

URLCC

ITU-R

eMBB

3GPP

5G

W I R E L E S S

mMTC

ITU-T



Contents

Preface xxiv

Part I Overview

Chapter 1: Cellular Networks: Concepts and Evolution 2

1.1 Evolution of Cellular Networks	2
1.2 Cellular Network Concepts	4
Cellular Organization	4
Network System Elements	10
Operation of Cellular Systems	11
1.3 First Generation (1G)	14
1.4 Second Generation (2G)	15
Time-Division Multiple Access	16
GSM Architecture	17
Radio Link Aspects	20
1.5 Third Generation (3G)	20
Code-Division Multiple Access	21
3G Architecture	21
1.6 Fourth Generation (4G)	24
E-UTRAN	26
LTE-Advanced Transmission Characteristics	29
Evolved Packet Core	33
1.7 Key Terms and Review Questions	34
Key Terms	34
Review Questions	35

Chapter 2: 5G Standards and Specifications 36

2.1 ITU-R and IMT-2020	37
International Mobile Telecommunications	38
Capabilities	38

Usage Scenarios	41
Use Cases	43
Evaluation	46
2.2 ITU-T and IMT-2020	57
Requirements	57
Network Slicing	59
Fixed Mobile Convergence	61
IMT-2020 Core Network Framework	62
2.3 3GPP	65
3GPP Releases	69
Detailed Requirements	70
2.4 Key Terms and Review Questions	72
Key Terms	72
Review Questions	73
2.5 References and Documents	74
References	74
Documents	74
Chapter 3: Overview of 5G Use Cases and Architecture	76
3.1 5G Use Cases	77
ITU-R	77
5G Americas	77
NGMA Alliance	79
3.2 NGMN 5G Architecture Framework	81
Layered Functionality	82
Network Slicing	85
3.3 3GPP 5G Architecture	87
5G Core Network Architecture	87
Radio Access Network Architecture	94
Session Establishment	97

3.4 Key Terms and Review Questions	98
Key Terms	98
Review Questions	99
3.5 References and Documents	99
References	99
Documents	100

Part II Use Cases and Applications

Chapter 4: Enhanced Mobile Broadband	102
4.1 eMBB Deployment Scenarios	103
Indoor Hotspot	104
Dense Urban	105
Rural	105
4.2 eMBB Performance Characteristics	105
Data Rate Requirements	106
Spectral Efficiency Requirements	107
Latency Requirements	108
Mobility Requirements	108
System Requirements	109
4.3 Smart Office: An Indoor Hotspot Use Case	110
4.4 Dense Urban Information Society: A Dense Urban Use Case	111
4.5 Radiocommunication Systems Between Train and Tracksides: A Rural eMBB Use Case	113
Elements of RSTT	114
Applications of RSTT	114
Broadband Connectivity for Passengers	116
Linear Cell Architecture	117

4.6	Key Terms and Review Questions	118
	Key Terms	118
	Review Questions	118
4.7	References and Documents	119
	References	119
	Documents	119
Chapter 5: Massive Machine Type Communications		122
5.1	mMTC Performance Requirements	123
5.2	The Internet of Things	124
	The Scope of the Internet of Things	125
	Things on the Internet of Things	128
	Components of IoT-Enabled Things	128
	Constrained Devices	129
	IoT and Cloud Context	130
5.3	Relationship Between mMTC and the IoT	133
5.4	Relationship Between mMTC and NB-IoT and eMTC	135
	Comparison of NB-IoT and eMTC	135
	Low-Power Wide Area (LPWA)	136
5.5	Smart Agriculture	138
	Model of IoT Deployment	138
	Use Cases	140
	Precision Crop Management	145
	Network Performance Requirements	150
5.6	Smart Cities	150
	Smart City Use Cases	151
	ICT Architecture for Smart Cities	158
	Network Performance Requirements	160

5.7	Key Terms and Review Questions	162
	Key Terms	162
	Review Questions	163
5.8	References and Documents	163
	References	163
	Documents	164
Chapter 6: Ultra-Reliable and Low-Latency Communications		166
6.1	URLLC Performance Requirements	167
	Latency	168
	Mobility	169
	Reliability	170
	NGMN Definitions	170
6.2	URLLC Use Cases in Emerging Mission-Critical Applications	171
	Industrial Automation	171
	Ground Vehicles, Drones, and Robots	171
	Tactile Interaction	171
	Augmented Reality and Virtual Reality	172
	Emergencies, Disasters, and Public Safety	176
	Urgent Healthcare	176
	Intelligent Transportation	177
6.3	URLLC Applications Based on Performance Requirements	178
	Higher Reliability and Lower Latency	179
	Higher Reliability, Higher Availability, and Lower Latency	180
	Very Low Latency	181
	Higher Accuracy Positioning	181
	Higher Availability	181
	Mission-Critical Services	182

6.4 Industry 4.0	182
Factory Automation Architecture	183
Application Areas	185
Use Cases	187
Performance Requirements	189
6.5 Unmanned Aircraft System Traffic Management	191
UTM Architecture	192
5G Performance Requirements for UTM	194
6.6 Key Terms and Review Questions	197
Key Terms	197
Review Questions	198
6.7 References and Documents	199
References	199
Documents	199

Part III 5G NextGen Core Network

Chapter 7: Software-Defined Networking	202
7.1 Evolving Network Requirements	203
7.2 The SDN Approach	204
Modern Network Requirements	204
SDN Architecture	204
Characteristics of Software-Defined Networking	207
7.3 SDN Data Plane	207
Data Plane Functions	207
Data Plane Protocols	209
7.4 OpenFlow	209
Flow Table Structure	211
Flow Table Pipeline	214
The Use of Multiple Tables	216

Group Table	217
OpenFlow Protocol	217
7.5 SDN Control Plane	218
Control Plane Functions	218
Southbound Interface	220
Northbound Interface	220
Eastbound Interface	223
Westbound Interface	224
7.6 SDN Application Plane	225
Application Plane Architecture	225
Northbound Interface	226
Network Services Abstraction Layer	226
Network Applications	226
User Interface	230
7.7 Key Terms and Review Questions	230
Key Terms	230
Review Questions	231
7.8 References and Documents	231
References	231
Documents	232
Chapter 8: Network Functions Virtualization	234
8.1 Background and Motivation for NFV	235
8.2 Virtual Machines and Containers	236
Hypervisor	236
Architectural Approaches	238
Container Virtualization	241
8.3 NFV Concepts	242
Simple Example of the Use of NFV	245
NFV Principles	246
High-Level NFV Framework	246

