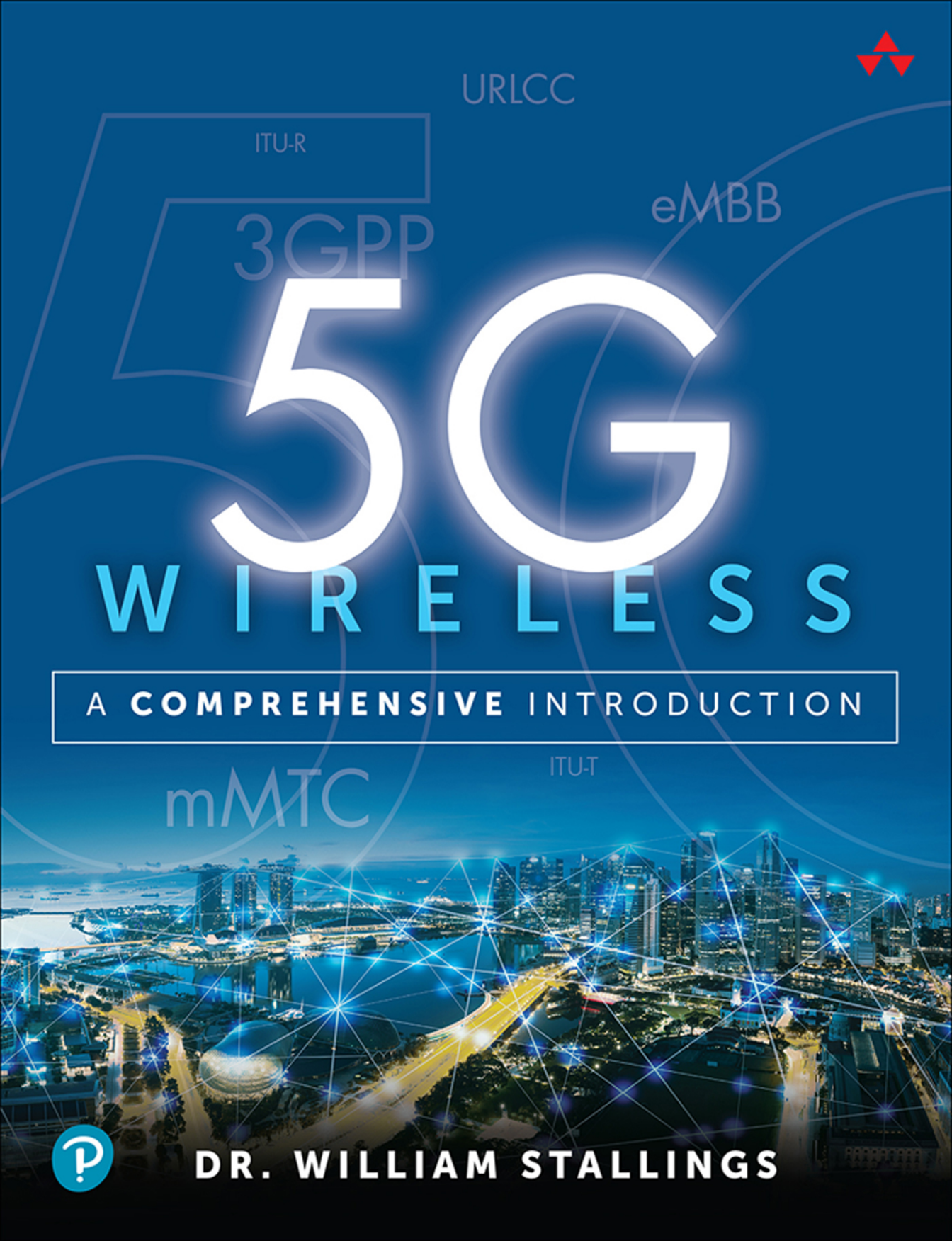
URLCC
ITU-R
eMBB
3GPP
5G
WIRELESS
A COMPREHENSIVE INTRODUCTION
ITU-T
mMTC
DR. WILLIAM STALLINGS

