B. Shortened TTI and HARQ RTT Timeline

With the above queueing model in mind, one simple and straightforward method to meet the URLLC QoS is to reduce the durations of TTI and HARQ RTT so that more HARQ retransmissions are allowed to achieve high reliability. Reducing the TTI duration includes using fewer OFDM symbols in one TTI and shortening the OFDM symbols by increasing the subcarrier spacing.[1] Note that short OFDM symbols also enable more efficient pipeline processing that lead to a more tightened HARQ RTT timeline. The RTT timeline reduction, however, mostly depends on the hardware processing capability. One example of the shortened TTI and HARQ RTT is presented in Fig. 2, where RTT is equal to three two-symbol TTIs.

System-level simulation results in Fig. 3 show the system capacity [2] of URLLC under two different HARQ RTT timelines and hard latency requirements, where all URLLC users are deployed at the worst case cell edge with averaged -3-dB signal-to-interference-and-noise ratio (SINR) and assumed to have Poisson packet arrival traffic. We observe that the system capacity and the minimally achievable latency can be substantially improved with shorter HARQ RTT because 1) less time is needed to have enough HARQ retransmissions to meet the reliability target; and

[1]It is noteworthy that the scalable numerology is a new design tool for 5G. In 3GPP, it is agreed that numerology is scalable with subcarrier spacing (SCS) scaled by the power of 2, i.e., $2^n \times 15$ kHz ($n = 1, 2, ...$). In addition, multiple OFDM numerologies multiplexed in the same frequency range should also be supported.

[2]Recall that the URLLC capacity herein refers to the outage capacity rather than the ergodic capacity, since every URLLC user in the serving cell needs to meet the service requirements. For example, if in downlink we assume that all URLLC users have the same packet arrival rate and the maximum packet arrival rate at which all URLLC users meet the service requirements is x packets/second/user, where there are y users in the serving cell and z-bit packet payload is assumed, then the URLLC system capacity is xyz b/s/cell.

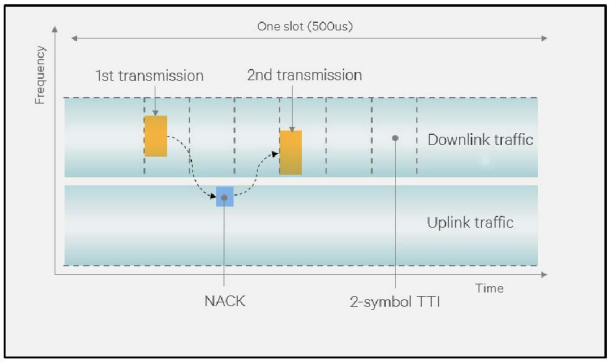

Fig. 2. ***HARQ timeline for URLLC transmissions in FDD system (RTT = 3 two-symbol TTIs).***

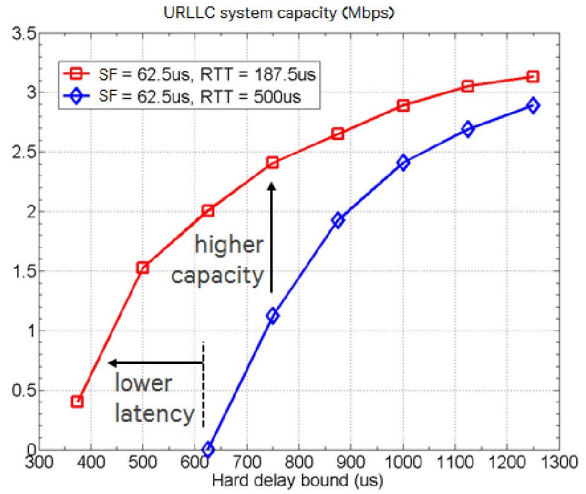

Fig. 3. ***The URLLC system capacity under different hard latency requirements and HARQ RTT timelines (in microseconds). Simulation assumptions include system bandwidth of 20 MHz, 3GPP 3-D urban macrochannel model, Poisson traffic, and the reliability requirement of 99.99%. The term SF stands for subframes that is equivalent to TTI.***

2) URLLC packets have wider delay margin to tolerate more queueing delay before their deadlines.

To obtain qualitative insights on the relationship between the URLLC capacity and the hard latency requirement in Fig. 3, we consider an $M/M/m/k$ queueing model: the first letter M indicates the nature of the arrival process, e.g., M stands for memoryless, which herein refers to Poisson packet arrival process with rate λ; the second letter M indicates the nature of the probability distribution of the service time, e.g., M stands for exponential distribution, which herein refers to the exponential service time with mean $1/\mu$ second (i.e., the time to decode a data packet after multiple HARQ retransmissions) and can be approximated as a geometric distribution[3]; the third letter m indicates the number of allowed concurrent transmissions, which is proportional to the system bandwidth available to the URLLC services; the last letter k indicates that if an arriving packet observes k outstanding packets in the system for some $k > m$, both being queued and undergoing HARQ retransmissions, it will be dropped from the network. The value of k increases with the hard latency requirement, say d milliseconds, in the sense that if an arriving packet sees d-millisecond worth of packets awaiting in the system, it will surely miss its deadline and will be discarded from the network. In this queueing model, the probability p_{block} of dropping packets, i.e., the loss of system reliability, is given by [31]

$$p_{\text{block}} = \left(G p_0 \frac{m^m}{m!}\right) \rho^k = \Theta(\rho^k), \quad \rho = \lambda/(m\mu)$$

where G is a constant, p_0 is the probability that the system is empty, $(1/\mu)$ is the mean service time, and λ is the Poisson arrival rate. From this equation, we have $\lambda = \Theta(\sqrt[k]{p_{\text{block}}})$. Then, the URLLC capacity λ_{URLLC} is the admitted arrival rate in the $M/M/m/k$ queue

$$\lambda_{\text{URLLC}} = (1 - p_{\text{block}})\lambda \approx \lambda = \Theta\left(\sqrt[k]{p_{\text{block}}}\right) = \Theta\left(p_{\text{block}}^{(1/d)}\right)$$

where the first approximation follows from $p_{\text{block}} \ll 1$ and d is the latency requirement (i.e., latency bound of the service requirement). This formula implies that λ_{URLLC} is an increasing function of the latency requirement d with diminishing return because $p_{\text{block}} < 1$. This theoretical analysis explicitly reflects the delay-capacity scaling curve in Fig. 3.

Although reducing RTT can always help increase the capacity, shortening TTI is not always beneficial. The drawback of short TTI duration includes more control overhead which reduces the data capacity for URLLC, although this can be this can be partially alleviated by grant-free transmissions (e.g., semipersistent scheduling). Furthermore, shortening TTI by reducing the number of OFDM symbols implies that more resource blocks (RBs) need to be allocated in the frequency domain for a packet transmission to achieve a target BLER, in which the loss of trunking efficiency adversely affects the outage capacity of URLLC. Similarly, shortening TTI by reducing the duration of OFDM symbols (or, equivalently, increasing the subcarrier spacing) implies that fewer RBs are available in the frequency domain, a fact that will definitely cause more queueing effect. Overall, the TTI and RTT durations should be chosen carefully to optimize these performance tradeoffs. The TTI durations of 2, 4, 7, or 14 OFDM symbols have been agreed in 3GPP standardization.

C. Wideband Allocation for URLLC

The traffic demand of URLLC services may be sporadic in some use cases. For instance, telesurgery requires the network to support timely, reliable, and sporadic haptic feedback enabled by sensors located on the surgical equipment. Another use case could be the sporadic safety message exchange among automated driving vehicles when an emergence event occurs. In such cases, the system resources cannot be fully utilized (to be seen later). Thus motivated, in 5G systems efficient multiplexing between URLLC and eMBB becomes important to maximize the overall system spectral efficiency. One approach is to partition the system bandwidth statically or semi-statically and serve the URLLC and eMBB traffic by FDM. However, we next present a queueing analysis and system-level

[3]This is only an approximation in a sense that a packet is continuously serviced in the queueing model, but the resources are made available in between HARQ retransmissions of the packet in the wireless network.

simulations to show that statically or semi-statically reserving bandwidth for URLLC yields low spectral efficiency and resource utilization.[4] Instead, wideband allocation for URLLC is desired, a motivation for the dynamic multiplexing that will be discussed later.