8.4	NFV Benefits and Requirements	248
	NFV Benefits	248
	NFV Requirements	249
8.5	NFV Reference Architecture	250
	NFV Management and Orchestration	251
	Reference Points	251
8.6	NFV Infrastructure	252
	Container Interface	253
	Deployment of NFVI Containers	256
	Logical Structure of NFVI Domains	257
	Compute Domain	258
	Hypervisor Domain	260
	Infrastructure Network Domain	261
8.7	Virtualized Network Functions	263
	VNF Interfaces	263
	VNFC-to-VNFC Communication	264
	VNF Scaling	266
8.8	SDN and NFV	267
8.9	Key Terms and Review Questions	269
	Key Terms	269
	Review Questions	270
8.10	References and Documents	271
	References	271
	Documents	271
Chapter 9: Core Network Functionality, QoS, and Network Slicing		272
9.1	Core Network Requirements	273
	Network Operational Requirements	273
	Basic Network Requirements	275

9.2 Core Network Functional Architecture	279
Tunneling	281
PDU Session Establishment	283
Policy Control Function	287
9.3 Quality of Service	291
QoS Capabilities	291
QoS Architectural Framework	294
QoS Classification, Marking, and Differentiation	297
3GPP QoS Architecture	298
QoS Parameters	299
QoS Characteristics	301
Standardized 5QI-to-QoS Characteristic Mapping	303
9.4 Network Slicing	307
Network Slicing Concepts	308
Requirements for Network Slicing	310
Identifying and Selecting a Network Slice	312
Functional Aspects of Network Slicing	313
Generic Slice Template	314
9.5 SDN and NFV Support for 5G	315
9.6 Key Terms and Review Questions	319
Key Terms	319
Review Questions	319
9.7 References and Documents	320
References	320
Documents	320
Chapter 10: Multi-Access Edge Computing	322
10.1 MEC and 5G	323
10.2 MEC Architectural Concepts	324

10.3	ETSI MEC Architecture	327
	Design Principles	328
	MEC System Reference Architecture	329
	Related Elements	332
10.4	MEC in NFV	333
	MEC Components Implemented as VNFs	333
	MEC Components Replaced by NFV Components	335
	MEC System-Level Components	335
10.5	MEC Support for Network Slicing	335
10.6	MEC Use Cases	336
	Consumer-Oriented Services	337
	Operator and Third-Party Services	342
	Network Performance and QoS Improvements	344
10.7	3GPP Architecture for Enabling Edge Applications	349
	EDGEAPP Functional Architecture	350
	Synergized Mobile Edge Cloud Architecture	352
10.8	Key Terms and Review Questions	353
	Key Terms	353
	Review Questions	353
10.9	References and Documents	354
	References	354
	Documents	355

Part IV 5G NR Air Interface and Radio Access Network

Chapter 11: Wireless Transmission	356
11.1 Channel Capacity	357
Nyquist Bandwidth	357
Shannon Capacity Formula	358

11.2	Line-of-Sight Transmission.....	361
	Refraction	362
	Optical and Radio Line of Sight.....	363
11.3	Line-of-Sight Transmission Impairments	364
	Attenuation	364
	Free Space Loss	365
	Path Loss Exponent in Practical Cellular Systems.....	368
	Noise	369
	The Expression E_b/N_0	372
	Atmospheric Absorption.....	374
	Multipath	375
	Refraction	376
11.4	Fading in the Mobile Environment.....	376
	Multipath Propagation	376
	The Effects of Multipath Propagation	377
	Types of Fading.....	378
	The Fading Channel	379
11.5	Millimeter Wave Transmission for 5G	381
	Propagation Impairments.....	381
	Implications for 5G	385
11.6	Key Terms and Review Questions.....	387
	Key Terms	387
	Review Questions	388
11.7	References.....	389
ANNEX 11A: Decibels and Signal Strength.....		389
Chapter 12: Antennas		394
12.1	Channel Correction Mechanisms.....	394
	Adaptive Equalization.....	395
	Diversity Techniques.....	396

Adaptive Modulation and Coding	398
Forward Error Correction	398
12.2 Introduction to Antennas	399
Radiation Patterns	399
Antenna Types.....	401
Antenna Gain.....	406
12.3 Multiple-Input/Multiple-Output (MIMO) Antennas	407
MIMO Principles	408
Multiple-User MIMO	410
12.4 Advanced Cellular Antennas	411
Evolution of Cellular Antennas.....	411
Beamforming.....	412
Active Antenna Systems.....	416
Massive MIMO	417
12.5 Key Terms and Review Questions.....	420
Key Terms	420
Review Questions	421
12.6 References.....	422
Chapter 13: Air Interface Physical Layer	424
13.1 Modulation Schemes	424
Modulation.....	425
Amplitude-Shift Keying.....	425
Frequency-Shift Keying	426
Phase-Shift Keying	426
Quadrature Amplitude Modulation	433
$\pi/2$ -BPSK.....	437
13.2 OFDM, OFDMA, and SC-FDMA	438
Orthogonal Frequency-Division Multiplexing	438
OFDM Implementation	441
Difficulties of OFDM	444

Orthogonal Frequency-Division Multiple Access	445
Single-Carrier FDMA	446
13.3 Waveforms and Numerologies	448
Numerology	448
CP-OFDM	449
DFT-S-OFDM	450
13.4 Key Terms and Review Questions	451
Key Terms	451
Review Questions	452
13.5 References and Documents	453
References	453
Documents	453
ANNEX 13A: Multiplexing and Multiple Access	453
Multiplexing	453
Multiple Access	455
Chapter 14: Air Interface Channel Coding	456
14.1 Transmission Errors	457
Error Burst	457
Data Transmission Channels	458
14.2 Forward Error Correction	460
Block Error Correction	460
Hamming Distance	462
14.3 Parity-Check Matrix Codes	466
Encoding	469
Error Detection	470
Error Correction	470
14.4 Low-Density Parity-Check Codes	471
Code Construction	472
Error Correction	473
Encoding	476

14.5	Polar Coding	476
	Polar Encoder	477
	Synthetic Channels	479
	Decoding	483
14.6	3GPP Channel Coding Specification.	484
	Quasi-Cyclic Low-Density Parity-Check Codes.	485
	Polar Coding with CRC.	486
14.7	Hybrid Automatic Repeat Request	488
14.8	Key Terms and Review Questions	490
	Key Terms	490
	Review Questions	490
14.9	References and Documents	491
	References.	491
	Documents	492
Chapter 15:	5G Radio Access Network	494
15.1	Overall RAN Architecture	494
15.2	RAN–Core Functional Split.	499
	RAN Functional Areas.	500
	Core Functional Areas	501
15.3	RAN Channel Structure	502
	Logical Channel.	503
	Transport Channels	503
	Physical Channels	504
15.4	RAN Protocol Architecture	505
	Air Interface Protocol Architecture.	505
	Channel Structure	508
	RAN–Core Network Interface Protocol Architecture	509
	Xn Interface Protocol Architecture.	511
15.5	NG RAN Transport Network.	512

