5GMF White Paper

5G Mobile Communications Systems for 2020 and beyond

Version 1.01

July 4, 2016

The Fifth Generation Mobile Communications Promotion Forum

Contents

Scope 1
1. Introduction 2
2. Objectives 7
3. Market and User Trends related to 5G 8
3.1. Shift from PCs to Devices such as Smartphones and Tablets and wearable terminals 8
3.2. Increase in Location-Based Services 9
3.3. Forefront of a new way of building human relations with a focus on women 10
3.4. Introduction of the Sharing Economy 11
3.5. Introduction of Artificial Intelligence and Robots 12
3.6. Self-Driving Vehicles 13
3.7. Internet of Things (including industry, wearables, and agriculture) 15
3.8. Changes in the Work Style 17
3.9. Acceleration of Fintech Services 18
3.10. Penetration of Peer to Peer Service 19
4 Traffic Trend 20
4.1 General 20
4.2 Communication traffic growth and traffic nature trend 20
4.2.1 Communication traffic growth 20
4.2.2 Communication traffic nature 22
5 Cost Implications 25
5.1 General 25
5.2 Costs per communication traffic aspect 26
5.3 User density perspective 29
5.4 Daily dynamics aspect 35
5.5 Capital investment aspect 39
5.6 Conclusion 39
6. 5G Key Concept 42
6.1 Key Concepts of 5G 42
6.2 5G key technical aspects 43
6.2.1 General 43
6.2.2 Advanced Heterogeneous network 44

6.2.3 Network Softwarization and Slicing 45
6.3 5G Typical Use Cases 46
6.3.1 Ultra-reliable and low latency communications..46
6.3.2 Massive Connection 48
6.3.3 eMBB enhanced Mobile Broadband (Data rate, Capacity, Mobility) 48
7. Typical Usage Scenarios of 5G.. 50
7.1 Four representative typical usage scenarios 50
7.2 Case studies of Typical Usage Scenarios 51
7.3 Dynamic approach 72
8 Requirements for 5G.. 75
8.1 High level requirements..75
8.2 Requirements related to 5G radio access network..75
8.2.1 Definitions of the requirements 75
8.2.2 List of 5G RAN requirements and their mapping to use cases..78
8.3 Requirements for 5G networks..78
9. Spectrum Implications 81
9.1 Concept for 5G spectrum..81
9.2 Below 6GHz..83
9.2.1 Roles of bands below 6GHz..83
9.2.2 Technical Implementation and Challenges..84
9.2.3 Current spectrum allocation and its plan in Japan, below 6GHz 85
9.2.4 Spectrum identified for IMT below 6GHz in WRC-15..86
9.3 Above 6GHz..86
9.3.1 Roles of bands above 6GHz..86
9.3.2 Preferred frequency ranges/bands 86
9.3.3 Technical implementation issue and Challenges..92
9.3.4 WRC-19 Agenda Item 1.13 94
10 Overview of 5G Technologies 102
11 5G Radio Access Technologies 103
11.1 General 103
11.2 Overview of 5G radio access network.. 103
11.3 RAN related technical works update 104
11.3.1 General.. 104
11.3.2 Information of technical works related to modulation or coding scheme...... 104

11.3.3 Information of technical works related to multiple access scheme, duplex scheme ... 110
11.3.4 Information of technical works related to MIMO or multiple antenna technologies ... 113
11.3.5 Information of technical works related to RAN deployment or is control schemes ... 116
11.3.6 Information of technical works related to certain use cases or applications 121
11.3.7 Information of technical works related to energy saving nature ... 122
11.3.8 Information of technical works related to RAN virtualization ... 124
11.3.9 Other information of technical works related to '5G' RAN ... 125
12. Network Technologies for 5G ... 129
12.1 Technology focus area ... 129
12.2 Network softwarization ... 131
12.2.1 General definition ... 131
12.2.2 Network softwarization in 5G ... 132
12.2.3 Information Centric Network (ICN) enabled by network softwarization ... 139
12.3 Management and Orchestration ... 145
12.3.1 Overview ... 145
12.3.2 Approaches for 5G network management ... 147
12.4 Fronthaul and Backhaul ... 167
12.4.1 Overview ... 167
12.4.2 Fronthaul technologies ... 179
12.4.3 Backhaul technologies ... 181
12.5 Mobile Edge Computing (MEC) ... 187
12.5.1 Overview of MEC ... 187
12.5.2 Application of MEC ... 193
13. Conclusion ... 205
Annex Future Business and Services ... 208

Scope

This white paper addresses the results of studies carried out by the Fifth Generation Mobile Communications Promotion Forum (5GMF) in Japan. As a result of the study, the white paper proposes two key concepts of 5G and two main key technologies required to realize these key concepts.

The scope of the study also includes market and user trends, traffic trends, cost and spectrum implications, typical usage scenarios, and requirements of 5G. Radio access technologies and network technologies of 5G are also addressed. In the Annex, the perspectives of future business and services are introduced for reference.

1. Introduction

Japan led the world in mobile feature phone services, such as e-mail services via the internet, wide variety of information services by websites specifically designed for mobile phone screens, TV broadcasting service by One-Seg, and electronic payment services. With the advent of the 4th generation mobile communication system (4G), the people of Japan have gained access to a nationwide mobile broadband network. As users of smartphones are growing rapidly in Japan, services with rich content, such as HD video, e-books, music, and video games are widely provided. With these cutting-edge services, Japan has one of the most mature mobile communication markets which are able to enjoy the world leading mobile services.

As these new content-rich services have become more popular year after year, internet traffic has sharply increased along with the need for more network capacity and higher speeds. The services being offered are also diversifying, and both human-to-human and human-to-device communication is increasing. As network and sensor technology advances, device-to-device communication, what is called, the Internet of Things (IoT) is also expanding worldwide, leading to a further increase in traffic. This facilitates changes in ICT services for entertainment, transportation, industry/verticals, and emergency and disaster relief. Examples include artificial intelligence and adorable robots that assist people in their home and work lives, autonomous vehicles like unmanned taxis as well as vehicles that can provide mobility for senior citizens, and wearable devices that collect and analyze vital data to assist in health and medical services. These are just some of the services that are expected to be implemented in the near future as these trends continue to accelerate. However, the current 4G technologies, as well as its extension, may limit the growth of mobile services, especially when considering the needs of the 2020s. In order to accommodate the rapid growth with sufficient capacity and speed, there is strong global interest for research and development of the 5th generation mobile communication system, known as 5G.

The Fifth Generation Mobile Communications Promotion Forum (5GMF) was established in Japan on September 30, 2014 to actively promote 5G study in line with trends both in Japan and abroad based on a roadmap on 5G implementation policy published by the government of Japan. This white paper discusses the expected many new uses of ICT in the 5G era by various industries, as well as the new businesses and markets that will be created and the expectations of the fuller lifestyle that it will bring to people everywhere. 5GMF has collected in this white paper the opinions and ideas of experts in industry, academia and government concerning their views of the future of applications, networks, and wireless technology related to 5G in order to provide a clear goal for the development of 5G.

The description contents in each chapter are indicated easily in the following.

- **Market and User Trends of ICT (Chapter 3)**

This chapter, in addition to gathering information by industries promoting ICT services on their customers, broken down by age group, type of content, and type of device, attempts to predict future trends in order to understand what the communication environment will be, and thus what mobile communication services will be in demand, in 5G era.

- **Traffic Trend (Chapter 4)**

This chapter provides an analysis of the latest communication traffic trends. For the past years, considerable increase of communication traffic has been observed and several estimation studies consistently forecast that the increase would be continued to the next decade. In addition, new traffic nature different from ever happened one could come in considering new traffic types generated in variety of use cases with variety of 'connected things' or 'connected services'.

- **Cost Implications (Chapter 5)**

This chapter discusses the cost of mobile communication systems and analyzes from the perspective of several '5G' related use cases. The fundamental cost implications of '5G' were analyzed in [1] where every element of a mobile communication system was analyzed in terms of CAPEX or OPEX. The analyses were made with a focus on the domestic market of Japan, in light of demographic and survey data as well as local market indexes. Since the market in Japan is one of the leading markets in the world, these case studies may be of use when considering markets in other locations around the world.

- **Key Concepts in 5G (Chapter 6)**

This chapter proposes two key concepts for 5G: "Satisfaction of End-to-End (E2E) quality" and "Extreme Flexibility." "Satisfaction of E2E quality" means providing every user access to any application, anytime, anywhere, and under any circumstance. "Extreme Flexibility" is the communications system which will allow 5G networks to always achieve E2E quality.

This chapter identifies two key technologies necessary to support the wide range of use cases expected in the 5G era through "Extreme Flexibility". The first is an "Advanced Heterogeneous Network", which will include multiple technologies far in advance of previous heterogeneous networks. The second is "Network Softwarization and Slicing", which will make networks easier to upgrade and maintain.

In addition, using the ITU-R vision report M.2083-0 as a base, typical use cases (high reliability, ultra-low latent communications, large scale communications, advanced mobile broadband) with examples of what technology and requirements will be needed to make these use cases a reality is discussed.

- **Typical Usage Scenarios of 5G (Chapter 7)**

This chapter considers future market trends and user trends discussed in section 3, this section first surveys some examples of new usage scenarios, which are expected to realize by 5G, and categorizes them into four facets; 1) Entertainment, 2) Transportation, 3) Industries/Verticals, and 4) Emergency and disaster relief.

It further analyses the usage scenarios and develops the list of required capabilities of individual usage scenarios. It finally provides key items of 5G capabilities for deriving overall 5G requirements in Chapter 8.

The section also gives an insight of "dynamic approach" into nature of 5G capabilities which must dynamically change corresponding to the wide variety of 5G usage scenarios.

- **Requirements for 5G (Chapter 8)**

This chapter describes the requirements related to radio access network, front-haul/backhaul and communication networks. 5G systems should include "Extreme Flexibility", in order to satisfy the end-to-end quality required in each use scene even in extreme conditions. End-to-end context in the ICT environment includes not only

UE-to-UE, but also UE-to-Cloud, which implies that the technology focus on flexibility extends beyond 5G radio technology to the backbone networks.

- **Spectrum Implications (Chapter 9)**

To realize the "Extreme Flexibility" of 5G, it is necessary to utilize all frequency bands, including both the lower ranges (below 6GHz) and the higher ones (above 6GHz), while considering the different characteristics of each frequency band.

The first section of this chapter will describe the roles of both lower bands and higher bands, and the following section will focus on the evaluation of preferable frequency bands in the range between 6 and 100GHz. The results came from a study that includes three stages of evaluation, i.e. 5G intra-system, inter-system, harmonization point of view, respectively. The resulting preferred bands from the results of Stage 2 are then discussed.

- **Overview of 5G Technologies (Chapter10)**

This chapter overviews the following two chapters, in which several key technical enablers are discussed.

- **5G Radio Technologies (Chapter11)**

This chapter discusses promising radio access technologies in order to realize 5G system. The subsections contain information on the latest radio access technologies embraced in [1] or newly introduced technologies. The 5G communications system should be constructed by selecting, combining or modifying these technologies in order to make 5G systems work in each use case.

- **Network Technologies for 5G (Chapter12)**

This chapter describes network technologies for 5G. Based on the guiding concept "Network Softwarization", which elaborates the overall transformation trend including Network Functions Virtualisation (NFV) and Software Defined Networking (SDN), technology focus area is identified as the result of study in the network architecture group of 5GMF. The brief description of the area and the associated technical issues are described in the following sections.

- **Future business and services (Annex)**

This annex introduces the perspectives of future business and services for reference, using market trends and future capabilities.

Reference

[1] "Mobile Communications Systems for 2020 and beyond," ARIB 2020 and Beyond Ad Hoc Group White Paper, Oct.2014 (URL: http://www.arib.or.jp/ADWICS/20bah-wp-100.pdf).

2. Objectives

The primary objective of this white paper is to identify the key concepts and key technologies to realize 5G. It is also the objective of this paper to send messages and proposals proactively to the outer world surrounding 5G, including industries and potential partners to create the future society together through 5G technologies and ultimately to promote and to stimulate 5G development.

3. Market and User Trends related to 5G

This chapter describes the result of survey on the current various services realized by ICT and analysis of the market and user trend for each service, and our consideration of the forecast of services as the introduction of 5G.

3.1. Shift from PCs to Devices such as Smartphones and Tablets and wearable terminals

Internet traffic initially began to increase as the number of PCs connected to the internet increased. The emergence of reasonable flat-rate internet connection services lead to rich content, such as video, which lead to further and further increase in internet traffic.

The past few years, however, internet traffic has increased with the dramatic rise of the use of smartphones, especially among young people. These devices, unlike PCs, allow people to be connected to the internet 24 hours a day with something they hold in their hands. While delivering video and images to smartphones has contributed to the increase of internet traffic, just as it did with PCs, the rise of social media has also led to an increase of internet traffic. (See Fig. 3.1)

Smartphones have become indispensable for young people now that they are being used by those in their teens and twenties to strengthen their relationships with each other. This generation will bring this communication style with them as they enter the workforce by the 2030s. 5G, which is being introduced in 2020, will be fully implemented by then, meaning most people will have a 5G compatible smart phone in their possession. When we consider the use scene for 5G, which will have a maximum possible speed of 10Gbps, we will need to consider that this generation will be the main users of these services. (ITU-R M.2083 states that 5G will require a minimum user experience data rate of 100Mbps)

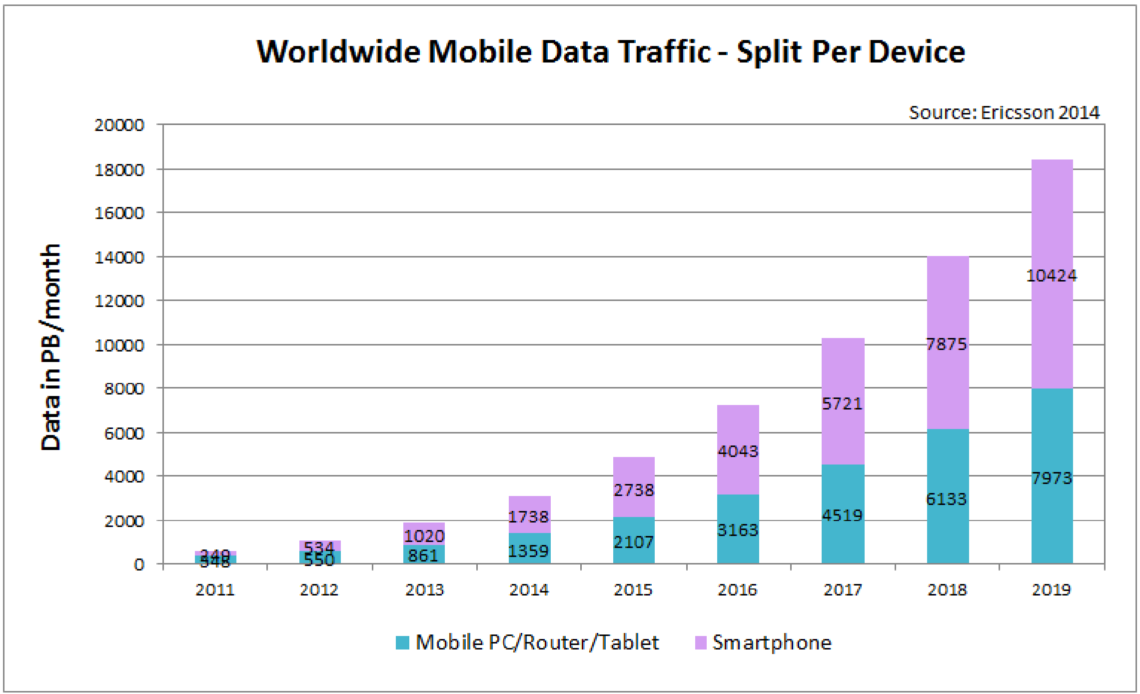


Fig.3.1 Worldwide Mobile Data Traffic – Split Per Device

Source: Ericsson 2014

3.2. Increase in Location-Based Services

Both the private and public sectors are developing services that use GPS and digital maps, which have become important parts of today's society. These services are expected to continue to develop and evolve as they being to use high speed mobile and cloud based services enabled by 5G.

For example, current digital maps are modified to when new information is delivered to a device. The best example is Google Maps, whose smartphone application can not only be used while walking, but has also become popular as an alternative to a dedicated car navigation system. (See Fig. 3.2) In the future, it won't only be people who are using electronic maps, however, but also self-driving vehicles will be able to function when high speed data transmission allows for real-time information updates. When this becomes a reality, digital maps will be able to be dynamically updated, including information on traffic jams and road construction. The ultra-low latency of 5G will enable these maps to be dynamically updated in real-time.

Municipalities also need hazard maps that can be updated in real-time to be used in times of disasters or evacuations. 5G will also assist in creating maps that will change in real time in response to disaster information, just like the dynamic maps self-driving vehicles will use.

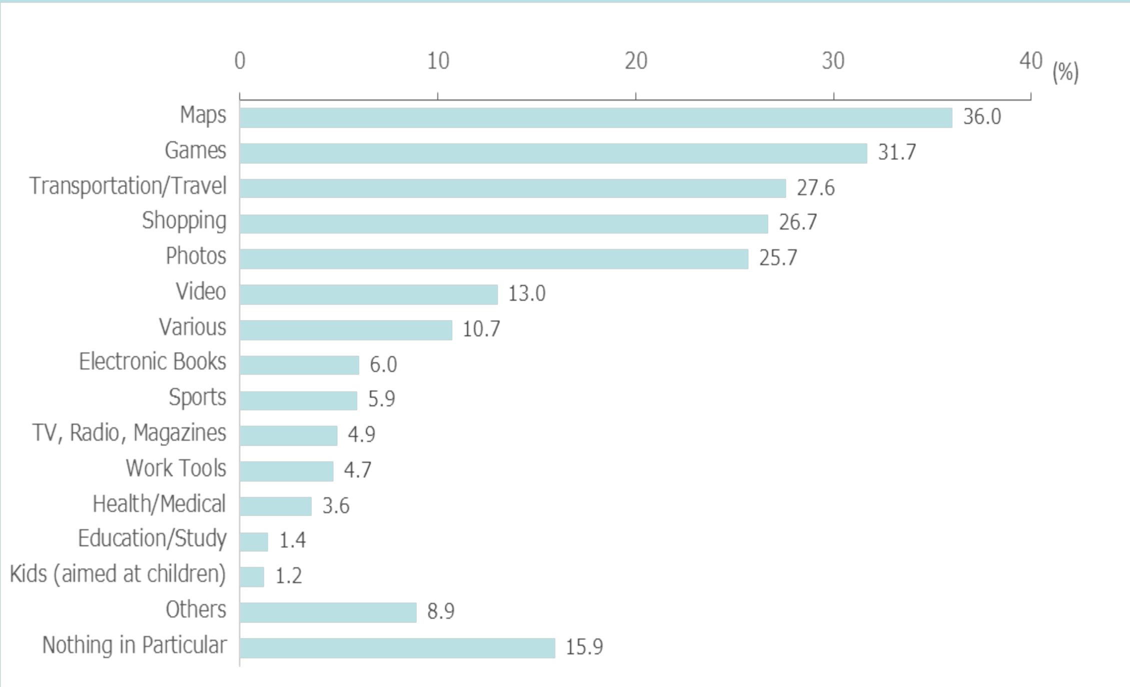


Fig.3.2 application usage

3.3. Forefront of a new way of building human relations with a focus on women

The main use of smartphones today is the exchange of real-time information between people through social networking services (SNS), becoming a tool that supports human communication. Smartphones themselves, devices that are never out of reach, are used more by women than men and have become indispensable devices for women and young people today. This is different from when PCs were the main way to connect to the internet, when more internet users were men rather than women. Therefore, internet applications will change as the typical user also changes.

Women are more likely than men to create their own content, for example taking and posting photos, on social media. Rather than sharing them publicly on blogs, however, they are more inclined to share them with only their friends on social media sites such as Facebook[1] and Instagram[2], or with only their family and close friends on LINE[3], one of the most popular real time messaging service in Japan, or through email. Teenage girls are able to stay closely connected with their friends, knowing exactly what each other is doing, for 24 hours a day, as well. Many issues, such as protecting the privacy of individuals when sending their information across data networks, will require a deep understanding of how these applications are used. In the end, we can say these women who are using smartphones to stay connected to many different online communities are at the forefront of a new way of building human relations. (See Fig. 3.3)

The way teenage women are using smartphones gives us a glimpse into the future of their use. Already, we can see that SNS is providing value to individuals through the use of financial technology. In terms of identity, an individual can participate in online communities using several SNS accounts, allowing them to make mistakes while trying out new identities and following new possibilities. These changes are breaking down old systems and moving society towards the building something new.

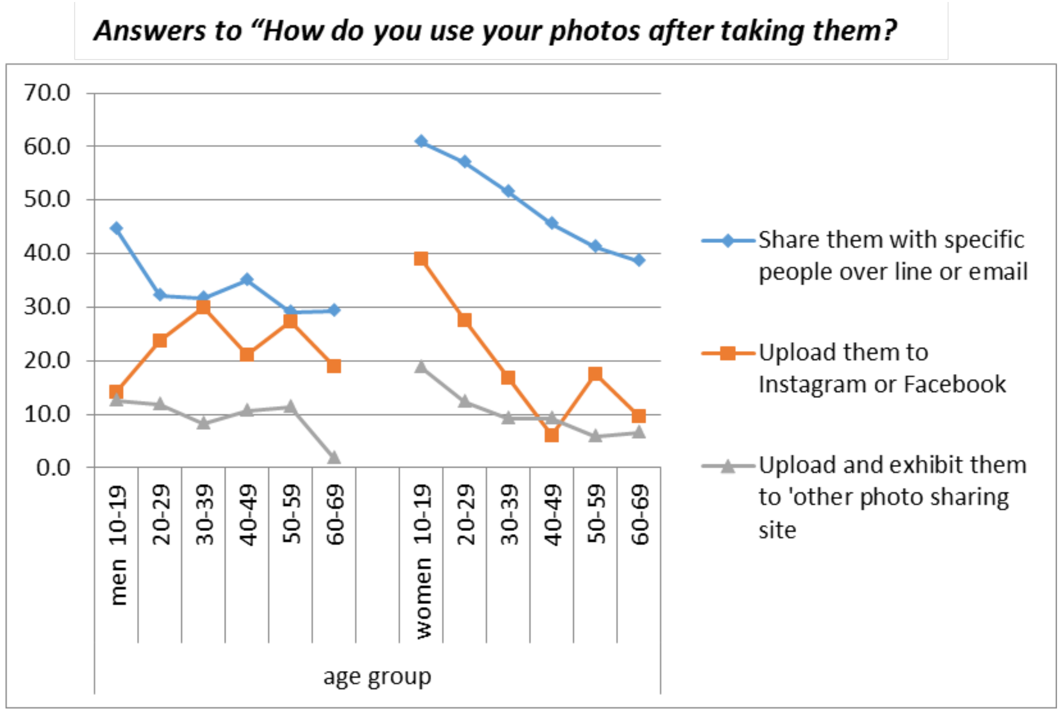


Fig.3.3 Photo usage

3.4. Introduction of the Sharing Economy

With the advent of 4G technology, it has become normal for people to share content. Among young people, buying and borrowing of things directly from each other has become common as well. With 5G, services dedicating to sharing not only information but real objects will become the norm. The destructive new entrants to the taxi service market, Uber[4], allows for cars to be shared in a neighborhood in a way that is cheaper than taxis. LINE also began a new service which allows people to discover nearby taxis through its system. During the Tokyo 2020 Olympic and Paralympic Games, the ban on using private residences for lodging will be lifted so that they can be used as part of the hotel infrastructure. Once 5G goes online, people will be able to search in real time for objects that other people can lend to them. For example, if someone needs to drive a car, they will be able to look for a car that is available and then borrow it for a specific period of time. (See Fig. 3.4)

Now, people who are looking for information on something they want use individual auction sites. With 5G and mobile edge computing technology, the base for this search becomes the edge cloud, allowing people to search the entire world for their needs. Mobile edge computing technology will be the representative service in this sharing age.

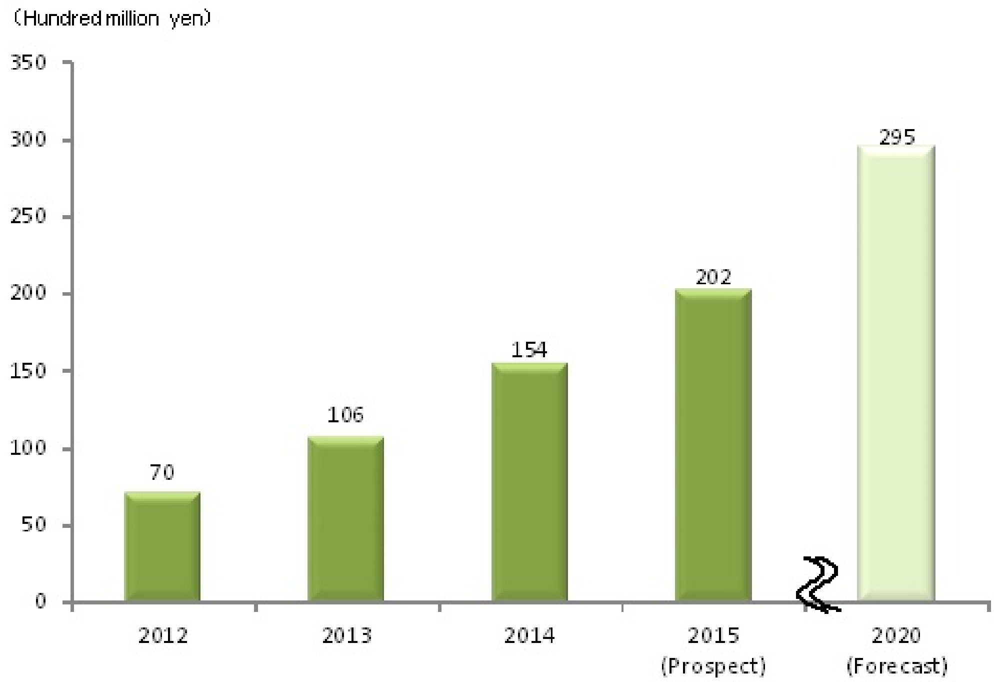


Fig.3.4 Transition and Forecast of Car Sharing Market Size

Source : YANO Research

3.5. Introduction of Artificial Intelligence and Robots

It has become normal for smartphones to be controlled by voice commands such as OK Google and Hey Siri. With the increasing ability of Artificial Intelligence (AI) applications, which reside in cloud services, simple conversations needed for information services and telephone help desks can be done by AI. Recently, Softbank helper representatives have started to use humanoid robots. These humanoid robots will be able to be powered by AIs residing in the cloud as opposed a CPU inside the robot. It is expected that the low latency of 5G will connect AIs and robots, allowing them to communicate with human beings in real-time. It is also expected that, in addition to humanoid robots, specialized household robots to be used for cleaning and daily chores will also become parts of people's daily lives. In terms of industrial machinery, there are now remote controlled drones and robots. The ability for cloud based AI to pilot drones will greatly depend on a stable connection to a network, which will be provided by 5G's ultra-low latency. 5G's capabilities will be used greatly as the use of robots become as common as smartphones around the world. (See Fig. 3.5)

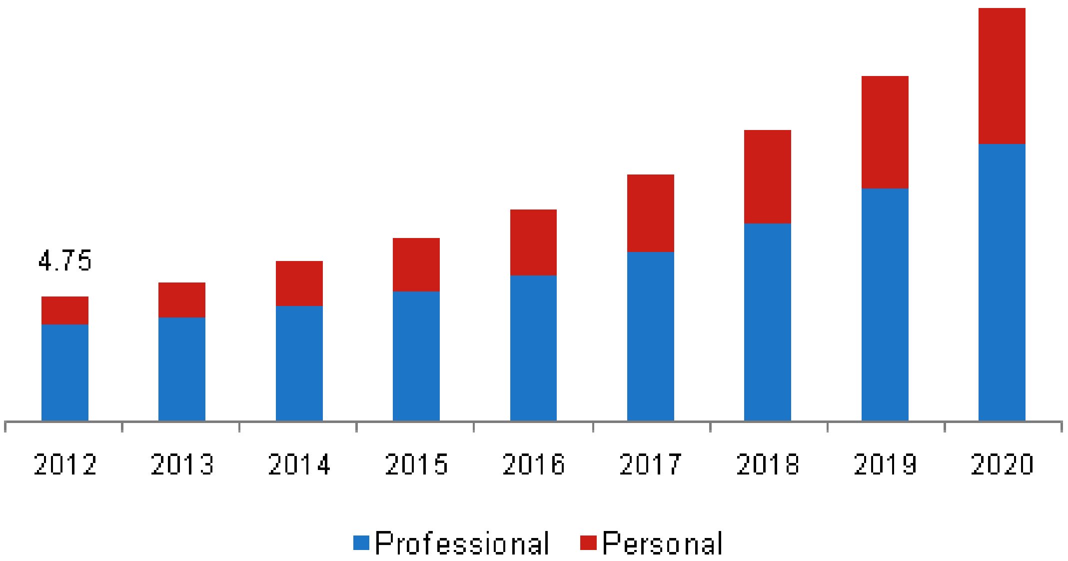


Fig.3.5 Global service robotics market by application (USD Billion), 2012 - 2020

Source : Grand View Research

3.6. Self-Driving Vehicles

Automated driving can be divided into four levels. Level one is that the car has break assistance or cruise control functions to assist the human driver when necessary. Level four, on the other hand, is that the car is driving itself without any help of the human driver. As the levels increase, the need to be connected to a network increases as well. In the future, all cars will be connected to a network. Even automated driving at level one, which does not seem to be necessary to be connected to a network to assist drivers, will benefit from being connected to a network which will allows for navigation systems to be dynamically updated and breaks operations to be updated with data about the car itself. Level two and level three automated systems, since they sometimes require the car to drive autonomously, will need to be connected to a network in order to understand local road information quickly and make judgements about the driver and the car in order to make proper decisions. Since these decisions can be made in the cloud, data needs to be quickly transmitted between the car and the cloud, meaning 5G's low latency will be an important factor. In the end, however, any delay in the connection between the base station and the car will not be a concern, since the connection between the data center where information will be processed and the car is expected to be an end to end delay guaranteed network.

5G can be used to implement automated driving services in levels one to three. In order to have fully automated level four driving cars, 5G will be used for people who don't drive. For example, a level four automated driving car can be ordered with a smartphone. These vehicles, examples include driverless taxis or elderly care pick up service vehicles, will bring the customer to their destination. While enjoying the drive to their destination, the customer is then freed from operating the car and can enjoy the free time provided to them on their trip. (See Fig. 3.6) From 2020 when automated cars will be allowed on expressways, the organizations that manage expressways may provide automated car users with entertainment such as films to enjoy during their trips, —— A car moving on an expressway, in order for a 4K movie to be delivered, will use 5G handover services. In these ways, the commute time inside automated cars will also utilize 5G services.