1) Queueing Analysis: Assume some bandwidth is reserved for URLLC. The URLLC traffic model is assumed to be Poisson. The QoS requirements include the delivery of fixed-size data packets over the air interface within the hard latency bound of L milliseconds and the system reliability of $(1 - p_{\text{loss}})$, e.g., $L = 1$ and $p_{\text{loss}} = 10^{-5}$ [13]. The amount of reserved bandwidth decides how many URLLC transmissions can be packed in the frequency domain in one TTI. From these assumptions, we consider an $M/D/m/m$ queueing model: the first letter M refers to the Poisson arrival process; the second letter D stands for the deterministic distribution, which here means that the over-the-air delay is a fixed value (e.g., assuming all URLLC transmissions can be decoded successfully in one transmission and there is no need for HARQ retransmissions nor acknowledgment/negative-acknowledgment (ACK/NACK) feedback); the third letter m denotes the maximum number of concurrent transmissions in one TTI; and the last letter m indicates that packets that have positive queueing delay are dropped from the system due to missing a stringent hard deadline.[5] In this queueing model, the loss of system reliability is the Erlang-B formula [31, Sec. 3.4.3]

$$p_{\text{loss}} = \frac{(\lambda/\mu)^m/m!}{\sum_{n=0}^{m}(\lambda/\mu)^n/n!} \tag{1}$$

where λ is the Poisson arrival rate in the unit of packets per TTI and $(1/\mu)$ is the fixed service time in the unit of TTIs where we can let $\mu = 1$ here. In addition, assume k out of m resources are used per TTI. Based on the $M/D/m/m$ queueing model, the system utilization, defined as the time-average proportion of the allocated resources, is given by

$$\frac{1}{m}\sum_{k=1}^{m}\frac{k\cdot(\lambda/\mu)^k/k!}{\sum_{n=0}^{m}(\lambda/\mu)^n/n!}. \tag{2}$$

Fig. 4 shows the relationship among the loss of system reliability, packet arrival rate, and resource utilization according to (1) and (2). We observe that increasing the packet arrival rate yields more severe queueing effect and thus the loss of system reliability. As the reliability requirement is tightened, we must reduce the traffic load with lower resource utilization in order to continue meeting the QoS requirement. Fig. 5 shows the maximally supported arrival rate (i.e., system capacity) and the resource utilization under the requirement of $p_{\text{loss}} = 10^{-5}$. As the number of allowed concurrent transmissions (i.e., the available bandwidth for URLLC) decreases, the system capacity decreases exponentially with low resource utilization. As an example, when the number of concurrent transmissions is reduced from 16 to 8 to 4, the maximally supported arrival rate changes from 4.2 to 1.0 to 0.1 packets per TTI, respectively. If we consider the spectral efficiency over the available bandwidth to be the ratio of the achievable packet arrival

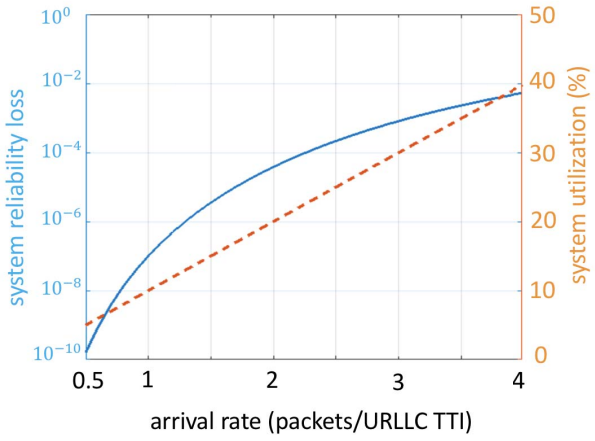

Fig. 4. *The relationship among the loss of system reliability, resource allocation, and the packet arrival rate in an M/D/m/m queueing model with m = 10. The hard latency requirement is implied in the queueing model, in which queueing delay is disallowed and leads to packet drop.*

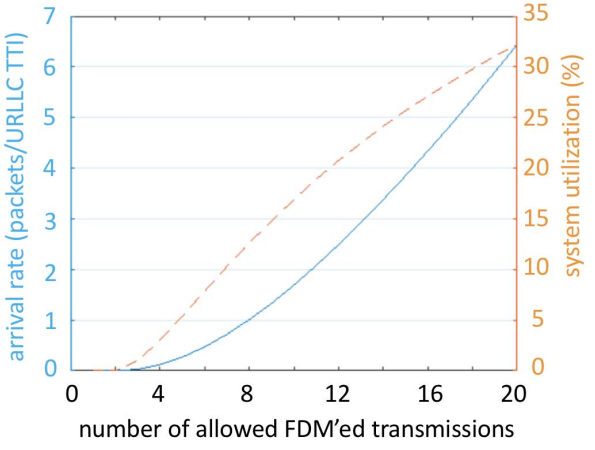

Fig. 5. *Maximally supportable Poisson arrival rate and the resource utilization under the reliability of $p_{loss}=10^{-5}$ in an M/D/m/m queueing model. The "x-axis" refers to the first "m" that is the number of allowed simultaneous data transmissions and scales with the reserved bandwidth for URLLC.*

[4]The resource utilization is measured by the ratio of averaged used/scheduled resource over the total available resource in terms of RBs. If the number of available RBs is less than the number of required RBs for a transmission, no scheduling will be conducted.

[5]Different from the $M/M/m/k$ queueing model in Section II-B, an $M/D/m/m$ queueing model is considered here to get a simple and closed-form expression of packet-loss probability in (1).

Table 1 URLLC System Capacity and Spectral Efficiency Under Different Available System Bandwidth and the Hard Latency Requirements. The Reliability Requirement Is 99.999%

	Hard latency requirement		
System bandwidth	500μs	750μs	1ms
5MHz	0Mbps	0Mbps	0Mbps
10MHz	0.28Mbps (0.03bps/Hz)	3.94Mbps (0.39bps/Hz)	6.2Mbps (0.62bps/Hz)
20MHz	10.7Mbps (0.54bps/Hz)	15.8Mbps (0.79bps/Hz)	16.9Mbps (0.85bps/Hz)

rate to the number of concurrent transmissions, then the spectral efficiency is improved from $0.1/4 = 0.025$ to $1.0/8 = 0.125$ to $4.2/16 = 0.2625$ as the number of concurrent transmissions is doubled. It indicates that the spectral efficiency of the URLLC traffic super-linearly increases with the available bandwidth.

2) System-Level Simulations: To verify our theoretic queueing analysis, we perform system-level simulations on the FDD downlink. There is one URLLC serving cell with 20 eMBB neighboring cells in a wrapped-around model. Overhead of the control channels [physical downlink control channel (PDCCH) and physical uplink control channel (PUCCH)] is not modeled. The gNB in the serving cell periodically sends reference signals to the randomly distributed URLLC users for channel estimation, and the users report the channel state information to the gNB for the MIMO operation, e.g., computation/selection of precoding matrix. The scheduling policy is delay based and focuses on providing EGOS to the users to optimize the outage capacity. A packet is dropped from the system when its deadline is expired. The serving cell is subject to full-buffer intercell interference from all neighboring cells. All URLLC users are assumed to have the same Poisson traffic load. Recall that the system capacity for URLLC is defined as the largest arrival rate at which the hard latency and reliability requirements are satisfied for all users, including the cell-edge ones. A range of arrival rates is swept to find the system capacity. See the Appendix for the details of simulation assumptions.

Table 1 shows the URLLC system capacity under different available bandwidth and the hard latency requirements with the target reliability requirement of 99.999%. The system capacity scales super-linearly and the spectral efficiency improves with the available bandwidth as predicted in the above queueing analysis. The URLLC capacity under the reserved bandwidth of 5 MHz is negligible because the lack of RBs in the frequency domain requires more HARQ retransmissions and causes severe queueing effect so that the hard latency and reliability requirements cannot be met simultaneously. Meanwhile, Table 2 shows how sensitive the capacity is with respect to the reliability requirement. For instance, at 500-μs latency requirement the capacity is reduced by 44% when the target reliability increases from 99.99% to 99.999%. This is due to the fact that the URLLC capacity is essentially determined by the tail events (e.g., deep fades). Table 3 shows the resource utilization of the reserved bandwidth when the URLLC users are fully loaded at the capacity-achieving points shown in Table 1. The resource utilization is generally low, especially over the narrow bandwidth, and cannot be increased by admitting more traffic because the QoS requirement will be violated. Therefore, making the wide system bandwidth available is beneficial to the practical deployment of the URLLC services.

Table 2 The URLLC System Capacity and Spectral Efficiency Under Different Reliability Requirements. The System Bandwidth Is Assumed to Be 20 MHz

	Hard latency requirement		
Target reliability	500μs	750μs	1ms
99.99%	19Mbps (0.95bps/Hz)	20.3Mbps (1.02bps/Hz)	21.3Mbps (1.07bps/Hz)
99.999%	10.7Mbps (0.54bps/Hz)	15.8Mbps (0.79bps/Hz)	16.9Mbps (0.85bps/Hz)

Table 3 Resource Utilization of the Available Bandwidth at the Capacity-Achieving Points in Table 1

	Hard latency requirement		
System bandwidth	500μs	750μs	1ms
5MHz	0%	0%	0%
10MHz	2.3%	32.4%	50.9%
20MHz	43.3%	64%	68.6%

D. Dynamic Multiplexing of eMBB and URLLC

Following the above discussions, the eMBB and URLLC services, when sharing the same radio spectrum, need to be dynamically multiplexed over the entire system bandwidth to maximize the spectral efficiency of both services. As an example in Table 3, the URLLC services have the maximum resource utilization of 50.9% under the system bandwidth of 10 MHz and the latency requirement of 1 ms. The 49.1% of the system resources are wasted under the static FDM between eMBB and URLLC. If dynamic multiplexing is applied, the eMBB capacity could jump from zero to 20.52 Mb/s following the calculation of 4.18 b/s/Hz $\times$ 10 MHz $\times$ 0.491 = 20.52 Mb/s, where 4.18 b/s/Hz is the eMBB full-buffer spectral efficiency on the downlink [32].