15.6 Integrated Access and Backhaul.....	517
IAB Architecture	519
Parent/Child Relationship.....	520
IAB Protocol Architecture.....	520
15.7 Key Terms and Review Questions.....	523
Key Terms	523
Review Questions	524
15.8 References and Documents.....	525
References.....	525
Documents	525

Part V Appendixes

Appendix A: Review Questions and Solutions	528
Appendix B: Glossary	594
Appendix C: Acronyms	604
Index	608

Sample pages

Chapter 3

Overview of 5G Use Cases and Architecture

Learning Objectives

After studying this chapter, you should be able to:

- Present an overview of the types of use cases enabled by 5G
- Discuss the user experience and performance requirements of use cases in various categories
- Present an overview of the NGMN 5G architecture
- Present an overview of the 3GPP 5G core network architecture
- Present an overview of the 3GPP 5G radio access network architecture

This chapter presents a high-level view of the concepts underlying 5G. Section 3.1 provides an overview of the types of use cases enabled by 5G. This section presents three taxonomies or groupings of use cases developed by three different organizations. These different perspectives enable you to develop a solid grasp of the types of use cases that are emerging with 5G and their chief characteristics.

Section 3.2 presents a 5G architecture framework developed by the Next Generation Mobile Networks (NGMN) Alliance. This framework is widely accepted within the telecommunications industry. The framework provides an excellent high-level rendering of the overall 5G system and enables you to understand the overall structure of 5G.

Section 3.3 presents architecture models for the 5G core network and radio access network developed by the 3rd Generation Partnership Project (3GPP), which is the organization developing the technical specifications for 5G. The diagrams presented in this section illustrate the framework being used for the development of detailed technical specifications and enable you to see the context of the separate technical aspects of 5G that are described in subsequent chapters and to see how the various technical elements of 5G fit together.

3.1 5G Use Cases

There is a reciprocal cause–effect relationship between use cases and the capabilities being designed into 5G networks. As users and organizations have come to depend increasingly on wireless and Internet-based applications, these users have come to expect more capability, in terms of data rate and range of features, from wireless networks. These expectations create a demand that justifies the enormous investment required to deploy 5G networks. Conversely, 5G networks are the logical next step in the evolution of cellular networks, and the new capabilities provided by 5G enable users to envision a wide range of new applications, termed *use cases*, in the standards and specifications documents.

This section looks at examples of use cases, or applications, for 5G compiled by three different organizations: ITU Radiocommunication Sector (ITU-R), 5G Americas, and the Next Generation Mobile Networks (NGNM) Alliance. The three compilations provide a good sense of the possibilities enabled by 5G.

ITU-R

ITU-R Report M.2441 describes 16 emerging use cases for IMT-2020, as shown in Figure 3.1. M.2441 provides a description of a number of specific applications encompassed by each use case, as well as a summary of technical capabilities of the IMT systems needed to support each use case.

Machine-type communication	PPDR	Transport applications	Utilities
Industrial automation	Remote Control	Surveying and inspection	Healthcare
Sustainability/environmental	Smart City	Wearables	Smart Homes
Agriculture	Media and entertainment	Enhanced personal experience	Commercial Airspace UAS

PPDR = Public Protection and Disaster Relief
UAS = Unmanned Aerial Systems

FIGURE 3.1 Emerging 5G Use Cases (ITU-R)

Chapter 2, “5G Standards and Specifications,” briefly describes each of the 16 use cases.

5G Americas

5G Americas is an industry trade organization composed of leading telecommunications service providers and manufacturers. The organization’s mission is to advocate for and foster the advancement

and full capabilities of LTE (Long Term Evolution) wireless technologies and their evolution to 5G throughout the ecosystem's networks, services, applications, and connected devices in the Western Hemisphere. 5G Americas is affiliated with 3GPP as a market representation partner.

The 5G Americas document *5G Services & Use Cases* (November 2017) describes 16 use cases and provides a framework for mapping use cases to 5G capabilities. Figure 3.2, from the 5G Americas document, shows the framework, which has two dimensions. One dimension consists of the three usage scenarios similar to those defined by ITU-R and described in Chapter 2.

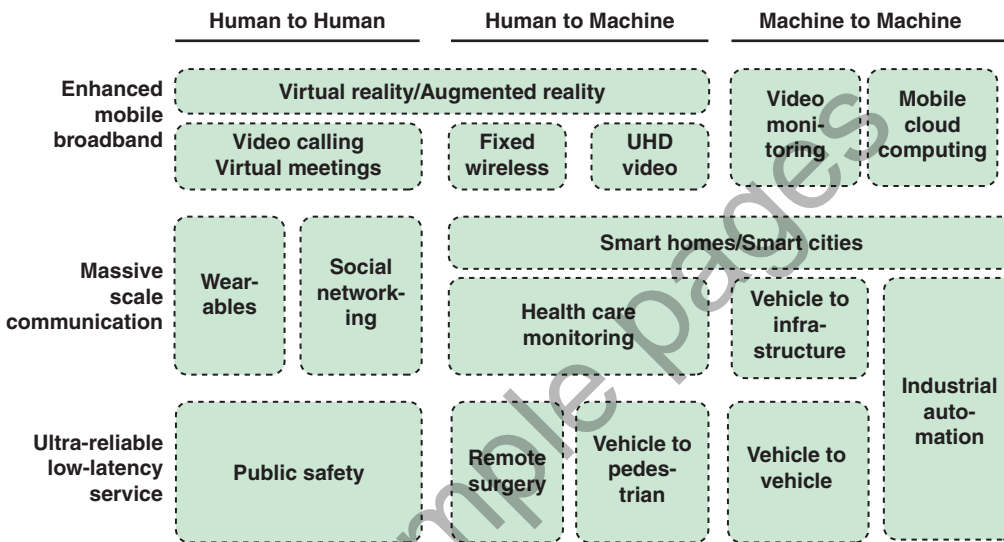


FIGURE 3.2 Some 5G Use Cases Grouped by Type of Interaction and the Range of Performance Requirements (5G Americas)

The three scenarios, based on their performance requirements, are:

- **Enhanced mobile broadband (eMBB):** These use cases generally have requirements for higher data rates and better coverage.
- **Massive scale communication:** These use cases generally have requirements to support a very large number of devices in a small area and, therefore, very high device density.
- **Ultra-reliable low-latency service:** These use cases have very strict requirements on latency and reliability and are also referred to as ultra-reliable and low-latency communications (URLLC).