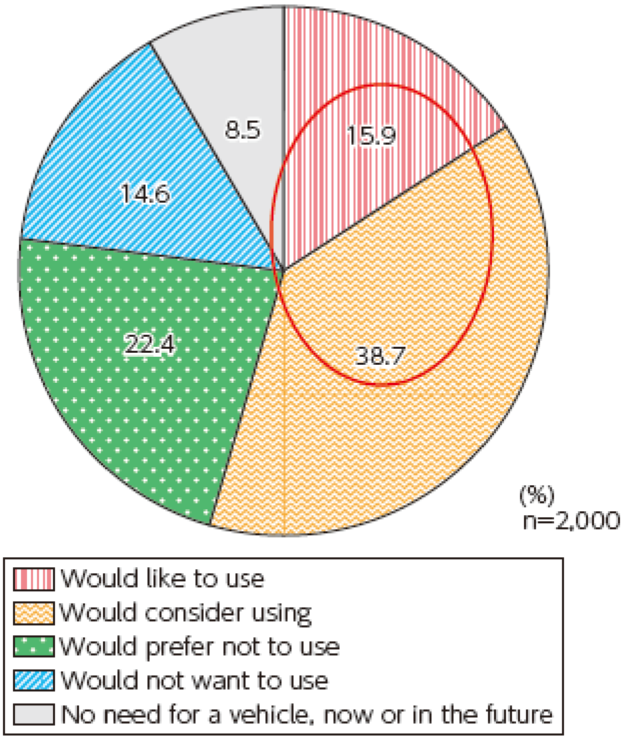


Fig.3.6 Intention to use autonomous vehicles

3.7. Internet of Things (including industry, wearables, and agriculture)

The Internet of Things (IoT) will be one of the foundations of the 5G, but a large amount of objects will not be directly connected to mobile networks via 5G, but they will be connected by a mix of radio services. For example, NB-LTE (Narrow Band LTE), introduced with 3G, can provide 5G IoT-like services. Active Tags, BLE (Bluetooth Low Energy), and Wi-Fi will all provide access points for other IoT services. Germany has proposed an Industry 4.0 with an ICT that is engineered for industry. This does not mean only factories, but for a whole global supply chain, from procuring raw materials to completing finished goods, all tracked in real time. In this situation, the location of goods, whether on container ships, trucks, or trains, and where they are going will be known using the power of 5G. Active and passive RFID tags will be placed on both raw materials and finished goods, with access points on individual pallets and containers, all connected by 5G.

As for individuals who use Bluetooth keyboards and headsets, as well as smart watches and other devices that have the ability to track vital data, these will be connected to 5G through smartphones. These wearable devices will be connected to 5G smartphones and will be working in tandem with Bluetooth devices in order to collect and store data in the cloud, which can be used with cloud based applications. (See Fig. 3.7)

Currently in the agricultural industry, which is driven mostly by manual labor, ICT has been used to mainly expand the number of sales channels. That can now be expanded in order to increase productivity, by setting up field sensors to track variables such as the composition of the soil, moisture and rain levels, and solar radiation. In addition to keeping track to this data for planting, this information can be used when selling products, thereby adding more value to the crops. 5G will not be left out of these IoT devices, as well as it will be used to connect the various field sensors together.

There are many uses for IoT, but there are major differences in how people use IoT versus how people use smartphones and mobile phones. IoT systems should have a longer life for two aspects. First aspect is that IoT system should have longer life as the wireless systems. ICT often changes, and 5G will eventually change when the next generation comes along. However objects connected to the IoT by 5G services do not follow this lifecycle. They will have backward compatibility from 5G back to 4G, and will be able to connect to the next generation mobile system, as well. Another aspect is that objects connected to the IoT should consume very low energy to work in the environment that IoT will be deployed. IoT objects may be deployed in places that cannot access a stable power supply of 100v/200v, and so they will need to run on battery power. In this instance, 5G modules will have to be minimized in order to conserve power.

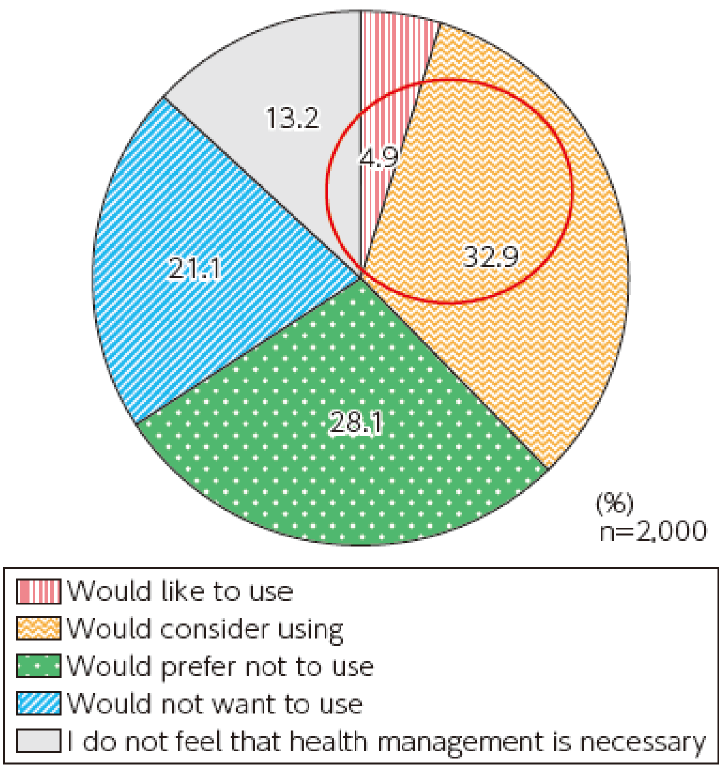


Fig.3.7 Intention to use health management services that rely on wearable devices

3.8. Changes in the Work Style

Holding video meetings and connecting to people via email is now possible with the internet, so the need to live in urban areas has decreased. Additionally, there is no need for office managers to assign individual desks for each person. As mobile networks are established, this kind of work environment will increase, leading to large changes within companies.

In an aging society, many people will need to assist their elders, meaning many people may be required to leave a job to take care of someone at home. The lack of day care facilities for children will also force many parents to reluctantly leave their full time jobs to take care of their children. However, highly skilled workers will be able to use 5G networks to work from home and thus return to their jobs in these situations. In addition, although now it is normal to work at one job for one company, it will become normal to be able to work for several companies using the tools provided by 5G. These tools can be used to help reverse the trends in Japan towards a society with a low birth rate and an aging population which now exist today. (See Fig. 3.8)

As mobile workplaces using 5G spread, people will no longer need to be at work early in the morning. People will be released from their commutes, using that time to work and relax. This will also affect other industries infrastructure, as well. Since rush hour commuters will decrease, railway company costs for maintaining capital infrastructure will decrease. Not only railway companies, but roads and office and other large scale infrastructure systems will be used less than before. With populations continuing to decrease outside of large cities, a new Japan can be built using mobile networks that avoids this increasing centralization of people in cities.

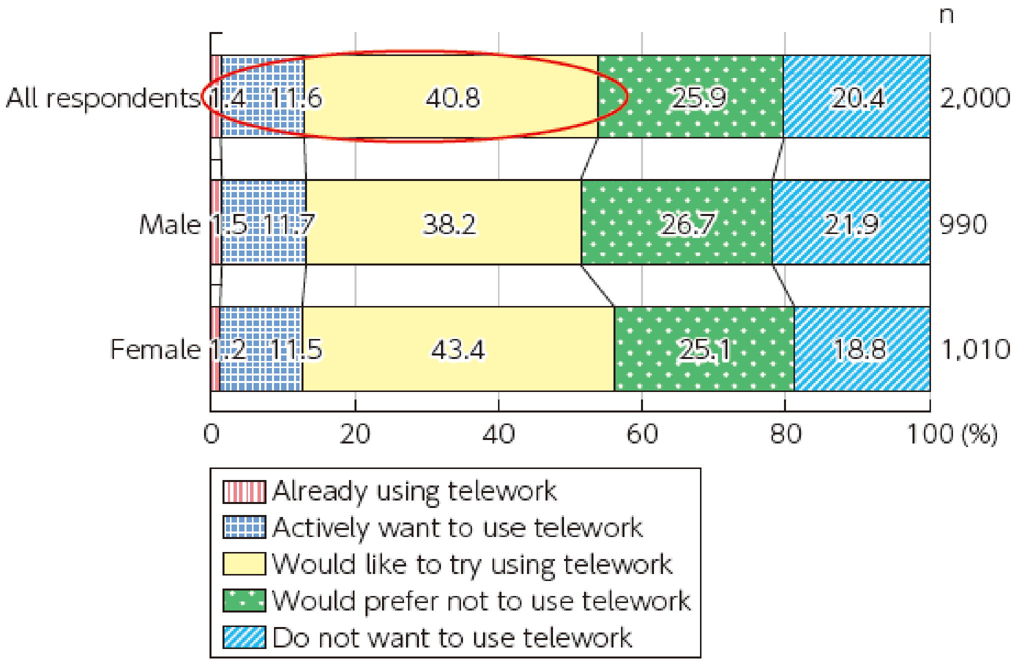


Fig.3.8 Intention to use telework among employed people (by gender)

3.9. Acceleration of Fintech Services

Finance and technology companies are working together to provide new financial services for both individuals and companies. These services include those for payments, remittances, asset management, investments and lending. On the technology side, AI technology such as distributed financial transactions using blockchain technology and AIs using deep learning are moving ahead. Blockchain technology will not only dramatically decrease the infrastructure costs for financial institutions but also create opportunities for new, unknown services and players to appear.

Once 5G is established, mobile payments and mobile remittances will be a part of everyday life as the use of paper money will continue to decrease. 5G infrastructure, build to withstand any disaster, will be able to continue to function, and with these services will allow donation money to be collected and sent directly to disaster areas, benefiting those communities in need. Additionally, AIs which are analyzing equity investments can be contacted any time, acting as investment advisors, on mobile devices and any changes in the investment environment can quickly be relayed to the investor and adjustments made accordingly in real time. This advice can be acted on in real time, with 5G's low latency, though an individual's device, providing positive benefits to individuals.

3.10. Penetration of Peer to Peer Service

From 2001 one major change in the high speed broadband internet was the quick rise of peer to peer file sharing software, like BitTorrent or Winny, and with it a corresponding rise in broadband traffic. More recently, virtual currencies using blockchain technology like bitcoin have not only impacted financial services, but also the greater field of IoT devices, smart contracts, copyright issues concerning digital rights management, and authentication processes.

Blockchain technology is an extension of peer to peer technology, consisting of nodes that use a consensus system in order for each node to act on its own. For example, one use case is the deployment of a large amount of IoT sensors. Each sensor is connected to the client server by a contract agreement, but each node can act independently, increasing efficiency while decreasing overall system costs. These multiple nodes can work together, becoming the building blocks for a stable 5G network, creating a situation where the computer and network are not separate but are acting together as one,

Note

[1] Facebook
A social networking service on the managed internet provided by Facebook, Inc.

[2] Instagram
A free online mobile photo-sharing, video-sharing, and social networking service developed Facebook.

[3] LINE
A social networking service provided by LINE Corporation.

[4] Uber
A car allocating website and application provided by Uber Technologies Inc.

4 Traffic Trend

4.1 General

This chapter provides an analysis of the latest communication traffic trends. For the past years, considerable increase of communication traffic has been observed and several estimation studies consistently forecast that the increase would be continued to the next decade. In addition, new traffic nature different from ever happened one could come in considering new traffic types generated in variety of use cases with variety of 'connected things' or 'connected services'.

As the consequence, it could be concluded that 5G mobile communications system should handle these enormous increasing communication traffics as well as new traffic nature due to new traffic types in proper and efficient manner.

4.2 Communication traffic growth and traffic nature trend

4.2.1 Communication traffic growth

The general trends up to the year 2014 were analyzed in [1]. The data collected consisted of the details on communication traffic including, wired (or fixed) communications. Figure 4.2.1-1 represents communication traffic growth in Japan since the year 2015 [2]. As can be seen, downstream fixed communication traffic has shown enormous growth in recent years. Upstream fixed communication traffic has also relatively increased. Mobile communication traffic, both upstream and downstream, has also shown a large increase in recent years. A certain portion of mobile communication traffic has come at the expense of fixed communication traffic.

The increasing rate of communication traffic within the last twelve month period is depicted in Figure 4.2.1-2. Increase of mobile communication traffics, both upstream and downstream, has become rather stable without having lost the three-fold increase in growth over the previous three years. Fixed communication traffic also shows an increase in demand.

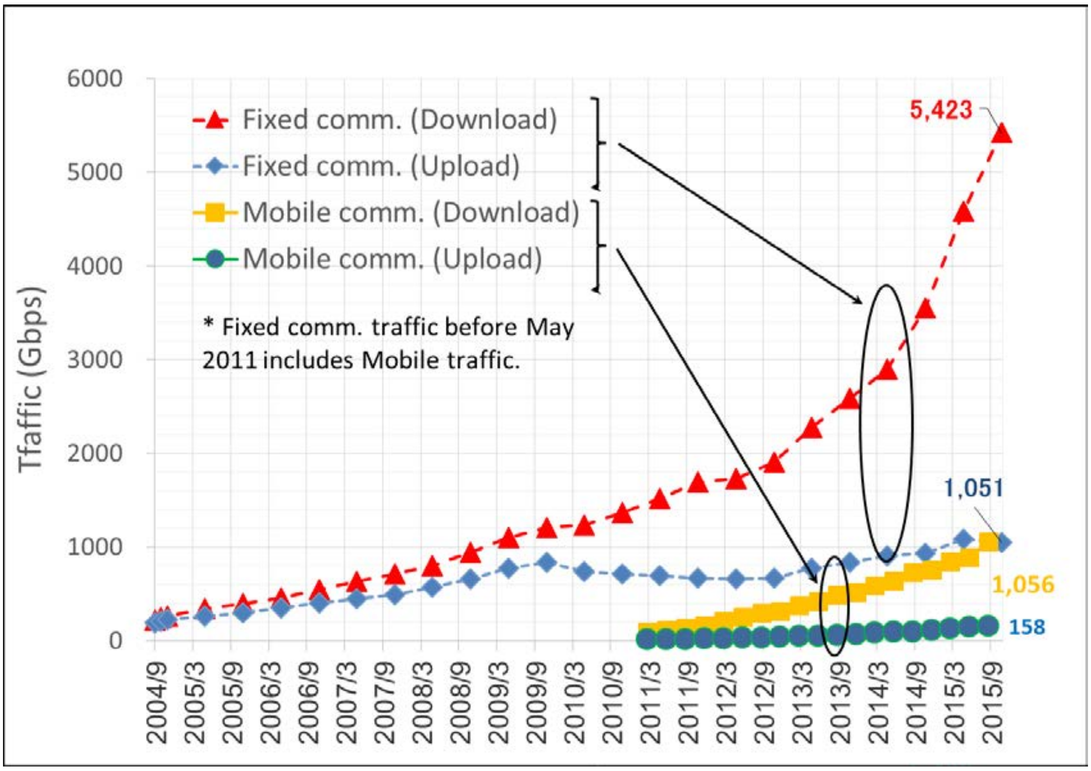


Figure 4.2.1-1 Communication traffic growth in Japan [2][3]

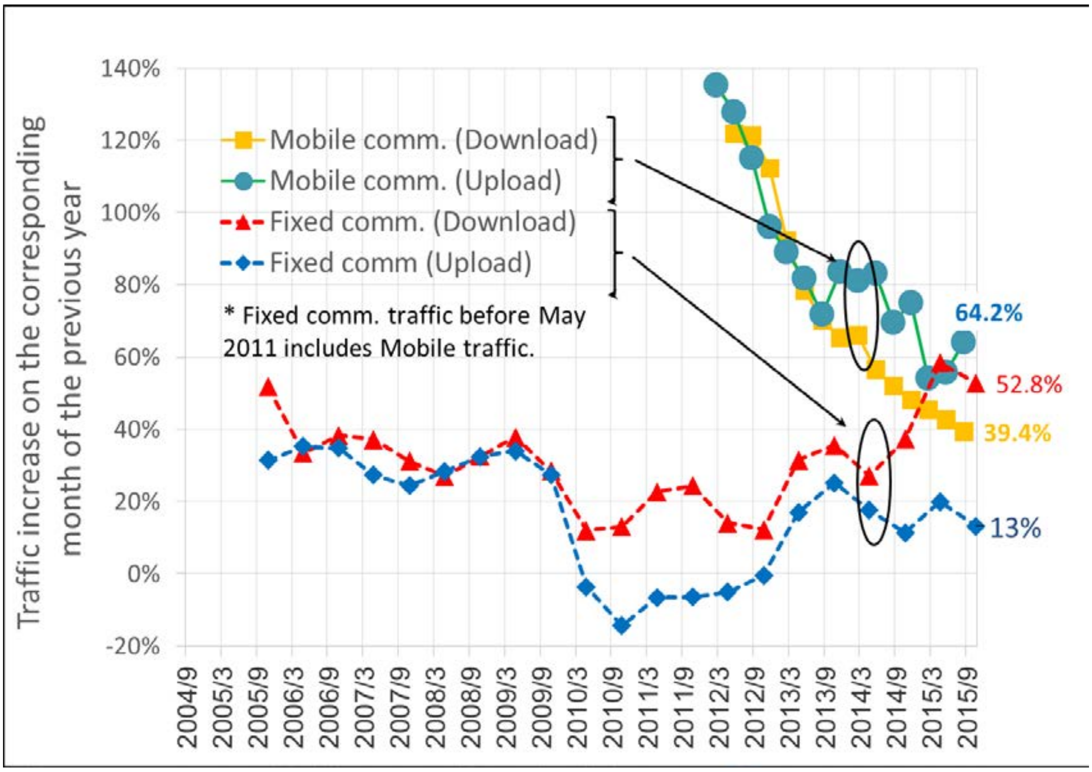


Figure 4.2.1-2 Communication traffic growth rates in Japan [2][3]

Global (IMT) traffic estimation to 2030 was captured in [4] in which three estimation results from different entities were summarized. As an example, a chart in its chapter 5.3, representing the cumulative compound annual growth rate (CAGR) relative to 2010, is shown in Figure 4.2.1-3. The chart forecasts considerable increase of communication traffic towards year of 2030. Similar estimation or observation for global traffic trends can be found in other documents such as [5].

Although this increase in the rate of communication traffic may be affected by a variety of economic or social factors though, the fundamental trend will be generally maintained over the next decade. Thus 5G mobile communications system should be prepared to handle this enormous increase in communication traffic properly.

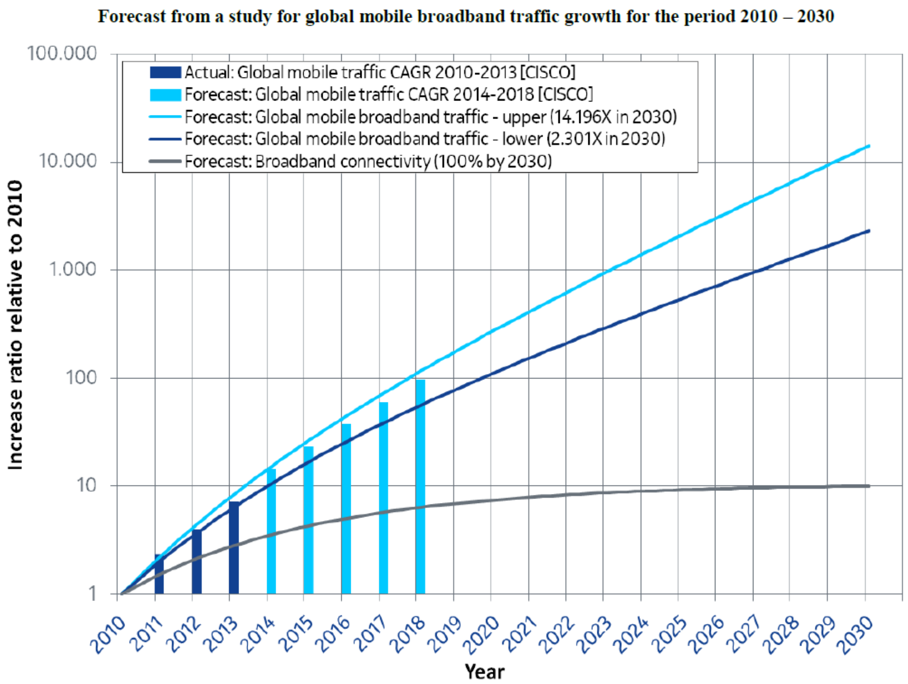


Figure 4.2.1-3 A global traffic estimation to 2030 (in [4], involving original data in [6])

4.2.2 Communication traffic nature

In terms of the types of communication traffic being handled, ordinary voice communication traffic has been relatively stable, even as it gradually decreases, as shown in Figure 4.2.2-1 [7]. On the other hand, data traffic between objects directly, e.g. IoT traffic, has been drastically increasing. (Figure 4.2.2-2 [8])

As has been discussed in the previous section on quantitative increase of communication traffic, nature of communication traffic will also be changing as a variety of use cases related to 'connected things' increases.

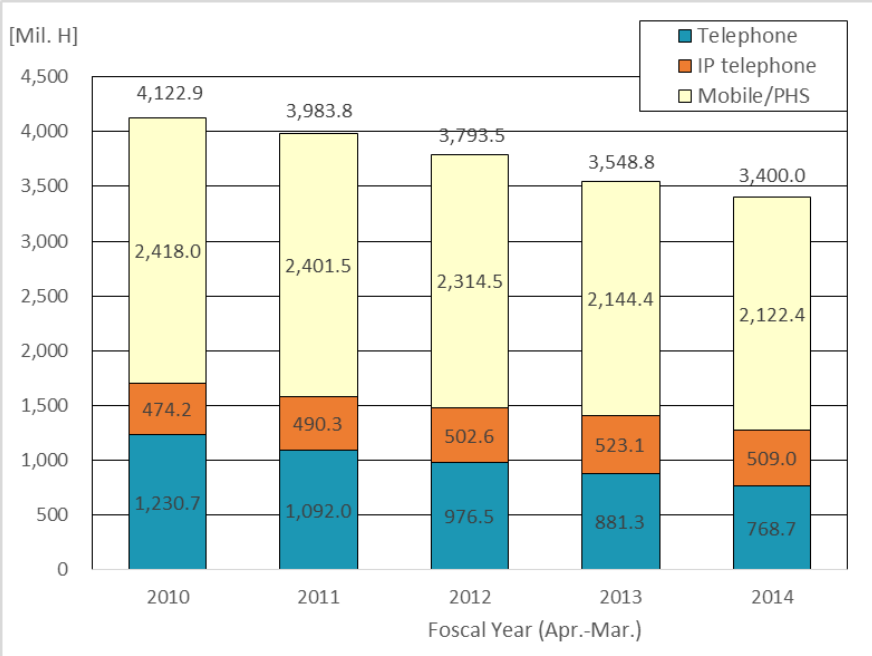


Figure 4.2.2-1 Voice communication traffic trend [7]

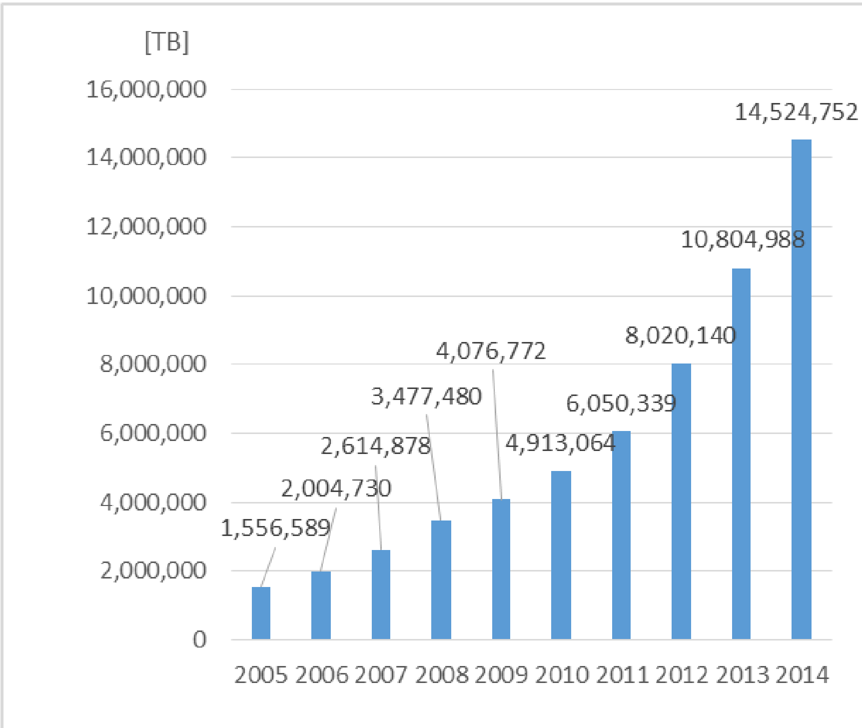


Figure 4.2.2-2 Data traffic (Service industry, ICT, transport, real estate, money & securities, commercial services, utilities, construction and manufacturing) [8]

References

[1] "Mobile Communications Systems for 2020 and beyond," ARIB 2020 and Beyond Ad Hoc Group White Paper, Oct.2014.

[2] "Status of the mobile communications traffic of Japan (Sep. 2015)," Information and Communications Statistics Database, Ministry of Internal Affairs and Communications of Japan, Nov. 2015.

[3] "Aggregation and Provisional Calculation of Internet Traffic in Japan (as of Nov., 2015)," Ministry of Internal Affairs and Communications of Japan, Mar. 2016.

[4] "IMT Traffic estimates for the years 2020 to 2030," ITU-R Report M.2370-0, July 2015. (http://www.itu.int/pub/R-REP-M.2370)

[5] "Ericsson mobility report," Ericsson AB 2015. (http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf)

[6] "Global Mobile Data Traffic Forecast Update," Cisco Virtual Networking Index, February 2015. (http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html)

[7] "Voice communication traffic trends," Ministry of Internal Affairs and Communications of Japan, Dec. 2015.

[8] "Traffic of big data flow estimation and investigations on usage of the big data," Ministry of Internal Affairs and Communications of Japan, 2015.

5 Cost Implications

5.1 General

This chapter discusses the cost of mobile communication systems and analyzes from the perspective of several ‘5G’ related use cases. The fundamental cost implications of ‘5G’ were analyzed in [1] where every element of a mobile communication system was analyzed in terms of CAPEX or OPEX. The analyses were made with a focus on the domestic market of Japan, in light of demographic and survey data as well as local market indexes. Since the market in Japan is one of the leading markets in the world, these case studies may be of use when considering markets in other locations around the world.

Section 5.2 presents a case study on communication traffic as it relates to mobile broadband. While existing mobile communication systems such as LTE has already experienced many of the situations discussed, more sophisticated and enriched services including device to device communications are expected in the age of 5G. Section 5.3 discusses scenarios related to coverage in sparsely populated areas. As with communication traffic, considerable efforts have been made up to now and existing mobile communication systems cover more than 99.97% of the total population of Japan. The number of people in Japan that is not covered by current mobile communication systems is estimated to be less than 39 thousand [2]. Accordingly, the expansion of service areas themselves would be about on-going improvement and the task of ‘5G’ in this regard is to be able to provide reasonable services at a reasonable cost even in sparsely populated areas. As in the first use case, devices deployed in sparse manner that provide device to device communications should also be taking into account. Section 5.4 considers the dynamics of communication traffic. The case study presented is the enormous flow of commuters in the mornings and evenings in the context of daytime population density vs. night time population density. The ‘5G’ system needs to be able to cope with a large variation of population density, including a large daily flow of commuters. Since people may carry more mobile devices than ever in the ‘5G’ era, the ratio of daytime and nighttime communication traffic may become larger and the volume of mass communication traffic along commuting routes may rapidly increase, as well.

Since one important framework of ‘5G’ is to be able to cover a variety of use cases in a cost effective manner, a ‘5G’ system should be designed to be as flexible and scalable as possible. Accordingly, in most of the cases (especially use cases specific to ‘5G’), it would be useful to apply specific technologies, configurations or operating method suitable to these certain use cases and combine these elements into a unified communication system, rather than to seek generic, common and robust technologies covering each and every use cases overall.

5.2 Costs per communication traffic aspect

As described in the previous section, the volume of mobile communication traffic has been increasing rapidly thanks to the expansion of mobile broadband applications. This growth will continue over the next decade. In this section, mobile communication system costs are analyzed in the context of traffic volume versus revenue of the mobile communication operators as well as users' expense.

The charts in Figure 5.2-1 estimate annual traffic volume using data from [3]. Simple linear interpolation is applied to the original estimated mobile communication traffic for every three months' interval and then added up to derive the estimated annual traffic volume. The results, summarized in Figure 5.2-2, show growth in traffic, both overall and on a per subscriber basis traffic (also derived from [3]). As can be seen, overall traffic grew five times between the years of 2011 and 2013 and four times during that same time period on a per subscriber basis.

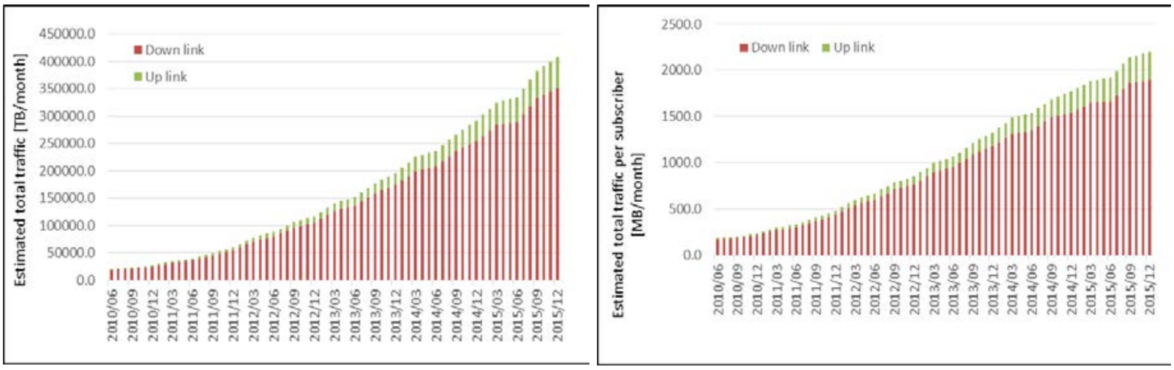


Figure 5.2-1 Estimated communication traffic growth in Japan (Derived from [3])

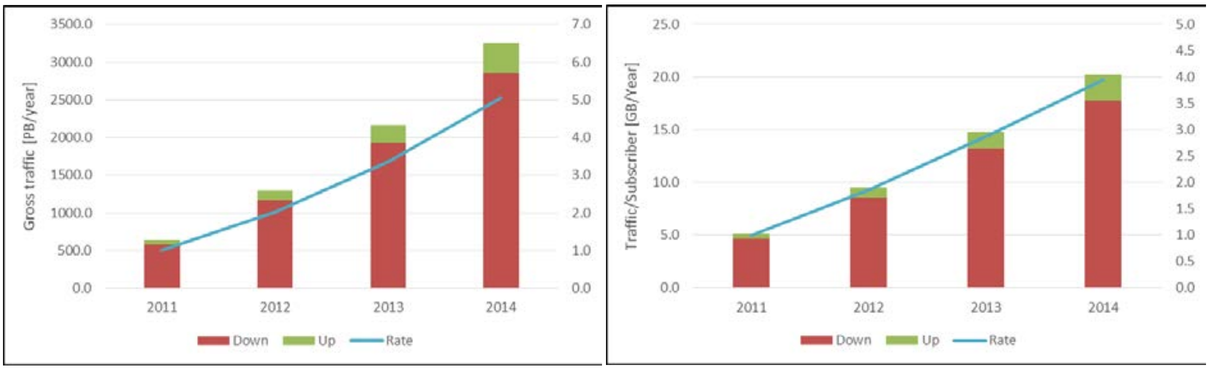


Figure 5.2-2 Estimated communication traffic growth rates in Japan (Derived from [3])

In the meantime, user expenditures are fairly stable as shown in Figure 5.2-3 (Derived from [4]). The total average expenditure of households has been gradually increasing However; rate of increase is around 50% (1.5 times) over the 10-year time frame between 2004 and 2014. Expenditures in one-person households have remained flat at around 4,000 yen over the last decade.

Figure 5.2-4 (Derived from [4]) represents the increasing rate of household expenditures for mobile communication services from 2011 to 2014. The rate increased about 10% over this period. Compared to growth of traffic over the same of the same period, which was four times per subscriber, the increase in the rate of the expenditure could be considered rather modest.