In the time domain, URLLC requires short TTI duration, e.g., two OFDM symbols or 71.43 μs under the subcarrier spacing of 30 kHz, to be the smallest scheduling time unit to reduce both the HARQ RTT and the scheduling delay as discussed in Section II-B. For the eMBB traffic that typically involves bulk data transfer, long TTI duration, e.g., 500 μs, is preferred to minimize control overhead and exploit the coding gain of long codewords to improve the data capacity and spectral efficiency. Due to the different scheduling granularity, eMBB and URLLC services need to be dynamically multiplexed in the time domain as well. See Fig. 2 for an example. When URLLC packets (yellow) arrive at the gNB during an ongoing eMBB transmission, they cannot wait until the end of the eMBB burst/slot to be scheduled because of the hard latency requirement. In this

case, the gNB may suspend the eMBB transmission to free up resources to transmit the URLLC data in the next URLLC TTI, which is referred to as the puncturing mechanism. In such cases, an indication mechanism is used for the gNB to inform the punctured eMBB user about the location of the preempted resources to improve its decoding performance. In 3GPP, it has been agreed that eMBB users will monitor preemption indication at least on the slot level. Having multiple monitoring opportunities in a slot allows eMBB users to flush the HARQ buffer faster, which improves the decoding delay and reduces the maximum HARQ buffer size for the eMBB users.

In summary, dynamic multiplexing of eMBB and URLLC services in both time and frequency domains is desired. This is a flexible scheme that will be able to handle the mixed yet unknown demand of both services.

E. URLLC on the Uplink

The systems design for the URLLC services on the uplink is more challenging than the downlink case for several reasons.

First, the user equipment (UE) may be power limited on the uplink so that it is harder to satisfy the high reliability requirement, especially for the users that are temporarily at the cell edge or in deep fade. Moreover, in typical uplink transmission protocols, a user initiates a hand-shaking procedure to request resources from the gNB for initial uplink data transmissions, referred to as the grant-based method. Here, a gNB performs dynamic scheduling and allocates orthogonal system resources to the users to exploit the multiuser diversity gain and maximize the system capacity. The initial hand-shaking procedures need to be sufficiently reliable for the gNB to detect the presence of data that are pending for transmission at the user side, which puts a heavy burden on the design of uplink control channels. The hand-shaking procedure also incurs additional delay overhead that reduces the delay budget for the actual data transmissions. An alternative approach for uplink data transmissions is the grant-free approach, where each user semi-statically allocates some system resources when it is connected to the network. Upon new packet arrival, users use the reserved resources for the initial uplink transmission, avoiding the overhead of hand-shaking procedures in the grant-based method. In the grant-free method, users may not have good channel state information when initiating grant-free transmissions, and resources need to be reserved conservatively to the users to reduce the expected block error rate. To improve overall resource utilization, the same system resources may be reserved for multiple users. Nonorthogonal multiple-access (NOMA) method is needed when users that share the same resources seek to perform initial transmissions at the same time.

Despite the challenges outlined above, the systems design principles discussed in the URLLC downlink are applicable to the uplink case. Table 4 presents the system-level simulation results for the uplink URLLC transmissions under idealized assumptions (see [33] for more details). Similarly to the downlink, the system capacity grows super-linearly and the spectral efficiency improves with the available bandwidth. Dynamic multiplexing of eMBB and URLLC services is desired on the uplink as well, where the gNB can preempt ongoing eMBB uplink bursts by explicit signaling to free up resources for the URLLC transmissions.

Table 4 System Capacity, Spectral Efficiency, and Resource Utilization in the Case of Uplink URLLC Transmissions

System BW (MHz)	System capacity (Mbps)	Spectral efficiency (bps/Hz)	Resource utilization
5	4.51	0.9	51.5%
10	12.39	1.24	65.3%
20	28.16	1.41	74.4%

As discussed in this section, the shortened HARQ technique is a key enabler to achieve a certain level of reliability and latency. In some cases, however, the latency due to retransmissions still exceeds the target if the target latency bound is extremely tight. One alternative solution beyond the PHY/MAC layers could be packet duplication (PD) without HARQ retransmissions at the packet data convergence protocol (PDCP) layer. The fundamental idea underlying PD is to transmit multiple instances of the same packet over different uncorrelated channels such that the receiver can exploit the channel diversity to achieve high reliability. An overview of PD technique and the architecture enhancements related to URLLC beyond PHY/MAC layers in 3GPP NR RAN can be found in [34].

To end this section, we remark that the above PHY/MAC design establishes the foundation for NR URLLC. Additional design aspects including coding techniques (e.g., polar codes, low-density parity-check codes), advanced MIMO and NOMA, each of which could improve a particular functionality of the system, should also be studied for the integrity of the systems design. From the 3GPP standardization perspective, besides the aforementioned fundamental PHY/MAC design, the support for URLLC in Release 15 also includes new channel quality indicator (CQI) table to achieve 10^{-5} BLER, new modulation and coding scheme (MCS) table with lower spectral efficiency entries for URLLC, and physical uplink shared channel (PUSCH) repetition for UL semi-persistent scheduling (SPS). In future releases, the technical items to study will include eMBB and URLLC multiplexing in UL using UL preemption, carrier aggregation (i.e., FDD+FDD or FDD+TDD), coordinated multipoint (CoMP) enhancement for industrial automation, power optimization, NOMA and SPS enhancements if control channel capacity becomes bottleneck, UE/gNB processing time reduction by hardware optimization, and design enhancements tailored to various URLLC use cases/QoS from SA1 and SA2.[6]

[6]SA is the 3GPP Technical Specification Group on service and system aspects. There are subgroups on different topics in which SA1 and SA2, respectively, refer to services and architecture.

III. LTE URLLC SYSTEMS DESIGN

As an alternative candidate for the solutions to tactile internet, high-reliability and low-latency LTE has attracted extensive effort since 3GPP Release 13 [28]. Furthermore, different aspects of latency reduction have been addressed in the Release 14 version of the LTE standard [29]. Unlike the NR URLLC which allows for the introduction of an entirely novel air interface design, the LTE URLLC enhancements must be backward compatible. In this regard, the LTE URLLC follows the same numerology (i.e., subcarrier spacing and symbol duration) as the legacy LTE, which only supports 15-kHz subcarrier spacing. Note that the same numerologies can significantly simplify the multiplexing of the legacy LTE and the URLLC LTE. Although this constraint may separate the paths of their evolution or their usage for different applications, the main design ideas considered for supporting URLLC services under the two technologies (i.e., NR and LTE) are almost identical. In this section, we present an overview of the possible PHY/MAC layer enhancements for LTE URLLC.

Similar to NR URLLC, TTI shortening is a simple and effective way to reduce the air interface latency and to speed up the PHY and MAC procedures. However, as we have to maintain the same numerology as the legacy LTE, the only way to shorten the TTI length is to reduce the number of symbols in one subframe or TTI,[7] e.g., from 14 symbols (i.e., 1-ms TTI) to only a few symbols, e.g., two symbols. Due to the TTI shortening, the HARQ and channel state information (CSI) feedback timelines can be significantly reduced. This enhancement, the same as NR URLLC short TTI and RTT design, leads to: 1) more retransmission opportunities within the hard delay bound; 2) a more accurate link adaptation (or rate control) that improves the link reliability; and 3) faster DL and UL scheduling.

To demonstrate the benefit of short TTI with legacy LTE numerology, we investigate two system configurations where CSI feedback periodicity is proportionally scaled with the TTI size. The LTE baseline (14-symbol TTI) has CSI update of every 5 ms and the one-slot (seven-symbol) TTI-based low latency has CSI update of every 2.5 ms. We compare the average cell loads in Table 5. [8] As can be observed, the seven-symbol low-latency LTE exhibits lower cell load compared with the LTE baseline. For example, for 1-Mb burst size the cell load measured through TTI utilization drops from 31% to 24%. This observation leads to the potential for substantial increase in system capacity as the system can support more traffic before exhibiting instability.

Table 5 Average Cell Load (TTI Utilization Ratio)

	100kb burst size and 0.1s burst interval	1Mb burst size and 1.0s burst interval
LTE baseline (5ms CSI update)	0.43	0.31
One-slot TTI (2.5ms CSI update)	0.40	0.24

Besides very tight delay budget (e.g., 1-ms over-the-air delay) for some use cases for URLLC, it is noteworthy that there are also use cases where the delay budget is in the order of tens of milliseconds while still requiring ultrareliable packet delivery (e.g., 10^{-5} or lower PER). Such cases make it possible to utilize a longer TTI such as one-slot TTI or 1-ms TTI with relaxed HARQ processing time. In addition, a longer TTI also makes it easier to accommodate users who are far away from the base station, i.e., a longer TTI length allows for enhancing the uplink coverage.