The other dimension characterizes use cases based on whether they involve human-to-human, human-to-machine, or machine-to-machine communication. As indicated in Figure 3.2, a use case may correspond to more than one type of interaction or more than one usage scenario. This depiction clarifies the requirements for the various use cases.

NGMA Alliance

The Next Generation Mobile Networks (NGMN) Alliance is an association of mobile operators, vendors, manufacturers, and research institutes. It is an open forum whose goal is to ensure that the standards for next-generation network infrastructure, service platforms, and devices meet the requirements of operators and, ultimately, that they will satisfy end user demand and expectations. The NGMN Alliance complements and supports standards organizations by providing a coherent view of what mobile operators require. NGMN is affiliated with 3GPP as a market representation partner.

NGMN's *5G White Paper* (February 2015) defines 24 use cases intended as representative examples of the applications for 5G and meant to highlight the diversity of performance requirements that 5G networks must satisfy. Figure 3.3, from the white paper, illustrates the framework for positioning use cases in a way that clarifies requirements. At a high level, use cases are grouped into eight use case families. These families are roughly similar to the three usage scenarios defined by ITU-R but at a greater granularity. Each family reflects the dominant characteristic of the use cases in that family.

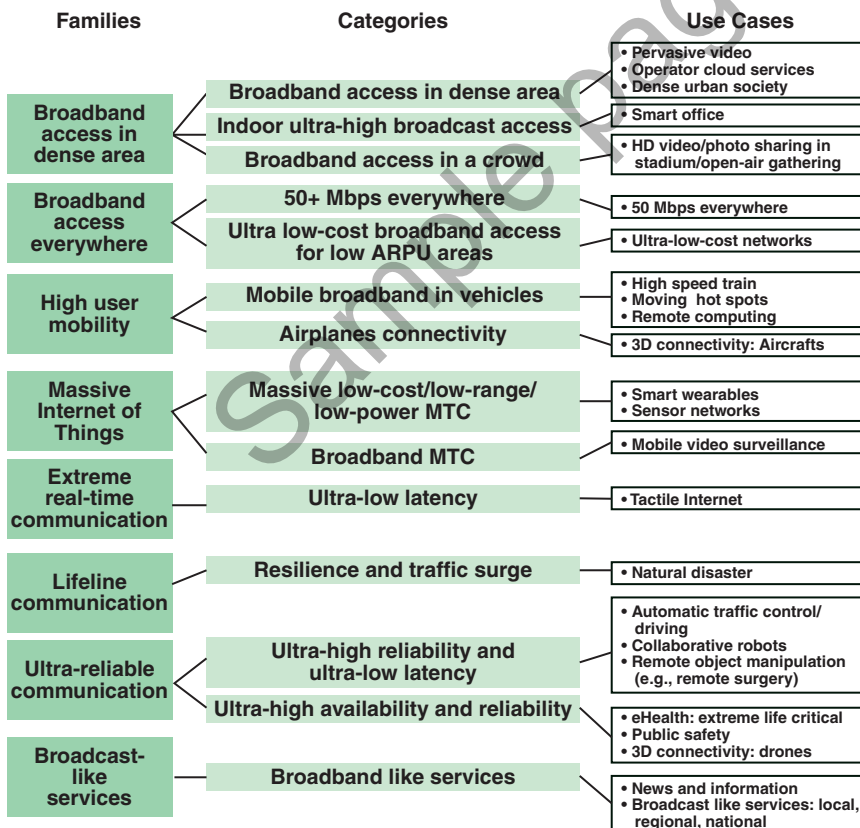


FIGURE 3.3 Use Case Categories Definition (NGMN)

Each family is in turn divided into a number of categories. The categories represent distinct types of demands on the 5G network in terms of user experience requirements and system performance requirements. For each use case category, one set of requirement values is given, which is representative of the extreme use cases(s) in the category. As a result, satisfying the requirements of a category leads to satisfying the requirements of all the use cases in this category. Tables 3.1 and 3.2, from the white paper, summarize the requirements.

TABLE 3.1 User Experience Requirements

Use Case Category	User-Experienced Data Rate	E2E Latency	Mobility
Broadband access in dense areas	DL: 300 Mbps UL: 50 Mbps	10 ms	On demand: 0–100 km/h
Indoor ultra-high broadband access	DL: 1 Gbps UL: 500 Mbps	10 ms	Pedestrian
Broadband access in a crowd	DL: 25 Mbps UL: 50 Mbps	10 ms	Pedestrian
50+ Mbps everywhere	DL: 50 Mbps UL: 25 Mbps	10 ms	0–120 km/h
Ultra-low-cost broadband access for low average revenue per user (ARPU) areas	DL: 10 Mbps UL: 10 Mbps	50 ms	On demand: 0–50 km/h
Mobile broadband in vehicles (cars, trains)	DL: 50 Mbps UL: 25 Mbps	10 ms	On demand: Up to 500 km/h
Airplanes connectivity	DL: 15 Mbps per user UL: 7.5 Mbps per user	10 ms	Up to 1000 km/h
Massive low-cost/long-range/low-power machine-type communication (MTC)	Low (typically 1–100 kbps)	Seconds to hours	On demand: 0–500 km/h
Broadband MTC	See the requirements for the broadband access in dense areas and 50+ Mbps everywhere categories		
Ultra-low latency	DL: 50 Mbps UL: 25 Mbps	< 1 ms	Pedestrian
Resilience and traffic surge	DL: 0.1–1 Mbps UL: 0.1–1 Mbps	Regular communication: not critical	0–120 km/h
Ultra-high reliability and ultra-low latency	DL: 50 kbps–10 Mbps UL: a few bps–10 Mbps	1 ms	On demand: 0–500 km/h
Ultra-high availability and reliability	DL: 10 Mbps UL: 10 Mbps	10 ms	On demand: 0–500 km/h
Broadcast-like services	DL: Up to 200 Mbps UL: Modest (e.g., 500 kbps)	< 100 ms	On demand: 0–500 km/h