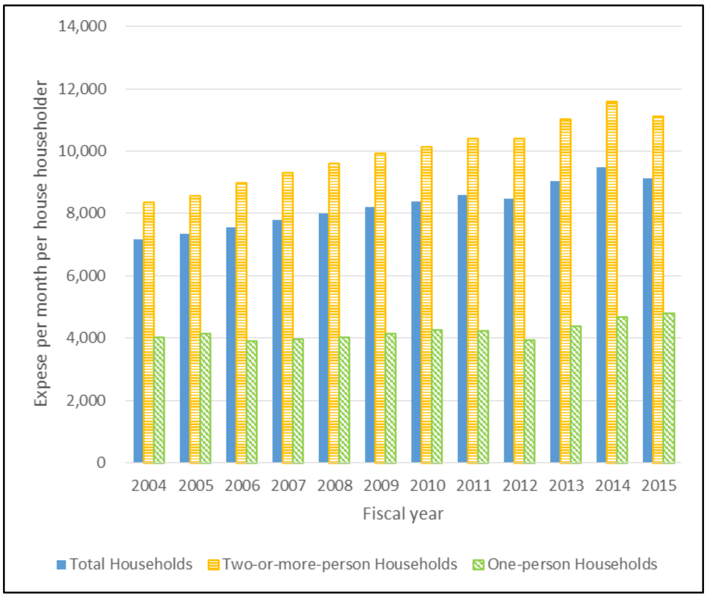


Figure 5.2-3 Householders expenditure for mobile communication services (Derived from [4])

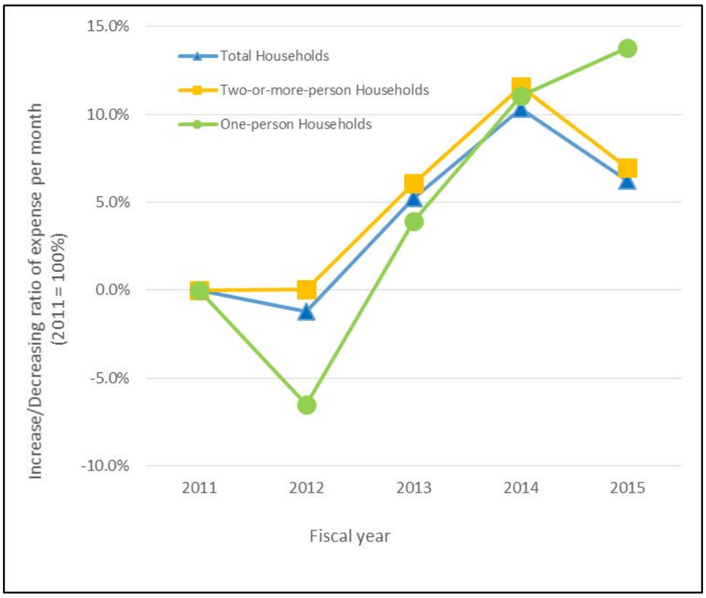


Figure 5.2-4 Relative Householders expenditure for mobile communication services (2011 = 1.0) (Derived from [4])

As a counterpoint to households' expenditures, Figure 5.2-5 (Derived from [5]) represents the total sales of mobile communication operators in Japan. While revenue from data communications services largely increased by a factor of two, income from voice services declined, keeping overall sales revenue constant.

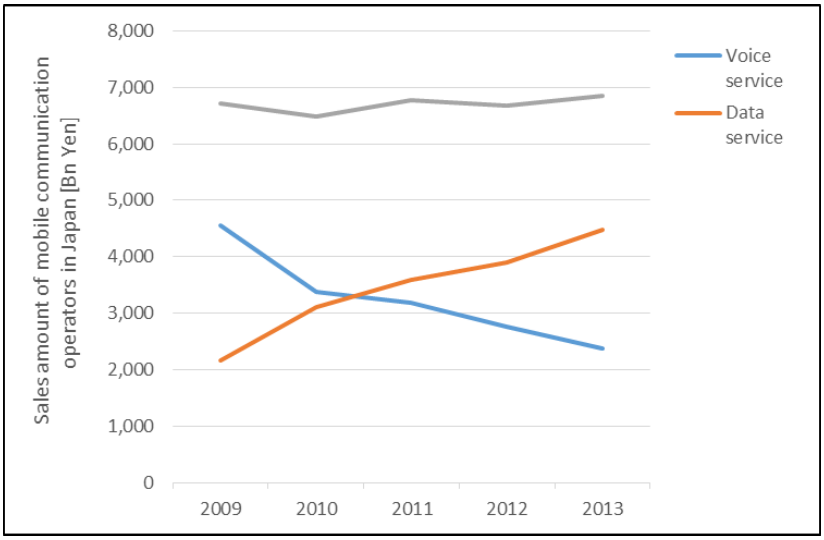


Figure 5.2-5 Sales amount of mobile communication operators in Japan (Derived from [5])

In conclusion, it can be observed that the increase in data traffic in mobile communication services has not contributed to the income of mobile communications operators'. While the increase in communication traffic will continue for the next decade, there is an upper boundary of the capacity for mobile communication systems due to the physical upper limit of communication resources (such as frequency spectrum, frequency efficiency of the radio access technologies etc.). Accordingly, one of the fundamental factors of '5G' will be to provide wider bandwidth services utilizing wider frequency spectrum, for example, without any with only a reasonable cost increase, at most, compared to the new value provided by it.

5.3 User density perspective

This section describes the effects density has on the cost of mobile communication systems. As has been analyzed for optical fiber communication lines and mobile communication networks in [6] and [7], the average cost per individual contract increases as user density is decreases (see Figure 5.3-1 and Figure 5.3-2).

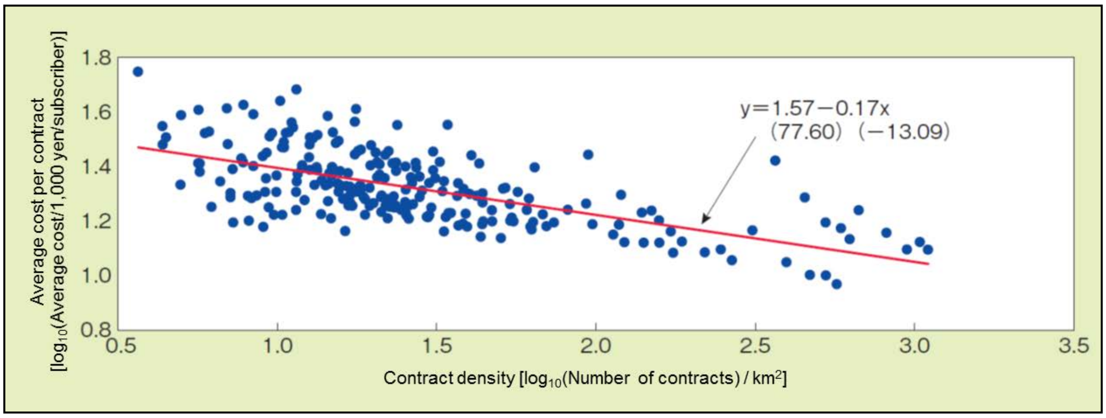


Figure 5.3-1 Contract density and cost of optical fiber networks (English translation of [6])

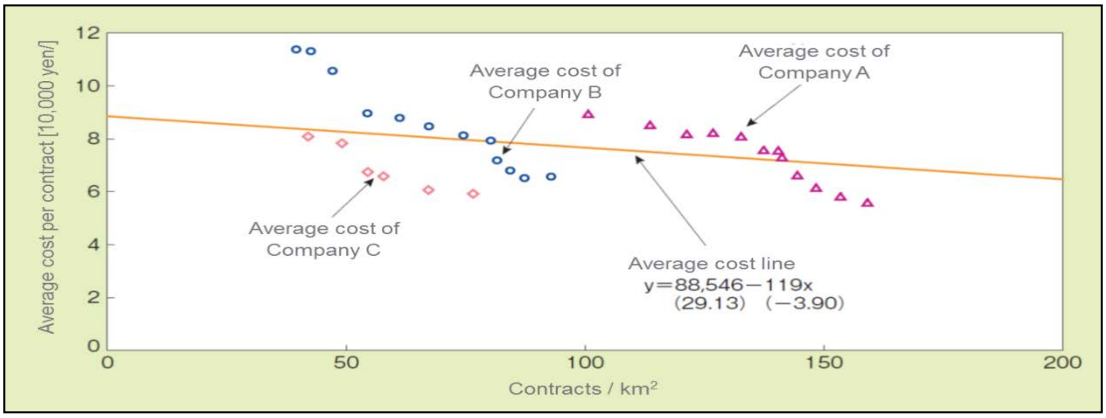


Figure 5.3-2 User density and cost of mobile communication networks (English translation of [6], Summary in English in [7])

In the charts below, approximated costs are expressed by the following equations:

For the average cost per contract of optical fiber lines;

.(1)

Where represents common logarithm of the average cost (unit in 1,000 yen) divided by number of the contracts in question and represents common logarithm of number of contracts divided by the area (unit in square kilo meter).

For the average cost per contract of mobile communication networks;

.(2)

Where represents the average cost (unit in 10,000 yen) divided by number of the contracts in question and represents density of the contracts (unit in per square kilometer).

This can be shown by using the population density of Japan, which is expected to reflect the density of the contractor or subscribers discussed above. To understand the dynamics in daytime and nighttime, two types of demographical statistics i.e. daytime populations and nighttime populations, are used.

Table 5.3-1 (Derived from [8] and [9]) shows the classification of municipal governments within Japan. According to statistics given in [9], there are 1892 municipal governments in the nation as of 2015. In terms of population, municipalities span a few hundred people to several hundred thousand people. In terms of size, the smallest municipality has an area of less than five square kilometers while the largest ones can have an area of more than two thousand square kilometers. This wide range of populations and sizes affects deployments cost as well as operational costs of the communication systems.

Table 5.3-1 Classification of local governments in Japan (Derived from [8] and [9])

Population range (p: Poplulation per local government)	Number of local governments	Gross						Average (per local government)						Population density	
		(a) Area [km^2]	Population					(a') Area [km^2]	Population						
			(b) Daytime	(c) Night time	(d) Delta = (a) - (b)	(e) Daytime / Night time = (a) / (b)			(b') Daytime	(c') Night time	(d) Delta = (a') - (b')	(e) Daytime / Night time = (a') / (b')		(f) Daytime = (b) / (a)	(g) Night time = (c) / (a)
p ≥ 500,000	20	3,060.9	12,513,780	9,985,144	2,528,636	125.3%		153.0	625,689	499,257	126,432	125.3%		4,088.3	3,262.2
500,000> p ≥ 300,000	58	18,144.2	22,147,083	21,337,950	809,133	103.8%		312.8	381,846	367,896	13,951	103.8%		1,220.6	1,176.0
300,000> p ≥ 200,000	74	14,463.4	18,347,162	17,465,208	881,954	105.0%		195.5	247,935	236,016	11,918	105.0%		1,268.5	1,207.5
200,000> p ≥ 100,000	243	44,083.0	34,390,935	36,099,366	-1,708,431	95.3%		181.4	141,526	148,557	-7,031	95.3%		780.1	818.9
100,000> p ≥ 50,000	284	61,282.0	19,986,351	21,163,849	-1,177,498	94.4%		215.8	70,374	74,521	-4,146	94.4%		326.1	345.4
50,000> p ≥ 30,000	246	54,540.7	9,591,857	10,156,000	-564,143	94.4%		221.7	38,991	41,285	-2,293	94.4%		175.9	186.2
30,000> p ≥ 10,000	468	82,976.5	8,562,351	9,173,970	-611,619	93.3%		177.3	18,296	19,603	-1,307	93.3%		103.2	110.6
10,000> p	499	94,399.7	2,517,833	2,675,865	-158,032	94.1%		189.2	5,046	5,362	-317	94.1%		26.7	28.3
Total	1892	372,950.4	128,057,352	128,057,352	0	100.0%		197.1	67,684	67,684	0	100.0%		343.4	343.4

Table 5.3-2 and Table 5.3-3 resorts the municipal governments in Table 5.3-1 in terms of daytime population and area and daytime population density and area respectively. Figure 5.3-3 and Figure 5.3-4 is a three dimensional visualization of the previous tables. It can be observed that the mode of the counts in the daytime population (Table 5.3-2 and Figure 5.3-3) is found at population range less than 10,000 and the area of the organizations are between 200 and 500 km^2. It can be inferred from this that a considerable number of relatively wide and sparse local municipalities in rural areas are dominant in terms of the counts of these organizations. Looking at the daytime population density in Table 5.3-3 and Figure 5.3-4, the number of densely populated municipalities with relatively small sizes becomes visible. These wards (Ku) or cities (Shi) form densely populated metropolises. It should also not be overlooked that the mode is still located at sparsely populated yet relatively large municipalities which have a population density of less than 50 people per square kilometers.

Table 5.3-2 Counts of local governments in Japan (Area vs. Day time population) (Derived from [8] and [9])

Area ＼ Population	10,000> P	30,000> P ≥ 10,000	50,000> P ≥ 30,000	100,000> P ≥ 50,000	200,000> P ≥ 100,000	300,000> P ≥ 200,000	500,000> P ≥ 300,000	P ≥ 500,000	Total
A ≥ 1,000	5	5	3	6	5	1	2	0	27
1,000> A ≥ 500	38	30	24	32	27	9	11	2	173
500> A ≥ 200	123	104	72	59	34	11	18	3	424
200> A ≥ 100	111	99	55	46	31	8	5	1	356
100> A ≥ 50	103	85	38	53	31	13	8	5	336
50> A ≥ 20	75	90	33	38	58	22	10	5	331
20> A ≥ 10	25	35	18	33	44	9	3	4	171
10> A	19	20	3	17	13	1	1	0	74
Total	499	468	246	284	243	74	58	20	1892
							(A: Area in km^2, P: Daytime Population)		

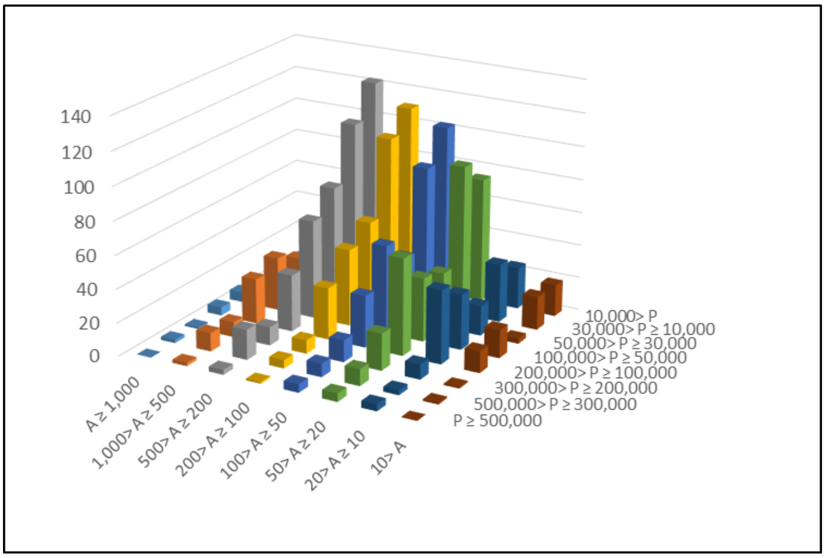


Figure 5.3-3 Counts of local governments in Japan (Day time populations and areas) (Derived from [8] and [9])

Table 5.3-3 Counts of local governments in Japan (Area vs. Day time population density) (Derived from [8] and [9])

Area ＼ Population	50> PD	100> PD ≥ 50	200> PD ≥ 100	500> PD ≥ 200	1,000> PD ≥ 500	2,000> PD ≥ 1,000	5,000> PD ≥ 2,000	PD ≥ 5,000	Total
A ≥ 1,000	15	5	4	3	0	0	0	0	27
1,000> A ≥ 500	73	34	31	29	4	2	0	0	173
500> A ≥ 200	152	83	71	70	29	18	1	0	424
200> A ≥ 100	86	58	56	93	34	23	6	0	356
100> A ≥ 50	42	33	57	73	54	36	29	12	336
50> A ≥ 20	8	17	26	70	50	39	57	64	331
20> A ≥ 10	2	2	3	12	18	18	47	69	171
10> A	2	1	2	3	5	10	18	33	74
Total	380	233	250	353	194	146	158	178	1892
						(A: Area in km^2, PD: Daytime Population Density)			

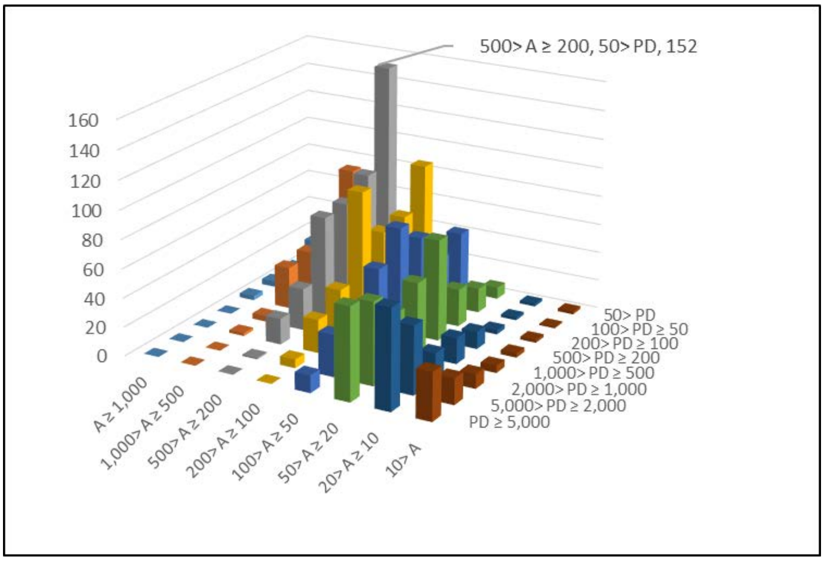


Figure 5.3-4 Counts of local governments in Japan (Area vs. Day time population) (Derived from [8] and [9])

By sorting municipalities in descending order of daytime population density and comparing each municipality's size as well as daytime population to the total area of all municipalities in Japan, Figure 5.3-6 provides a 'portfolio' of these organizations in the context of the daytime population density. The population density curve has a lolled-S shape. The daytime population gradually decreases while the size of the corresponding area increases as the daytime population density decreased. Though there are fluctuations in every sample, the approximated dashed lines show the overall tendency.)

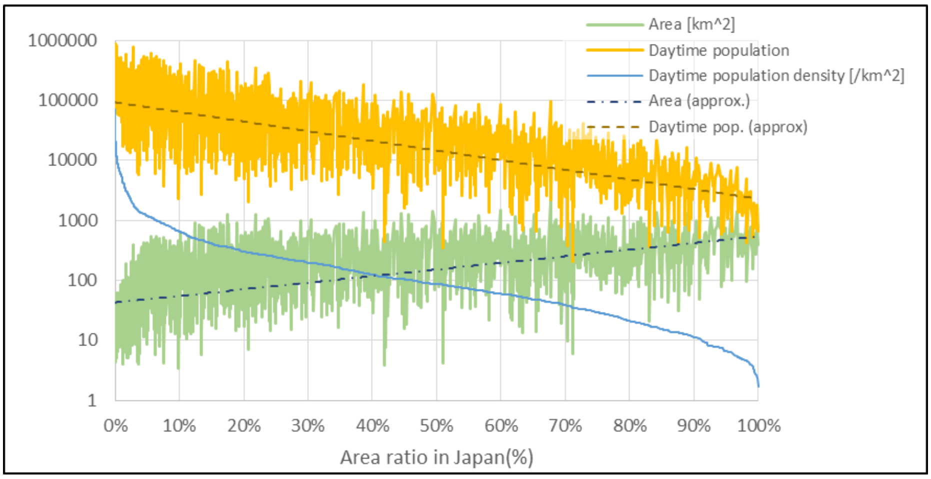


Figure 5.3-5 Profile of local municipalities (Area, Population and Population density) (Derived from [8] and [9])

Figure 5.3-6 depicts the cumulative daytime population curve against the cumulative area of corresponding municipalities with the descending order of the daytime population density. It can be observed that 95% of the total population of the nation spends the daytime in municipalities which encompasses 50.1% (see point (a) in the figure) of land of the nation. In these areas, the daytime population density is higher than 85 people per square kilo meter (point (b)). In the case of only half of the total population, i.e. 50% of the nation's gross population, the area where they spend their daytimes corresponds to only 3.7% of the total area (point (c)) and the population densities of these areas is larger than 1,396 people per square meter (point (d)).

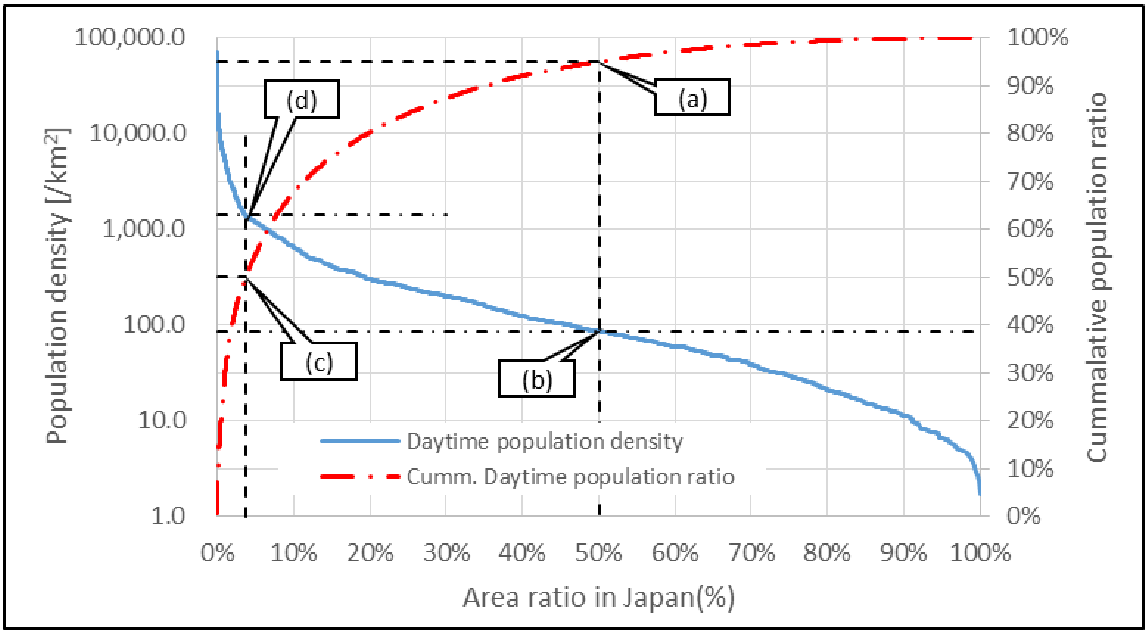


Figure 5.3-6 Cumulative area vs. Population density and Cumulative population (Derived from [8] and [9])

As an experiment, costs were estimated for the case which mobile networks are operated in the areas in which daytime population densities are less than 200/km^2 was derived using equation (2) [6]. The results are expressed in Figure 5.3-7 (Derived from [6], [8] and [9]). Operation cost increase constantly as the coverage of the area increases, reaching 1,000 billion yen when the network covers areas which daytime population density is down to 50/km^2.

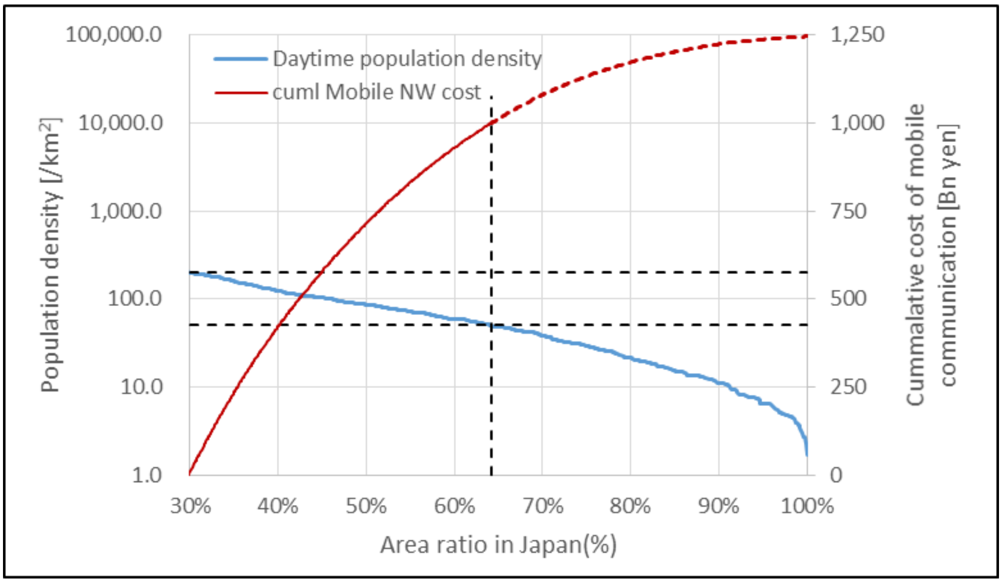


Figure 5.3-7 Cumulative cost of mobile communication networks of areas of population density between 50 to 200 (Derived from [6], [8] and [9])

The last experiment uses equations derived from existing mobile networks, resulting in enormous costs. Therefore, when considering a '5G' system, an efficient technologies/deployment approach could be applied when the system covers a sparsely populated area in the nation.

5.4 Daily dynamics aspect

An example of daily dynamics of population densities during the day and at night in local municipalities within the Tokyo metropolitan area is shown in Figure 5.4-1 (Derived from [8] and [9]). The figure captures the thirty densest areas in the region, showing the enormous changes in population densities between the daytime and nighttime. In terms of population density changes, the most extreme case within a municipality (Chiyoda-ku in Tokyo) has a 17 times higher population density during the daytime period compared at night. Existing mobile communication networks have been able to handle these large population density swings in dense, urban areas.

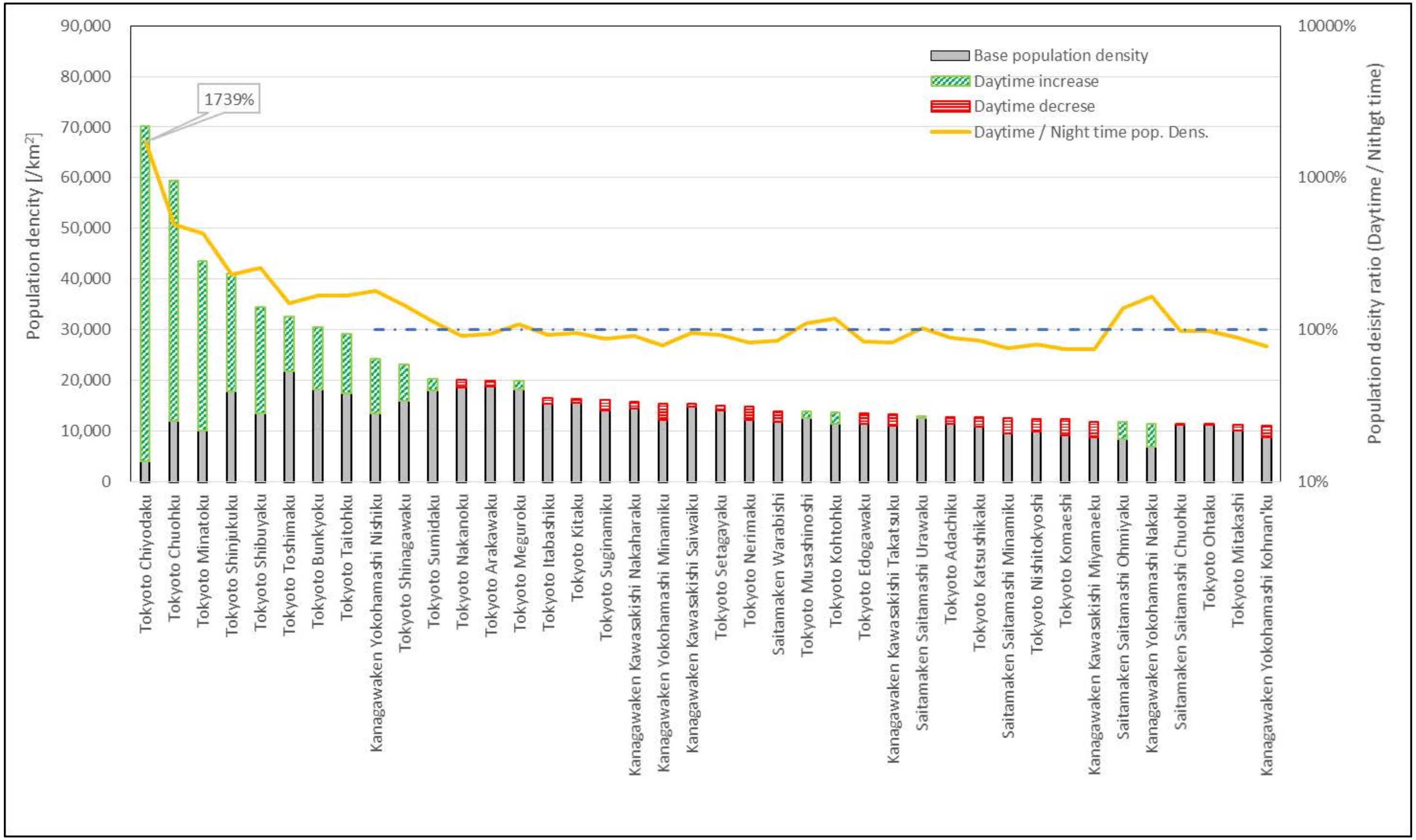


Figure 5.4-1 Dynamics of population density (day time and night time in Tokyo metropolitan area) (Derived from [8] and [9])

Figure 5.4-2 (Derived from [8] and [9]) shows the nature of daily population dynamics nationwide by sorting out daytime populations (not densities) and nighttime populations in each local municipality by their daytime populations in descending order. The chart shows that half of the whole population lives and commutes within 10% of the total landmass of the nation, and the most significant dynamics is observed within 10% of the whole population, who lives and commutes within 1% of the total mass.

As the population portfolio shows the net dynamics and because of the bidirectional nature of commuting, the actual flow of the people could be larger than these increases or decreases of daytime and night time populations express. Data showing the actual numbers of commuters can be found in Table 5.4-1[10] where the transportation flow within the Tokyo metropolitan area is summarized. According to this analysis, more than 7,000,000 people are moving in the area on a daily basis. The same survey was taken in the Kansai area around Osaka and Kyoto area as well as in Chukyo area around Nagoya. According to the analysis in [10], total number of commuting people is 2,450,000 and 728,000 respectively.

As has been already pointed out, these numbers reflect the current situation and express how existing mobile communication systems have already been able to cope with them. However, considering growing volume for communication traffic as demand increases as well as a future where many more personal objects will be equipped with communication capabilities, the '5G' system should be capable to handle double or triple the traffic flow as necessary every morning and evening, which would require certain technical breakthroughs.