Similar to the coexistence issue of 5G NR network, a 5G LTE network may have URLLC users and legacy LTE users coexisting in an FDM or a time-division multiplexing (TDM) manner. This is illustrated during a downlink transmission in Fig. 6. Since the first LTE 3GPP Release, each user may be scheduled on a per RB basis. For URLLC, each user may be scheduled on a per URLLC block basis, where each URLLC block consists of multiple RBs in frequency and one symbol in time balancing between scheduling flexibility and control overhead. One key difference between a 1-ms legacy LTE TTI and a two-symbol URLLC TTI is that for the latter, the scheduling decision is made more frequently (per 1 ms versus per two-symbol TTI). Consequently, a newly scheduled two-symbol TTI transmission may overlap with an earlier scheduled ongoing 1-ms-based transmission. Some careful scheduling and standardization can be used to minimize or to handle such a collision. As an example, legacy broadcast physical downlink shared channel (PDSCH), intended for all UEs of a given cell,

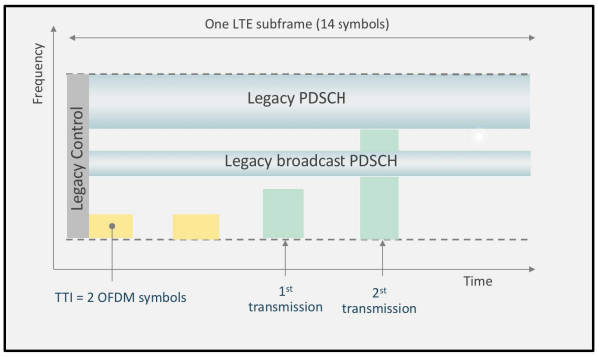

Fig. 6. ***Multiplexing of two-symbol TTI and 1-ms TTI in downlink.***

[7] One subframe or TTI of 1 ms in LTE contains 14 symbols, or equivalently, two consecutive slots each of seven symbols.

[8] We conducted the system-level simulations assuming a macrocellular layout with 21 wrap-around cells utilizing the model defined in [30]. There are ten UEs per cell sharing the total of 10-MHz bandwidth, and each UE is receiving a Poisson burst traffic over the extended typical urban (ETU) channel model where the maximum Doppler frequency is 3 Hz. In particular, we consider two Poisson traffic configurations: 1) the burst size of 100 kb with 0.1-s average burst interval; and 2) the burst size of 1 Mb with 1-s average burst interval.

may be given a higher priority compared with user-specific URLLC scheduling (e.g., the green URLLC user in Fig. 6).

Although TTI shortening sets the foundation for supporting the URLLC services, to meet their stringent requirements, additional enhancements should be introduced in both downlink and uplink directions. Unless otherwise specified, our discussion below will focus on two-symbol-based TTI.

In the downlink direction, it is essential to consider the reliability of both the downlink control and downlink data channels [15].

1) Downlink control: Providing the control/scheduling in every few OFDM symbols, reducing the downlink control information (DCI) payload size (e.g., by semi-statically setting some parameters), and reducing the DCI false detection rate are among the potential schemes for improving the reliability of the downlink control transmission. In particular, legacy PDCCH can fit in one or more symbols at the beginning of each subframe. The basic construction block for PDCCH is a resource element group (REG), which consists of four frequency-consecutive resource elements. Such a design allows for frequency-distributed operation for PDCCH, which is robust against various channel conditions. In order to ensure fast scheduling, the shortened physical downlink control channel (sPDCCH) may be present in every one or two symbols. For a cell specific reference signal (CRS)-based RB set, sPDCCH length could be one or two symbols. For demodulation reference signal (DMRS)-based sPDCCH, it is always two symbols. The fact that legacy PDCCH can fit in one symbol makes it a natural candidate for sPDCCH. However, it is possible that the control payload size for sPDCCH may be different than that of the PDCCH. For example, under the relaxed resource allocation granularity, sPDCCH may have a smaller control payload size. Therefore, some detailed design parameters for sPDCCH can be different than those for PDCCH. What is more, in Release 15, the DCI repetition from the network perspective is supported. In particular, the gNB can transmit the same DCI multiple times (in multiple short TTIs) to ensure reliability. However, from the UE perspective, once one DCI is detected, the UE will ignore other DL DCIs received within the repetition window.[9]
2) Downlink data: The legacy PDSCH has multiple transmission modes, which are widely classified based on the number of antennas on transmit and receive sides. In LTE downlink, the transmission modes include single antenna mode, transmit diversity mode, open-loop spatial multiplexing, closed-loop spatial multiplexing, and multiuser MIMO. Some modes are based on CRS for demodulation, while other modes are based on a user-specific DMRS. Furthermore, while some modes are limited to a single-layer operation, other modes can support multilayer operation, in an open-loop or a closed-loop manner. For the subslot or slot physical downlink shared channel (sPDSCH), different CRS and DMRS transmission modes are supported as well. In addition, different TTI lengths may be further considered. For example, in addition to a two-symbol TTI-based operation mode, a one-slot TTI may be configured as another mode for a user, especially when the user is in relatively unfavorable channel conditions requiring longer transmission duration to meet link budget limitations. Furthermore, providing the spatial diversity by accommodating a large number of antennas at the base station and/or UEs, the channel frequency diversity via transmitting the same TB across different component carriers and temporal diversity via repeatedly transmitting the same TB across multiple TTIs (i.e., HARQ-less retransmission) can be considered for improving the reliability of the downlink data channel. Moreover, coordinating adjacent cells to protect the resources assigned to URLLC services against intercell interference is another design dimension that can be explored.

Similarly, the uplink transmissions should be properly optimized to meet the URLLC requirements. In addition to some of the solutions mentioned above which are applicable for uplink transmissions as well, some enhancements specific to uplink control and data channels need to be considered [15].

1) Uplink control: The legacy physical uplink control channel (PUCCH) has a duration of 1 ms or two slots, utilizing one RB in each slot of the subframe. The two RBs are mirror mapped in frequency to maximize frequency diversity gain. To ensure single-carrier (SC) waveform property, each symbol is dedicated to either control data or reference signal for the control data. The number of reference symbols for PUCCH depends on the PUCCH format, where multiple PUCCH formats are supported targeting different channel conditions and uplink control payload sizes. For the shortened PUCCH (sPUCCH), similar to the DCI repetition, sPUCCH repetition could be one way to enhance the UL control coverage.
2) Uplink data: The legacy PUSCH also follows SC–FDM waveform, and can support either a single-layer or a multilayer transmission. In each slot, there are six symbols dedicated for data and one symbol dedicated for reference signals. Similar to sPUCCH, a careful analysis is necessary to ensure that the shortened physical uplink shared channel (sPUSCH) can achieve a reasonable tradeoff among link budget efficiency, overhead for reference signals, and flexi-

[9]Repetition window refers to the number of consecutive TTIs over which the same transport block (TB) is sent multiple times.

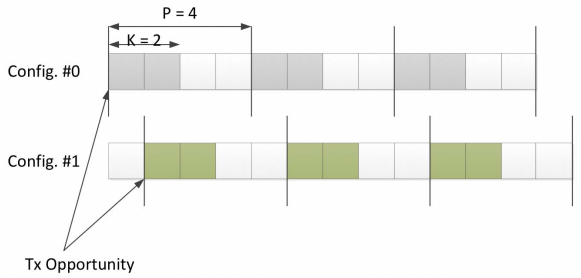

Fig. 7. *Two SPS configurations (config.) for a UE with P = 4, K = 2 and offset of one TTI. The transmission (Tx) opportunity is also illustrated.*

bility in operation. It is worth noting that the TTI duration for uplink is not necessarily the same as that of the downlink for a user, given the fact that the impact of shortened TTI can be different in downlink and uplink. Furthermore, in Release 15, UL SPS with repetition is allowed. Multiple "active" SPS configurations can be configured for a UE on each serving cell. Different configurations are shifted with respect to each other, and each of them has parameters P (periodicity) and K (number of repetitions). One example is illustrated in Fig. 7. Depending on when the packet arrives, the data packet can be sent on one of the configurations starting with the first TTI out of K.

Besides the aforementioned enhancements on LTE control and data channels, the uplink power control schemes should also be revisited to guarantee the successful reception of the UL packets sent from the cell-edge users. To end this section, we summarize in Table 6 the key features of NR and LTE URLLC systems design discussed in this paper. Note that other features such as power optimization, NOMA, and multiconnectivity, which herein are not thoroughly discussed, are not captured in this table. But these features should not be ignored for the integrity of NR and LTE URLLC solution.

IV. URLLC SECURITY

In this section, we discuss another important aspect of URLLC design, information security, which is undoubtedly of importance to tactile internet. We first provide an overview on the 5G security landscape, then discuss two major possible attacks and their respective countermeasure from a PHY/MAC perspective.