TABLE 3.2 System Performance Requirements

Use Case Category	Connection Density	Traffic Density
Broadband access in dense areas	200–2500/km ²	DL: 750 Gbps/km ² UL: 125 Gbps/km ²
Indoor ultra-high broadband access	75,000/km ² (75/1000 m ² office)	DL: 15 Tbps/km ² (15 Gbps/1000 m ²) UL: 2 Tbps/km ² (2 Gbps/1000 m ²)
Broadband access in a crowd	150,000/km ² (30,000/stadium)	DL: 3.75 Tbps/km ² (0.75 Tbps/stadium) UL: 7.5 Tbps/km ² (1.5 Tbps/stadium)
50+ Mbps everywhere	400/km ² suburban 100/km ² rural	DL: 20 Gbps/km ² suburban UL: 10 Gbps/km ² suburban DL: 5 Gbps/km ² rural UL: 2.5 Gbps/km ² rural
Ultra-low-cost broadband access for low ARPU areas	16/km ²	16 Mbps/km ²
Mobile broadband in vehicles (cars, trains)	2000/km ² (500 active users per train × 4 trains or 1 active user per car × 2000 cars)	DL: 100 Gbps/km ² (25 Gbps per train, 50 Mbps per car) UL: 50 Gbps/km ² (12.5 Gbps per train, 25 Mbps per car)
Airplanes connectivity	80 per plane 60 airplanes per 18,000 km ²	DL: 1.2 Gbps/plane UL: 600 Mbps/plane
Massive low-cost/long-range/low-power MTC	Up to 200,000/km ²	Not critical
Broadband MTC	See the requirements for the broadband access in dense areas and 50+ Mbps everywhere categories	
Ultra-low latency	Not critical	Potentially high
Resilience and traffic surge	10,000/km ²	Potentially high
Ultra-high reliability and ultra-low latency	Not critical	Potentially high
Ultra-high availability and reliability	Not critical	Potentially high
Broadcast-like services	Not relevant	Not relevant

3.2 NGMN 5G Architecture Framework

NGMN introduced a framework for the 5G architecture in its 2015 white paper and subsequently elaborated the framework in a later document, *5G End-to-End Architecture Framework* (August 2019). The NGMN framework emphasizes the need for modular network functions that could be deployed and scaled on demand to accommodate various use cases in an agile and cost-efficient manner. The NGMN approach is built on the concept of the softwarization of 5G networks. In essence, **softwarization**

is an overall approach for designing, implementing, deploying, managing, and maintaining network equipment and/or network components through software programming.

Four approaches to softwarization are important in 5G networks and reflected in the NGMN model:

- **Software-defined networking (SDN):** An approach to designing, building, and operating large-scale networks based on programming the forwarding decisions in routers and switches via software from a central server. SDN differs from traditional networking, which requires configuration of each device separately and relies on protocols that cannot be altered. Chapter 7, “Software-Defined Networking,” covers SDN.
- **Network functions virtualization (NFV):** The virtualization of compute, storage, and network functions by implementing these functions in software and running them on virtual machines. Chapter 8, “Network Functions Virtualization,” covers NFV.
- **Edge computing:** A distributed information technology (IT) architecture in which client data is processed at the periphery of the network, as close to the originating source as possible.
- **Cloud-edge computing:** A form of edge computing that offers application developers and service providers cloud computing capabilities as well as an IT service environment at the edge of a network. The aim is to deliver compute, storage, and bandwidth much closer to data inputs and/or end users. Chapter 10, “Multi-Access Edge Computing,” covers cloud-edge computing.

Layered Functionality

Figure 3.4 illustrates the NGMN architecture framework. The architecture comprises three layers and an end-to-end (E2E) management and orchestration entity.

The **infrastructure resource layer** consists of the physical resources and system software of a fixed-mobile converged (FMC) network. Figure 3.4 shows in detail the core network portion, which includes these types of devices:

- **Cloud nodes:** These nodes provide cloud services, software, and storage resources. There are likely to be one or more central cloud nodes that provide traditional cloud computing service. In addition, cloud-edge nodes provide low latency and higher-security access to client devices at the edge of the network. All of these nodes include virtualization system software to support virtual machines and containers. NFV enables effective deployment of cloud resources to the appropriate edge node for a given application and given fixed or mobile user. The combination of SDN and NFV enables the movement of edge resources and services to dynamically accommodate mobile users.
- **Networking nodes:** These are IP routers and other types of switches for implementing a physical path through the network for a 5G connection. SDN provides for flexible and dynamic creation and management of these paths.

- Access nodes:** These provide an interface to radio access networks (RANs), which in turn provide access to mobile user equipment (UE). SDN creates paths that use an access node for one or both ends of a connection involving a wireless device.

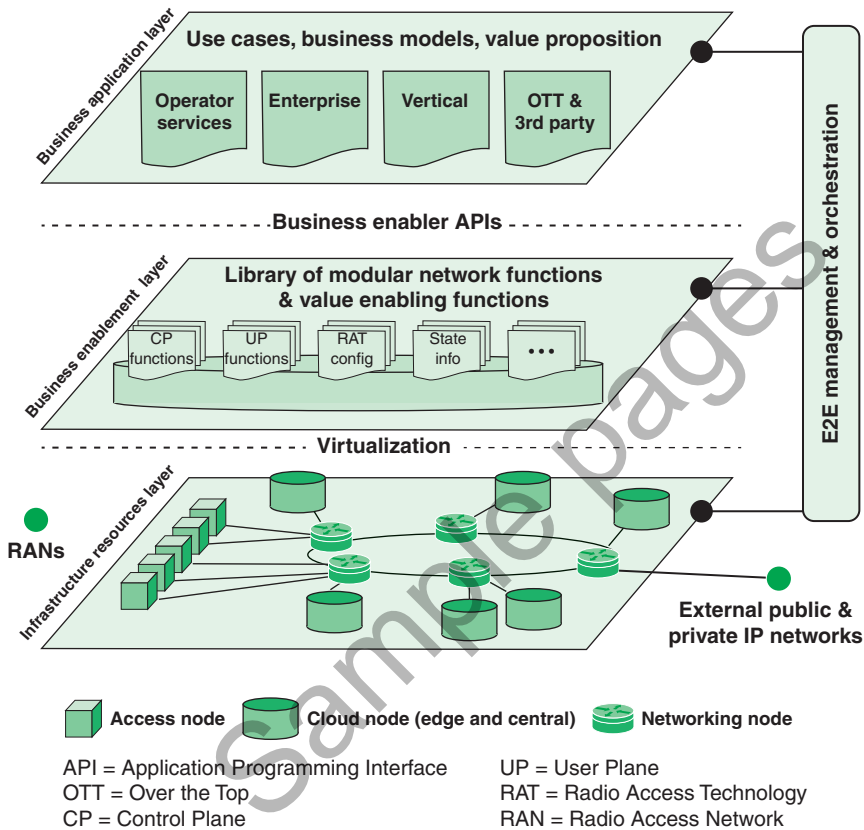


FIGURE 3.4 5G Architecture (NGMN)

Hardware and software resources at the infrastructure resources layer are exposed to higher layers and to the end-to-end management and orchestration entity through relevant application programming interfaces (APIs). Performance and status monitoring as well as configurations are intrinsic parts of such an API.