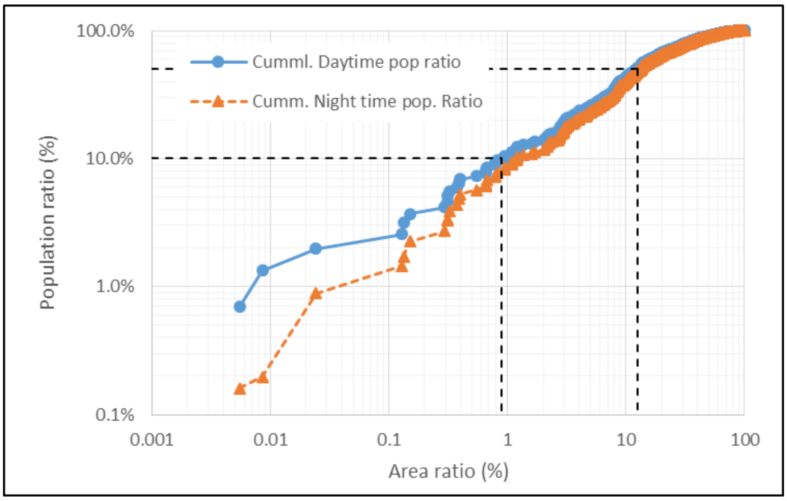


Figure 5.4-2 Dynamics of population (day time and night time in Tokyo metropolitan area) (Derived from [8] and [9])

Table 5.4-1 Transportation flow in Tokyo metropolitan area (English translation from [10])

(Unit: 1000/day, one-say direction)

Start \ Destination		Tokyo			Kanagawa pref.				Saitama pref.			Chiba pref.			Ibakaki pref.	Other prefectures	Area total
		Metropolitan	Other area	Subtotal	Yokohama-sh	Kawasaki-shi	Other area	Subtotal	Saitama-shi	Other area	Subtotal	Chiba-shi	Other area	Subtotal			
Tokyo	Metropolitan	1,767	109	1,876	64	38	21	123	21	45	66	16	43	59	3	0	2,127
	Other area	585	219	804	33	23	22	78	5	16	21	1	5	6	1	1	911
	Subtotal	2,352	328	2,680	96	61	43	200	26	61	87	18	48	65	4	1	3,038
Kanagawa	Yokohama-shi	446	32	478	247	64	80	391	2	4	6	2	4	6	0	0	880
	Kawasaki-shi	239	23	262	41	46	18	105	1	2	3	1	2	3	0	0	372
	Other area	246	46	292	151	40	132	323	0	2	2	1	2	3	0	0	620
	Subtotal	930	101	1,031	439	150	230	819	3	8	11	4	8	12	0	0	1,872
Saitama	Saitama-shi	183	7	190	3	3	1	7	34	34	68	1	5	6	0	0	273
	Other area	622	42	664	10	7	3	20	98	149	247	2	19	21	2	1	953
	Subtotal	805	49	854	14	10	4	28	131	183	314	3	24	27	2	1	1,226
Chiba	Chiba-shi	116	2	118	3	2	0	5	0	2	2	24	28	52	0	0	176
	Other area	607	12	619	10	7	1	18	8	18	26	65	136	201	6	0	871
	Subtotal	723	14	737	12	9	2	23	8	20	28	89	164	253	7	0	1,049
Ibakaki pref.		69	3	72	1	0	0	1	3	2	5	4	10	14	4	0	98
Other prefectures		10	4	14	0	1	0	1	2	2	4	0	0	0	0	0	19
Area total		4,889	499	5,388	562	231	279	1,072	173	277	449	118	253	371	17	2	7,301

5.5 Capital investment aspect

As an additional observation related to costs, capital investment by mobile operators is shown in Figure 5.5-1 (Derived from [12]) in which several investment peaks are observed during every ten-year period. A 30% increase in capital investment can be seen between the years of 2009 to 2014. Since this is an investment for future businesses, it should reflect the long term trends of the market and peaks in the chart should be consistent with a period of deploying new mobile communication systems while the other times would periods where operators are improving their system capacities.

During the time when '5G' will be deployed, a certain amount of capital investment will be needed. However, the way 5G will be deployed should be efficient enough to provide the reasonable service quality and quantity at a reasonable cost as has been discussed in the previous sections.

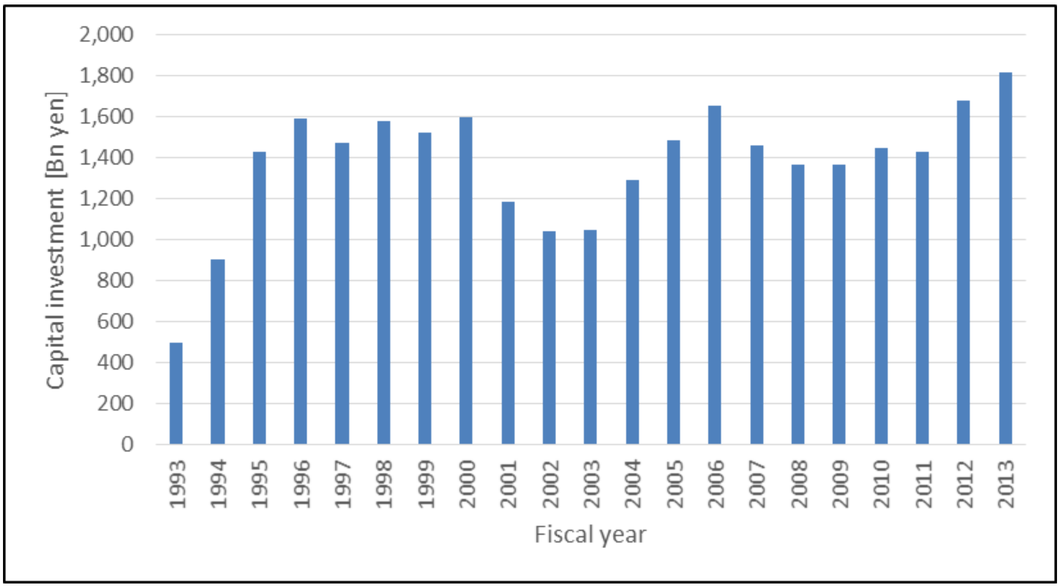


Figure 5.5-1 Capital investment of mobile communication operators in Japan (Derived from [12])

5.6 Conclusion

This chapter reviewed the costs of the mobile communication systems and analyzed it from several '5G' use case perspectives. The case studies considered included looking at costs in relation to communication traffic, as related to mobile broadband and scenarios related to coverage over sparsely populated areas and the dynamics of communication traffic in areas where population density varied throughout the day.

First, the conclusion as related to communication traffic, suggests that '5G' provide wider bandwidth services utilizing, for example, wider frequency spectrum, that would not increase costs dramatically, or even at all, even with the new value provided in terms of new services.

Second, experiments estimating operating costs when deploying mobile communications systems in sparsely dense suggests utilizing an efficient technologies/deployment approach as the '5G' system covers more sparsely populated areas of the nation.

Finally, the dynamics of a widely changing population density as it relates to communication traffic were studied, by calculating the deltas between daytime and nighttime populations in the Tokyo metropolitan area. Traffic as people commute into and out of these areas during the day can double or triple and the '5G' system should be ready to serve these kinds of dynamic population shifts over the course of one day. It should also be noted that all of these above mentioned use cases should also take into account new deployment scenarios that would provide device to device communication as well as the number of wearable devices carried by everyone.

Since an important framework for '5G' is the ability to cover a variety of use cases in a cost effective manner, a '5G' system should be designed to be as flexible and as scalable as possible. Accordingly, in most cases, especially use cases specific to '5G', it would be useful to apply technologies, configurations or operating methods suitable to these specific use cases and combine these elements into a unified communication system, rather than to seek generic, common and robust technologies covering all the use cases in a general manner

References

[1] "Mobile Communications Systems for 2020 and beyond," ARIB 2020 and Beyond Ad Hoc Group White Paper, Oct.2014 (URL: http://www.arib.or.jp/ADWICS/20bah-wp-100.pdf).

[2] "Report on the Ideal Development of Mobile Phone Base Stations," Study Group on the Ideal Development of Mobile Phone Base Stations (MIC), Mar. 2014 (In Japanese) (URL: http://www.soumu.go.jp/main_content/000281538.pdf).

[3] "Status of the mobile communications traffic of Japan (Sep. 2015)," Information and Communications Statistics Database, The Ministry of Internal Affairs and Communications, Nov. 2015 (In Japanese) (URL: http://www.soumu.go.jp/johotsusintokei/field/data/gt010601.xls).

[4] "Results of Households expenditure research," Statistics Bureau, Ministry of Internal Affairs and Communications, 2004-2015 (In Japanese) (URL: http://www.stat.go.jp/data/joukyou/12.htm).

[5] "Basic Survey on the Information and Communications Industry," The Ministry of Economy, Trade and Industry, and Statistics Bureau, Ministry of Internal Affairs and Communications, 2004-2015 (In Japanese) (URL: http://www.soumu.go.jp/johotsusintokei/statistics/statistics07b.html).

[6] "Annual report of on the Japanese Economy and Public Finance 2013," Cabinet Office

Government of Japan, Jul. 2013 (In Japanese) (URL: http://www5.cao.go.jp/j-j/wp/wp-je13/index_pdf.html).

[7] "Annual report of on the Japanese Economy and Public Finance 2013," Cabinet Office Government of Japan, Jul. 2013 (Summary in English) (http://www5.cao.go.jp/keizai3/2013/0723wp-keizai/summary.html).

[8] "Population based on place of working or schooling (Daytime population), Population based on place of usual residence (Nighttime population), Rate of daytime population to Nighttime population - Shi, Ku, Machi and Mura (2010)," Statistics Bureau of Japan, 2010-2015 (In Japanese) (URL: http://www.stat.go.jp/data/kokusei/2010/kihon4/zuhyou/syuyou.xls).

[9] "Area of Shi, Ku, Machi and Mura in each Ken," Geospatial Information Authority of Japan, 2015 (In Japanese) (URL: http://www.gsi.go.jp/KOKUJYOHO/MENCHO/201510/opening.htm).

[10] "Report of analysis on the eleventh Metropolis Transportation Census (2010)," Ministry of Land, Infrastructure, Transport and Tourism, Mar. 2013 (In Japanese) (URL: http://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000047.html).

[11] "Research on investment and retirement of fixed asset of private enterprises (2015)," Cabinet Office Government of Japan, Jul. 2013 (In Japanese) (URL: http://www.esri.cao.go.jp/jp/sna/data/data_list/jyokyaku/files/h26/tables/H26hyou3-1.xls).

[12] "Radio industry year book 2015," Association of Radio Industries and Businesses, Nov. 2015 (In Japanese) (URL: http://www.arib.or.jp/johoshiryo/nenkan/nenkan.html).

6. 5G Key Concept

6.1 Key Concepts of 5G

End-to-end (E2E) quality required by applications and/or users will be far more diversified in the 5G era than what we have seen in the preceding generations. For example, the ITU-R Vision recommendation [1] illustrates a number of usage scenarios in which the capabilities required are not identical but diverse depending on their E2E quality expected.

Fig. 6.1-1 represents potential 5G applications mapped on a domain of the quality in user experience by the quantity of data. Some attractive services and more user friendly utilities will emerge as new applications by means of innovative technologies deployed in 5G.

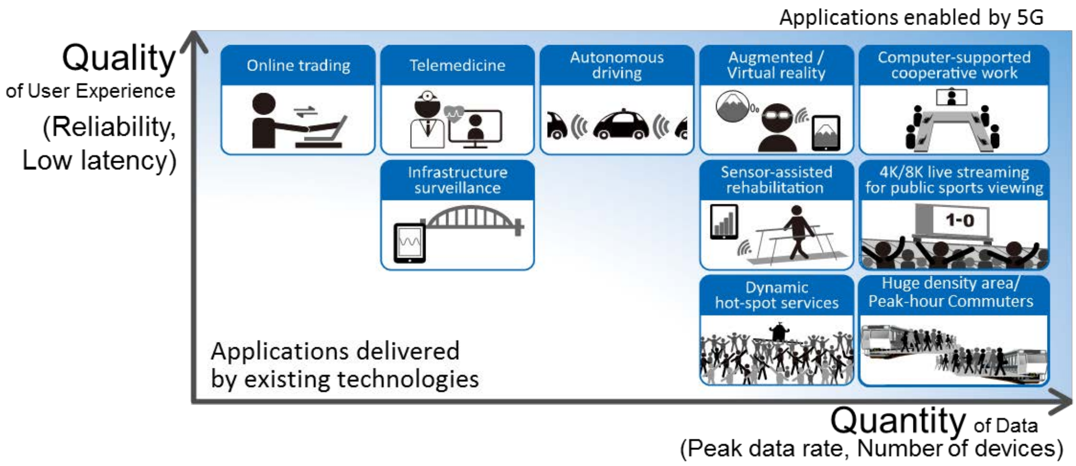


Fig. 6.1-1 Potential 5G applications

5GMF believes that one of the two key concepts of 5G consists in “Satisfaction of E2E quality” required in all usage scenarios, making users feel satisfied with the quality, whatever applications are used anytime, anywhere. Achieving “Satisfaction of E2E quality” in 5G is an essential goal that differentiates 5G from preceding generations, which were designed based on “best-effort” scenario.

Another key concept that 5G systems should have is the ultimate “Extreme Flexibility”, in order to satisfy the E2E quality required in each use scene in a flexible manner, even if it is in the extreme.

In the 5G era, the E2E quality required by users will be far more diversified than in the preceding generation systems and the dynamic range will also be much wider depending on locations, spaces, and other factors. These are the requirements specific for 5G, totally different from those for preceding best-effort based systems, which opens up a new area of innovations in order to design such systems in an economically viable manner.

In the previous generation systems, radio access networks were regarded as dominant bottleneck which determines the E2E quality of mobile applications and services, since the performance of radio access networks were limited by a number of constraints, such as radio propagation characteristics, available bandwidth, handset power, mobility, and so forth.

In the 5G era, however, it is expected that most of these constraints will be greatly relaxed by the advancement of radio technologies. The performance of radio access networks alone is no longer a sole bottleneck and that of core networks should also be taken into account to satisfy E2E quality. Therefore, the technologies for radio access and core networks should be jointly studied and developed on an equal basis in order to realize the feature of the “Extreme Flexibility”.

One of the typical capabilities mandating close interwork between radio access network and core network is the low latency. Neither radio access network nor core network alone cannot realize the E2E low latency, since the latency is determined by the overall path length spanning over both networks.

6.2 5G key technical aspects

6.2.1 General

5G is expected to satisfy E2E quality of services in wider range of use cases in flexible, secure and efficient manner. It is necessary that radio access and core networks should work jointly to realize the “Extreme Flexibility”. In this white paper, the two key technologies are identified as follows in order to support the “Extreme Flexibility”;

- Advanced heterogeneous network
- Network softwarization and slicing

In the following sub clauses, these key technologies are illustrated.

6.2.2 Advanced Heterogeneous network

The term 'Heterogeneous network' could have several interpretations or definitions depending on the context used. In some wireless communication networks, the term refers to network consisting of smaller cells laid over a larger cell in order to increase their system capacity by offloading traffic from a single large cell to these smaller cells [1].

In case of 5G, the idea of 'Heterogeneous network' should be enhanced to involve more than the idea described above and represents configuration of communication networks that organize its entire elemental network portions to serve variety of use cases.

In [2], importance of 'heterogeneous network' integrating multiple Radio Access Technologies (RAT) existing, such as 2G, 3G, LTE, W-LAN with 5G RAT(s) was pointed out in order to achieve efficient utilization of higher and wider frequency spectrum beyond 6GHz in a cost effective manner. Considering the new use cases foreseen in 2020s especially, 5G RAT(s) should cover most of these cases efficiently and cooperate with legacy RAT(s) in the networks.

This white paper proposes that the scope of integrated radio access networks be largely extended to include multiple technologies shown above and that the network realizing the heterogeneity far beyond that in the previous heterogeneous network be called 'Advanced heterogeneous network'.

As has been described in the preceding sections, 5G should support wide range of services. Accordingly it would not be the best and efficient way to establish single technical solution that could serve all the range of the requirements for every use case of 5G. Instead, it would be reasonable to adopt proper radio communication technology with proper parameters as a unified system depending on the use cases required.

References:

[1] "Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN," TS36.932 (Ver.12.1.0), 3rd Generation Partnership Project, Mar.2013.

[2] "Mobile Communications Systems for 2020 and beyond," 2020 and Beyond Ad Hoc Group, Association of Radio Industries and Businesses, Oct.2014.

6.2.3 Network Softwarization and Slicing

Network Softwarization is an overall transformation trend for designing, implementing, deploying, managing and maintaining network equipment and/or network components by software programming, exploiting the natures of software such as flexibility and rapidness all along the lifecycle of network equipment/components. The industry effort on Network Functions Virtualisation (NFV) and Software Defined Networking (SDN) are integral part of this transformation.

The basic concept of the Network Softwarization is “Slicing” as defined in [ITU-T Y.3011], [ITU-T Y.3012]. Slicing allows logically isolated network partitions (LINP) with a slice being considered as a unit of programmable resources such as network, computation and storage. Considering the wide variety of application domains to be supported by 5G or IMT-2020 network, it is necessary to extend the concept of slicing to cover a wider range of use cases than those targeted by the current SDN/NFV technologies, and the need to address a number of issues on how to utilize slices created on top of programmable software defined infrastructure.

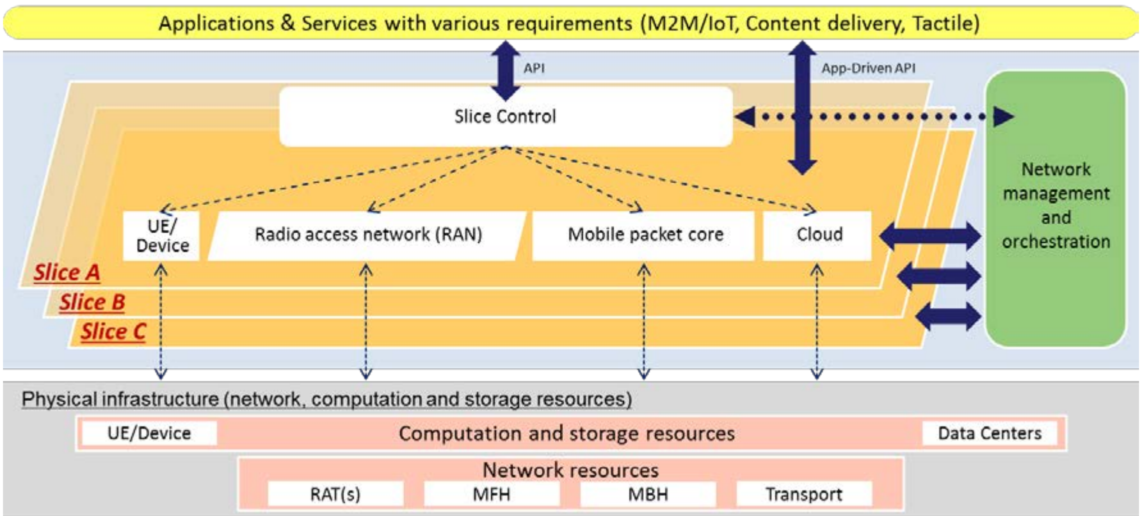


Fig. 6.2-1 Network softwarization view of the 5G systems

Fig. 6.2-1 illustrates the network softwarization view of 5G systems, which consist of a couple of slices created on a physical infrastructure by the “network management and orchestration”. A slice is the collection of virtual or physical network functions connected by links to create an end-to-end networked system. In this figure, the slice A consists of radio access network (RAN), mobile packet core, UE (User

Equipment)/device and cloud, each of which is collection of virtual or physical network functions. Note that the entities are shown rather symbolically and links are not described in Fig. 6.2-1 for simplicity. The "network management and orchestration" manages the life cycle of slices: creation, update and deletion. It also manages the physical infrastructure and virtual resources, which are abstraction of physical resources. The physical infrastructure consists of computation and storage resources that include UE/devices (e.g. sensors) and data centers, and network resources that include RATs, MFH, MBH and Transport. It should be noted that both computation/storage resources and network resources are distributed and are available for creating virtual network functions.

Network softwarization will greatly improve flexibility in design, implementation, deployment, operation and maintenance of network functions and components, and increase velocity of service delivery by making the best use of programmability. In addition, application of "Slicing" will increasing efficiency and dynamicity of 5G systems, since it enables just-in-time assembly of network functions and components for service delivery in concert with arrangement of advanced heterogeneous networks.

6.3 5G Typical Use Cases

This section addresses typical 5G use cases and enhancements required for individual usage scenarios, based on the IMT Vision recommendation ITU-R M.2083-0, which classifies typical 5G usage scenarios into following three use cases.

6.3.1 Ultra-reliable and low latency communications

The 5G should support not only human communications but also applications for non-human equipment, including machines, vehicles, sensors and etc. Some applications in this category will be more stringent to delay and loss of information than other applications or those in the preceding generation systems and will require that packets should be delivered to the other end in a specified period certainly. They will call for capabilities such as lower latency and higher reliability than in the preceding generation system.

The radio access networks, core networks and other part of the networks, which constitute E2E networks, should work closely to satisfy these E2E quality. For example, in order to achieve required E2E latency, distribution of latency budget to each constituent part of networks, i.e., handset, radio access network, fronthaul/backhaul, core network should be considered.

Typical use scenarios in this use case include wireless control of industrial manufacturing or production processes, remote medical surgery, distribution automation in a smart grid, transportation safety, etc.

Designers of mission critical applications will focus on end-to-end quality provided by 5G systems. In a typical arrangement of such applications, the end-to-end is comprised not only of radio access networks but also of terminals, fronthaul/backhaul networks, mobile packet core and inter-service provider networks and data-centers. This implies that the end-to-end quality depends on the quality provided by both radio part and wired part of networks, and in contrast, that 5G systems should have the capability to tailor the end-to-end by organizing functions and connectivity so as to satisfy the requirements of mission critical applications.

Mobile Edge Computing (MEC) is the concept to provide an IT service environment at a location considered to be the most lucrative point in mobile networks, characterized by proximity, ultra-low latency and high bandwidth.

One of the mission critical applications includes that requires low latency. It includes the maintenance and control of devices, instruments and equipment in factories, remote control of construction equipment and delivery robots, distance medicine, autonomous driving. Such applications require a close feedback system where the information from sensors that capture status of working environment is transmitted to a control function that makes decisions for reaction, and the commands that realize the reaction is conversely transmitted to the actuators that execute the commands. The overall propagation time of the system is of interest to the designers, which is usually required to be in a range of tens of milliseconds. Designers may consider where to place the control function to meet application's requirements: Usually it is considered to place it in a proximity to the sensors and actuators physically.

6.3.2 Massive Connection

As shown in Chapter 7 “Typical usage scenarios of 5G”, in order to cover applications for non-human equipment in addition to human communications, specific capabilities are required by these applications. Those capabilities will include area coverage expansion to non-resident area, cell radius expansion, and massive connections in order to accommodate as many equipment as possible in the system and so on. In order to attain this objective, the system should be designed so as to accommodate numerous equipment in an efficient manner, while the data volume generated by the equipment may be relatively small as compared with signaling traffic in some cases. Also the system should be designed to reduce cost and power consumption of devices. This use case include infrastructure monitoring, sensor network, etc.

6.3.3 eMBB enhanced Mobile Broadband (Data rate, Capacity, Mobility)

This use case will require increasingly improved and seamless user experience as compared with the preceding generation systems. As stated in the previous section, the 5G systems should aim at providing sufficient user experienced data rate in every circumstance. Typical use scenarios in this use case will include enjoying a sports game in a stadium more vividly by watching the video, video communication employing augmented reality, virtual reality technologies.

The increase of the data rate will broaden the opportunity for supporting various high-quality streaming applications. The characteristics of such streaming applications depend on average bandwidth, end-to-end latency, and possible latency fluctuations, and so forth. In transport networks, we have sufficient knowledge and experiences to control such metrics. It is expected to explore how we control streaming applications in 5G radio access, and how we design end-to-end networking system so as to fulfill the E2E quality.

This use case requires enhancement of fundamental network capabilities sufficiently to satisfy user requirements without making users feel frustrated, which will be made possible by removing constraints and restrictions imposed by the network of succeeding generations. Examples of augmentation and enhancement to the capabilities are illustrated below.

－Peak data rate and system capacity

The eMBB usage scenario will require improvement of performance in terms of peak data rate and system capacity to satisfy user experience in new applications.

Availability of spectrum bandwidth is one of the mandatory requirements for 5G, since spectrum is a constraint to limit peak data rate and system capacity. This constraint should be relaxed by securing bandwidth sufficiently wide to provide peak data rate required by users and/or applications and to accommodate as many users as possible.

－Mobility

Maximum vehicular speed at which 5G systems can provide sufficient QoE should be enhanced to cover all surface vehicles to appear around 2020, including Magnetic Levitation train known as "linear motor cars" in Japan.

－Quality improvement by multi-antenna technologies

In the 5G era, beamforming technology is effective to improve quality by concentrating transmission power in a small area and/or on a moving user(s). Beamforming is based on multi-antenna technologies to control large number of antenna elements, which is more adaptable to higher frequency band.

In the practical usage scenario, requirements for 5G systems are changing constantly, varying with time, locations, applications in use, user distribution and other factors. It is foreseen that 5G systems should accommodate numerous users and/or applications of various requirement, therefore, the dynamic range of requirement will be much wider than in the previous generations.

Typical examples of usage scenarios are described in Chapter 7 of this white paper.

7. Typical Usage Scenarios of 5G

Based on the considerations on future market trends and user trends discussed in Chapter 3, this chapter first illustrates some examples of new usage scenarios, which are envisioned for 5G, and categorizes them into four facets; 1) Entertainment, 2) Transportation, 3) Industries/Verticals, and 4) Emergency and disaster relief.

Further analysis on the usage scenarios clarifies the list of required capabilities of individual usage scenarios. It finally provides key items of 5G capabilities for deriving overall 5G requirements in Chapter 8.

This chapter also gives an insight of "dynamic approach" into nature of 5G capabilities which may dynamically change corresponding to the wide variety of 5G usage scenarios.

7.1 Four representative typical usage scenarios

The feature of 5G capabilities is, among others, peak data rate of more than 10 Gbps, mobility of more than 500 km/h, latency of 1 ms, number of connected devices per cell of 10 thousand, capacity per unit area of 1000 times larger than that of 4G and furthermore significant reduction of power consumption. This section introduces 4 typical usage scenarios by using comprehensive illustrations (see Fig. 7.1-1) in order for readers to grasp a clear picture of 5G usage scenarios.

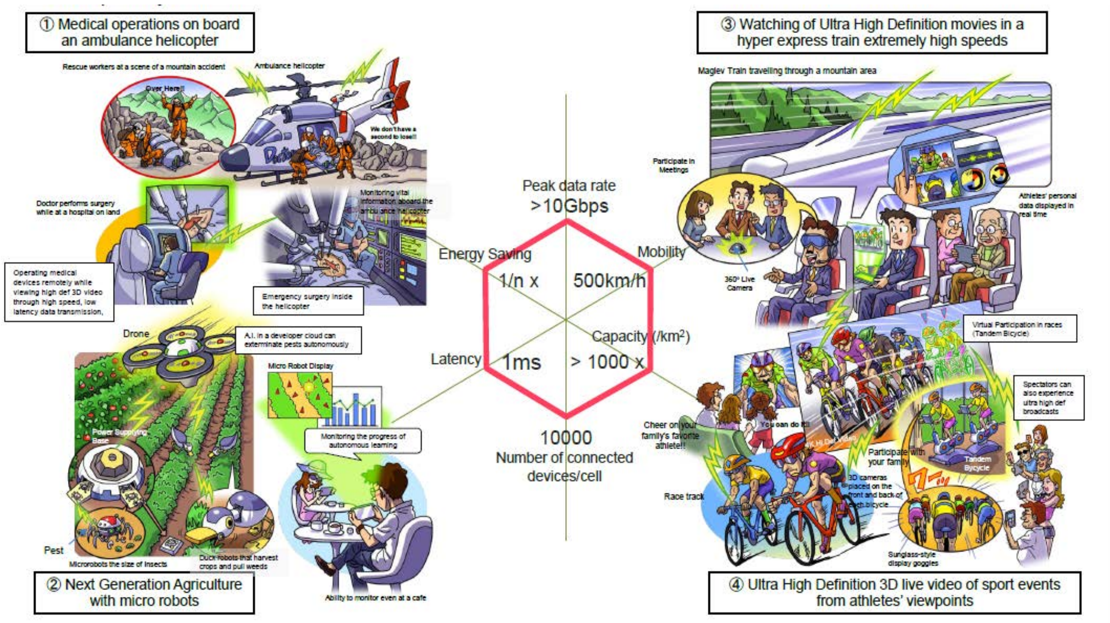


Fig. 7.1-1 5G capability

(1) Medical operations on board the ambulance helicopter

This usage scenario requires both higher peak data rata and low latency. It also assumes robust 5G communication link even in disaster areas.

(2) New generation smart agriculture by using micro robots

This usage scenario shows a 5G application to smart agriculture by using 5G's capability of low power consumption, enabling extremely longer time duration of data communication with extremely long life of battery.

(3) Watching of Ultra High Definition movies in a hyper express train at extremely high speed

This usage scenario indicates that 5G makes it possible that passengers in a hyper express train at extremely high speed can watch and enjoy ultra high definition movies.

(4) Enabling users' experience by Ultra high definition 3D live video of sport events from sport player's viewpoints

This usage scenario requires i) broadband live video uploads from sport event areas/courses, ii) massive video connections to audiences in a stadium, iii) low latency communication for audience/users to participate in virtual sport race.

7.2 Case studies of Typical Usage Scenarios

7.2.1 Entertainment

Fig. 7.2-1 Watching sports games

In this section, the usage scenarios that provide a person with unique and/or advanced experiences to enjoy leisure time when watching sports games in stadium, playing games and going for travels. It ranges from enhanced real experiences to fully virtualized experiences. Ultra-high definition moving pictures and high fidelity acoustics will be extensively utilized. Comfortable communication environment even in highly congested area will be provided and advanced technologies to allow smooth remote collaboration will be equipped.

Usage Scenario #1	Enhanced real experience entertainment (Shared experiences and virtual reality experience)
Overview	(1) Experience sharing scenario (a) Users watch 3D video of an event, for example a sporting event, from multiple viewpoints through cooperation with other fans by sharing their videos. Users are then able to watch the even from any viewpoint they wish. (b) Fans going to and leaving a stadium, for example at a soccer match, share information and experiences with other fans on the train by using their smartphones. For this purpose, a 5G system needs to support high data transmission so that many users, in this case soccer fans, in a single train car can simultaneously watch high definition video and/or exchange a huge amount of data. (c) High definition video communication while watching a soccer match at a sold-out soccer stadium (both upstream and downstream) (2) Simulated Experiences Scenario (a) An environment where users can always see exhibitions in crowded museums. (b) Family members discuss their plans while on a sightseeing trip using streaming arbitrary viewpoint video. Since the streaming video e provides arbitrary viewpoints, the family can view their sightseeing routes virtually from their desired angle. (c) While on a sightseeing drive, a traffic accident occurs at an upcoming intersection, resulting in a major traffic jam. An arbitrary viewpoint video and other related information from the accident location are distributed automatically. The family is able to download more video from different angles as well as other related information. They can consider viable alternative routes, taking advantage of this up-to-date information. (3) Virtual Reality Scenario (a) Outdoor real time gaming created by a virtually real visual sphere.

In the scenarios (1)(a), (2) (a) to (c) and (3)(a), arbitrary viewpoint video is assumed to be a 5G application. Arbitrary viewpoint video is a video system which simultaneously transmit videos taken from multiple angle (typically 6 angles) which is combined on the terminal side so that users can enjoy seeing an object from an angle they like.

The arbitrary viewpoint video enables;

(i) Users to be able to see and confirm video from an arbitrary angle in real-time on their mobile terminals.

(ii) Users to be able to see an object from an arbitrary angle in 3D space on their terminal, by being able to access multiple cameras which video-tape an object from a different angle.

(iii) Therefore, users are able see an object from an angle that any camera operator would not be able to shoot in real time through processing video data from different mobile terminals over a 5G network.

Enabling technologies such as AR/VR technologies, high precise time synchronization, and huge data synchronization technologies (several tens of msec precision for synchronization among video cameras, AR/VR display and game machines) will need several hundreds of msec of processing time to display video taken from multiple cameras as well as high speed data transmission at 60 Gbps from cameras to a BBU edge server. Video data distribution from the BBU to individual's terminals will have data rate of 6T bit/s maximum.