5G Wireless

A Comprehensive Introduction

Dr. William Stallings

Addison-Wesley

Boston • Columbus • New York • San Francisco • Amsterdam • Cape Town
Dubai • London • Madrid • Milan • Munich • Paris • Montreal • Toronto • Delhi • Mexico City
São Paulo • Sydney • Hong Kong • Seoul • Singapore • Taipei • Tokyo

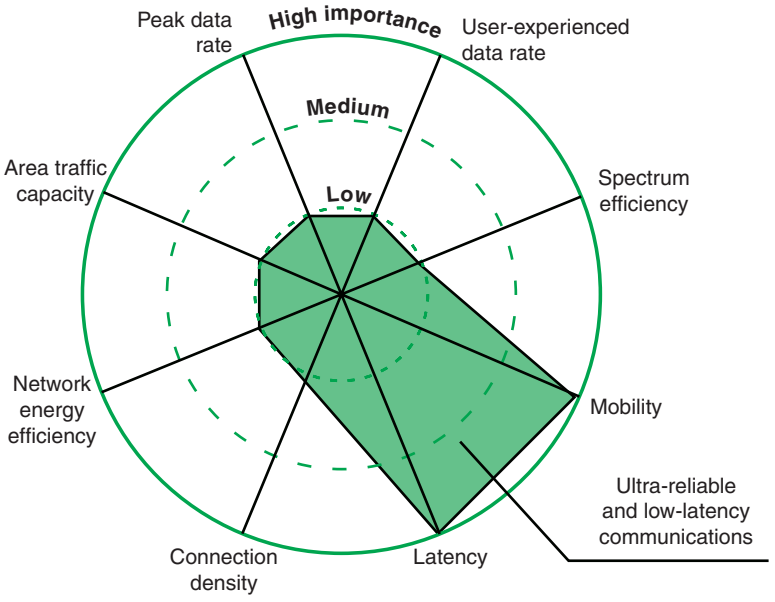

FIGURE 6.1 The Relative Importance of Key Capabilities in the URLLC Usage Scenario

Latency

ITU-R Report M.2410 breaks the latency requirement into two parts:

- **User plane latency:** This is the contribution by the radio network to the time from when the source sends a packet to when the destination receives it (in ms). It is defined as the one-way time it takes to successfully deliver an application layer packet/message from the radio protocol Layer 2/3 service data unit (SDU)[4] ingress point to the radio protocol Layer 2/3 SDU egress point of the radio interface in either uplink or downlink in the network for a given service in unloaded conditions, assuming that the mobile station is in the active state. The minimum requirement (i.e., the maximum allowable value) is 1 ms, assuming unloaded conditions (i.e., a single user) for small IP packets (e.g., 0 byte payload + IP header), for both downlink and uplink.
- **Control plane latency:** This refers to the transition time from a most battery-efficient state (e.g., Idle state) to the start of continuous data transfer (e.g., Active state). The minimum requirement is 20 ms.

4. In a packet, the SDU is data that the protocol transfers between peer protocol entities on behalf of the users of that layer's services. For lower layers, the layer's users are peer protocol entities at a higher layer; for the application layer, the users are application entities outside the scope of the protocol layer model.

User plane latency, however, is only one component that a UE experiences, as illustrated in Figure 6.2. The end-to-end (E2E) latency is generally defined as the time it takes from when a data packet is sent from the transmitting end to when it is received at the receiving entity (e.g., Internet server or other device). The measurement reference is the interface between Layers 2 and 3. It is also referred to as one-trip time (OTT). It includes the user plane latency in one direction, transport network delays, and application processing time. A related measure is round-trip time (RTT), which is the time from when a data packet is sent from a source device until an acknowledgement or response is received from the destination device. Unfortunately, E2E latency is sometimes equated to RTT latency in the literature, even in some 3GPP documents. However, the implication in most standards and specification documents is that E2E latency refers to one-way latency, not round-trip latency.