The **business enablement layer** is a library of all functions required within a converged network in the form of modular architecture building blocks, including functions realized by software modules that can be retrieved from the repository to the desired location, and a set of configuration parameters for certain parts of the network (e.g., radio access). The functions and capabilities are called upon request

by the orchestration entity, through relevant APIs. For certain functions, multiple variants might exist. For example, there may be different implementations of the same functionality that have different performance or characteristics. The different levels of performance and capabilities offered could be utilized to differentiate the network functionality much more than in today's networks (e.g., to offer mobility functions such as nomadic mobility, vehicular mobility, or aviation mobility, depending on specific needs). Specific types of components at this layer include:

- **Control plane functions:** These are modules that implement the control signaling functions within the network, as well as control signals associated with SDN and NFV.
- **User plane functions:** These modules deal with the exchange of user data over a connection.
- **Radio access technology (RAT) configuration:** These functions facilitate the configuration of elements in the RAN, including base stations.
- **State information:** State information is split from functions and nodes and managed separately. This includes the state of network connections and RAN radio channels.

With the use of SDN and NFV, 5G networks can support flexible functions and capabilities. The functions at the business enablement layer can tailor connections, configurations, and resource deployment for each use case.

The **business application layer** contains specific applications and services that support the following users of the 5G network:

- **Mobile network operator:** A mobile network operator is a wireless telecommunications organization that provides wireless voice and data communication for its subscribed mobile users. The operator owns or controls the complete telecom infrastructure for hosting and managing mobile communications between the subscribed mobile users with users in the same and external wireless and wired telecom networks, including radio spectrum allocation, wireless network infrastructure, backhaul infrastructure, billing, customer care, provisioning computer systems, and marketing and repair organizations. Mobile network operators are also known as wireless service providers, wireless carriers, and cellular companies.
- **Enterprise:** An enterprise is a business that offers services over the mobile network. These services include applications that run on mobile devices and cloud-based services that enable application portability across multiple devices.
- **Verticals:** An industry vertical is an organization that provides products and/or services targeted to a specific industry, trade, profession, or other group of customers with specialized needs. A vertical might provide a range of products or services useful in the banking industry or healthcare. In contrast, a horizontal provides products or services that address a specific need across multiply industries, such as accounting or billing products and services. Some 5G use cases are realized by standalone private networks managed by the vertical industry itself rather

than the network service provider (NSP). A good example is factory automation. In such cases, the vertical can own and control its own application packages and business application layer.

- **Over-the-top (OTT) and third parties:** OTT or third-party services can be defined as any services provided over the Internet and mobile networks that bypass traditional operators' distribution channel. Cooperation between the mobile network operator and the OTT involves providing quality of service (QoS) and latency attributes in network slices. Examples of OTTs include:
 - **Voice over Internet Protocol (VoIP):** Skype, Viber, etc.
 - **Short Message Service (SMS):** WhatsApp, Kakao Talk, Line, Telegram, etc.
 - **Apps:** Search portals, news portals, banking, weather, shopping, etc.
 - **Cloud services:** Dropbox, Google Drive, Apple iCloud, etc.
 - **Internet television (video streaming):** Netflix, Hulu, YouTube, Amazon Video, etc.

At the interface to the end-to-end management and orchestration entity, users at the business application layer can build dedicated network slices for an application or map an application to existing network slices.

The **E2E management and orchestration** entity is the contact point to translate the use cases and business models into actual network functions and slices. It defines the network slices for a given application scenario, chains the relevant modular network functions, assigns the relevant performance configurations, and finally maps all of this onto the infrastructure resources. It also manages scaling of the capacity of those functions as well as their geographic distribution. In certain business models, it could also possess capabilities to allow for third parties to create and manage their own network slices.

Network Slicing

Chapter 2 introduces the concept of network slicing. Figure 2.11 suggests that slices can divide classes of internal network functions, such as dividing eMBB from mMTC from URLLC. The NGMN framework suggests that slices could effectively partition networks in such a way that different classes of user equipment, utilizing their respective sets of radio access technologies, would perceive quite different infrastructure configurations, even though they would be accessing resources from the same pools.

NGMN defines a network slice as being composed of a collection of 5G network functions and specific RAT settings combined together for the specific use case or business model. Thus, a 5G slice can span all domains of the network—software modules running on cloud nodes, specific configurations of the transport network supporting flexible location of functions, and a dedicated radio configuration or even a specific RAT—as well as configuration of the 5G device.

Figure 3.5 illustrates three use cases highlighted in the NGMN white paper. The blacked-out core network resources represent resources not used to create the network slice.

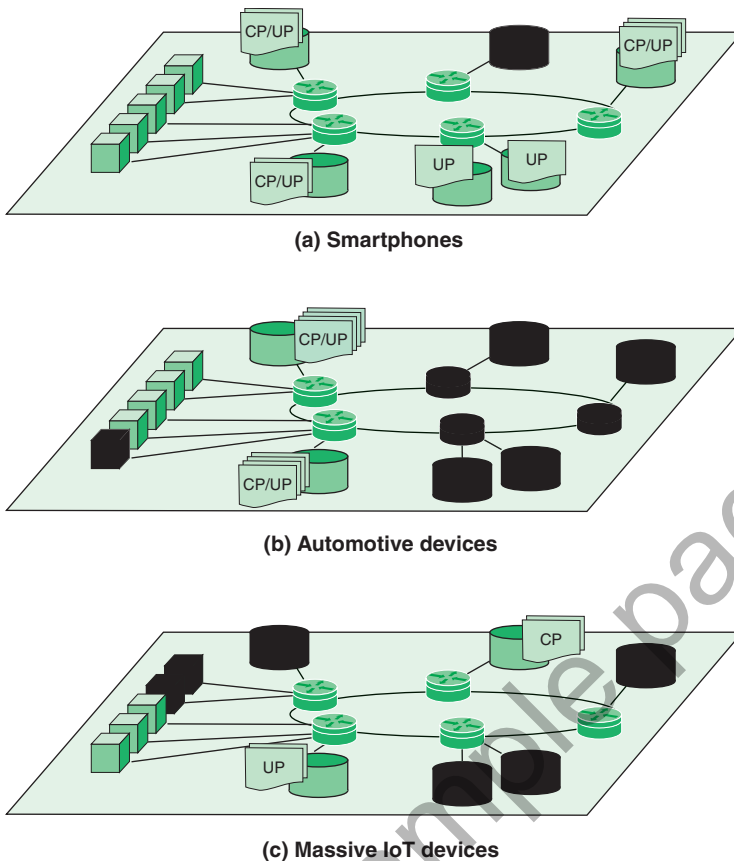


FIGURE 3.5 5G Network Slices Implemented on the Same Infrastructure

Cloud nodes that are part of the slice may include the following:

- Control plane (CP) functions associated with one or more user plane (UP) functions (e.g., a reusable or common framework of control)
- Service- or service category–specific control plane and user plane function pairs (e.g., user-specific multimedia application session)

The first network slice depicted in Figure 3.5 is for a typical smartphone use case. Such a slice might have fully fledged functions distributed across the network. The second network slice in Figure 3.5 indicates the type of support that may be allocated for automobiles in motion. This use case emphasizes the need for security, reliability, and low latency. A configuration to achieve this would limit core network resources to nearby cloud edge nodes, plus the recruitment of sufficient access nodes to support the use case. The final use case illustrated in Figure 3.5 is for a massive IoT deployment, such as a huge number of sensors. The slice can contain just some basic CP and UP functions with, for

example, no mobility functions. This slice would only need to engage the access nodes nearest the IoT device deployment.

Thus, as Figures 3.4 and 3.5 suggest, a layered architecture with emphasis on softwarization provides the flexibility needed to fully exploit the 5G infrastructure.

3.3 3GPP 5G Architecture

As discussed in Chapter 2, the 3rd Generation Partnership Project (3GPP) is the organization responsible for developing reports and specifications that define 5G networks. The project covers cellular telecommunications technologies—including the air interface, RAN, core network, and service capabilities—that provide a complete system description for 5G mobile telecommunications. The 3GPP specifications also provide hooks for fixed access to the core network and for interworking with 4G networks and networks outside the 3GPP specifications.

Hundreds of 3GPP reports and specifications related to 5G together describe an extraordinarily complex system. This book provides an overview that encompasses the scope of 5G technology and networks, based on the 3GPP documents. This section provides a “big picture” overview of 5G, summarizing the two key 3GPP documents that together describe the 5G architecture:

- **TS 23.501** (*Technical Specification Group Services and System Aspects; System Architecture for the 5G System (5GS); Stage 2 (Release 16), December 2020*): This document describes the core network architecture, together with its services and interfaces.
- **TS 38.300** (*Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2 (Release 16), December 2020*): This document describes the radio access network architecture, together with its services and interfaces.

5G Core Network Architecture

The 5G architecture model provides a framework within which detailed specifications can be developed.

Principles

TS 23.501 lists the following as the key principles for the architecture:

- Separate the user plane (UP) functions from the control plane (CP) functions to allow independent scalability, evolution, and flexible deployments (e.g., centralized location or distributed [remote] location).
- Modularize the function design (e.g., to enable flexible and efficient network slicing).
- Wherever applicable, define procedures (i.e., the set of interactions between network functions) as services so that their reuse is possible.

- Enable each network function (NF) and its network function services (NFS) to interact with other NF and its NFS directly or indirectly via a service communication proxy, if required. The architecture does not preclude the use of another intermediate function to help route control plane messages.
- Minimize dependencies between the access network (AN) and the core network (CN). The architecture is defined with a converged core network with a common AN–CN interface that integrates different access types (e.g., 3GPP access and non-3GPP access).
- Support a unified authentication framework.
- Support stateless NFs. A stateless NF separates an NF into a processing module and a data store module containing state information. This design results in a more agile NF in terms of scalability, resilience, and ease of deployment [KABL17].
- Support capability exposure. As mentioned in Chapter 2, this refers to the ability to provide relevant information about network capabilities to third parties.
- Support concurrent access to local and centralized services. To support low-latency services and access to local data networks, UP functions can be deployed close to the access network.
- Support roaming with both home-routed traffic as well as local breakout traffic in the visited public land mobile network (PLMN).

Roaming

The final item in the preceding list references some important concepts. A PLMN, often called a carrier, is a telecommunications network that provides mobile cellular services. A home PLMN for a given mobile phone subscriber is the PLMN that is contracted to provide cellular service to the subscriber. Roaming is the ability for a user to function in a serving network different from the home network, called the visited network. Roaming can occur internationally, in which case mobile subscribers get coverage and at least basic services similar to their domestic package from a network operator in another country. For 5G, any user device should work in any other country. With 4G and earlier, there are multiple technologies in use, and a cell phone from one country may not be able to operate on a network in another country. *National roaming* refers to the ability to move from one mobile operator to another in the same country.

Both the home and visited PLMN use the same network architecture, which defines protocols, services, and interfaces between the home PLMN and the visited PLMN. The 3GPP specifications support two models of operation: home routed and local breakout (LBO). In the home-routed roaming model, the subscriber's data traffic is serviced by the subscriber's home network, which gives the home network operator more control over the user's traffic and is preferable when the relationship between the two operators is not totally trustworthy. In the LBO model, the subscriber's data is serviced by the visited network; this model delivers more efficient routing in terms of bandwidth and latency. In the case of LBO, the home network owner loses control of the customer and has no role in delivering services to that user. The LBO model is used when there is a trusted relationship between the two operators.

An architecture that includes roaming adds complexity but no change to the basic network services and functionality. Therefore, this section focuses on the non-roaming case.

Architecture Diagrams

TS 23.501 contains a number of architecture diagrams from several different points of view and at varying levels of detail. All of the diagrams depict the architecture in terms of a number of interconnected network functions (NFs). An NF is a processing function in a network that has defined functional behavior and interfaces. A network function can be implemented either as a network element on dedicated hardware, as a software instance running on dedicated hardware, or as a virtualized function instantiated on an appropriate platform.

The interconnection between NFs is represented in two ways:

- **Service-based representation:** NFs within the control plane enable other authorized NFs to access their services. This representation illustrates how a set of services is provided/exposed by a given network interface. This interface defines how one network function within the control plane allows other network functions that have been authorized to access its services. This representation also includes point-to-point reference points where necessary.
- **Reference point representation:** This representation uses labeled point-to-point links to show the interaction that exists between two NFs or between an NF and an external functional module or network. The reference point representation is beneficial when showing message sequence charts. It shows the relationships between NFs that are used in the message sequence charts.