Table 6 Key Enablers for NR URLLC and LTE URLLC

	NR URLLC	LTE URLLC
Scalable numerology	yes	no
Shortened TTI by scaling SCS	yes	no
Shortened TTI by reducing the number of symbols	yes	yes
Shortened RTT	yes	yes
HARQ timeline optimization	yes	yes
Dynamic multiplexing	yes	yes
Grant-free transmission	yes	yes

A. Fifth-Generation Security Landscape

Service diversity in 5G systems implies that various types of devices with different capabilities would be connected to the 5G system and consequently the 5G system should be flexibly configured to adapt to meet different service requirements. This would require substantial security capability enhancements since traditionally 3GPP has taken a monolithic view of the security architecture. For example, in LTE, a device can only authenticate with the network based on the method of extensible authentication protocol (EPS), authentication and key agreement (AKA) [19] to establish a security context with the network. This context is then used to protect all subsequent communications with the network.

Fig. 8 shows access stratum (AS) and nonaccess stratum (NAS) security protocol layers in the device and network for LTE. The communication between a device and a base station occurs either over signaling radio bearers (SRBs) for control signal exchanges [e.g., radio resource scheduling (RRC)] or over the data radio bearers (DRBs) for user plane (UP) data transfer; and is protected using AS security based on keys provided by the core network to the RAN. The control plane (CP) signaling between the device and the core network (CN) is protected using the NAS security mechanism. There was no core network for user plane security defined for LTE as the user-plane data security relies on the AS security that terminates at a base station and the network domain security between the base station and CN nodes (e.g., serving gateway and PDN gateway).

3GPP has also introduced additional security features for new services in LTE since the initial release. For example, for battery efficiency of machine-type communication (MTC) or IoT devices of tactile internet, certain security features were introduced in the 3GPP systems in the context of control plane optimization (e.g., data over NAS [20], [22]), user-plane optimization (e.g., RRC suspend/resume [20], [23]), or battery efficient security design (e.g., BEST [21]). These efforts are some precursors of the needs for the additional/evolvable security models for the diverse services in 5G. Security enhancement for diverse services in 5G includes, but not limited to, supporting extensible authentication protocol (EAP) framework, end-to-end authentication for enhanced home network control, subscription identifier privacy, and introduction of security anchor to support security feature evolution and various network deployment scenarios. For example, as one of the most valuable new security features in 5G, a new authentication framework can be employed in which mobile operators can flexibly choose authentication credentials, identifier formats, and authentication methods

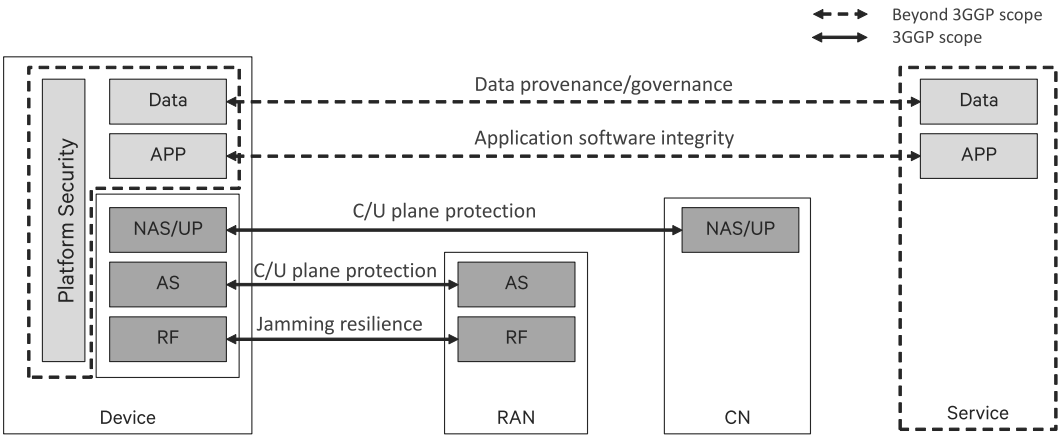

Fig. 8. ***Architecture of cellular system security (C/U plane stands for control/user plane).***

for subscribers. In 3GPP specifications, this is called the 5G authentication and key agreement (5G AKA) protocol and the extensible authentication protocol (EAP) framework. One can see that the flexibility in which different authentication protocols and credential types can be used without affecting intermediate nodes is the key feature of EAP. For the latest progress in 3GPP related to 5G security, we refer the interested readers to [24], which specifies the security architecture, i.e., the security features and the security mechanisms for the 5G system and the 5G core, and the security procedures performed within the 5G system including the 5G core and the 5G new radio.

For the new use cases in 5G such as critical services, the AS security may not be sufficient for the user-plane data protection because compromising a base station, which is susceptible to physical attacks, would break the security of the user-plane data of the connected devices. This is especially true for small cells that utilize high frequencies (i.e., millimeter-wave) as they need to be installed close to devices due to their smaller coverage. Implementing user-plane security at a gateway in the CN would enhance the data protection of user data that is essential for critical services.

Besides the AS and NAS security that is under 3GPP control, 5G systems need to be equipped with more intrinsic security functionalities tailored for IoT services. IoT devices such as sensors and actuators are usually deployed and operate in unattended environments and managed remotely. As such, these devices are susceptible to malicious attacks either launched locally (e.g., based on physical attacks) or remotely (e.g., by exploiting software vulnerabilities). To identify or withstand such attacks, the administrator of those devices should be able to verify the integrity of the application running in the device remotely. In other cases, the remote servers that collect data from the devices and process the data need to verify the data provenance which guarantees the data origin and its authenticity as well as the data governance that manages the access and processing privilege of the data. These 5G security features, in effect, rely on the strong security of the device platform that is preferably established based on the hardware root of trust (e.g., secure element) and is verifiable based on device authentication and attestation. In particular, the non-3GPP related security features that are currently managed by third party services are essential to mobile operators' networks to establish new services natively.

B. Availability Guarantees

While the 5G system aims at high reliability and low latency over the radio link, little investigation has been made whether such goals could be achieved in the presence of attacks against availability, especially radio jamming attacks. In general, it is extremely difficult to counter the radio jamming attacks due to the intrinsic vulnerability of the wireless medium. At the same time, it is difficult for an attacker to effectively jam a wide spectrum over the wide area that a cellular base station covers. However, this does not necessarily mean the cellular system is relatively safe against such jamming. A sophisticated attacker may exploit the radio resource scheduling mechanisms employed in the 3GPP system to maximize attack effects. We next discuss two types of attacks and also the approaches of countermeasure.

1) Jamming Attacks: Intelligent jamming attacks can be classified into two categories, namely control channel attacks and data channel attacks.

- Control channel jamming attack: Jamming uplink and/or downlink control signals can effectively deny packet transmissions and receptions over the data channels. For the uplink, PUCCH is used to transmit scheduling request (SR), ACK/NACK or CSI by a UE, hence jamming PUCCH will result in preventing/disrupting data transmission or triggering a radio link failure (RLF). In LTE, the PUCCH resources are

located at the edge of the system bandwidth and utilize narrow bandwidth. Hence, they can be easily jammed by an attacker if the serving gNB location is known. For the downlink, PDCCH is used to carry DCI such as downlink scheduling assignments or uplink scheduling grants. For a specific UE, jamming PDCCH would prevent the UE from determining that it has been scheduled. However, the effect of PDCCH jamming is relatively localized if the attacker's jamming signal power is limited and hence can only reach to nearby UEs. Nevertheless, PDCCH offers very useful information for the attacker who is targeting a particular set of UEs, since PDCCH contains information on the UE specific uplink and downlink RB allocation.

- Data channel jamming attack: Selective jamming of the shared data channels (i.e., PUSCH and PDSCH) targeting a set of UEs is possible by decoding PDCCH. Since the PDCCH contains information about the RBs allocated to the target UEs, an attacker can concentrate a jamming signal on those RBs in the data channels, thereby reducing power utilization.

Control channel jamming, especially against PUCCH, impacts all UEs connected to a base station. At the same time, these impacts manifest the presence of attacks which can be easily identified by the base station and would trigger an alarm or a reactive measure [e.g., by locating the jamming source(s)]. On the other hand, data channel jamming can be selectively launched against the target UEs, hence is much stealthier than the control channel jamming. Note that the entire RBs allocated to the target UEs do not need to be jammed. Damaging only a portion of RBs (or certain RBs) may be sufficient to disrupt services or at least significantly degrade the service qualities. Worse yet, its difficulty to detect enables the attacker to launch the attack persistently.