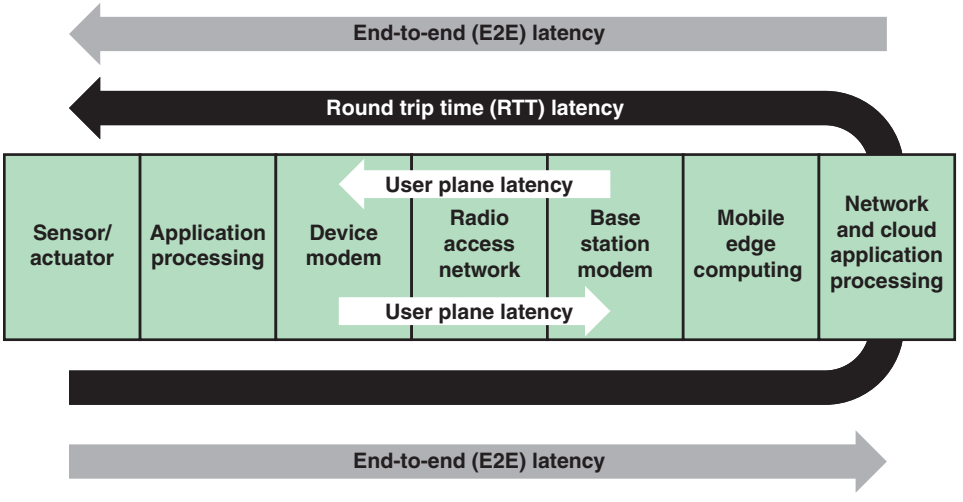

FIGURE 6.2 End-to-End Latency and Round-Trip Time Latency

To reduce the other components of E2E latency besides user plane latency, carriers are moving increasingly to an MEC strategy. Chapter 10, "Multi-Access Edge Computing," explores MEC in detail.

Mobility

Mobility is the maximum UE speed (in km/h) at which a defined quality of service (QoS) can be achieved. Mobility assumes that a seamless transfer between radio nodes—which may belong to different layers and/or radio access technologies (multilayer/multi-RAT)—can be achieved without dropping QoS below a defined threshold. The following classes of mobility are defined:

- **Stationary:** 0 km/h
- **Pedestrian:** 0–10 km/h

- **Vehicular:** 10–120 km/h
- **High-speed vehicular:** 120–500 km/h

M.2410 does not provide a specific measure of QoS. Report ITU-R M.2412 (*Guidelines for Evaluation of Radio Interface Technologies for IMT-2020*, October 2017) defines QoS as successful delivery of 99% of messages within 10 s.

Another aspect of mobility addressed in M.2410 is **mobility interruption time**, which is the shortest time duration supported by the system during which a UE cannot exchange user plane packets with any base station during transitions. This includes the time required to execute any radio access network procedure, radio resource control signaling protocol, or other message exchanges between the mobile station and the radio access network.

The minimum requirement for mobility interruption time is 0 ms. Thus, there should be no interruption of service when a moving UE switches from one base station to another.

Reliability

Reliability, though not mentioned in M.2083, is another vital parameter for URLCC and is included in M.2410. **Reliability** is defined as the probability of successful transmission of a Layer 2/3 packet within a required maximum time, which is the time it takes to deliver a small data packet from the radio protocol Layer 2/3 service data unit (SDU) ingress point to the radio protocol Layer 2/3 SDU egress point of the radio interface at a certain channel quality. The minimum requirement is $1–10^{–5}$ success probability of transmitting a Layer 2 protocol data unit (PDU) of 32 bytes within 1 ms for the urban macro-URLLC test environment.