Even with high efficiency video coding (HEDC), a transmission rate of 90 Mbps per angle is required for 5G radio networks. Driving on a highway, for example, will require a high throughput with high speed mobility. For example, 90 Mbps * 6 = 540 Mbps is required while moving with 100 km/h speed. On the other hand, in the use case of a traffic accident occurring at an intersection which results in a traffic jam, communications data will be transmitted under stationary or near stationary conditions. In this case, arbitrary viewpoint video will be

	transmitted to many vehicles, resulting in dense data traffic. Assuming that the width of a car lane is 3.5 m, the length of a vehicle is 5 m, and the distance between vehicles is 3 m, arbitrary viewpoint video traffic is estimated to be 540 Mbps / (3.5 m * 8 m) = 19 Mbps/m^2. If one out of every two vehicles uses arbitrary viewpoint video simultaneously, traffic density will be 9.6 Mbps/m^2. In the scenario (1)(b) above, the following radio capabilities will be required on a train: - Peak user throughput of 1Gbps for high speed broadband communications; - User mobility of 100km/h for providing stable communication; - Several thousand efficient user connections for broadband communications; - Capability to support simultaneous handover at a same timing for several thousands of users or alternative equivalent technology scheme/capability without a handover; - Cost-efficient highly flexible traffic control beyond "best effort service"; - Average user data rates of 2 Mbps for each user on a single train. This means that, assuming that there are 1000 passengers per train car, trains running with 1.6km of spacing between them and a rail width of 10 m, 2 Mbps x 1000/ (0.01 km x 1.6 km) = 125 Gbps/km^2 will be necessary. In the scenario (1)(c) above, the following radio capabilities are required: (i) Peak user throughput 1Gbps for high speed broadband communications; (ii) Stable radio communication at a low mobility of several km/h; (iii) Provision of several thousands of efficient connections for broadband communication users; (iv) Provision of random handover by several thousands of

	users; (v) Cost-effective flexible traffic control capability beyond traditional "best effort service"; (vi) Average user throughput of 2Mbps in a stadium. This means assuming stadium bench seats 1m wide and 0.5m depth, one 5G mobile user per every 10 people in attendance, the user density at the stadium is1 user/(0.0005 km x 0.0011km). Therefore, 2Mbps x 1000 user/(0.01km x 1km) x 1/4 = 400 Gbps/km will be required.

Required capabilities	Peak data rate	X
	User experienced data rate	X
	Latency	
	Mobility	X
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	X
	Area traffic capacity	X
	Others	

Usage Scenario #2	Dynamic Hot-Spot services
Overview	● User Scenes (examples): Size of data and voice traffic change dramatically in dynamic ways as population density rises and falls in one location on a single day. - Stadium attendance (Olympic games, football matches, etc.) - Concert attendance, fireworks viewing, festival goers □In the above cases related to entertainment, a specific location is crowded with people only during the event itself with almost nobody there on other days. In these hot spots, people enjoy uploading videos they have taken to be able to show their families at home and downloading message/music data or other audio/visual information. For example, Nx1,000 or Nx10,000 devices may be activated

	simultaneously with a high data rates (e.g.10M to 100Mbps/device) in a stadium or an outdoor ground only while an event is occurring. - Disaster refugees going back home, a sudden rush of people into or out of a station, and emergency calls in disaster scenes. □Dynamic hot spots will occur in the same way as the entertainment use scenes above, but only during an emergency situation after a disaster occurs. ● Shortage of the existing general network: - A solid network structure is used regardless of the user service or application type having diverse natures in network. - Solid transport routes are arranged in a fixed network structure, and specific functionalities are allocated to each physical server. - Network composition resources and the power activation rate are solidly fixed. ● Challenge: - Extreme scalable capability by the network Management & Orchestration driven scalable network. - Much large scale of dynamic range will be required in some transmission capabilities of 5G network. - Control of the life-management of network slices matched with services. - Depending on the targeted service traffic or condition of transmission lines, traffic is dynamically controlled by software at the slice level, including VNF elements structure, transport topology, E2E transmission line, and transmission bandwidth. - Infrastructure resources of mobile networks are logically scheduled for the use in timely manner at appropriate situations. In the case of idle situations, resources can be used for other networks or pooled to prepare for re-use. This type of resource management contributes to reduction of

	CAPEX and OPEX. ● View points - Scalable network with dynamic flexibility. - Connectivity of devices spreading in both low density and ultra-high density environments. - Network architecture with reliable connectivity and high quality service provision, even in high density environments created by a temporary or specific localized situation with a huge number of connections and a large amount of traffic on the network. - Efficient utilization of surplus network and power resources under low data or voice traffic conditions.

Required capabilities	Peak data rate	
	User experienced data rate	X
	Latency	
	Mobility	
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	X
	Others	Dynamic Flexibility

Usage Scenario #3	A large marathon
Overview	● A big marathon race held in a city has many sensors placed at every main intersection. In order to meet the environment conditions for holding the race, the city government collects information related to atmospheric pollution levels from the sensors through massive connection techniques. Some runners wear a runner' view cameras, and upload the high-definition video from the camera while running thanks to ultra-high speed data transmission techniques. After the marathon, runners can watch the high-definition video with their family or friends. Many people can watch the race with

	their smart phones even while along the roadside. The city also allocates many high-definition video cameras to the roadside, and delivers the video from these cameras to the marathon spectators in real-time. Thanks to the runners' positioning estimation techniques, spectator can choose to watch an individual runner. The enhancement of wireless communication technologies contributes many new, diverse ways to make a marathon more enjoyable and exciting. ● Another important point for organizers of a large marathon is taking care of the health of the runners. Even in a race with more than 30,000 participants can have their runners wear sensors to collect their vital data (e.g., heart rate) by massive connection techniques to be able to check their health in real time. If something happens to a runner's health and well-being during the race, a medical institution in the area will be immediately notified with the necessary information thanks to new access techniques without the need for scheduling to be granted. And, the information from high-definition cameras allocated to the roadside that were focused on that particular runner will be provided to the medical institution to support their diagnosis and care for him or her. ● And, after the marathon finishes, collected information from the sensors equipped by the runner can be structured as big data to assist and advance industries such as health care and sports equipment.	
Required capabilities	Peak data rate	X
	User experienced data rate	
	Latency	X
	Mobility	
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	X
	Others	

Usage Scenario #4	A trip on the shinkansen high speed train	
Overview	● A large number of passengers on a *shinkansen* train enjoy entertainment services, such as real time competitive games and watching live-streams with their smart phones or tablets ● Passengers are able to watch a smooth moving picture and are content with the quality despite being on a high speed train. ● Reduce power consumption of base stations and terminals respectively. ● Technology for high capacity, adaptive beamforming and group mobility are necessary. ● Similar cases include: - Cars on the highway (Especially a bus where a large number of passengers are in movement simultaneously) - Ships - Airplanes (when use of terminals is allowed even during in takeoff and landing)	
Required capabilities	Peak data rate	
	User experienced data rate	X
	Latency	
	Mobility	X
	Connection density	X
	Energy efficiency	X
	Spectrum efficiency	
	Area traffic capacity	X
	Others	

Usage Scenario #5	Content downloads by commuters
Overview	● A user can instantaneously download large-volume files when the user touches their mobile device to am HRCP (high-rate close proximity) access point, for example an

	automatic ticket gate. An example scenario: When the transmission rate is 2 Gbit/s, downloading time for a 30-minute 50 MB video file will be 220 msec. - Mitigates wireless traffic loads in 5G mobile networks, by downloading large-volume files at the HRCP access point. - Reduces power consumption on the mobile device, because wireless communication is not required while playing video, unlike streaming usage. ● Required technologies include: (1) high-rate multi-Gbit/s wireless transmission, (2) device management function that turns on the wireless module only during downloading, and (3) cache mechanisms for delivering the content file to the HRCP access point where the download will occur. - Radio access technologies using unlicensed bands will be employed for (1). - To realize (2) and (3), new management/control functions that interoperate with 5G mobile networks are needed.

Required capabilities	Peak data rate	X
	User experienced data rate	X
	Latency	
	Mobility	
	Connection density	
	Energy efficiency	X
	Spectrum efficiency	
	Area traffic capacity	
	Others	

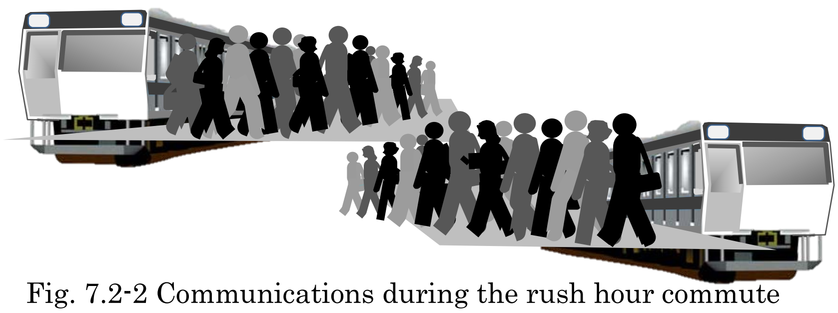


Fig. 7.2-2 Communications during the rush hour commute

Usage Scenario #6	Communications during the rush hour commute	
Overview	● In the Tokyo metropolitan area, the number of people commuting to work or school is increasing slowly, including 5.5 million railway passengers a day. These railway passengers when going through a terminal station create especially huge communication traffic. Shinjuku station, the largest terminal station in the Tokyo metropolitan area, has eleven railway lines and a train arrives for each line every two minutes during peak rush hour. Assuming 90% of the "accumulating passengers" use cellular phones, the number of phones exceeds 25,900. "Accumulating passengers" consist of (1) passengers getting on/off, (2) passengers staying on the train, and (3) people coming into/going off the station. ● Considering the area of Shinjuku station as 200m X 500m, the density of cellular terminals is 259,000 UE/km^2, and assuming user data rate in 2020 as 20Mbps, the communication traffic per km^2 reaches 5.18 Tbps/km^2.	
Required capabilities	Peak data rate	
	User experienced data rate	
	Latency	
	Mobility	
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	X
	Others	

7.2.2 Transportation

In this section, the user scenarios that provide comfortable experiences through advanced methods of transportation ranging from automobiles to high-speed magnetic levitated trains. It includes, for example, autonomous vehicles that are able to drive themselves without any intervention by a human at all, driver assisting services that provide comfortable rides by avoiding traffic jams or other obstacles, and computer-aided management of crowds during popular events. Novel intelligent mechanisms based on the combination of tremendous amount of data from advanced sensing technology and emerging artificial intelligence methodologies will greatly enhance conventional expectations.

Usage Scenario #7	Smart automobiles (driver assistance system)	
Overview	This system provides automobile collision avoidance at intersections with bad visibility. To monitor cars, bicycles, and people that are entering an intersection in real time, video cameras are placed at the intersection, and image processes are carried out with a low-latency application server which is placed at a base band unit. When intersection ingresses are detected, a detection result is created, consisting of an alarm and a video, and it is transmitted to automobiles through low-latency 5G networks. The automobiles that received the detection result automatically slow down while the alarm and the video are displayed on monitors. Also, this system predicts intersection ingresses by gathering traffic information from neighboring intersections.	
Required capabilities	Peak data rate	
	User experienced data rate	X
	Latency	X
	Mobility	X
	Connection density	
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	
	Others	

Usage Scenario #8	Behavior support in city
Overview	A large amount of environmental data is obtained from massive sensors installed in a city and user devices and is sent to edge servers and/or cloud servers. The data is then used for real-time human behavior support in shared audience use-scenes such as street/public space congestion and outdoor street events, as well as providing information tailored to the characteristics of an individual user, such as disability, age, and possession of luggage. For example: - An overview of the current situation in many places. At first, collecting data while using the network infrastructure of the system necessary to society, for example as a crime prevention system. This data can then be analyzed and used to support people's day to day lives by providing traffic information and people flow information, using color-displayed cars or a people density map. This information will reduce confusion during or after an event by indicating areas with less people in the event of a marathon, an *ekiden* relay race, or a fireworks display. - Provide information related to event venues in public places (citizen's marathon, a parade, etc.) through users' smart phones with a high image quality to provide highly realistic details. To ensuring privacy, the display can be changed to show people and vehicles or just show the people- or vehicle-density on the map stored only on the edge servers. - During a disaster, immediately provide safe evacuation routes tailored to individual user needs (e.g. their home location, physical fitness, possessions, clothes). To lessen the spread of confusion in the event of a disaster, provide general information on street, traffic and communication tools to the affected areas in an easy-to-understand form such as color-displayed density maps of cars or people by processing in edge servers. - Wheelchair driving support for walking disabilities.

	Characteristics of people with disabilities are diverse and building a uniform and general automatic operation and navigation system is difficult. Even when considering the roads a disabled user might want to use must consider issues such as the shortest route may not be selected if the route has an uneven road and the user has less muscular strength and less endurance than is necessary to use that route. In cloud servers, environmental data that the individual has collected is sent and shared to develop a database. In the edge servers, current (real-time) environmental data is collected. Finally, in order to provide information tailored to each person's behavior individual demographic data, physical fitness and judgment ability, is given to navigation and drive actuators (i.e., wheelchairs). Having an actuator drive work with minimum delay from when an event occurs is also effective to lessen risk. For example, in order to watch a street-event with a high-quality high-realistic sensation on a users' smart phone in a remote area, the network system is requested to have high-speed performance, with a peak-data-rate of 40 Gbit/s when transferring data from the street-side smart phones and fixed cameras to an edge server in BBU.

Required capabilities	Peak data rate	X
	User experienced data rate	
	Latency	X
	Mobility	X
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	X
	Others	

7.2.3 Industries/Verticals

Fig. 7.2-3 Remote control of agricultural machines

In this section, the usage scenarios described provide novel methods to enhance conventional ones used in verticals, such as manufacturing and agriculture. They will create additional value, by improving productivity, create new business models and new customer values. Applications of sensor networks, big data analysis, and low latency feedback for prompt actuation will develop new uses for robots, drones, instruments and machinery.

Usage Scenario #9	Robot Control
Overview	An environment with many robots moving about in an urban area, including transportation robots for delivery services, small passenger robots to ensure safe movement of people such as the elderly, children and those who are visually handicapped and unmanned aircrafts (drones) for emergency transportation of medical equipment and from the sky. These robots will move slowly (maximum 30km/h) in a wide range of areas including sidewalks with many pedestrians, roadways with many cars driving, and in the sky above them. In addition, these robots may change their positions if an area is crowded. When trouble or an accident occurs, an operator may control individual robots remotely, send an emergency avoidance operation instruction to robots in a specified area, or may request support for a robot that is having trouble. Examples of the use of 5G networks in this scenario include: - If a high resolution movie from a robot's camera is transmitted uncompressed for low latency in an emergency

	situation, the peak data rate required will be over 1Gbps. - A robot moves around 8cm per 10ms at 30km/h. If the distance between robots is reduced to about 30cm, communications with ultra-low delay in the order of msec will be required for safe and continuous movement of robots in the case of unexpected accidents. - In a normal situation use-case, it is assumed that the density of robots in a region of $100m^2$, such as at an intersection, is one robot per $1m^2$. When each robot generates an average 2Mbps traffic, the total traffic in the area is 2Mbps x 100robots / (0.01km x 0.01km) = $2Tbps/km^2$. Although this area is small, the density of the traffic causes a high load in and around this area. Another assumption can be that there is an average 20 robots at each intersection of a 90m grid road based on the Manhattan model. When each robot creates 1Gbps traffic, the total traffic in this area is 1Gbps x 20robots / (0.09km x 0.09km) = $2.4Tbps/km^2$. This traffic also causes a high load. - High speed unmanned aircraft will require stable, always-on communication connections over 1Gbps with an unterminated handover. The above robot use-case scenarios will be realized with other user's traffic. A 5G network system is required to satisfy the above requirements.

Required capabilities	Peak data rate	X
	User experienced data rate	
	Latency	X
	Mobility	X
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	
	Area traffic capacity	X
	Others	Group Mobility

Usage Scenario #10	Smart agriculture
Overview	Automated/autonomous driving/operations of agricultural machines, e.g. tractors and harvesting machines ● Remote control of agricultural machines, such as tractors. Remote control of tractors, soil cultivators, planters and/or harvesting machines without on-board operations/controls. The machines can be controlled both in close proximity of several tens of meters to as far away as several hundred kilometers. ● Remote monitoring and control by human, compared with fully autonomous driving of agricultural machines, requires low-latency or no codec, i.e. no information source coding. Therefore, large data rate requirements for transmitting monitoring video become necessary. Coding schemes, such as HEVC (high efficiency video coding), cannot be used due to its large coding latency. ● Remote control or autonomous driving of agricultural machines means an on-board human driver/operator is no longer required. This allows for high speed operation/driving of agricultural machines, as no human operator/drivers are onboard, removing the need of low speed operation/driving to ensure the safety of those operating the machines. This will further improve efficiency of agriculture work, since the rapid operation/driving of agricultural machines will reduce the overall operating/driving time while working. In this case, communication latency should be as low as possible. [IT agriculture] ● Agriculture work does not always require low latency is not always required. The ability to sustain massive connections, however, would be required. The machinery at a typical agricultural operation might include a water pump that would provide water to agricultural fields, a drainage water pump, an on/off machine of sprinkling water machine, an

	electric fan to prevent frost for farm products. Overall, there would be many devices that could be connected to a network. ● IT-led agriculture would require a periodical data collection system to collect small size data from water temperature sensors, anemometers, air temperature sensors, humidity sensors, daylight sensors, and soil humidity sensors. ● Big data collected from sensors would then be shared by a regional entity such as JA (Japan Agricultural cooperatives) ● Big data would also be processed at the point where the data is gathered and merged, e.g. averaging the data, eliminating abnormal values. Big data collected from the fields would also be used and shared by local agricultural experimental centers for species breeding.	
Required capabilities	Peak data rate	
	User experienced data rate	
	Latency	X
	Mobility	X
	Connection density	
	Energy efficiency	
	Spectrum efficiency	X
	Area traffic capacity	
	Others	

7.2.4 Countermeasures in emergency and disaster situations

Fig. 7.2-4 Enhanced Emergency call

In this section, the usage scenarios that provide countermeasures against emergency situations such as traffic accidents and sudden illnesses, or disaster situations caused by earthquakes, floods, fires and typhoons. These countermeasures are intended to support initial responses, confirming the safety of victims, providing evacuation guidance and assisting in rescue attempts.

<table>
<tr><td>Usage Scenario #11</td><td colspan="2">Anti-Crime System with Image Recognition</td></tr>
<tr><td>Overview</td><td colspan="2">Performs criminal searches and tracking based on videos /images captured by surveillance cameras or mobile phones carried by civilians while reporting to the police and contacting the families of the victims immediately.
Collects video information related to a location when having received a warning notification from a GPS-enabled anti-crime mobile phone, and determines presence or absence of any suspicious activity. If suspicious activity is confirmed the results of analysis are sent to the Cloud to notify the police.
Links to the location/time information and video information are confirmed using GPS and are combined with image recognition features to enable a better understanding of the presence or absence of suspicious activity and the situation at the time in detail.
The Network system requirements will include a real-time requirement (latency) of several hundred milliseconds from the time of notification from mobile phones or surveillance cameras and the detection of suspicious activity at an edge server to the report to the police terminal, and high-speed performance (Peak Data Rate) to transmit data from camera/mobile phones to the BBU edge server at 4Gbit/s at regular times and 5Gbit/s at the peak time.</td></tr>
<tr><td rowspan="7">Required capabilities</td><td>Peak data rate</td><td>X</td></tr>
<tr><td>User experienced data rate</td><td></td></tr>
<tr><td>Latency</td><td>X</td></tr>
<tr><td>Mobility</td><td>X</td></tr>
<tr><td>Connection density</td><td>X</td></tr>
<tr><td>Energy efficiency</td><td></td></tr>
<tr><td>Spectrum efficiency</td><td></td></tr>
</table>

	Area traffic capacity	X
	Others	

Usage Scenario #12	Enhanced Emergency Call, Large Scale Disaster Rescue Network
Overview	● Enhanced Emergency Call - Emergency calls to perform an automatic call originating with or without the consciousness of the injured person; - Supporting ambulances equipped with remote high quality, low latency video transmission communication to operate effectively; - The ambulance delivers the patient's vital data to a medical institution, including a high-definition video image, en route to the institution. ● Securing of traffic accident data - A rapid data uploading from the drive recorder as injured person rescuing supplement information at the scene of the traffic accident or as evidence information of the accident decision in court.

Required capabilities	Peak data rate	X
	User experienced data rate	
	Latency	X
	Mobility	X
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	X
	Area traffic capacity	X
	Others	Reliability

Usage Scenario #13	Emergency Calls for Earthquake/Tsunami
Overview	● Mobile networks will handle multiple originating calls to confirm people's safety or to facilitate urgent communications after an earthquake; ● Mobile networks will send specific warning information for a

	tsunami as well as specific evacuation course instructions to individuals; ● Setup of the substitute facilities for communication lines which will be damaged by a large-scale disaster and providing access to IP networks; ● High reliability / high quality / low latency communication systems for required rescue operations and remote medical operations; ● Automatic driving functions control abandoned cars left on a side of a road to assist in unmanned evacuation measures; ● High reliability / high resolution video / low latency communication systems to control unmanned remote robot heavy industrial machines to assist in road clearing to secure access to disaster areas by emergency teams; ● Triage information and communication systems for monitoring disease outbreaks among victims in relief camps; ● High reliable / high resolution video / low latency / Highway mobile communication systems connecting medical helicopters and medical institutions in order transfer information on seriously injured patients being brought to the institution; ● Establishment of communication systems to assist in the second and third stages of relief assistance, such as safety and location confirmation of refugees relief; ● Information and communication systems to support disaster relief headquarters, such as providing high definition video.

Required capabilities	Peak data rate	X
	User experienced data rate	
	Latency	X
	Mobility	X
	Connection density	X
	Energy efficiency	
	Spectrum efficiency	X
	Area traffic capacity	X
	Others	Reliability

7.3 Dynamic approach

- Not all 5G requirements are necessary to be simultaneously met in providing 5G services.
- From an economic viewpoint, 5G capabilities need to adapt to the wide and dynamic variations of 5G requirements for a particular time, space and situation, since required 5G capabilities in providing a service are dependent on each particular use case, the requirements of which differ from each other.
- 5G networks also need to be dynamically optimized to meet the dynamic variation of 5G requirements.
- A heterogeneous network is a promising approach for this optimization which will allow a 5G network to have the ability to systematically work together in different RATs, including new 5G RAT(s), which have different capabilities.
- Adaptive virtualizations, network slicing, and softwarization are crucial key factors to realize flexibility of end-to-end networks, as diverse services emerge with a wide range of traffic variation

Dynamic Approach	Dynamic adaptation of 5G networks to the dynamic variations of 5G use scenes. Behavior of a football fans while watching a match at a big stadium
Overview	● The usage scene describes one in which many fans arriving and leaving a stadium share information or experiences they have with each other by watching high-definition video over their smart phones ● In this scene, the following features are observed from a viewpoint of radio access networks and fixed networks; 1. 【Variation in time】Communication traffic rapidly increases in the few hours before the match begins until a few hours after the match is completed. This phenomenon is related to the movement of people. In addition, this communication traffic depends on the amount of time those people to arrive at the stadium from their home as well as the applications they use en route. For example, on the train car on their way to the stadium fans may exchange e-mails and Short Message

	Service (SMS) texts with each other. They also may access the web in order to get information regarding their favorite football teams and players on the way to the stadium while on their way back from the stadium they may enjoy watching replay videos of most exciting scenes in the soccer match, even sending video clips over the smart phones, for example when their team makes a last minute goal to win the match. 2. 【Geographical variation】The rapid increase in communication traffic mentioned above also would vary depending on which transport station or stop fans use on their way to and from the stadium, including whether or not they are passing by a main trunk road. It also depends on what kind of applications they use and at what geographical point they use them. 3. 【Variation due to an event to be held】Communication traffic significantly varies depending on whether or not a match takes place. Traffic will tend to be light when there is no event or match but will become very heavy when a match or event does occur. 4. 【Burst variation due to a collective movement 】Public transportation such as trains and buses will carry many passengers, which will cause extremely heavy communication traffic when it passes by a geographical point. 5. 【Variation due to the nature of the event on the event day】The amount of communication traffic in case of an exciting match is significantly larger from than in the case of a boring one. The communication traffic also varies according to weather conditions, i.e. a clear day or rainy day. The kinds of applications used also vary according to weather conditions and the nature of the event. ● Radio communication networks up to 4G have been designed to satisfy the maximum value of the communication traffic when an event is to be held. In cases when communication

	traffic exceeds the system's pre-designed value, radio systems do not always guarantee communication quality during these periods of extremely heavy traffic according to the best effort service concept. Radio systems have employed temporal base stations in order to off-load communication traffic and minimize the degradation of its communication quality. ● In 5G networks, it is economically very difficult to design the network in order to satisfy estimated maximum amount of communication traffic, due to the wider variety of applications used, the larger amount of communication traffic, the wider variation of communication traffic and the overall more ubiquitous communication traffic compared with 4G networks. ● In order to overcome these issues, 5G networks need to introduce dynamic traffic control schemes to solve the five issues mentioned above. ● As for issue four mentioned above, in addition to the dynamic traffic control of macro and micro cells and use of wired network slicing and softwarization, 5G systems need to take a new technological approach for "group mobility use scenario" in which access points are installed within a train/bus for use by mobile terminals, such as smartphones, and radio links between access points and a fixed base station on the ground to convey aggregated traffic originally arising from the mobile terminals. ● Furthermore, 5G systems need to introduce a new smooth handover scheme which considers "users traffic line" and in which, for example, access points are set along with a road from the railroad station to the stadium. ● 5G networks need to handle communication traffic which dynamically varies by introducing "flexibility" into the network as a whole at an affordable cost.

8 Requirements for 5G

8.1 High level requirements

This chapter describes the requirements related to radio access network, front-haul/backhaul and communication networks.

5G systems should include “Extreme Flexibility”, in order to satisfy the end-to-end quality required in each use scene even in extreme conditions. End-to-end context in the ICT environment includes not only UE-to-UE, but also UE-to-Cloud, which implies that the technology focus on flexibility extends beyond 5G radio technology to the backbone networks.

8.2 Requirements related to 5G radio access network

8.2.1 Definitions of the requirements

The definitions of the requirements related to 5G RAN (Radio Access Network) are given in the following sub-clauses. Subset of the ‘candidate’ requirements may be applied to each use case. Qualitative or Quantitative requirements will be given in later stages of the study considering corresponding use cases as well as applicable technologies.

(1) Bandwidth

Bandwidth or sum of bandwidths that can be supported by a 5G RAT in order to provide a radio communication link between transmission entities to receiving entities should be defined. Scalable bandwidth which is the ability of 5G RAT to operate with different bandwidth allocations could be defined. The bandwidth may be supported by single or multiple RF carriers. The width of the bandwidths should be defined quantitative manner such as the minimum bandwidth supported or the maximum bandwidths supported.

The purpose of this requirement is to define bandwidth that 5G RAT to utilize. Plural bandwidths of different widths may be defined in conjunction with use cases considered.

(2) Transmission/ Reception Point (TRP) spectrum efficiency

Transmission / Reception Point spectral efficiency is defined as an aggregate throughput of all users divided by the channel bandwidth divided by the number of transmission/reception point. The aggregate throughput can be defined as the number of correctly received bits, i.e. the number of bits contained in the service data units (SDUs) delivered to Layer 3 over a certain period of time.

The purpose of this requirement is to define spectral efficiency of a 5G RAT. Either or both of peak spectral efficiency or xth percentile (x to be defined) spectral efficiency could be defined.

The peak spectral efficiency gives the maximum spectral efficiency achieved under ideal conditions.

The 'xth' of xth percentile spectral efficiency could be fifth for example and to define the boundary of spectral efficiency that 'xth percentile device of the whole devices could be served by a pair of TRP or a set of TRPs in question.

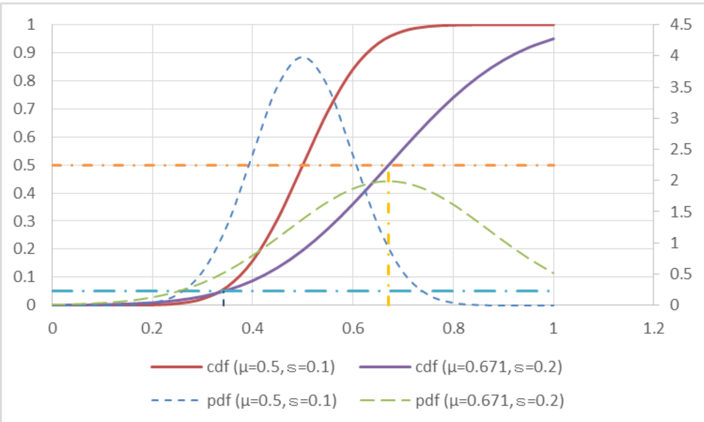


Figure 8.2-1 Example of variety of different normalized distributions and CDFs

(3) Latency

Latency could be defined either for control plane feature or user plane feature. In case control plane (C-Plane) latency is given, it could be measured as the transition time from different connection modes, e.g., from idle to active state. Ultimate requirements may be defined with actual mode states (to be defined) or use scenes (to be defined) between which 'control plane (transfer) latency' will be defined. In case user plane latency (also known as transport delay) is given, it could be defined as the one-way transit time between an SDU packet being available at the IP layer in the TRPs (the user device and the base station) and the availability of this packet (protocol data unit, PDU) at IP layer in the TRPs. User plane packet delay could include delay introduced by associated protocols and control signaling assuming the user terminal is in the active state.

(4) Mobility

Mobility requirements are given as the maximum moving speed of a user device (terminal) at which the device can provide certain quality of communication link to a TRP (aka a base station).

(5) Mobility interruption time

The mobility interruption time could be defined as the time duration during which a user device cannot exchange user plane packets with any base station.

(6) Energy efficiency

Energy efficiency could be defined for TRPs of 5G RAT as their energy consumption ratio (the increase of energy consumption) between no or limited user traffic cases to fully traffic loaded operation cases. For the devices, energy efficiency could be defined as their operational lifetime.

(7) Peak data rate

Peak data rate could be defined as the maximum data rate that a user device transmits or receives under an ideal condition.

(8) User experienced data rate

User experienced data rate could be linked to the CDF of xth percentile user spectral efficiency and the bandwidth for the data transmission.

(9) Area traffic capacity

Area traffic capacity corresponds to the total traffic throughput served per geographic area (in bit/s/m^2). This can be linked to the TRP spectral efficiency and can be derived for a particular use case or deployment scenario based on the achievable TRP spectral efficiency, network deployment (e.g., TRP (site) density) and bandwidth.

(10) Connection density

Connection density is defined as the numbers of mobile device per area that can be connected to the system.

(11) Reliability

The reliability can be defined as a success rate or success probability of data transmission over a certain period of time and gives the reliability of the communication link under certain conditions defined.

(12) Coverage

Coverage is defined as cell range expansion functionality.

8.2.2 List of 5G RAN requirements and their mapping to use cases

The mapping between use cases described in section [7] and the requirements are summarized in Table 8.2.2-1 Mapping between Use cases and 5G RAN requirements Table 8.2.2-1.

Table 8.2.2-1 Mapping between Use cases and 5G RAN requirements

Required Items	eMBB	URLLC	mMTC	Remarks
Bandwidth	X	X	X	
TRP spectral efficiency	X		X	
Peak data rate	X			
Area traffic capacity	X			
Connection density			X	
Latency	X	X		
Coverage			X	
Mobility	X			
Mobility interruption times	X	X		
Energy efficiency	X		X	
Reliability		X		

Note: 'x' denotes corresponding requirement in its row should be applied to the use case in its column. Applying relaxed or general requirements to the use cases that are not denoted by 'x' is not precluded.

8.3 Requirements for 5G networks

5G networks need extreme flexibility in order to support various applications and services with largely different requirements.

In 5G networks, it is necessary to consider end-to-end application quality and enablement through network softwarization platform.

Mobile networks will need to have the following capabilities in end-to-end connections, because diverse usage scenarios are anticipated towards 2020 and beyond:

- As broadband data traffic will continue to increase while the traffic volume varies dynamically, networks is required to have larger traffic transport capability.
- As traffic volume will vary greatly time by time depending on service type and usage scenario, networks will required to have flexible scalability.
- As a large increase of connections is foreseen due to the rapid emergence of IoT/M2M devices, networks will be required to accommodate those packets traffic characteristics which have different statistics from the other services.
- As the advent of ultra-low latency control services is anticipated in such services as tactile communications, real-time M2M, V2X, and AR, network structures will be required to be capable of lower latency data transmission.
- As various network access technologies are anticipated, the transport network needs to have wide adaptability to those access network connections.
- As some types of service devices may move faster across a wider range than the 4G use cases, mobile networks will be required to have a capability of tracking and connecting those devices with seamless communication in between the service areas.
- Lower energy consumption is expected for 5G mobile networks which has those capabilities above.