2) Potential Countermeasures: In this section, we present potential countermeasures to the aforementioned attacks. The effects of PUCCH jamming can be mitigated in a couple of ways. First, spreading PUCCH to a wider bandwidth would make it more difficult for an attacker to jam the PUCCH. Second, scheduling PUCCH along with PUSCH would randomize the PUCCH region, thereby disallowing narrow band PUCCH jamming. In fact, 3GPP Release 9 introduced simultaneous PUCCH and PUSCH transmission by multiplexing PUCCH with data on PUSCH. However, sending PUCCH over PUSCH is not enough. As already mentioned, a UE is allocated with RBs on PUSCH in PDCCH that can be decoded by anyone in the same cell including the attacker. Therefore, to randomize PUCCH effectively, it is necessary to hide resource allocation to UEs at least related to the PUCCH. Such information hiding can be achieved with PDCCH encryption. One way to achieving it is to encrypt each DCI using a shared key between a base station and a UE. Alternatively, the whole PDCCH may be encrypted using a group key shared between a base station and a set of UEs. The former is more secure but requires more processing since each DCI for a UE in PDCCH is encrypted using a separate key. Whereas, the latter is relatively less secure but requires less processing as the entire PDCCH is encrypted using a single key. At least, the group-key-based encryption can prevent an unauthorized attacker (e.g., who has not authenticated to the network and hence does not have a secure connection with the base station) from reading the PDCCH and may sufficiently isolate the attacker(s) with strict group membership management. For example, if jamming protection is intended only for a certain set of UEs (e.g., associated with certain critical services), a base station provisions those UEs with a group key with which the PDCCH is encrypted. Of course, the PDCCH encryption, by hiding the resource scheduling information, would also offer resistance against data channel jamming attacks aiming at selective targets.

In summary, tactile internet requires a high level of safety and privacy. The off-the-shelf 4G LTE security design may not fulfill this requirement, while 5G technology has great potentials to enhance its security from both PHY/MAC layer and upper layer. The aforementioned 5G security discussion can be viewed as the first step to address the security issues of tactile internet. Based on this baseline security design, more sophisticated novel methods of security can be developed, including even the usage of hardware specific attributes such as biometric fingerprints [3].

V. CASE STUDY OF TACTILE INTERNET: FACTORY AUTOMATION

The tactile internet will have profound impact on automation in industrial applications like smart factories. In this section, we study a use case of 5G-enabled factory automation (FA) to show the potentials and identify the challenges to establish the 5G-based tactile internet in industrial automation.

Typically, factory automation requires closed-loop control applications between 0.25 and 50 ms and demands extremely low PER $\leq 10^{-9}$ [16]. Closed-loop control applications on the field level are executed by machines, each of which consists of a master control unit and a set of slaves. Thereby, a machine is typically built up of locally fixed slaves distributed around some tens of meters from the master, e.g., electric drives in a printing press or most numerical control applications of machine tools [18]. Traditionally, these applications have been handled by wireline bus systems [17] which constantly suffer from high installation and replacement cost of wiring as well as maintenance-prone support for moving sensors/actuators (S/As). In the recent years, a paradigm shift has set in to migrate FA communications from wireline to wireless, a fact that significantly drives the establishment of 5G-enabled tactile internet in factories.

In what follows, we present a 5G-based PHY/MAC design framework tailored to an industrial closed-loop

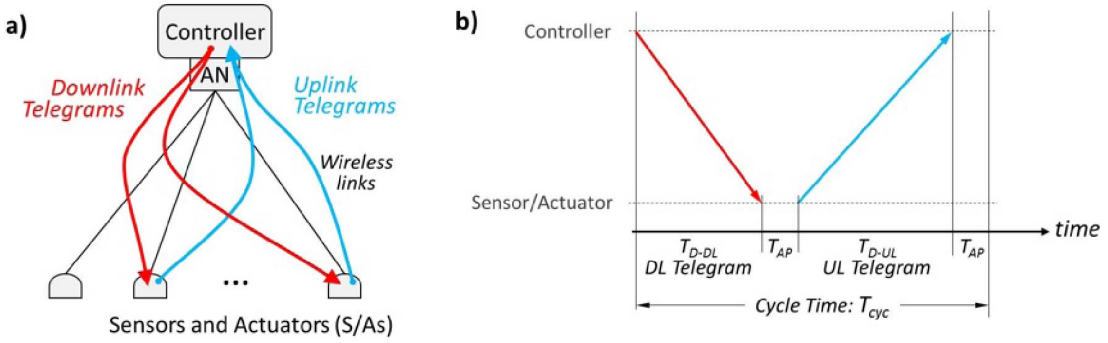

Fig. 9. *FA application using wireless URLLC. (a) Topology: controller exchanges downlink and uplink telegrams with 50 sensors/actuators. The access node (AN) is assumed to be collocated with the controller. (b) Application-layer time constraints per cycle (T_{AP} refers to application-layer processing time) [35].*

control application, printing machine, which today almost fully exploit the system capacity of Fast-Ethernet-based wired field-bus systems and can be regarded as a representative NR URLLC candidate in the FA space. In particular, as shown in Fig. 9, one printing machine is assumed to support 50 printing heads, which move over a distance of a few meters at moderate speeds and exchange telegrams with a controller. Within a cycle of 2 ms, each printing head receives one DL telegram within T_{D-DL} ms and replies with an UL telegram within T_{D-UL} ms, where each telegram carries a payload of 30 bytes. Typically, this DL-UL information exchange is periodic (e.g., one closed-loop transmission per link between controller and S/A in 2 ms). The tolerable PER per telegram is assumed to be 10^{-9}.

A. Fifth-Generation-Enabled PHY/MAC Design for Printing Machine

First, we claim that our following baseline design will be specific to TDD systems as it has been a common technique employed in unlicensed microwave transmission bands, such as 2.4 and 5.8 GHz. As discussed in the preceding sections for mobile broadband, (asynchronous) HARQ technique or its variants is of importance to achieve high reliability. Given the strict latency bound in printing machine applications, however, the HARQ turn-around delay could be substantially increasing the packet failure probability due to the packet-wise deadline violation. Thus motivated, our baseline design assumes that both UL and DL telegrams use the entire available time budget for data transmissions. Moreover, as a fairly starting point, one can assume the same payload, constellation, bandwidth, and coding rate for both UL and DL data transmissions, leading to 1-ms telegram for each direction by equally dividing the 2-ms control cycle.

Since NR URLLC allows configurable numerology, we choose OFDM-subcarrier spacing of 30 kHz. This indicates that the 1-ms time slot accommodates 28 OFDM symbols each with normal cyclic prefix; 27 of them are allotted to the payload data while the last symbol remains idle for MAC, and application layer processing. If assuming the machine-wise bandwidth of 20 MHz, then each link (controller to/from one sensor/actuator) requires 12 subcarriers (or resource element) per OFDM symbol to achieve a selected coding rate around 1:2. The detailed resource allocation is shown in Table 7.

Since the traffic of printing machine is strictly periodic with fixed payload size, our baseline design is assumed to be grant free (i.e., no control channels) to save resources for improving the reliability of data channel. In other words, when operating at high duty cycle, in our design, the MAC layer applies a semi-static resource allocation scheme for all links. That is, the machine-wise bandwidth demand is consistently reserved for 18 MHz without guard band, or roughly 20 MHz, when guard band is included. In practice, factory halls could consist of tens or hundreds of such machines, implying that multiple printing machine systems are operated in close vicinity. In such an environment, a frequency reuse scheme needs to be introduced. For a reuse factor of 1:4, which closely matches that of wireless sensor–actuator network (WSAN)-FA,[10] the

[10]The industrial, scientific, and medical radio band (ISM band) of 80 MHz is used for the existing wireless interface for sensors and actuators (WISA). In particular, the 80 MHz is partitioned into seven disjoint subbands, each containing 11 hop frequencies with channel spacing of 1 MHz [36]. Then, WISA employs frequency hopping over these hop frequencies.

Table 7 Resource Analysis for the Baseline Printing Machine PHY/MAC Design [35]

UL-and DL slot duration	1ms
Subcarrier spacing	30kHz
Number of OFDM symbols per slot	28
Number of OFDM symbols per slot for data	27
Resource elements per OFDM symbol per link	12
Resource elements per link	324
Resource elements for channel estimation per link	30
Resource elements for data per link	294
QPSK-coded data bits per link	588
Application bits per link including CRC	240
Resulting code rate	240:588
Bandwidth per link	360kHz
Bandwidth for 50 links	18MHz

factory-wide bandwidth demand increases to 80 MHz with guard band. It is noteworthy that this semi-static resource allocation scheme overcomes the bottleneck of the over-the-air scheduling signals in the 5G-based tactile internet, which itself requires even higher reliability than data channels to ensure the success of scheduling grant reception.

In the setting of one printing machine supporting 50 S/A devices in 20-MHz bandwidth, simulations indicate that this baseline design (with the selected code rate) requires a high degree of diversity (around 8–16 branches) to achieve PER at a level of 10^{-9} with reasonable SINR values (around 12–15 dB), a result that is aligned with those reported in [6] and [16]. Such degree of diversity can be achieved by combining spatial dimensions, e.g., a multiantenna MIMO system, with frequency dimensions, e.g., transmissions over an extended bandwidth or frequency hopping. Very recently, a bluetooth-based wireless solution, termed ENCLOSE, has been reported in [37]. In the setting of one access point supporting up to 120 devices in 80-MHz bandwidth, it is reported that ENCLOSE can provide a packet delivery ratio of 100% for an SNR threshold of up to 25 dB. Instead of leveraging the 5G framework, ENCLOSE adopts IEEE 802.15.1 (bluetooth) specifications at the physical layer, and optimizes time-division multiple access (TDMA), FDD, and frequency hopping at the MAC layer. Moreover, ENCLOSE uses cooperative multiuser diversity by requiring certain devices to act as relay for devices requiring retransmissions in uplink and/or downlink.