NGMN Definitions

The 2016 NGMN white paper breaks the URLLC requirements into three broad use case families:

- Ultra-low latency
- Ultra-high reliability and ultra-low latency
- Ultra-high availability and reliability

Table 6.1 shows the performance requirements for these three families.

TABLE 6.1 Performance Requirements for Different Varieties of URLLC

Key Performance Indicator (KPI)	Ultra-Low Latency	Ultra-High Reliability and Ultra-Low Latency	Ultra-High Availability and Reliability
User-experienced data rate	DL: 50 Mbps UL: 25 Mbps	DL: 50 kbps–10 Mbps UL: a few bps–10 Mbps	DL: 10 Mbps UL: 10 Mbps
E2E latency	< 1 ms	1 ms	10 ms

Key Performance Indicator (KPI)	Ultra-Low Latency	Ultra-High Reliability and Ultra-Low Latency	Ultra-High Availability and Reliability
Mobility	Pedestrian	On demand, 0-500 km/h	On demand, 0–500 km/h
Device autonomy	> 3 days	Not critical	> 3 days (standard) Up to several years for some critical MTC services
Connection density	Not critical	Not critical	Not critical
Traffic density	Potentially high	Potentially high	Potentially high

6.2 URLLC Use Cases in Emerging Mission-Critical Applications

A URLLC white paper from 5G Americas (*New Services & Applications with 5G Ultra-Reliable Low-Latency Communications*, November 2018) provides a useful way of understanding the wide variety of URLLC use cases, by focusing on emerging mission-critical applications that have demanding reliability and latency requirements. These are described in this section.

Industrial Automation

The area that has perhaps received the most attention as an application area that requires URLLC support is the smart factory or industrial automation. This application area is typified by extremely demanding KPIs for 5G communication links between sensors, actuators, and controllers. Section 6.4 examines this area in more detail.

Ground Vehicles, Drones, and Robots

This application area refers to remotely controlled mobile devices and robots. Such devices are in common use in factory applications and are also deployed in other contexts, such as smart agriculture. One area of particular interest is unmanned aircraft traffic management; this topic is examined in Section 6.5.

Tactile Interaction

Tactile interaction refers to a level of responsiveness that works at a human scale. For example, remote healthcare or gaming applications may require very low round-trip times to convince human senses that the perceived touch, sight, and sound are lifelike.

These use cases involve interaction between humans and systems, where humans wirelessly control real and virtual objects, and the interaction requires a tactile control signal with audio or visual feedback. Robotic controls and interaction include several scenarios with many applications in

manufacturing, remote medical care, and autonomous cars. The tactile interaction requires real-time reactions on the order of a few milliseconds. Remote surgery, discussed later in this chapter, is perhaps the most demanding use case.

Table 6.2 gives typical values of KPIs for tactile Internet applications.

TABLE 6.2 Key Performance Indicators for Tactile Internet Applications

KPI	Value
Traffic volume density	0.03–1 Mbps/m^2 (cell radius 100 m^2)
Experienced user throughput	UL: 0.3–1 Mbps
Latency	User plane latency less than 2 ms
Availability	> 99.999%
Reliability	> 99.999 % for healthcare or remote driving/manipulation 95% for remote gaming or remote augmented reality

Figure 6.3, from an *ITU Technology Watch Report* [ITU14], illustrates a typical latency budget.