There are several advantages to this form of architectural representation. The modular structure provides a framework for developing detailed specifications for each NF. The service-based interfaces and reference points provide a framework for developing detailed specifications of the interaction between NFs in terms of data formats, protocols, and service calls. In addition, the detailed interface specifications promote interoperability between different hardware/software providers. Finally, this type of architecture definition provides a way of ensuring that all 5G functional and service requirements are satisfied.

Service-Based System Architecture

Figure 3.6, based on TS 23.501, depicts the overall non-roaming 5G service-based architecture (SBA) for the core network. The functional components of the control plane of the network are network functions (NFs) that offer their services to any other applicable NFs via a common framework of interfaces accessible to all NFs. Network repository functions (NRFs) allow every NF to discover the services offered by other NFs present in the network; network exposure functions (NEFs) expose capability information and services of the 5G core network NFs to external entities. This model aims to maximize the modularity, reusability, and self-containment of network functions and to foster the ability to grow flexibly while taking advantage of network functions virtualization and software-defined networking.

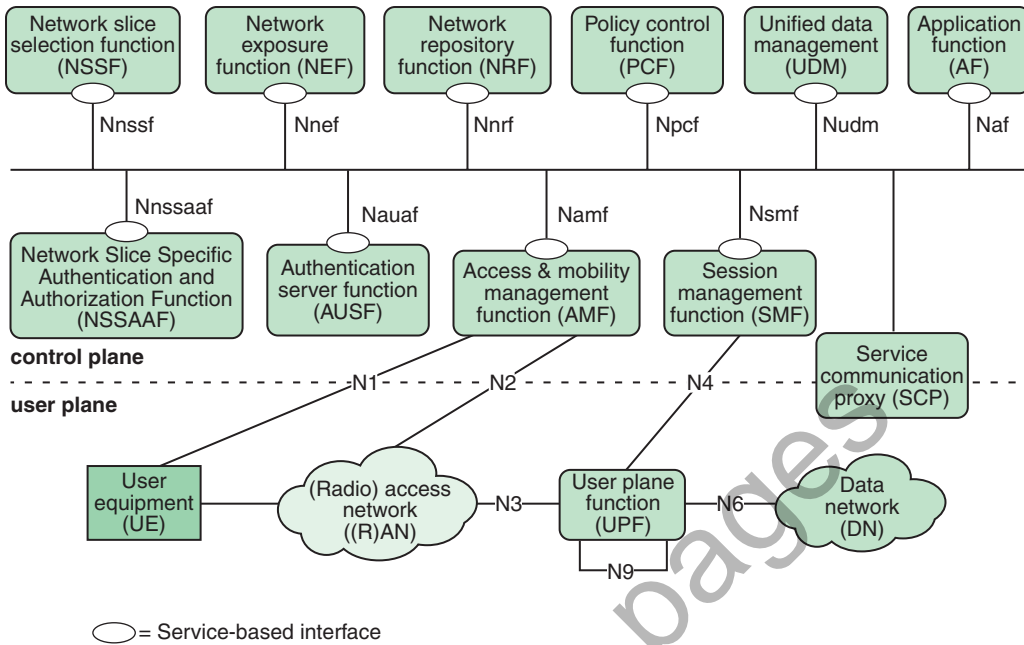


FIGURE 3.6 Non-Roaming 5G System Architecture

The figure includes the following NFs and other modules:

- **Authentication server function (AUSF):** Performs authentication between UE and the network.
- **Access and mobility management function (AMF):** Receives all connection- and session-related information from the user equipment (UE) (N1/N2) but is responsible only for handling connection, registration, reachability, and mobility management tasks. All messages related to session management are forwarded to the session management function (SMF).
- **Network exposure function (NEF):** Provides an interface for outside applications to communicate with the 5G network to obtain network-related information in the following categories:
 - **Monitoring capability:** Allows an external entity to request or subscribe to UE-related events of interest. The monitored events include a UE’s roaming status, UE loss of connectivity, UE reachability, and location-related events.
 - **Provisioning capability:** Allows an external entity to provide information about expected UE behavior to the 5G system (e.g., predicted UE movement, communication characteristics).
 - **Policy/charging capability:** Handles QoS and charging policy for the UE, based on a request from an external party.

- **Analytics reporting capability:** Allows an external party to fetch or subscribe/unsubscribe to analytics information generated by the 5G system.
- **Network repository function (NRF):** Allows NFs to register their functionality and to discover the services offered by other NFs present in the network.
- **Network slice selection function (NSSF):** Selects the set of network slice instances to accommodate the service request from a UE. When a UE requests registration with the network, AMF sends a network slice selection request to NSSF with information on the preferred network slice selection. The NSSF responds with a message that includes a list of appropriate network slice instances for the UE.
- **Network slice-specific authentication and authorization (NSSF):** Performs authentication and authorization specific to a slice.
- **Policy control function (PCF):** Provides functionalities for the control and management of policy rules, including rules for QoS enforcement, charging, and traffic routing. PCF enables end-to-end QoS enforcement with QoS parameters (e.g., maximum bit rate, guaranteed bit rate, priority level) at the appropriate granularity (e.g., per UE, per flow, per protocol data unit [PDU] session).
- **Session management function (SMF):** Responsible for PDU session establishment, modification, and release between a UE and a data network. A PDU session, or simply a session, is an association between the UE and a data network that provides a PDU connectivity service. A PDU connectivity service is a service that provides for the exchange of PDUs between a UE and a data network.
- **Unified data management (UDM):** Responsible for access authorization and subscription management. UDM works with the AMF and AUSF as follows: The AMF provides UE authentication, authorization, and mobility management services. The AUSF stores data for authentication of UEs, and the UDM stores UE subscription data.
- **User plane function (UPF):** Handles the user plane path of PDU sessions. This function is described subsequently.
- **Application function (AF):** Provides session-related information to the PCF so that the SMF can ultimately use this information for session management. The AF interacts with application services that require dynamic policy control. The AF extracts session-related information (e.g., QoS requirements) from application signaling and provides it to the PCF in support of its rule generation. An example is the IP multimedia subsystem (IMS), which may interface with the PCRF to request QoS support for VoIP calls.
- **User equipment (UE):** Allows a user access to network services. An example is a mobile phone. For the purpose of 3GPP specifications, the interface between the UE and the network is the radio interface.
- **(Radio) Access Network ((R)AN):** Provides access to a 5G core network. This includes the 5G RAN and other wireless and wired access networks.