B. Advanced Design Challenges and Opportunities

The aforementioned PHY/MAC baseline design for the 5G-based tactile internet in FA is fairly simple, yet sufficient to the real-world implementation in factories. In this section, we present two major issues on implementation and discuss the possible solutions/opportunities to address these issues. More detailed discussions and simulation results can be found in [35].

1) Single-Band Dilemma: As analyzed above, for a reuse factor of 1:4, one needs 80-MHz system bandwidth with high degree of diversity to fulfill the stringent reliability and latency requirements. If using licensed band, the associated cost could be significant, undoubtedly resulting in numerous obstacles to the successful commercialization of this 5G-based tactile internet in factories. If using unlicensed band, however, the installation must be compliant with regulatory requirements. In other words, the transmitter must perform a clear channel assessment (CCA) prior to transmission and back off, when the channel is found busy. CCA can be conducted on the first OFDM symbol of each slot, which slightly reduces the number of symbols available for data. To avoid latency from back-off, the factory needs to explicitly ensure that the unlicensed band is not used by other unlicensed-band technologies. Even in a restricted factory environment, this is still almost impossible to guarantee.

One solution could be the opportunistic dual-band design. That is, the transmission switches to licensed band in case the interference or resource contention jeopardizes reliable transmission in unlicensed band. To minimize licensed-band usage, band selection is conducted for each telegram. For instance, if interference occurs with a detected power level above a preconfigured threshold, the transmitter (controller or actuator/sensor) will switch to licensed band for the transmission of this telegram. However, there exist a few shortcomings that need to handle in our design as follows.

- CCA conducted by the transmitter may miss hidden nodes, whose interference is only visible to the receiver. Even if the receiver conducts CCA and reliably and instantaneously conveys the outcome to the transmitter, the telegram still remains vulnerable to non-yielding interferers that begin their transmission after CCA.
- The interference power threshold must be set very carefully. If it is low, i.e., a few decibels above noise, to achieve the stringent PER target, the licensed band cost could be substantially high. On the other hand, if the threshold is high, e.g., 20 dB above noise as used in WiFi, the behavior could be like non-yielding interferes and thus the reliability is not guaranteed.
- Before switching to the licensed band, the transmitter (controller or actuator/sensor) needs to send a silencing signal to suspend the ongoing licensed band traffic. Otherwise, such transmission will collide with the ongoing NR or LTE traffic on licensed band. However, the signaling design on the silencing signal is a challenging topic. This silencing signal needs to be highly reliable and instantaneous. Roughly speaking, the reliability requirement can be largely achieved by boosting the transmission power and a direct silencing mechanism (rather than a routed silencing via base station) to the licensed-band users is strongly desired.

2) Shadow Fades: Given the moderate mobility of printing heads, shadow fades may occur due to either the bulk printing arms or the presence of unpredictable obstacles. As a result, the required high degree of diversity could not be achieved from only MIMO technique and/or multifrequency dimensions. In this case, one has to secure coordinated multipath connections to a printing head, or equivalently, use the multipath spatial diversity. For instance, each printing head can be associated with at least two geographically separated controllers, where the controller are connected via Ethernet-based backhaul which allows almost instantaneous information exchange. By doing so, one has to consider at least the following.

- The optimization of controller deployment needs to be well studied. The solution will definitely depend on the real-world applications and layouts of factories. Furthermore, the optimized deployment could be static with one-time installation or dynamic based on the online optimization updates (in smart factories).

- The mechanism of association with multiple controllers has to carefully deal with the possibly increased intercell interference. This problem is tightly coupled with the deployment optimization. Namely, the mitigation of intercell interference with multicontroller association could be one necessary constraint in the controller deployment optimization.

VI. CONCLUSION

Motivated by the numerous applications in the fields such as automation, gaming, education, healthcare, and smart home/city, the interest in tactile internet has been rapidly growing over the recent years. Coincidentally, the emerging 5G and the evolving LTE technologies have been attempting to underpin the ultrareliable and low-latency applications, which can be viewed as promising wireless solutions making the real-world implementation of tactile internet come true.

In this paper, we have presented the up-to-date PHY/MAC design for NR/LTE URLLC in 5G technologies. To sum up, in NR URLLC, the scalable numerology (e.g., subcarrier spacing, TTI, RTT) enables NR to have the capability of offering services with different reliability and latency requirements. In particular, shortening TTI and HARQ RTT can significantly improve the system capacity under URLLC requirements. Furthermore, wideband allocation for URLLC, and thus the dynamic multiplexing of URLLC and other traffics are highly desirable as the URLLC system capacity is super-linearly increasing with respect to the available bandwidth. We have provided the theoretical queueing analysis and system-level simulations to support these systems design choices which are being contributed into 3GPP standards. On the other hand, in LTE technology, the URLLC design has to follow the same numerology as legacy LTE due to the consideration of backward compatibility. Similar to NR URLLC design, shortening TTI by reducing the number of OFDM symbols in one TTI also substantially improves the system capacity. In addition, enhancements on the current LTE control and data channels can facilitate LTE to fulfill the URLLC requirement. To address the security concerns on tactile internet, we have also discussed the 5G security landscape and PHY/MAC design to open up several areas for future research. In the end, we have investigated a particular application of tactile internet, factory automation, in which we have provided a 5G-based baseline systems design and discussed the challenges and opportunities down the road. ■

APPENDIX

SYSTEM-LEVEL SIMULATION ASSUMPTIONS FOR TABLES 1–3

To find the URLLC capacity in Tables 1–3, we need to sweep over the Poisson arrival rate, where the capacity is defined as the maximal Poisson arrival rate such that the reliability and latency requirements of all UEs are satisfied.

Layout	single macro layer; hexagonal grid with 21 cells wrapped around
Inter-gNB distance	500m
Carrier frequency	2GHz
System bandwidth	$\{5, 10, 20\}$MHz
Channel model	3D UMa
Transmission power	gNB: 49dBm power amplifier scaled with bandwidth; UE: 23dBm
Antenna configuration	2Tx/2Rx (X-pol)
gNB antenna height	35m
gNB antenna gain and connector loss	8dBi
gNB/UE receiver noise figure	5/9dB
Traffic model	eMBB: full-buffer; URLLC: Poisson with 32-byte packets.
UE distribution	URLLC: 22 UEs in the serving cell with 50% indoor and 50% outdoor. eMBB: one UE in each neighboring cell.
Subcarrier spacing	60KHz
Cyclic prefix duration	normal cyclic prefix
TTI/RTT duration	2/6 OFDM symbols
System reliability	99.99% or 99.999%
Hard latency	$\{500, 750, 1000\}\mu$s
MIMO	2×2 single user MIMO

REFERENCES

[1] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[2] *The Tactile Internet, ITU-T Technology Watch Report*, ITU-T, Geneva, Switzerland, 2014.

[3] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[4] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile Internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 82–89, Apr. 2017.

[5] K. M. Stanney, R. S. Kennedy, and K. Kingdon, *Virtual Environment Usage Protocols*. Boca Raton, FL, USA: CRC Press, 2002.

[6] O. N. C. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1190–1195.

[7] T. H. Szymanski, "Securing the industrial-tactile Internet of Things with deterministic silicon photonics switches," *IEEE Access*, vol. 4, pp. 8236–8249, 2016.

[8] J. G. Andrews, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[9] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.

[10] *Technical Specification Group Services and System Aspects: Service Requirements for the 5G System*, document TS 22.261, 3GPP, 2017.

[11] E. Dahlman *et al.*, "5G wireless access: Requirements and realization," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, Dec. 014.

[12] *Study on New Radio Access Technology Physical Layer Aspects*, document 38.802, 3GPP.

[13] *Study on Scenarios and Requirements for Next Generation Access Technologies*, document 38.913, 3GPP.

[14] K. Hosseini, S. Patel, A. Damnjanovic, W. Chen, and J. Montojo, "Link-level analysis of low latency operation in LTE networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[15] A. Damnjanovic, W. Chen, S. Patel, Y. Xue, K. Hosseini, and J. Montojo, "Techniques for enabling low latency operation in LTE networks," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2016, pp. 1–7.

[16] B. Holfeld *et al.*, "Wireless communication for factory automation: An opportunity for LTE and 5G systems," *IEEE Commun. Mag.*, vol. 54, no. 6, pp.