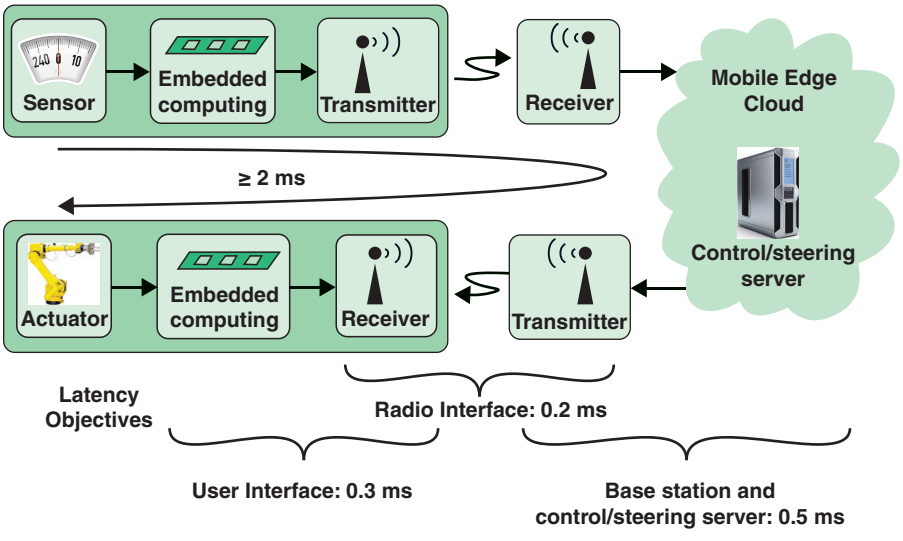

FIGURE 6.3 Example of a Latency Budget of a System for the Tactile Internet

Augmented Reality and Virtual Reality

Augmented reality (AR) and virtual reality (VR) tend to have relatively high data rate requirements. Some specific use cases also have URLLC requirements. An NGMN paper (*Verticals URLLC Use Cases and Requirements*, July 2019) lists three AR/VR examples with URLLC requirements: augmented worker, 360 panoramic VR view video broadcasting, and AR and MR cloud gaming.

Augmented Worker

Augmented work is work that integrates digital technologies into the industrial environment to improve how work is done. Augmented work is appropriate for situations when it is not cost-effective or even possible to fully automate tasks, but it is desirable to augment the capabilities of the human worker. A good example is a task such as equipment repair where access is difficult (e.g., a hazardous environment) or where the expert is physically distant. In such a case, a remote worker can be equipped with an AR headset and some sort of tactile interface for remote control. Sensor information from the remote target location in the form of audio, video, and haptic (tactile) feedback enables the remote operator to control actuators at the target location to achieve the required work. Figure 6.4, from the NGMN paper *Verticals URLLC Use Cases*, illustrates this arrangement.

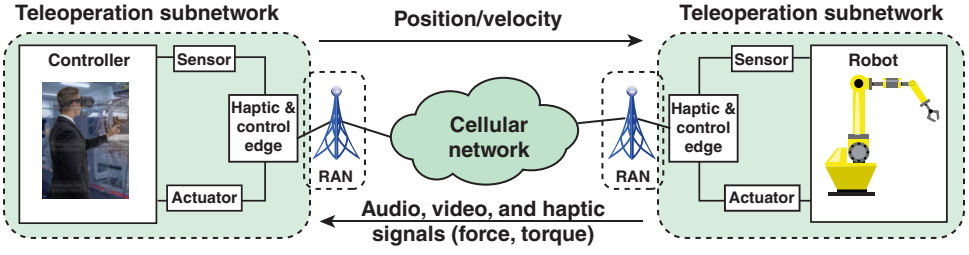

FIGURE 6.4 Augmented Worker

The NGMN document lists the following communications service requirements for this use case:

- **End-to-end latency:** 10 ms
- **End-to-end reliability:** 99.9999%
- **Positioning:** Indoor positioning service with horizontal positioning accuracy better than 1 m, 99% availability, heading < 10 degrees, and latency for positioning estimation < 15 ms for moving UE with speed up to 10 km/h
- **Other requirements:** Application-level requirements:
  - The AR application should have easy access to different context information (e.g., information about the environment, production machinery, the current link state).
  - The (bidirectional) video stream between the AR device and the image processing server should be encrypted and authenticated by the 5G system.
  - Real-time data processing is required.

5G network architecture requirements:

  - There is no need for dynamic scalability.
  - Mobility at standard values is needed.