Given those aspects, 5G networks will be desired to support the composition of multiple slices, and the control and management of slices over RATs, fronthaul/backhaul, and other fixed network elements in the end-to-end path. Besides, mobile networks should have a sufficient level of scalability in terms of functions, capabilities, and components.

An end-to-end scenario for latency design

Latency is the most susceptible quality that needs careful design of overall networks. Figure 8.3-1 shows a typical end-to-end scenario based on the current mobile network architecture. The end-to-end user data path from UE to a Server is divided into 11 different segments by focusing on the major network functions involved.

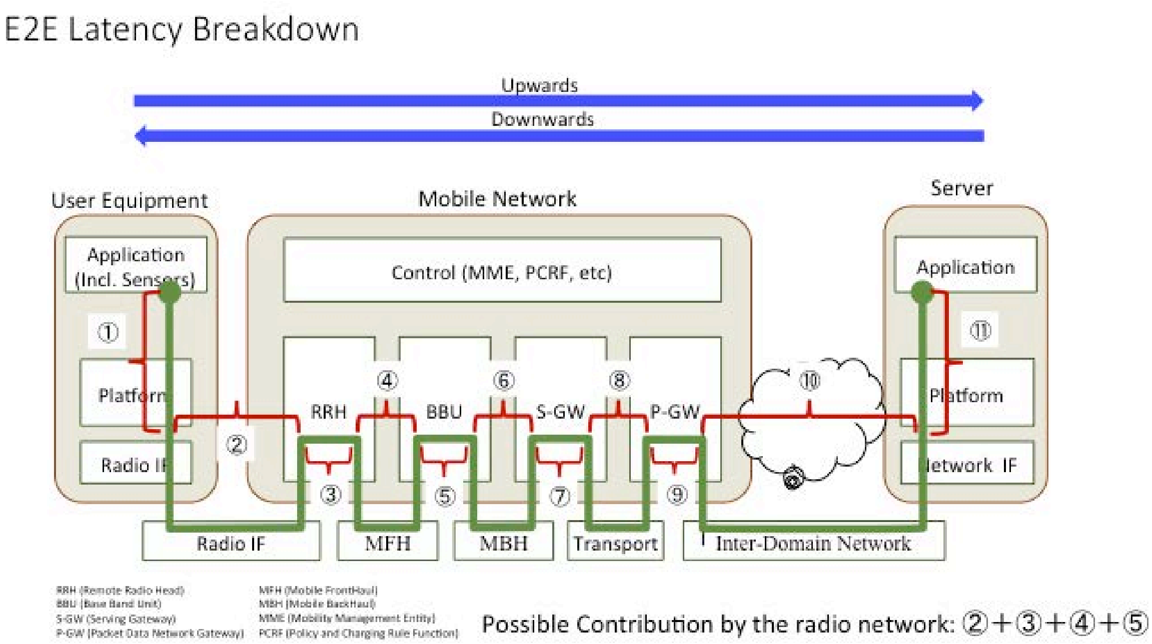


Figure 8.3-1 An end-to-end scenario based on the current mobile (LTE) network

Estimations or guarantees of latency in some form may help in designing a service. Reduction of latency in radio access segment may also help in enlarging the room for design choices. The capability of placing network functions in selected places in the overall network may also help to complete service design, in order to satisfy end-to-end requirement on the latency.

This implies that 5G network architecture must have "Extreme flexibility" in which it must be able to design networks to satisfy user requirements such as latency, capacity, throughput, device connections, and to execute network functions and services at any part along the end-to-end communication considering appropriate level of reliability, security, cost, energy consumption.

9. Spectrum Implications

To realize the "Extreme Flexibility" of 5G, it is necessary to utilize all frequency bands, including both the lower ranges (below 6GHz) and the higher ones (above 6GHz), while considering the different characteristics of each frequency band.

The first section of this chapter will describe the roles of both lower bands and higher bands, and the following section will focus on the evaluation of preferable frequency bands in the range between 6 and 100GHz. The results came from a study that includes three stages of evaluation, i.e. 5G intra-system, inter-system, harmonization point of view, respectively. The resulting preferred bands from the results of Stage 2 are then discussed.

9.1 Concept for 5G spectrum

As shown in Chapter 4, mobile communication traffic has been increasing at a high rate in recent years. This growth is predicted to continue into the next decade, especially in consideration of the addition of new traffic types generated by the variety of new use cases of 5G. This high traffic growth leads to the conclusion that the spectrum below 6 GHz, both currently used and future plans, targeted for exclusive licensing to mobile broadband operators could be exhausted before the end of the decade. Thus, there is a high demand to be able to utilize new spectra above 6GHz on top of the current spectra already in use for the mobile communication. The allocation of new spectra is not only essential for the mobile industry to expand its business opportunities, but also key factor in enabling other industries to be able to offer their customers new services and improve their productivities utilizing 5G. Without allocating new spectra, serious economic losses will occur in countries.

Industry, academia and regulatory agencies are already aware of these facts and challenges and they are working tirelessly to tackle the spectrum scarcity problem in a variety of ways. Some examples include:

- Improving spectral efficiency by employing such techniques as higher modulation and coding schemes, and spatial multiplexing in the physical (PHY) layer as well as reducing control overhead in the medium access control (MAC) layer;
- Reducing cell sizes by using less transmit power and reusing the spectrum intelligently, and more recently overlapping deployments of small cells over macro cell with interference coordination amongst cells, in other words, Heterogeneous Network, which could notably mitigate outages at the cell edge of smaller cells;

- Optimizing media content (i.e. compression) for transporting it more efficiently over the network;
- Pursuing the use of higher frequency bands above 6 GHz, which has garnered significant attention recently, even with technical challenges in terms of path loss and cost of RF components,;
- Offloading traffic to wireless LAN (WLAN), which practically improves as much capacity of the mobile network as WLAN offloading offers.

Among those efforts above, the radio communication industry is investing heavily in development of higher frequency bands above 6 GHz, including mmWave. This is primarily due to the availability and use of wider bandwidth, which will be become necessary for the enhanced mobile broadband usage scenario and increasing traffic.

In addition to the enhanced mobile broadband usage scenario, the increasing momentum of the Internet of Things (IoT) means the infrastructure for 5G will have additional requirements, including supporting massive connectivity and ultra-reliability with ultra-low latency, on top of the long-pursued challenge to enhance peak data rate. Some IoT use cases may also require larger coverage area with lower power consumption than today.

These new use cases will require the utilization of all frequency bands, including both the lower ranges of the spectrum below 6GHz and higher ranges above 6GHz, depending on the different characteristics of each frequency band. It is especially important to have Heterogeneous Network deployment across lower and higher frequency bands be common in the 5G era. This new concept of spectrum deployment will give 5G systems extreme flexibility in networks and services.

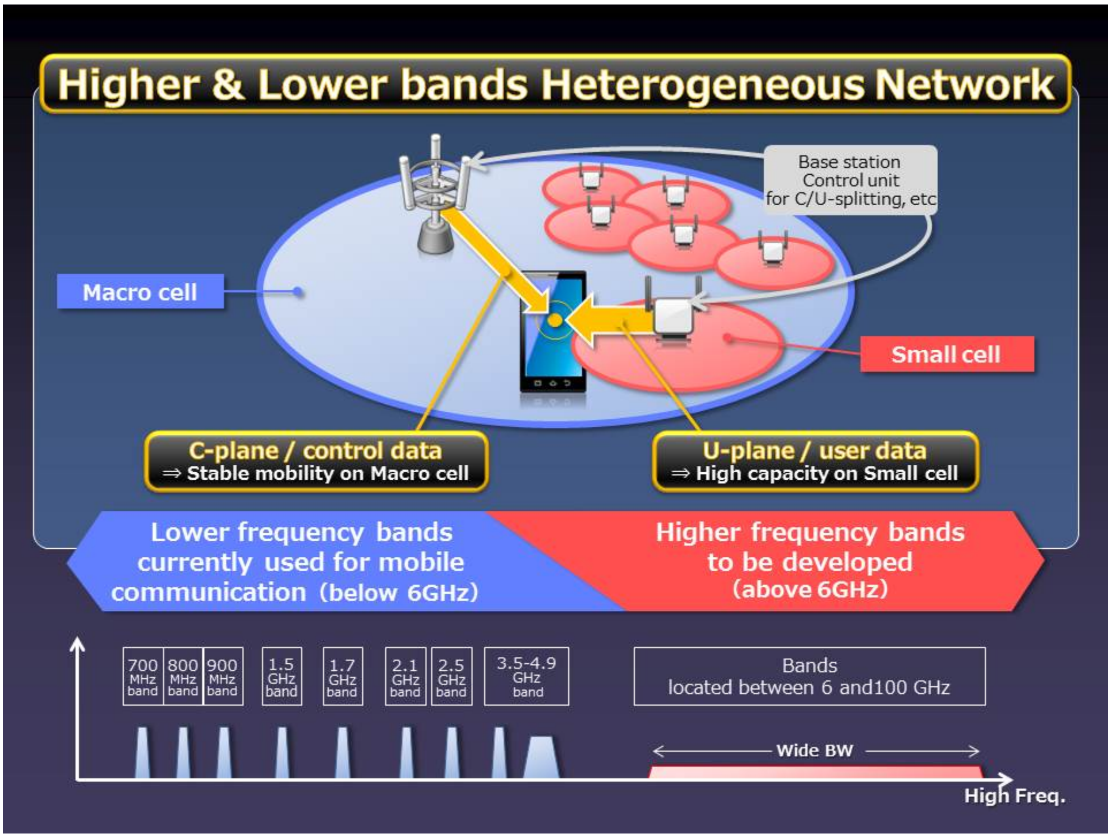


Fig. 9.1-1 Heterogeneous Network in 5G across lower and higher frequency bands

9.2 Below 6GHz

9.2.1 Roles of bands below 6GHz

Diverse demands and larger scale of performance ranges will be required in future radio networks in order to provide mobile users with a variety of services at any occasion, anytime and anywhere as needed.

In the lower frequency bands below 6GHz, radio wave propagation has the following inherent characteristics:

- Lower propagation loss due to the distance
- Larger diffraction and reflection effects, which can cover shadowed areas such as behind buildings or hilly terrain
- Better penetration from outdoor to indoor through building walls or window glass

Thus, the lower frequency bands below 6GHz are more suitable for wider and contiguous coverage. These benefits also assist the implementation of seamless mobility for tracking mobile devices.

On the other hand, the available bandwidth is relatively narrow in these bands. In addition, a larger cell radius results in the possibility of the multi-path propagation with a longer delay. A larger cell radius also results in the possibilities of larger interferences.

Taking the above characteristics into account, the bands below 6GHz will still play a very important role in the 5G era by providing:

- Conventional types of services and deployments;
- New IoT/M2M services, especially for low bit rate and low power consumption in the wide field area;
- Carriers for C-plane in C/U-split Heterogeneous Network deployment;
- Fundamental cells of 5G network for universal service continuity, where higher frequency small cells are not deployed

It should also be noted that bands below 6GHz are important for the backward compatibility and roaming support.

Services in 5G specific higher bands may be limited in some areas on location by location basis, while 4G/3G service areas are spread widely and contiguously. In order to ensure seamless service for mobile devices even in such a composite cell structure, user devices should be able to work accessing either 3G/4G or 5G radio network. This means backward compatible operation will be required for future mobile networks. In order to be backward compatible, conventional lower frequencies will be required to work as the fundamental band even within the structure of a 5G radio access network.

In addition, considering the frequency arrangement commonly used worldwide according to the already identified spectra for IMT by ITU-R under the international harmonization, the bands below 6GHz are thought to be suitably applicable for international roaming, at least at the beginning of the 5G era.

9.2.2 Technical Implementation and Challenges

Since the bands below 6GHz have been used in 3G/4G systems, technical challenges for utilizing these bands are the same as 3G/4G and therefor do not pose a challenge for 5G.

In principle, it seems there are no fundamental issues left in order to utilize these bands. However, there are some difficulties and challenges in implementation. One difficulty is the complexity of mobile devices needed to support the diverse frequency arrangements. As described in “ARIB 2020 and Beyond Ad Hoc Group White Paper”[9-1], the number of the logical frequency bands specified in 3GPP has been increased to more than 40 bands including both FDD and TDD[9-2]. Furthermore, the different combinations of carrier aggregation have been increasing over time. Small fragmented bands are another difficulty to overcome to use the wider bandwidth required in 5G.

It will also be necessary to develop mitigation techniques and coordination schemes for frequency sharing with incumbent radio systems for the new frequency bands to be introduced towards the year 2020.

9.2.3 Current spectrum allocation and its plan in Japan, below 6GHz

Fig.9.2.3-1 shows current allocation for IMT, Broadband Wireless Access (BWA) and Personal Handy-phone System (PHS) in Japan, which includes:

- 700M, 800M, 900M, 1.5G, 1.7G, and 2GHz bands for LTE/LTE-Advanced (FDD)
- 1.9GHz band for PHS
- 2.5GHz band for BWA
- 3.5GHz band for LTE-Advanced (TDD)

The total bandwidth allocated for these uses is approximately 640MHz.

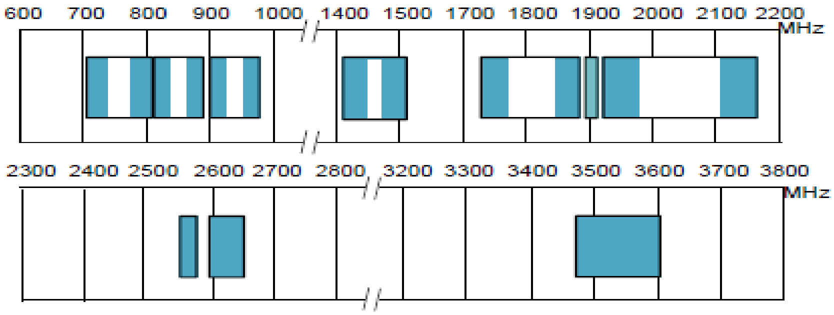


Fig. 9.2.3-1 Allocation for IMT, BWA and PHS in Japan (as of Dec. 2015) [9-3]

According to the report of the Commission on Radio Policy Vision in 2014[9-4], the total amount of spectrum bandwidth targeted for IMT, BWA, PHS and WLAN is 2700MHz for the frequency below 6GHz by the year 2020 in Japan. The candidate bands to be newly allocated below 6GHz include the 4GHz band (3.6-4.2GHz) and the 4.5GHz band (4.4-4.9GHz).

9.2.4 Spectrum identified for IMT below 6GHz in WRC-15

IMT bands have been identified in ITU Radio Regulation as a result of WRC (World Radiocommunication Conference).

In WRC-15 (Nov. 2015), among the bands proposed by Japan, 1.5GHz band (1427-1518MHz) which has already been allocated for mobile phone systems in Japan since 2G era, has been identified for IMT globally.

9.3 Above 6GHz

9.3.1 Roles of bands above 6GHz

Future 5G systems will need to support very high data throughput to cope with the growth of the data traffic and new and emerging usage scenarios, especially for enhanced mobile broadband scenarios. These scenarios may require contiguous and broader bandwidth than that of current mobile systems. Furthermore the development of technologies such as small cells, 3D beamforming and massive MIMO techniques may realize their full potential when applied to smaller wavelength, which is characteristic of higher frequency bands.

In order to address the challenges above, there has been ongoing research and development concerning the suitability of mobile broadband systems in frequency bands above 6GHz[9-5]. It is expected that the use of higher frequencies above 6GHz will become one of the key enabling components of 5G.

9.3.2 Preferred frequency ranges/bands

It is required to identify the frequency ranges or bands above 6GHz on which to deploy 5G systems. To facilitate this identification, an analytical investigation is being conducted to identify these preferred frequency ranges or bands for 5G.

9.3.2.1 Procedure of investigation

The process for the analytical investigation on the frequency bands between 6 and 100GHz consists of the three stages described in the figure below.

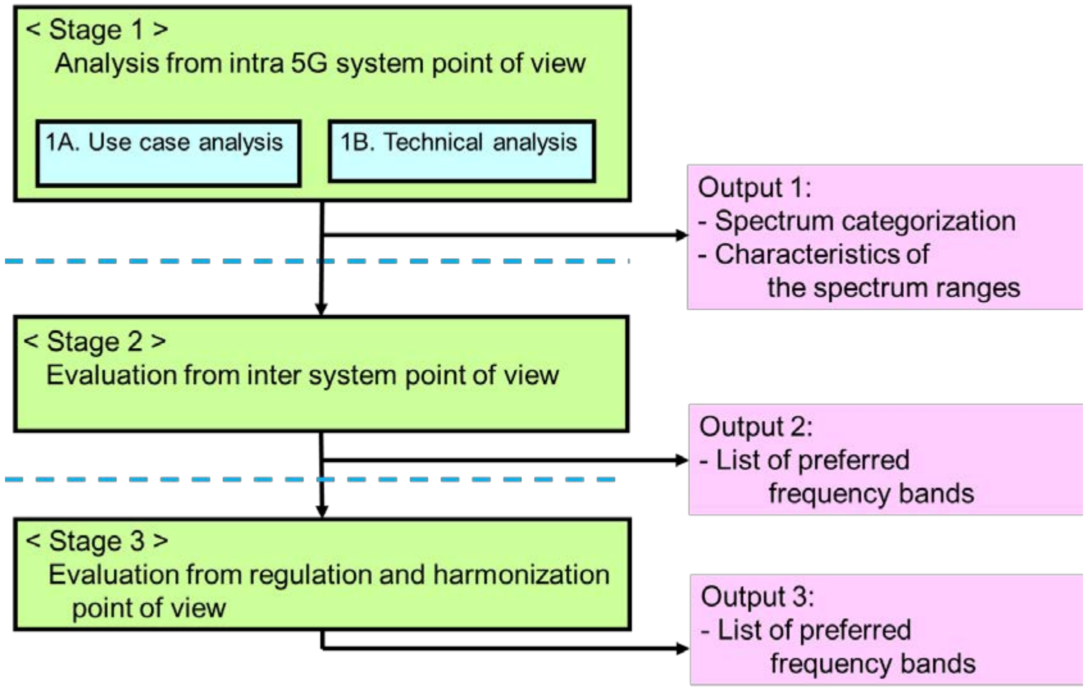


Fig.9.3.2.1-1 Process of spectrum analysis on above 6GHz for 5G

In Stage 1, the use case analysis and technical analysis were done in order to have technical characteristics of the frequency ranges 6 to 30GHz, 30 to 60GHz and 60 to 100GHz respectively.

In Stage 2, each frequency band was evaluated for 5G deployment while considering the incumbent radio systems already deployed in Japan on the frequency bands. This stage resulted in a list of preferred frequency bands

In Stage 3, the bands listed after the evaluation process in Stage 2 are being further evaluated in consideration of views from other countries and/or regions and spectrum harmonization. Stage 3 will result in another list of preferred frequency bands, which are potentially a subset of the Stage 2 output.

9.3.2.2 Stage1: Analysis from intra 5G system point of view

In this sub-section, results of the analysis for 6-100 GHz are shown, from the intra 5G system point of view, targeting to determine preferred frequency ranges from both a) bandwidth requirements coming from use case analysis for 5G era, and b) feasibility analysis on the technical aspects. Those analyses include general classification of the frequency band into several groups, and further investigation into coverage examples and typical deployment scenarios.

Firstly, in order to simplify discussion for frequency dependent characteristics, we classified the targeted frequency bands i.e. 6-100 GHz into 3 groups as follows:

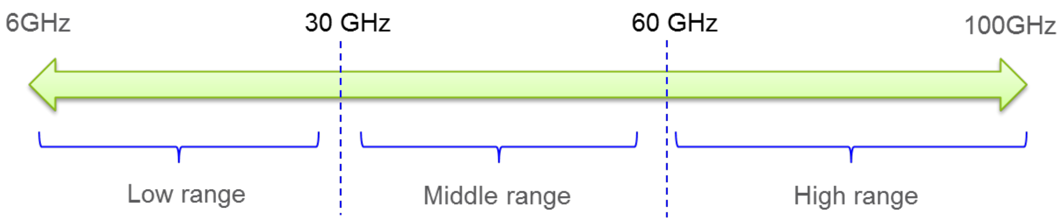
Low range 6-30 GHz / Middle range 30-60 GHz / High range 60-100 GHz

Secondly, we attempted to determine the bandwidth requirements per each spectrum range, based on use case analysis in the 5G era, which is summarized in Chapter 7.

Thirdly, we conducted a study of the following items for technical analysis.

- Frequency vs. distance attenuation
- Impact of rain attenuation
- Atmospheric influence
- Fractional bandwidth
- Antenna size / number of elements / antenna gain
- Devices (semiconductors, filters, connectors, etc.)

Finally, we concluded the study of stage 1 as described below (Fig.9.3.2.2-1).



Spectrum range	Low (6 – 30GHz)	Middle (30 – 60GHz)	High (60 – 100GHz)
Practical range of a contiguous spectrum bandwidth (Note 1)	Approx. 300MHz - 1.5GHz	Approx. 1.5GHz - 3GHz	Approx. 3 – 5GHz
Coverage example (Note 2)	Several 100m – Approx. 1km		Several 10m – Approx. 100m
Deployment scenario	Different scenarios for mobile communication are possible (Outdoor, Indoor, Outdoor to indoor, Hotspot and so on)		Scenarios for wider bandwidth and dense deployment (Indoor, Hotspot and so on)

(Note 1) These values are contiguous spectrum bandwidth considering fractional bandwidth of 5% with respect to the carrier frequency. The fractional bandwidth of 5% is derived by reviewing the existing 3GPP frequency bands. The values do neither represent required spectrum bandwidth (spectrum demand), nor imply any actual spectrum assignment which is subject to administrative authorities. This bandwidth is desired to be contiguous, in terms of the efficient use of spectrum and implementation (on the other hand, considering 5G applications (Mobile Broadband, M2M and so on), bandwidth of several 100MHz to several GHz is desired, however it is necessary to consider actually available bandwidth for 5G in each range).
(Note 2) The coverage values can vary depending on radio propagation condition, deployment scenario, applicable radio technologies and so on .

Fig. 9.3.2.2-1 Result of Stage 1

9.3.2.3 Stage2: Evaluation from inter system point of view

This sub-section shows the results of the Stage 2 evaluation on the preferred frequency bands from an inter system point of view. Inter system indicates considerations on the incumbent radio systems that are already deployed in Japan on the frequency bands. We evaluated the possibility of frequency sharing between IMT system and the incumbent radio systems in Japan on the frequency bands above 6GHz (6GHz to 100GHz). The evaluation was based on the following evaluation criteria decided within the 5GMF.

<Evaluation criteria of the frequency bands>

4 levels were set in order to express the results of the evaluation on possibility of frequency sharing as follows:

Level 1：No possibility for sharing

a) Bands listed in Footnote 5.340 of ITU-R Radio Regulation or in Footnote J107 of national allocation in Japan, where all emissions are prohibited.

b) Systems which are related to safety of human life and are always in use (e.g. Aeronautical radionavigation)

Level 2：Difficult for sharing

Level 3：Possible for sharing under certain conditions and worth considering for sharing

a) The incumbent radio system has already been shared with land mobile communication systems in other bands.

b) Sharing may be possible under certain operation conditions.

(The incumbent radio system is also operated by the mobile communication operator, e.g. wireless entrance for IMT, etc.)

c) Sharing is possible technically by introducing a certain sharing technology, mitigation technique, and/or geographical/allochronic isolation.

(e.g. The incumbent radio system is in use only when a disaster occurs, etc.)

Level 4：Possible for sharing

(e.g. No radio license is found in the public data base, or only radio stations for experimental or temporary operations are assigned.)

(Note) Unlicensed ISM (Industry - Science - Medical) bands were not evaluated such as 24-24.25GHz, and 57-66GHz.

Frequency bands in Levels 3 and 4 have a high possibility of frequency sharing and are therefore the preferred bands for 5G.

The results of the Stage 2 evaluation determining which frequency bands can be called level 3 or 4 are listed in the following tables. (Table 9.3.2.3-1, 9.3.2.3-2, 9.3.2.3-3)

Table 9.3.2.3-1 Results of Stage2 (6-30GHz)

Frequency Band (GHz)	Bandwidth (GHz)	Level of sharing possibility
5.925 – 7.25	1.325	3, 4
7.375 – 8.75	1.375	3, 4
10 – 10.5	0.5	3
10.55 – 10.68	0.13	3
10.7 – 11.7	1.0	3
14.5 – 15.35	0.85	3
15.4 – 21.4	6.0	3, 4
22 – 23.6	1.6	3
24.75 – 31	6.25	3, 4

Table 9.3.2.3-2 Results of Stage2 (30-60GHz)

Frequency Band (GHz)	Bandwidth (GHz)	Level of sharing possibility
31 – 31.3	0.3	4
31.5 – 42.5	11	3, 4
45.3 – 47	1.7	4
47 – 50.2	3.2	3, 4
50.4 – 52.6	2.2	3, 4
54.25 – 57	2.75	3

Table 9.3.2.3-3 Results of Stage2 (60-100GHz)

Frequency Band (GHz)	Bandwidth (GHz)	Level of sharing possibility
66 – 76	10	3, 4
81 – 86	5	3
92 – 100	8	4

All results of evaluation are shown in the following Figures (Fig. 9.3.2.3-1 to Fig. 9.3.2.3-4).

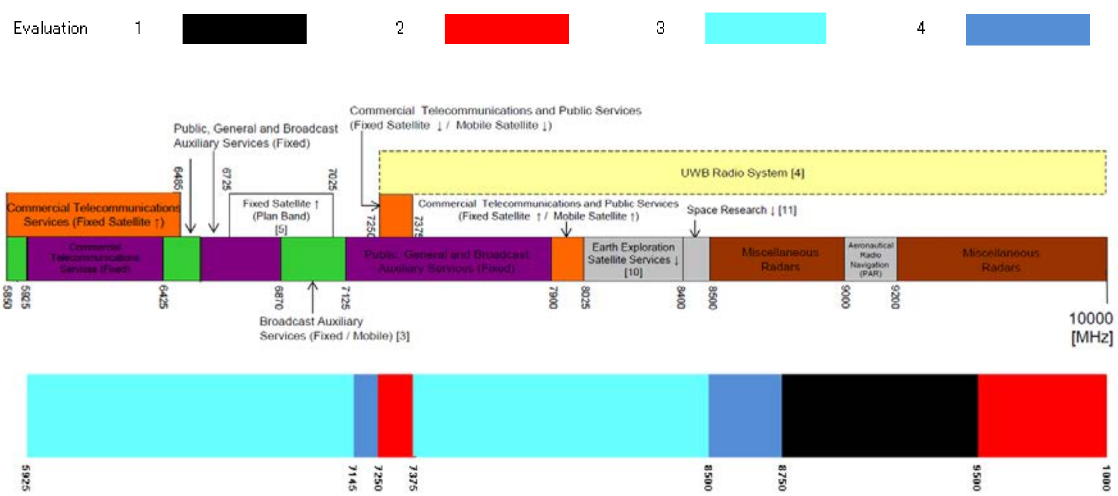


Fig. 9.3.2.3-1 Result of Stage 2 (6-10GHz)

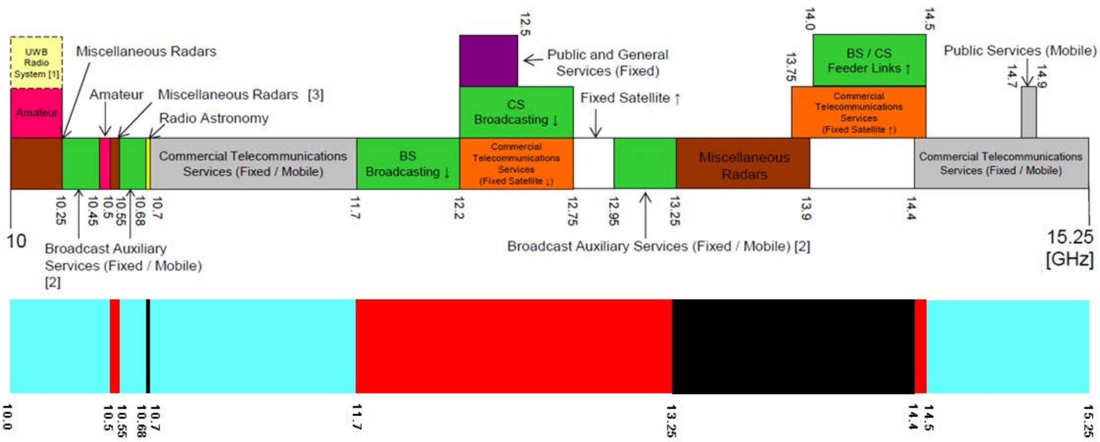


Fig. 9.3.2.3-2 Result of Stage 2 (10-15.25GHz)

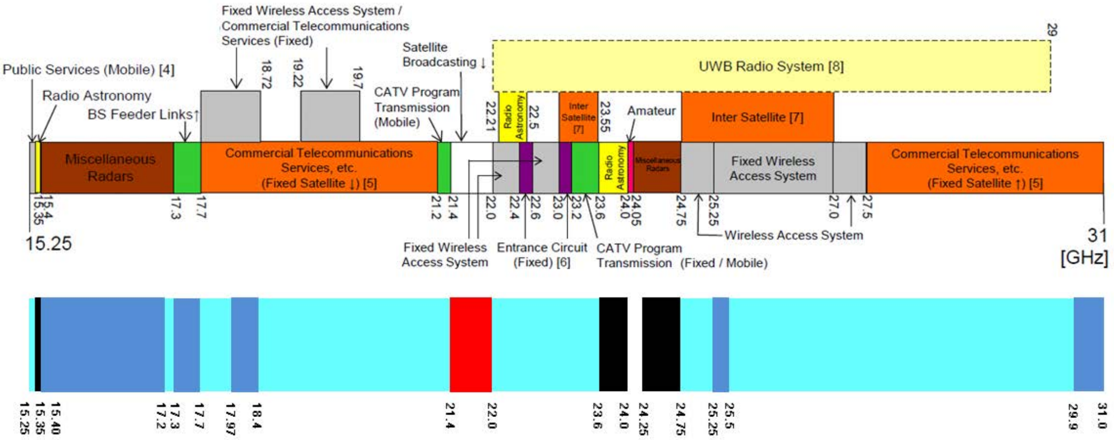


Fig. 9.3.2.3-3 Result of Stage 2 (15.25-31GHz)

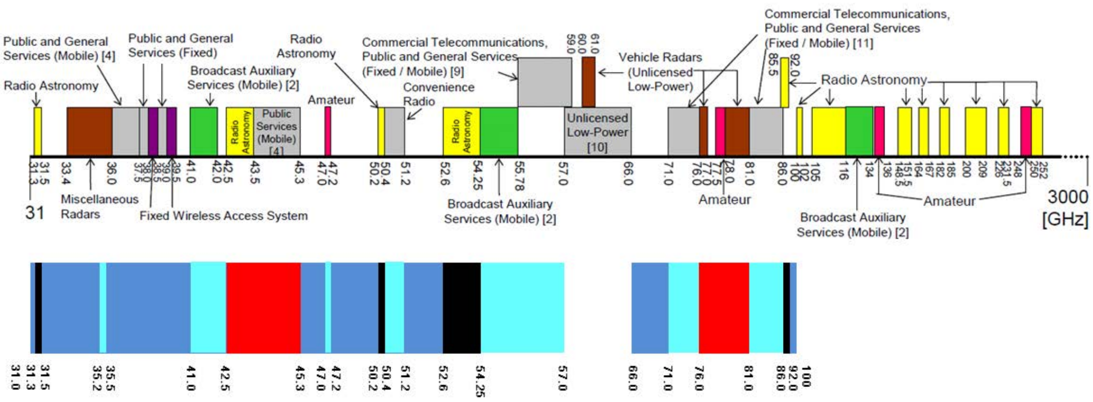


Fig. 9.3.2.3-4 Result of Stage 2 (31-100GHz)

9.3.2.4 Stage3: Evaluation from regulation and harmonization point of view

5GMF will continue further evaluation from the point of view of international harmonization. The spectrum ranges in WRC-19 Agenda Item 1.13 (see section 9.3.4) will be also taken into account.