36–43, Jun. 2016.
[17] R. Zurawski, *Industrial Communication Technology Handbook*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2015.
[18] A. Frotzscher *et al.*, "Requirements and current solutions of wireless communication in industrial automation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2014, pp. 67–72.
[19] *3GPP System Architecture Evolution (SAE); Security Architecture*, document 3GPP TS 33.401.
[20] *General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access*, document 3GPP TS 23.401.
[21] *Battery Efficient Security for Very Low Throughput Machine Type Communication (MTC) Devices (BEST)*, document 3GPP TS 33.163.
[22] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description*, document 3GPP TS 36.300.
[23] *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*, document 3GPP TS 36.331.
[24] *Technical Specification Group Services and System Aspects; Security Architecture and Procedures for 5G System (Release 15)*, document 3GPP TS 33.501.
[25] A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, and J.-P. Seifert, "Practical attacks against privacy and availability in 4G/LTE mobile communication systems," in *Proc. 23rd Annu. Netw. Distrib. Syst. Secur. Symp.*, Feb. 2016.
[26] [Online]. Available: https://www.theguardian.com/us-news/2015/feb/19/nsa-gchq-sim-card-billions-cellphones-hacking
[27] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–6.
[28] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2 (Release 13)*, document 3GPP TS36.300.
[29] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2 (Release 14)*, document 3GPP TS36.300.
[30] *Evolved Universal Terrestrial Radio Access (E-UTRA) Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, document 3GPP TR36.814.
[31] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.
[32] *Updated Sub6 DL Full-Buffer KPI Evaluation for eMBB*, document R1-166390, Qualcomm Incorporated, 3GPP RAN1 Meeting #86, Gothenburg, Sweden, 2016.
[33] *UL URLLC Capacity Study and URLLC eMBB Dynamic Multiplexing Design Principle*, document R1-1705624, Qualcomm Incorporated, 3GPP RAN1 meeting #88bis, Spokane, WA, USA, 2017.
[34] J. Rao and S. Vrzic, "Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis," *IEEE Netw.*, vol. 32, no. 2, pp. 32–40, Mar./Apr. 2017.
[35] G. Hampel, C. Li, and J. Li, "5G ultra-reliable low-latency communications in factory automation leveraging licensed and unlicensed bands," *IEEE Commun. Mag.*, to be published.
[36] G. Scheible, D. Dzung, J. Endresen, and J. E. Frey, "Unplugged but connected: Design and implementation of a truly wireless real-time sensor/actuator interface," *IEEE Ind. Electron. Mag.*, vol. 1, no. 2, pp. 25–34, Jul. 2017.
[37] A. Aijaz, "ENCLOSE: An enhanced wireless interface for communication in factory automation networks," *IEEE Trans. Ind. Informat.*, to be published.

ABOUT THE AUTHORS

Chong Li (Member, IEEE) received the B.E. degree in electrical engineering from Harbin Institute of Technology, Harbin, China and the Ph.D. degree in electrical engineering from Iowa State University, Ames, IA, USA.

He was a Staff Engineer at the Corporate R&D Department, Qualcomm Technologies, Inc., Bridgewater, NJ, USA. He is a co-founder of Nakamoto & Turing Labs, Inc. and Adjunct Assistant Professor at Columbia University, New York, NY, USA. He holds over 200 U.S. patents (granted and pending). His research interest includes information theory, machine learning, networked control and communications, and PHY/MAC design for advanced communication systems.

Dr. Li serves as a reviewer and technical program committee member for most prestigious journals and conferences in communications and control society. He received the MediaTek Inc. and Wu Ta You Scholar Award, the Rosenfeld International Scholarship, and the Iowa State Research Excellent Award.

Chih-Ping Li received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, CA, USA.

He is currently a Senior Systems Engineer at the Corporate R&D Department, Qualcomm Technologies, Inc., San Diego, CA, USA, where he has been working on the system design, analysis, and evaluation of ultrareliable and low-latency communications in the fifth-generation wireless networks since 2014. From 2011 to 2014, he was a Postdoctoral Associate at the Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, where he conducted research on stochastic optimization and dynamic control in data and wireless networks.

Kianoush Hosseini received the B.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2008, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 2010 and 2015, respectively.

Since 2015, he has been with Qualcomm Technologies Inc., San Diego, CA, USA, where he is working on the design and standardization of LTE and 5G NR ultrareliable and low-latency communication networks.

Dr. Hosseini is the recipient of the Edward S. Rogers Sr. Graduate Scholarship in 2008 and 2010, the Queen Elizabeth II Graduate Scholarship in Science and Technology in 2015, and the Shahid U.H. Qureshi Memorial Scholarship in 2015.

Soo Bum Lee received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, South Korea, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA.

He was a Senior Research Engineer at LG Industrial Systems from 1997 to 2003, where he developed distributed control systems (DCS). He is a Systems Security Architect at Qualcomm Technology, Inc., San Diego, CA, USA, where he has been working on 5G and WiFi security since 2013. Prior to joining Qualcomm, he was a Research Systems Scientist at CyLab, Carnegie Mellon University, Pittsburgh, PA, USA. His research focused on next-generation secure and available networks and network distributed denial-of-service (DDoS) defense mechanisms.

Jing Jiang received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2007.

Since 2007, he has been with Qualcomm Inc., San Diego, CA, USA, working on wireless systems, focusing on 4G LTE and 5G NR research and product developments.

Dr. Jiang is the recipient of the 2006 Best Paper Award from the IEEE Signal Processing for Data Storage society for the paper "Iterative soft-input–soft-output decoding of Reed–Solomon codes."

Wanshi Chen received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA.

He is currently 3GPP TSG RAN1 Chairman, where under this position, he has successfully managed a wide range of 3GPP TRG RAN1 long-term evolution (LTE) and new radio (NR) sessions. He has over 17 years of experiences in telecommunications in leading telecom companies including operators, infrastructure vendors, and chipset vendors. He has been with Qualcomm Inc. San Diego, CA, USA, since 2006 and is responsible for LTE and NR research, design, and standardization. From 2000 to 2006, he was with Ericsson for 3GPP2 related system design, integration, and standardization. He also worked for China Mobile between 1996 and 1997 for wireless network maintenance and optimization.

Gavin Horn received the Ph.D. degree in electrical engineering from California Institute of Technology, Pasadena, CA, USA.

He is a Principal Engineer in Corporate Research at Qualcomm Inc., San Diego, CA, USA, leading the network architecture and security design for 5G. He has over 15 years' experience in communications and holds over 100 patents related to cellular technologies. He has contributed to many areas of 4G design including the femto cells, relay, WLAN internetworking, and traffic offload features. Prior to Qualcomm, he worked in video compression and readable content delivery.

Tingfang Ji received the B.Sc. degree from Tsinghua University, Beijing, China, the M.S. degree from the University of Toledo, Toledo, OH, USA, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2001.

He joined Qualcomm Inc. San Diego, CA, USA, in 2003, where he is currently a Senior Director of Engineering. He has been involved in system design of multiple advanced wireless systems including IEEE 802.20, 4G UMB, 4G LTE, and 5G NR. From 2008 to 2014, he took an active role in standardizing LTE-Advanced in 3GPP and served as a Vice Chairman of the Radio Working Group (RAN4). Since 2014, he has been leading the system design of 5G NR sub-6-GHz radio access technology, with a focus on key NR concepts such as low-latency frame structure, scalable numerology, UL-based mobility, and wide-area wideband TDD massive MIMO. His team also made instrumental contribution to an industry-first sub-6-GHz 5G NR compliant live demo in MWC 2017. Before joining Qualcomm, he was a member of the technical staff at Bell Labs. He holds more than 200 granted U.S. patents.

John E. Smee (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Queen's University, Kingston, ON, Canada, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA.

He is a Vice President of Engineering at Qualcomm Technologies Inc., San Diego, CA, USA. He joined Qualcomm in 2000, holds 63 U.S. patents, and has been involved in the system design for a variety of projects focused on the innovation of wireless communications systems such as CDMA EVDO, IEEE 802.11, 4G LTE, and 5G. His work involves taking advanced system designs and signal processing techniques from theory through design, standardization, implementation, and productization. He is currently a 5G project engineering lead in Qualcomm research and development group.

Dr. Smee was chosen to participate in the National Academy of Engineering Frontiers of Engineering program and is a recipient of the Qualcomm Distinguished Contributor Award for Project Leadership.

Junyi Li (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, and an MBA from the Wharton School at University of Pennsylvania, Philadelphia, PA, USA.

He is a Vice President of Engineering at Qualcomm Inc., Bridgewater, NY, USA. He was a key inventor of Flash-OFDM, arguably the first commercially deployed OFDMA-based mobile broadband wireless communications system. He holds over 500 U.S. patents and has more than 600 pending patent applications. He was a founding member of Flarion Technologies, a startup acquired by Qualcomm in 2006. Prior to that, he was with Bell Labs Research in Lucent Technologies. He is a coauthor of the book *OFDMA Mobile Broadband Communications* (Cambridge, U.K.: Cambridge Univ. Press, 2013).

Dr. Li received the Outstanding Electrical and Computer Engineers award from Purdue University in 2012.