9.3.3 Technical implementation issue and Challenges

Frequency bands above 6GHz have already been used for many years for satellite communications, radar applications, fixed services, short range & broadband WLAN (WiGig®), etc. Since many technologies have already been developed utilizing these higher bands and there seems no fatal issue to introduce 5G in these bands.

This will, however, be the first time a cellular system is deployed in these bands and there are still some technical issues and challenges to be solved from a deployment and/or implementation point of view.

9.3.3.1 Propagation losses

One major characteristic of higher frequency bands is propagation loss.

(a) Loss due to distance / Antenna gain

Propagation loss becomes larger as frequency becomes higher. Across the same distance, the loss increases in 6dB when the frequency is doubled.

On the other hand, the size of the antenna element becomes smaller as the frequency increases. This means that, in higher frequency bands, high antenna gain can be obtained by directional antennas, for example an array antenna, with a realistic physical size. Since antenna gain of same aperture size increases in 6dB when frequency is doubled, the increase of propagation loss can be compensated using array antennas or other similar antennas.

However, antenna beam width becomes narrower as antenna gain increases. Also, hardware complexity increases as the number of elements in an array antenna increases. Thus, there is a trade-off between antenna gain and coverage area.

(b) Loss due to gasses, vapor, rain and mist

In the case of several 10 meters to a couple of 100 meter coverage distance, most of these characteristics seem to be of little concern compared to the free space loss due to distance. The loss around 60GHz due to gasses and the loss due to heavy rain (>30mm/h) should be taken into account in certain deployments.

However, these attenuations vary as temperatures and air pressure changes. Although the variation is not very large, it should be noted that there may occur unwanted “cell breathing” due to air temperature and pressure which is not seen in lower frequency band.

Detailed figures of these characteristics are shown in ANNEX [9].

9.3.3.2 RF Devices and Components

(a) Power amplifier devices for transmitter

One of key devices used in higher frequencies are power amplifiers for transmitters, both in the base station and in the terminal.

Such devices for 6 to 100 GHz are already available, although cost reduction and performance improvements are still necessary.

(b) Filters and passive devices

Passive devices including filters give us another challenge in higher frequency bands.

Some of the existing filter types used in the frequency below 6GHz may not work for the frequency bands above 6GHz. The introduction of new types of filters like the miniaturized Surface Mount Device (SMD) type waveguide filter is also necessary

(c) RF connectors and cables

The wavelength at very high frequencies is comparable or even shorter compared to the physical size of RF connectors, which makes it difficult to have good impedance matching. In addition, the increase of AC resistance due to skin effects becomes more remarkable.

Besides improvements to RF connectors and cables, these effects can be avoided by careful design of equipment, e.g. using frequency down conversion so that RF connectors and/or RF cables are only inserted at IF (Intermediate Frequency), etc.

9.3.4 WRC-19 Agenda Item 1.13

The WRC-15 decided to consider the identification of frequencies for IMT in the frequency range of 24.25GHz-86GHz at WRC-19 under Agenda Item 1.13.

Table 9.3.4-1 shows comparison between frequency bands listed under WRC-19 AI1.13 and the result of Stage 2 evaluation (levels 3 and 4) in 5GMF.

5GMF will continue to evaluate all the frequency bands listed in this table.

Table 9.3.4-1 Frequency bands listed under WRC-19 AI1.13
and the result of 5GMF Stage 2 evaluation
(green: which have allocations to the mobile service on a primary basis
yellow: which may be considered for additional allocations to the mobile service on a primary basis)

(a) 6-30GHz

WRC-19 Agenda Item 1.13			5GMF Stage2 (levels 3 & 4)		
BW(GHz)	frequency band(GHz)		frequency band(GHz)		BW(GHz)
			5.925	7.250	1.325
			7.375	8.750	1.375
			10.000	10.500	0.500
			10.550	10.680	0.130
			10.700	11.700	1.000
			14.500	15.350	0.850
			15.400	21.400	6.000
			22.000	23.600	1.600
3.250	24.250	27.500			
			24.750	31.000	6.250

(b) 30-60GHz

WRC-19 Agenda Item 1.13			5GMF Stage2 (levels 3 & 4)		
BW(GHz)	frequency band(GHz)		frequency band(GHz)		BW(GHz)
			24.75	31.00	(repeated)
			31.00	31.30	0.30
			31.50	42.50	11.00
1.60	31.80	33.40			
3.50	37.00	40.50			
2.00	40.50	42.50			
1.00	42.50	43.50			
			45.30	47.00	1.70
1.50	45.50	47.00			
0.20	47.00	47.20	47.00	50.20	3.20
3.00	47.20	50.20			
2.20	50.40	52.60	50.40	52.60	2.20
			54.25	57.00	2.75

(c) 60-100GHz

WRC-19 Agenda Item 1.13			5GMF Stage2 (levels 3 & 4)		
BW(GHz)	frequency band(GHz)		frequency band(GHz)		BW(GHz)
10.00	66.00	76.00	66.00	76.00	10.00
5.00	81.00	86.00	81.00	86.00	5.00
			92.00	100.00	8.00

References

[9-1] ARIB 2020 and Beyond Ad Hoc Group White Paper, "Mobile Communications Systems for 2020 and beyond" Version 1.0.0, Oct. 2014

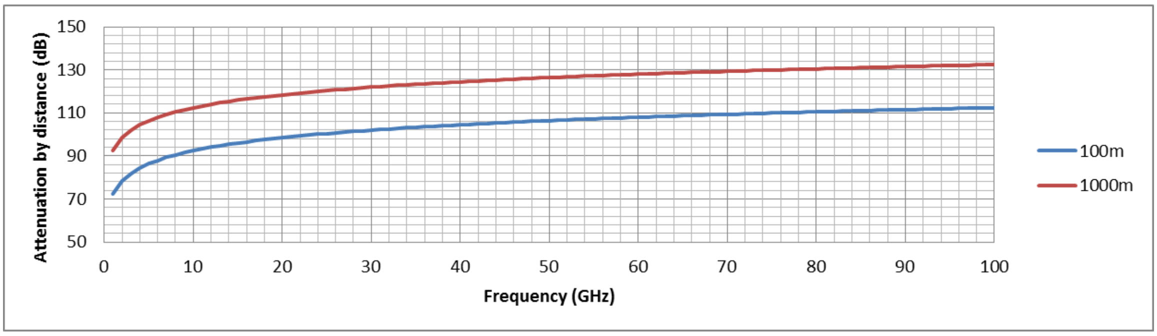
[9-2] 3GPP TS 36.101 V13.2.1, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 13)", Jan. 2016

[9-3] Ministry of Interal Affairs and Communications, Information and Communications Council #114 materials, Dec.11, 2015 (in Japanese)

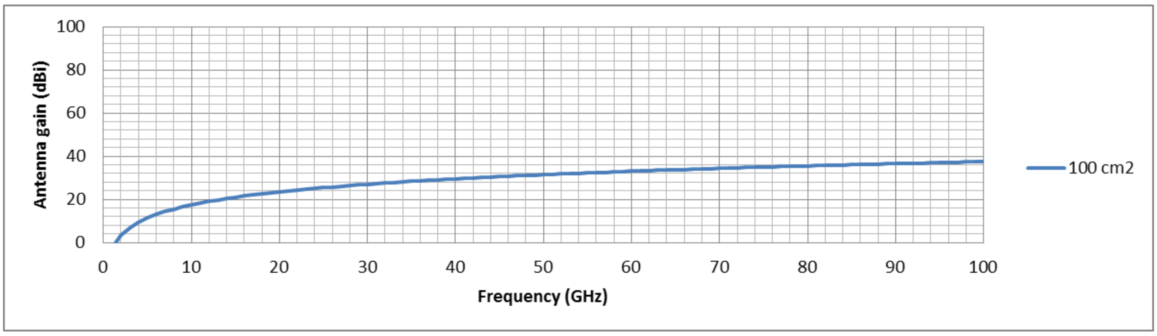
[9-4] Ministry of Interal Affairs and Communications, Report of the Commission on Radio Policy Vision, Dec. 26, 2014 (in Japanese)

[9-5] Report ITU-R M.2376-0, "Technical feasibility of IMT in bands above 6 GHz", July 2015

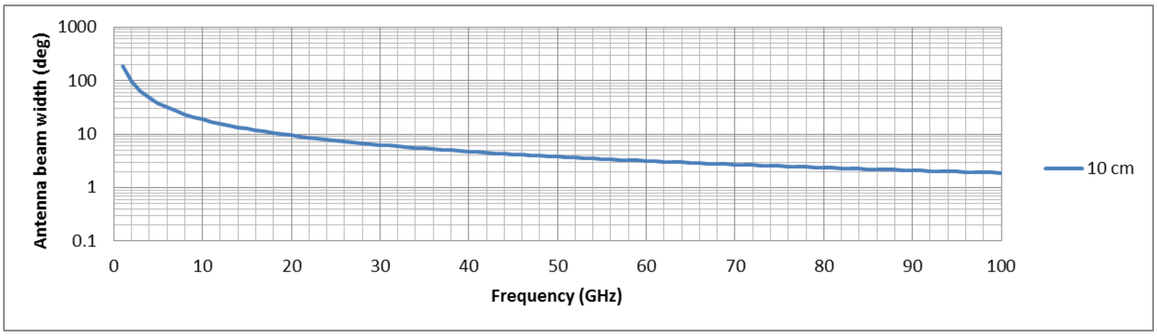
Annex [9] Characteristics of propagation loss above 6GHz



(a) Propagation loss in free space vs. frequency

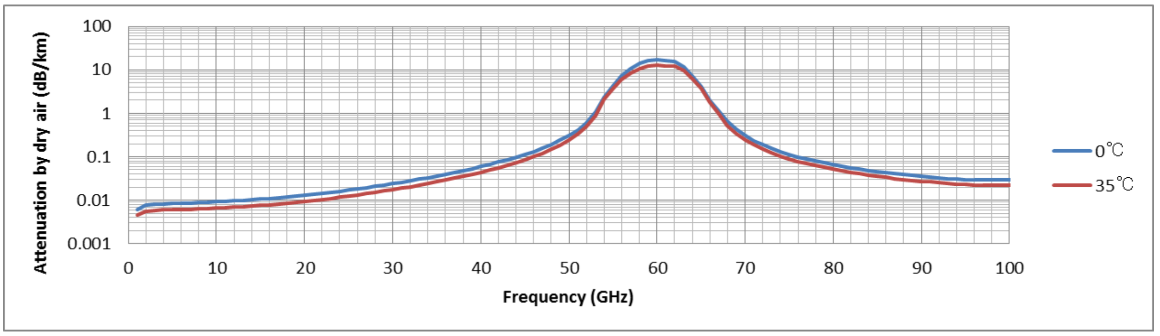


(b) Antenna gain of directional antenna with fixed aperture area

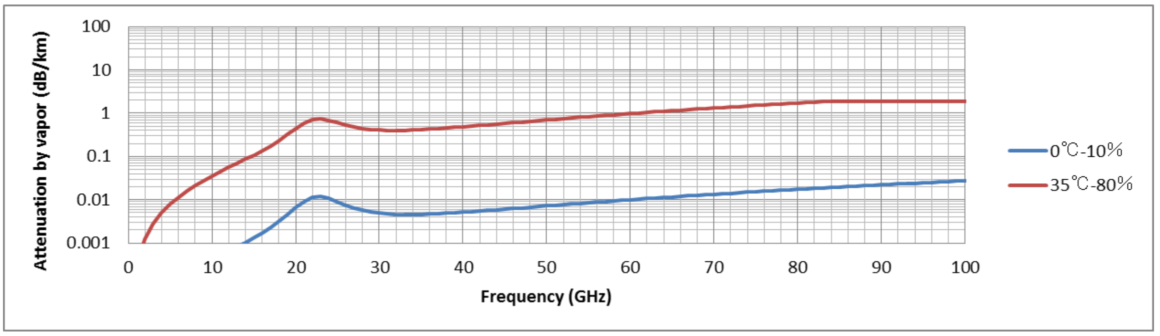


(c) Antenna beam width of directional antenna with fixed antenna width

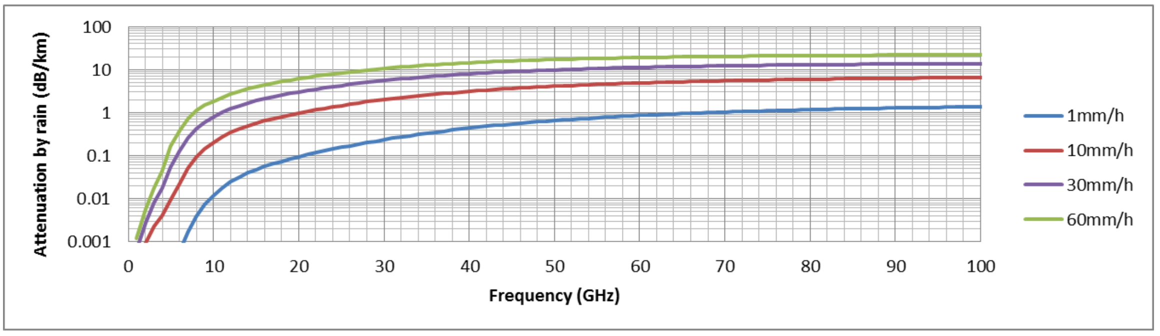
Fig.A[9]-1 Frequency dependency of propagation loss in free space and antenna characteristics



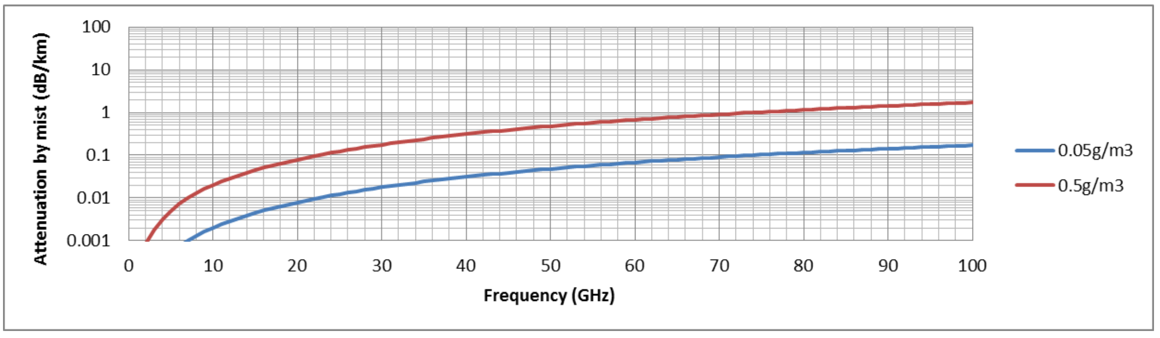
(a) loss due to gas (dry air)



(b) loss due to vapor



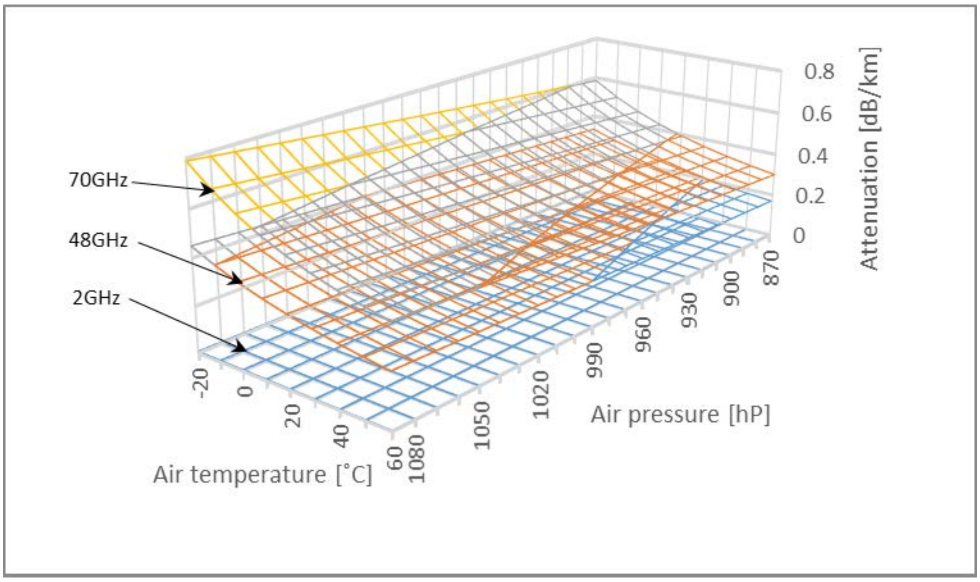
(c) loss due to rain



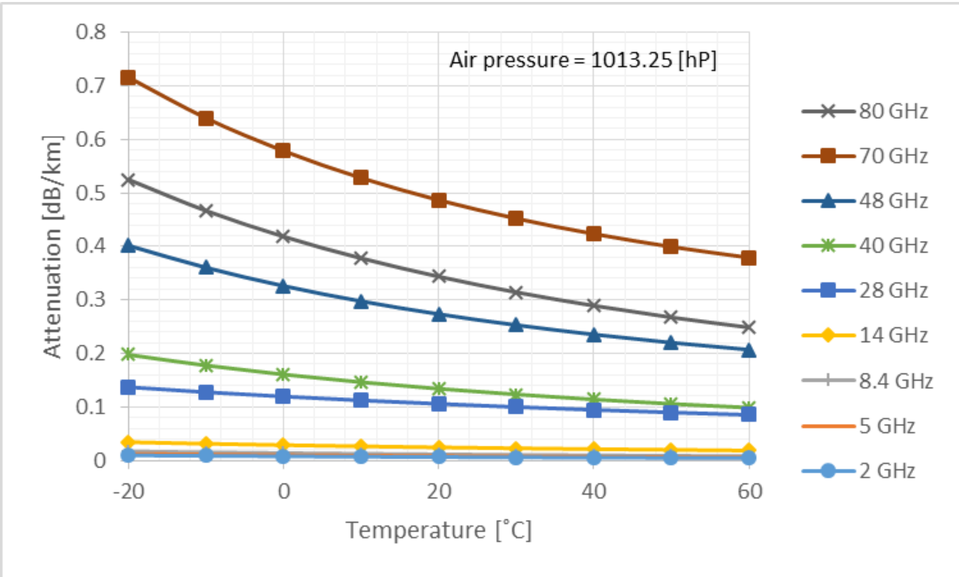
(d) loss due to mist

Fig.A[9]-2 Frequency dependency of propagation loss due to gas, etc.

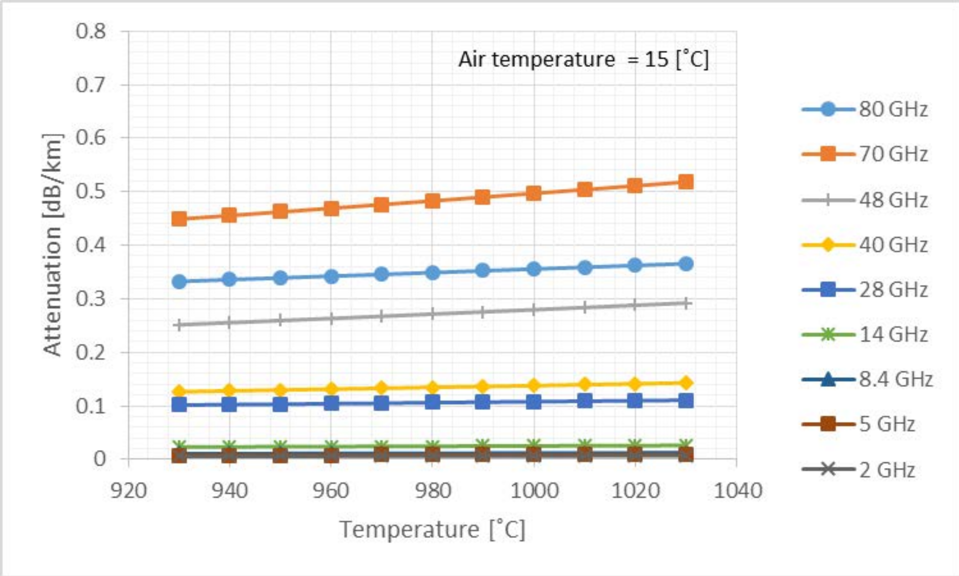
[A9-1][A9-2][A9-3]



(a) Attenuation due to atmospheric gasses (vs. air pressure and temperature)



(b) Attenuation due to atmospheric gasses vs. air temperature



(c) Attenuation due to atmospheric gasses vs. air pressure

Fig.A[9]-3 Detailed analysis of attenuation due to atmospheric gasses [A9-1]

References

[A9-1] ITU-R, "Attenuation by atmospheric gases", Recommendation ITU-R P.676-10, Sep. 2013

[A9-2] ITU-R, "Specific attenuation model for rain for use in prediction methods", Recommendation ITU-R P.838-3, Mar. 2005

[A9-3] ITU-R, "Attenuation due to clouds and fog", Recommendation ITU-R P.840-6, Sep. 2013

Terminology, abbreviations

Abbreviation	Unabbreviated term
3D	Three Dimensional
AC	Alternating Current
BW	BandWidth
C-plane	Control plane
FDD	Frequency Division Duplex
IMT	International Mobile Telecommunications
IoT	Internet of Things
ITU	International Telecommunication Union
ITU-R	ITU Radiocommunication sector
M2M	Machine To Machine
MIMO	Multiple Input Multiple Output
RF	Radio Frequency
TDD	Time Division Duplex
U-plane	User plane
WLAN	Wireless Local Area Network
WRC	World Radiocommunication Conference

10 Overview of 5G Technologies

In the following two chapters, several key technical enablers are discussed.

In Chapter 11, the latest updates concerning several radio access technologies are chosen and summarized. Each technology has specific advantages supporting or serving under certain foreseen use cases and could be an essential part of a heterogeneous network. This collected information would be useful and beneficial in consolidating an actual 5G communication system after carrying out relevant studies at the ITU-R or establishing technical standards at standardization organizations. The related baseline research can be found in [1].

Chapter 12 outlines the network technology in order to realize ‘extremely flexible networks’ which will be able to serve a variety of services having variety of demands in real-world markets. Four focus areas, namely ‘network softwarization’, ‘network operation and maintenance’, ‘network fronthaul and backhaul’, and ‘mobile edge computing’, have been considered. One major area of consideration is ‘network softwarization’ which includes penetrating Network Function Virtualization (NFV) / Software Define Networking (SDN), with the ultimate goal being to virtualize network functions. This chapter spans technology outlines and their use cases as well as technical challenges necessary to deploy a working network that is ‘extremely flexible’

Reference:

[1] “Mobile Communications Systems for 2020 and beyond,” ARIB 2020 and Beyond Ad Hoc Group White Paper, Oct.2014.

11 5G Radio Access Technologies

11.1 General

The following subsections discuss promising radio access technologies in order to realize 5G system. The subsections contain information on the latest radio access technologies embraced in [1] or newly introduced technologies. The 5G communications system should be constructed by selecting, combining or modifying these technologies in order to make 5G systems work in each use case.

11.2 Overview of 5G radio access network

The radio access network (RAN) and aggregated backhauls support the capabilities of data transport, radio transmission and reception. In the 5G era, these capabilities shall be enhanced to accommodate massive traffic capacity and device connectivity while providing enhanced quality of user experience.

As has been mentioned in the previous chapters, '5G' communications system should serve wide range of use cases. Depending on each of these use case, range of required capabilities to radio access technologies would be extremely different. Consequently '5G' communications system should be an intrinsic and genuine-type heterogeneous network which utilizes every proper radio access technologies according to required capabilities of the use case concerned. Intrinsic and genuine-type heterogeneous network would not be a simple 'overlaid cellular networks' aiming improved communication capacity anymore but it should be a consolidated communication system consist of functional elements tailored to each of the use cases and serve them in a suitable manner.

Lots of innovative technologies mentioned in the following sub sections will be introduced to improve the performance in the system for 2020 and beyond. Some of these technologies are illustrated in Fig. 11.2-1 [1].

It should be noted that the figure below illustrates candidate radio access technologies of '5G' and subject to further refinement.

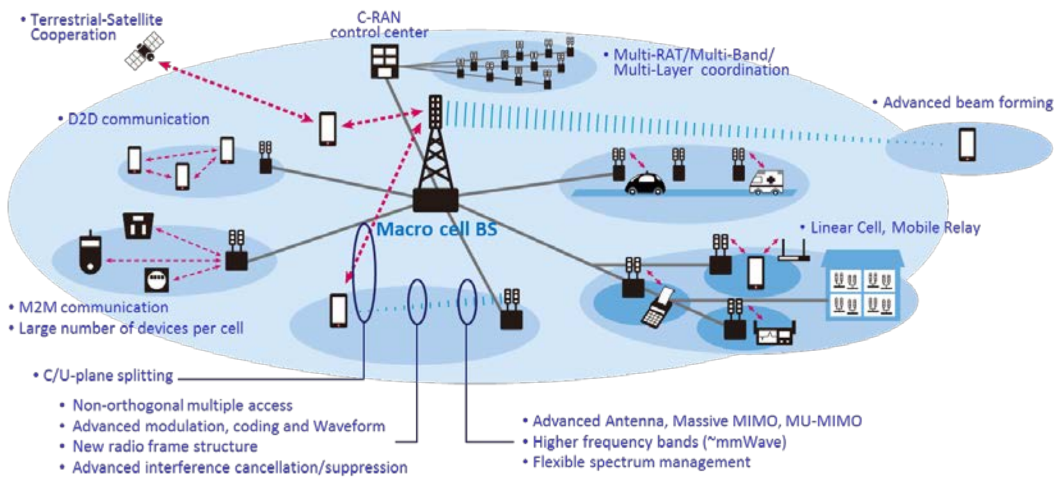


Fig. 11.2-1 Overview of '5G' RAN technologies

11.3 RAN related technical works update

11.3.1 General

Initial technical investigations were made in [1]. Since then, enormous progress has been made to consolidate '5G' RAN systems. The following sub sections describe some of the related technical achievements.

Note: The following sub sections do not intend to give an exhaustive list of technologies that will be used in a '5G' RAN system and their contents would further be reviewed considering the future study outcomes.

11.3.2 Information of technical works related to modulation or coding scheme

(1). OFDM-SSB-QAM [2][3][4][5][6]

This method belongs to orthogonal multiple modulation/ demodulation technologies, which is based on the analytic frequency form using the Hilbert transform. While the current OFDM uses DSB (double side band) carriers, this method uses a SSB (single side band) which separates one DSB into two of SSB carriers, so that the spectral efficiency is twice that of LTE/OFDM.

The schematic diagram below shows the principle of the technology (spectral structure) comparing OFDM and OFDM-SSB-QAM:

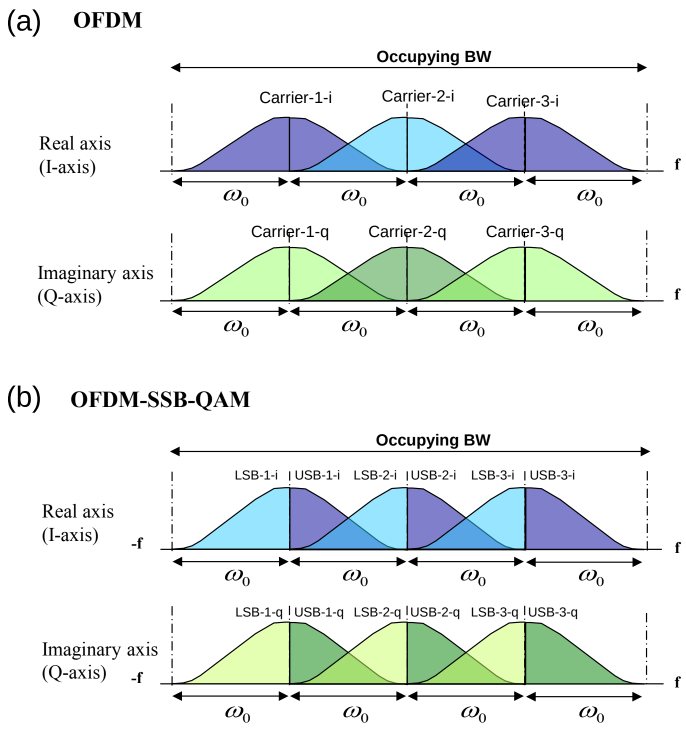


Fig. 11.3-1 Spectral structure of OFDM and OFDM-SSB-QAM

Architecture of the modulation & demodulation parts per partial block of one element (for four SSB carriers) are shown below. The inventive step is that local signals for quadrature demodulation are formed as SSB elements.

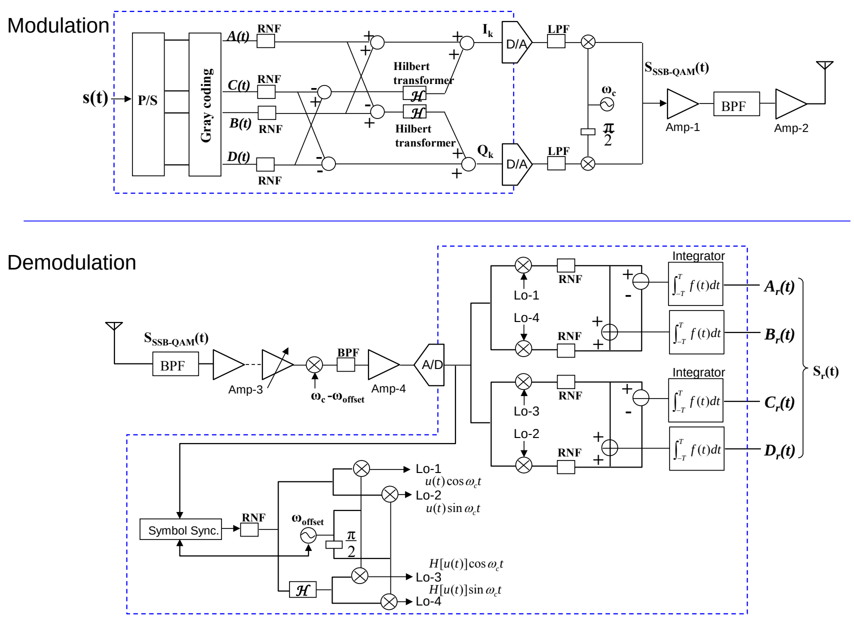


Fig. 11.3-2 Architecture of OFDM-SSB-QAM modulator and demodulator

SSB forming is carried out using a forming data table. Generating multicarrier and integration of demodulation are carried out using FFTs. These elements are very common in LTE/OFDM systems.

- The technology would be useful when applied to : eMBB
- Expected performance/features when applied:

This method not only provides double the spectral efficiency of LTE, but also takes over the whole access function built by LTE. Furthermore, this method is signal-processed in the baseband part closely, so that the spatial multiplication technologies; MIMO or NOMA can be adopted easily.

- Preconditions when applied:

SSB is said to be weak against frequency fluctuations like the Doppler shift effect, but its tolerance is the same when compared to OFDM. This solution method has the frequency synchronization as OFDM, as well.

Because both OFDM and this method are multicarrier systems, it is suitable to adopt this transformation into SC-FDMA when using millimeter wave bands.

(2). Time and frequency localized single carrier technology [7][8]

Insertion of zeros or a static sequence before DFT operation in DFT-s-OFDM can reduce out of band emission compared with the conventional DFT-s-OFDM. Fig. 11.3-3 shows a comparison between DFT-s-OFDM and DFT-s-OFDM with zero or static sequence. Fig. 11.3-4 demonstrates out of band suppression performance of DFT-s-OFDM with zero or static sequence. Maintaining the low peak to average power ratio of SC-FDMA, which is the standardized uplink waveform in LTE, the aforementioned technologies can reduce out of band emission compared to DFT-s-OFDM waveforms. The inserted zeros or static sequence can be used as a cyclic prefix, providing robustness against frequency selectivity in channels.

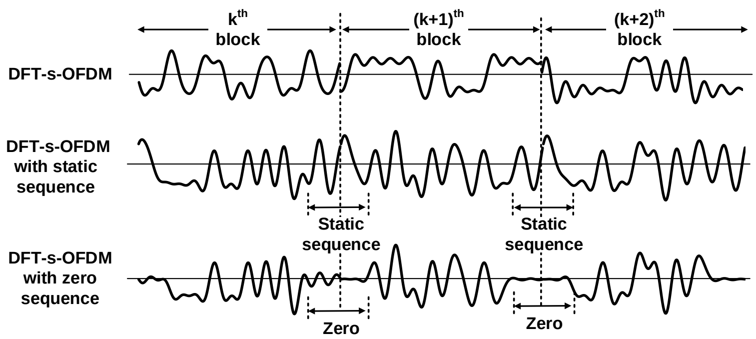


Fig. 11.3-3 DFT-s-OFDM with zero or static sequence insertion

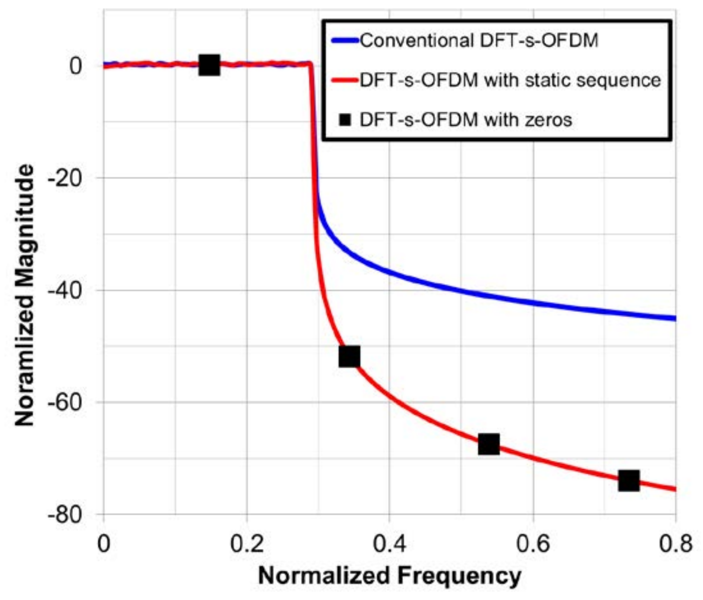


Fig. 11.3-4 Out of band suppression performance

・The technology would be useful when applied to : eMBB, mMTC, URLLC

Waveform technologies with flexible numerology are in demand. In millimeter and centimeter bands, waveforms with low PAPR are in demand to expand coverage without increasing linear region of a power amplifier. Both the number of connected devices and the frequency of asynchronous access are expected to increase due to the emergence of IoT applications. Out-of-band suppression to provide robustness against asynchronous access is also one of the key requirements for a 5G system.

・Expected performance/features when applied:

Low PAPR, low out-of-band emission, coverage expansion, saving cost for amplifiers ・

・Preconditions when applied:

Limited backoff, asynchronous access from UEs, coverage expansion for downlink and uplink.

(3). Filtered-OFDM (f-OFDM) [9][10][11][12]

f-OFDM can achieve desirable frequency localization while enjoying the benefits of CP-OFDM. This is attained by allowing the filter length to exceed the CP length of OFDM and designing the filter appropriately. Figure 3 of Ref. [11] (see Fig. 11.3-5) shows the baseband impulse response of the designed filter with bandwidth equal to 3 RBs. It can be seen that the main energy of the filter is confined within the CP length, and thus, its induced ISI is very limited.

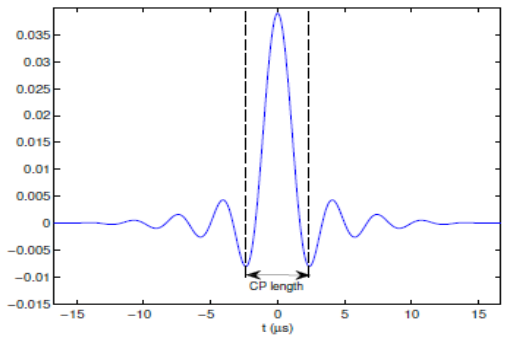


Fig. 11.3-5 Impulse response of the designed filter for f-OFDM with bandwidth equal to 3 RBs

- The technology would be useful when applied to : eMBB, URLLC, mMTC:

The f-OFDM scheme is widely applicable to diverse usage scenarios which are carried out through the conventional OFDM channel, with negligible ISI/ICI degradation impact. In addition, spectrum resources can be flexibly grouped on the f-OFDM resource block domain depending upon the traffic profile and the loading. That can be realized by the optimal radio parameters arrangement, which is suitable for the requirement of the associated application scenario.

- Expected performance/features when applied:

Because of the narrower strict band shaping of f-OFDM spectrum, additional sub-carriers can be allocated in the guard-band between two adjacent carrier bands on top of the conventional OFDM. This is beneficial in order to gain more spectrum efficiency and system capacity. filtered-OFDM supports diverse numerology, multiple access schemes, and frame structures based on the application scenarios and service requirements simultaneously. It allows co-existence of different signal components with different OFDM primitives. For example, three sub-band filters are used to create OFDM subcarrier groupings with three different inter-sub-carrier spacing, the OFDM symbol durations, and the guard times. By enabling multiple parameter configurations, f-OFDM is able to provide more optimum parameter numerology choice for each service group and hence better overall system efficiency.

Furthermore on the f-OFDM domain, the sliced sub-carrier resource blocks can be optimally allocated for the associated application devices in combination with the SCMA. Owing to the non-orthogonal coding scheme of SCMA, the scale of multiplexing access number can be enlarged significantly in low latency radio channel, while allowing grant-Free access connections.

- Preconditions when applied:

The f-OFDM is applicable frequency and deployment scenarios agnostically. Since f-OFDM has OFDM as its core waveform, it enjoys the desirable properties of OFDM while enabling immediate application of all existing OFDM-based designs. For instance, f-OFDM is MIMO-friendly and also its PAPR can be easily reduced using DFT precoding as in DFT-S-OFDM.

Also, "asynchronous" multiple access is possible with the proposed "filtered orthogonal frequency division multiple access (f-OFDMA)" / "filtered discrete-Fourier transform-spread OFDMA (f-DFT-S-OFDMA)", which uses the spectrum shaping filter at each transmitter for side lobe leakage elimination, and a bank of filters at the receiver for inter-user interference rejection.

(4). Polar code [13][14][15][16][17][18][19][20][21][22][23][24][25][26][27][28][29]

Polar code achieves very good channel quality and capacity with a simple encoder and a simple successive cancellation (SC) decoder even in cases where the code block size is larger. Polar codes have engendered significant interest and a lot of research has been done on code design and decoding algorithms. One of the most important decoding algorithms is SC-list decoding which can perform as well as an optimal maximum-likelihood (ML) decoding with an appropriate list size for moderate code block sizes.

・The technology would be useful when applied to : eMBB, URLLC, mMTC:

Polar coding is applicable to the 3 scenarios including eMBB, mMTC and uRLLC for providing better channel quality and reliability. The polar coding is effective and applied to both long bit service and short bit service data packets.

・Expected performance/features when applied:

Performance simulation have shown that Polar codes concatenated with cyclic redundancy codes (CRC) and an adaptive SC-list decoder can outperform turbo/LDPC (Low Density Parity Check) codes for short and moderate code block sizes. Polar code has better performance than the other codes currently used in the 4G LTE system, especially for short code lengths, thus it is considered as a desirable candidate for the FEC (Forward error correction) module in 5G air interface design.

Following effects can be also expected:

- For small packet (e.g. IoT, control channel), Polar Codes have 0.5-2dB gain comparing with Turbo Code used in LTE. (Page 14 of Ref.[14])
- No error floor, suitable for ultra-reliable transmission

- Low energy consumption

・Preconditions when applied:

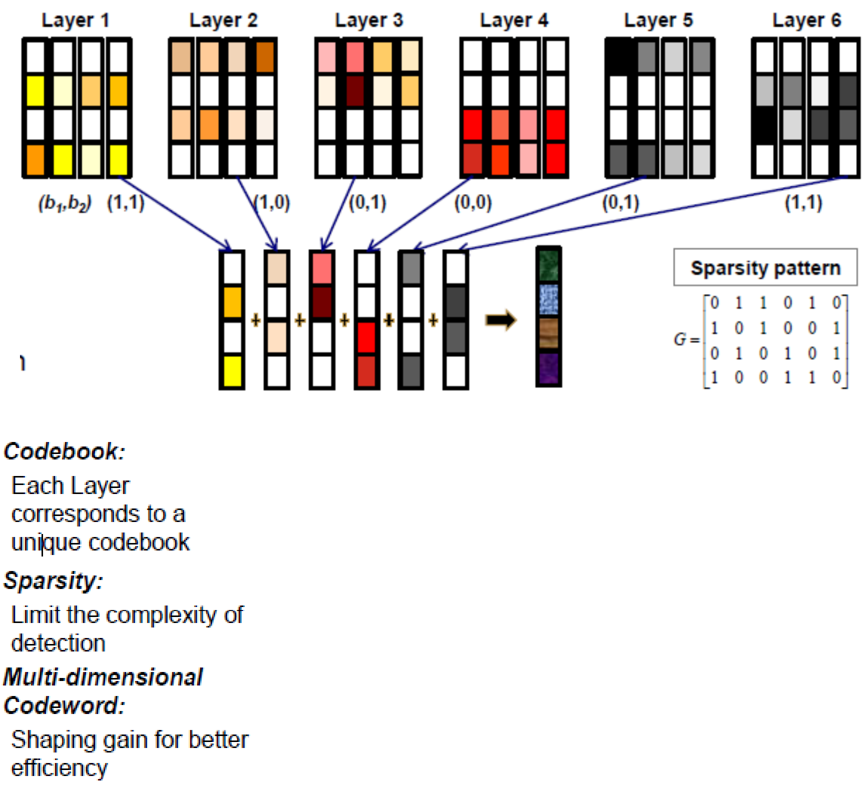
Polar code is an innovative FEC scheme to improve radio channel reliability. It is applicable and more effective to be combined with other radio channel technologies of new waveforms, multiplex access scheme, access protocols, frame structure, etc. in frequency agnostic.

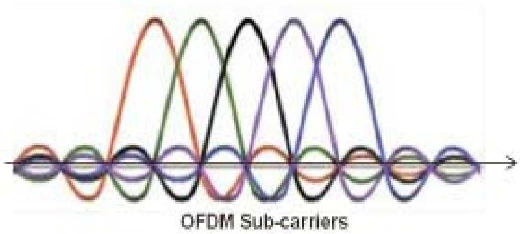
11.3.3 Information of technical works related to multiple access scheme, duplex scheme

(1). Sparse Code Multiple Access (SCMA) [30][31][32][33][34][35][36][37][38]

SCMA is introduced as a new multiple access scheme. In SCMA, different incoming data streams are directly mapped to codewords of different multi-dimensional cookbooks, where each codeword represents a spread transmission layer. Since the multiple SCMA layers are not fully separated in a non-orthogonal multiple access system, a non-linear receiver is required to detect the intended layers of every user. The sparsity of SCMA codewords takes advantage of the low complexity message passing algorithm (MPA) detector which achieves ML-like performance.

Additional technical information is available in [39], [40], [41], [42], [43] and [44].





Overlaid codewords with code domain resource sharing
Fig. 11.3-6 SCMA codebook mapping, encoding, and multiplexing

・The technology would be useful when applied to : eMBB, mMTC, URLLC

Massive connections with user devices become available via SCMA introduction. Long and short burst data packets on the devices are carried smoothly. It is also beneficial to achieve higher data throughput, compared with conventional OFDMA, under the same level of channel resource utilization with a smaller packet drop rate in small latency processing. (Ref.[36][37][38])

・Expected performance/features when applied:

Following improvements are expected compared with OFDMA:

- Multiplexing gain for massive connections.
- Grant-free multiple access that eliminates the dynamic request and grant signaling overhead, which is an attractive solution for small packets transmission in low latency connection.
- Robust with lower packet drop late, better BLER in link budget, higher throughput in loaded conditions.
- Some adaptive parameters can compromise among spectral efficiency, coverage, detection complexity, connectivity, and link budget, to adapt to different application scenarios.

・Preconditions when applied:

The SCMA scheme is theoretically applicable in frequency and deployment scenarios agnostically. User multiplexing can be realized without the need for full knowledge information of users' instantaneous channels. The spectrum efficiency is further enhanced if SCMA is used in conjunction with f-OFDM.

(2). Non-orthogonal multiple access (NOMA) [45][46]

In non-orthogonal multiple access (NOMA) with advanced receiver, multiple users can use the same time and frequency resource. In downlink NOMA, a base station multiplexes signals for users in power domain. In uplink NOMA, which is grant free access or scheduled access, multiple users' signals are spatially multiplexed at the base station.

・The technology would be useful when applied to : mMTC:

NOMA can increase the number of users who simultaneously transmit or receive data at the same resource.

・Expected performance/features when applied:

This technique can improve spectral efficiency since more users can transmit or receive data at the same resource compared to orthogonal multiple access, e.g. OFDMA or SC-FDMA. The number of users, which transmit or receive data simultaneously, also increases. As a result, a base station with NOMA can accommodate more users than orthogonal multiple access.

・Preconditions when applied

NOMA is suitable for the environment of massive users in both cases of grant free access and scheduled access. Grant free access causes NOMA interference which occurs statistically depending on the number of users and traffic condition and so on. However, the interference can be suppressed or canceled by advanced receiver (e.g., iterative canceller). In scheduled access, a base station can adequately select non-orthogonally multiplexed users based on their channel conditions if the base station accommodates massive users. NOMA may not be limited by the particular frequency band, but may be suitable for below 6GHz.

(3). Space Division Full Duplex [47]

Full duplex or STR (Simultaneous transmission and reception) is extremely challenging since very large TX/RX isolation is required. Space division full duplex utilizes spatially separated small transmission points (STPs) alongside with macro transmission points (MTPs). While the MTP serves DL to one or some terminals, the STP serves UL to other terminals, or vice versa simultaneously. MTP and/or STP may employ adaptive beamforming and successive interference cancellation (SIC) in order to reduce interference to acceptable level for receiving operation. Smart algorithms have to be developed since the selection of combination of STPs and terminals being served will have impact on the system performance.

・The technology would be useful when applied to : eMBB

・Expected performance/features when applied:

Following improvements are expected compared with OFDMA:

Ideally, cell capacity will increase by the factor of 2 compared to conventional duplex scheme.

・Preconditions when applied:

C-RAN (Centralized Radio Access Network) scenario in order to coordinate STPs.

11.3.4 Information of technical works related to MIMO or multiple antenna technologies

(1). Nonlinear Multi-User MIMO [48][49]

The linear precoding (LP) scheme is a general method for MU-MIMO. However, since most of the spatial resources at the BS are consumed to direct nulls, we cannot expect extra TX diversity.

On the other hand, nonlinear precoding (NLP) is a strategy realizing an inter user interference (IUI)-free DL transmission by canceling IUI observed by users in advance.

In NLP, user hierarchization is mandatory for practical IUI-precancellation (PC). Block triangulation (BT) is a hierarchization scheme which creates the system channel matrix as shown below:

User \ Stream	1	2	3	4	5
1	H_1B_1	O	O	O	O
2	H_2B_1	H_2B_2	O	O	O
3	H_3B_1	H_3B_2	H_3B_3	O	O
4	H_4B_1	H_4B_2	H_4B_3	H_4B_4	O
5	H_5B_1	H_5B_2	H_5B_3	H_5B_4	H_5B_5

Fig. 11.3-7 System channel matrix for nonlinear MU-MIMO

where Hi is channel matrix and Bi is the precoding vector for user i, enabling the IUI to be successively canceled.

The concept of NLP is illustrated in the following figure.

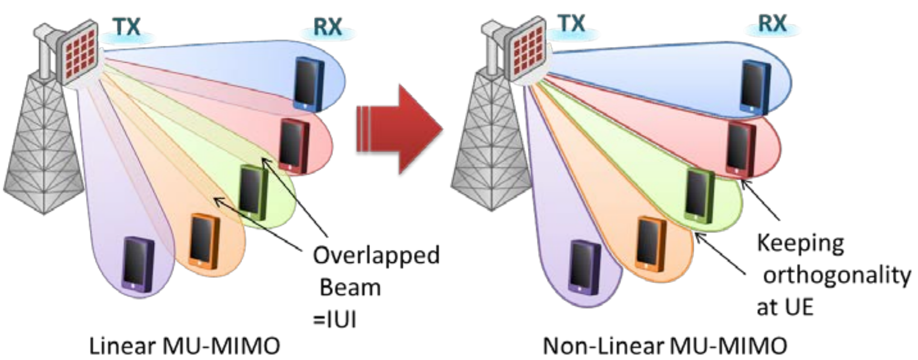


Fig. 11.3-8 Concept of Nonlinear precoding

・The technology would be useful when applied to : eMBB, URLLC

・Expected performance/features when applied:

Because this technology can decrease the dependency of throughput on user location, the system stability can be enhanced. Especially, the data rate of “near users” can be improved

・Preconditions when applied:

Dense BS deployment and cluster user distribution is assumed. High data rate system using simultaneous many-beams transmission, such as massive-MIMO, will be suitable.

(2). Multi-User (MU) MIMO with non-liner precoding, and massive MIMO [50][51][52]

MU-MIMO is an advanced antenna technology to improve spectrum efficiency by increasing the maximum throughput of multiple users on the cell with the limited resources of spectrum. In an experimental trial of MU-MIMO with downlink massive MIMO, a remarkable gain was verified by means of the RF channel calibration, dynamic UE selection scheduling, and non-liner precoding scheme employed in TDD reciprocity channel.

・The technology would be useful when applied to : eMBB:

In various scenarios of mobile broadband (MBB) services, higher throughput and higher spectrum efficiency are desired for wireless broadband applications such as data transfer utility and video services for multiple user terminals simultaneously in a cell coverage area.

・Expected performance/features when applied:

Following gains can be obtained by the MU-MIMO introduction:

- Spectrum efficiency
- Cell throughout, Cell capacity
- User device throughput

・Preconditions when applied:

Outdoor experimental trial was executed with following radio arrangement (Ref.[50]) and [51]):

- 64 Tx antenna in BS
- 24 UEs, with 2 Rx antenna per UE
- THP based non-linear precoding /EZF based liner precoding
- 2.3GHz band TDD, BW = 20MHz

44 bps/Hz was achieved with non-linear precoding scheme. (Fig.2 of Ref.[51])

In another trial of 73GHz mmWave MU-MIMO live demo, 20Gbps data rate was achieved to individual user. (Ref.[52]) The MU-MIMO is frequency diagnostic.

(3). Subarray Type Massive MIMO [53][54][55]

Massive MIMO is an effective approach for using higher spectrum because it can compensate for the large propagation loss in high frequency bands. However massive-MIMO needs many digital and analog components, resulting in higher costs and large energy consumption.

Hybrid beamforming is superior solution for reducing the complexity. Especially subarray type massive MIMO (see remarks) achieves two features, decreasing the number of digital and analog components, and superior transmit performance.

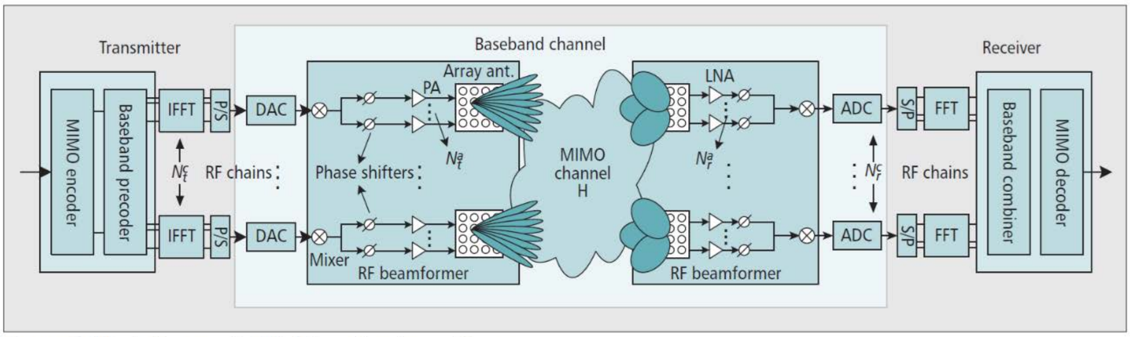


Figure 2. *Block diagram of a hybrid beamforming architecture.*

Fig. 11.3-9 Block diagram of a hybrid beamforming architecture

・The technology would be useful when applied to : eMBB:

In various scenarios of mobile broadband (MBB) services, higher throughput and higher spectrum efficiency are desired for wireless broadband applications such as data transfer utility and video services for multiple user terminals simultaneously in a cell coverage area.

・Expected performance/features when applied:

1. Limiting the number of digital (ADC and DAC) and analog components while also decreasing performance degradation, thus manufacturing cost and power consumption will decrease overall.
2. Realizing high-speed transmission in line of sight environments using multiple beams for one UE. Subarray type configuration takes advantage of the feature of low correlation among the subarrays.

・Preconditions when applied:

Radio frequency is above 6GHz because of limited space for mounting antenna elements, and lower frequency selectivity.

(4). Millimeter-wave beam multiplexing using inter-subarray coding [56]

The size of antennas with massive number of elements becomes reasonable range for millimeter-wave. However, in practical terms, since the same number of base-band processing as that of antenna elements would be required, equipment size and power consumption will increase. Hybrid beamforming is proposed in which leaves signal processing partially in the analog domain. When a sub-array configuration is used in the analog domain, the gain of each sub-array is limited and either sidelobe or obtainable beam width causes interference between the beams. The technique of inter-subarray coding enables generating non-interfering high gain beams which is suitable for massive MU-MIMO operation in millimeter wave.

・The technology would be useful when applied to : eMBB:

・Expected performance/features when applied:

Similar to full MIMO operation by reduced signal processing load.

・Preconditions when applied:

Millimeter wave where the size of antennas with massive number of elements becomes reasonable.

11.3.5 Information of technical works related to RAN deployment or is control schemes

(1). Beam based cell change procedure [57]

For the cell change, a UE measures a signal level/quality of each beam of surrounding base stations and reports the results to a macro cell. A selected base station uses several beams in addition to a selected beam for receiving access signal from the UE.

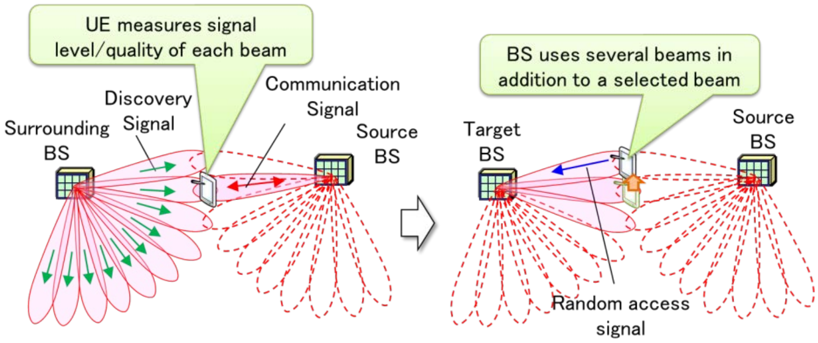


Fig. 11.3-10 Beam based cell change procedure

・The technology would be useful when applied to : eMBB:

・Expected performance/features when applied:

Fast and robust cell change.

・Preconditions when applied:

Dual connectivity.

(2). Linear Cellularization [58][59]

High-mobility scenarios related to land-mobile communications such as railways will generally should have service areas that are covered linearly. Although antennas placed alongside the road or track should direct high-gain beams to cover longer area especially in high-SHF or EHF bands, cellularization per antenna needs frequent handovers and to reuse several radio frequencies.

Linear cellularization, where we can virtually form a long linear cell by linearly-distributed antennas at an identical radio frequency tagged with the same cell ID, is one efficient solution for high-mobility scenarios. Longer cell range yields less handovers and less reuse radio frequencies, resulting in higher spectral efficiency user throughput and group mobility. In addition to fixed-beam antennas, massive antennas, spatial multiplexing and beam tracking can also be utilized over the linear cells, and they can bring further expansion in throughput and capacity to the high-mobility users.

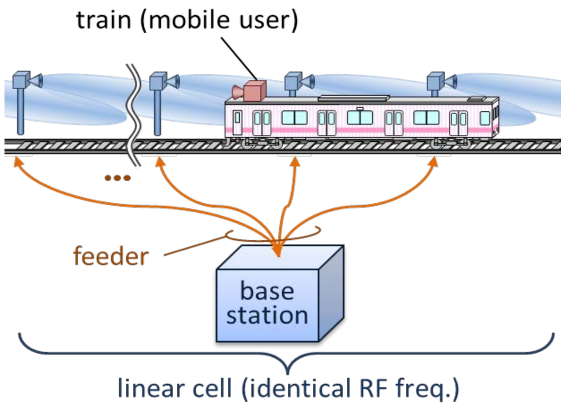


Fig. 11.3-11 Concept of linear cellularization

・The technology would be useful when applied to : eMBB, URLLC

・Expected performance/features when applied:

This technology can reduce handovers across cells and radio frequencies to form the entire service area. Therefore, it brings higher user throughput as well as simplification of system designing including radio frequency assignment.

・Preconditions when applied:

The technology is premised on linear and long area to be served, such as railway and highway, where antennas are placed alongside the area. High-mobility users such as trains and buses move in the same or opposite direction on the linear area.

(3). Throughput maximizing resource allocation for terminals with different QoS [60]

The channel aware resource allocation technique for terminals with different QoS can be achieved when resources are first allocated to terminals with low QoS constrains (e.g., best effort traffic in eMBB use case) taking into account a throughput/fairness trade-off. This is followed by assigning resources for tighter QoS constraints (e.g., delay-constraint traffic in URLLC use case) in order to satisfy QoS constraints (e.g., delay constraints). Allocating resources to terminals with tight QoS constraints after terminals with low QoS constraints is more efficient than allocating these terminals in priority (state-of-the-art approach) since the impact on the throughput of terminals with low QoS constraints can be assessed.

For example, a delay-constrained traffic scheduler can be converted into an online packet-oriented scheduler. This further allows for combining the resource gain metric with a resource preemption cost. Thus, the delay constraints are satisfied and the best effort throughput can be maximized while keeping an equal fairness before and after preemption.

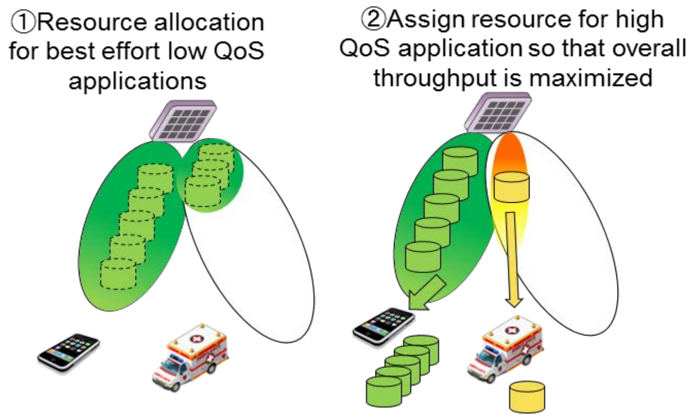


Fig. 11.3-12 Proposed scheduling algorithm

Fig. 11.3-13 A comparison of throughput gain, 70kbps assumed for delayed constrained high QoS applications, number of terminals for delay constrained and best effort applications is assumed to be 30 each

- The technology would be useful when applied to : eMBB, URLLC
- Expected performance/features when applied:

Thanks to fair resource allocation, QoS constraints are satisfied while higher throughput for users with low QoS constraints than in state-of-the-art approaches is achieved for a same fairness level between users with low QoS constrains. This fairness level can be set according to the needs.

- Preconditions when applied:

Terminals with different QoS constrains (e.g., best-effort traffic and delay-constraint traffic) must use same frequency band. A joint resource allocation is needed.

(4). Ultra High-Density Distributed Smart Antenna Systems [61]

In order to increase area traffic capacity, which is one of new KPIs of'5G', cell size has to be reduced. However, due to sever interference, achievable area traffic capacity is limited unless sophisticated, coordinated resource allocation technology is employed. In Ultra High-Density Distributed Smart Antenna Systems, transmission points are densely deployed and terminals will connect to the transmission points that can provide best transmission performance. Radio resource that is used by numbers of transmission points in the area is adaptively and coordinately controlled so that high throughput is achieved for terminals that are simultaneously transmitting data while avoiding interference between them. Beamforming technology may be employed at each transmission point or by coordinating multiple transmission points.

・The technology would be useful when applied to : eMBB

・Expected performance/features when applied:

Three times compared to 4G system with coordinated resource control.

・Preconditions when applied:

RAN (Centralized Radio Access Network) scenario is considered where numbers of transmission points are controlled by C-BBU (Centralized-Base Band Unit).

In order to effectively distribute transmission points, they have to be flexibly configurable and small size. Low SHF band (<6GHz) is considered in experiment and early deployment phase but the technology can be applicable in frequency bands above 6GHz.

(5). System control technologies with wireless LAN in multi-band and multi-access layered cells [62][63][64]

Combining heterogeneous wireless networks that cross licensed and unlicensed spectra is a promising way of supporting surges in mobile traffic. The unlicensed band is mostly used by wireless LAN (WLAN) nodes which employ carrier sense multiple access with collision avoidance (CSMA/CA). Since the number of WLAN devices and their traffic is increasing, the wireless resource of the unlicensed band is expected to become more depleted in 2020s. In such a wireless environment, the throughput could be extremely low and unstable due to the hidden terminal problem and exposed terminal problem. To solve this problem, one new channel access acquisition mechanism for systems control technologies in multi-band and multi-access layered cells is proposed [62][63][64]. This mechanism significantly reduces the impact of the hidden terminal problem in the unlicensed band by using licensed channel access. The information on the user data waiting at the transmitter is notified by using a licensed spectrum channel, so that the receiver under the hidden terminal problem can send a data request frame using an unlicensed spectrum channel and efficiently get data reception opportunities.

・The technology would be useful when applied to : eMBB

Application: High-throughput applications such as HD movies.

Location: High-density area such as stadium, shopping mall.

・Expected performance/features when applied:

Higher capacity in high-density cells.

・Preconditions when applied:

Multi-band and multi-access layered cells.

11.3.6 Information of technical works related to certain use cases or applications

(1). Deployment Strategies for Ultra-Reliable and Low-Latency Communication in Factory Automation [65]

[As conclusions of a study of deployment strategies for ultra-reliable and low-latency communication in factory automation, with a proper physical layer design exploiting diversity gain, it is possible to guarantee ultra-reliable communications with extreme low-latency down to the sub-millisecond. With such a design, full coverage can be provided for a 300 m x 300 m factory floor. For larger factory halls, where more base stations need to be deployed, interference becomes a limiting factor with reuse-1. Simulations have shown that if it is possible to keep the deployments under control, partial frequency reuse or frequency separated system can be more spectral efficient as it requires less bandwidth than a system in which the same frequency channels are fully reused among neighboring BS's.

Capacity evaluations have shown that it is possible to serve nearly 20 K devices with a reasonable antenna configuration and a bandwidth allocation. The system capacity is mainly affected by the diversity and the system bandwidth.

Furthermore, interference management techniques (e.g., ICI coordination) could be used to improve the system availability in terms of coverage and capacity.]

・The technology would be useful when applied to : [mMTC]

・Expected performance/features when applied:

[Capacity evaluations have shown that it is possible to serve nearly 20 K devices with a reasonable antenna configuration and a bandwidth allocation.].

・Preconditions when applied:

[With a proper physical layer design exploiting diversity gain]

(2). Millimeter-wave (60-GHz band) High-Speed Close Proximity Transmission Technology [66][67][68]

The Millimeter-wave High-Speed Close Proximity Transmission Technology uses wireless communications over a 60 GHz unlicensed band to enable instantaneous high-volume content transfer.

- Point-to-Point connection over millimeter-wave (60 GHz) band. Enable Point-to-Point high-speed wireless communication without interference from surrounding traffic.
- Close proximity (less than 10cm) data transfer. Prevents information leakage by close proximity transmission.

・The technology would be useful when applied to : [mMTC]

A user can instantaneously download large-volume files such as movies, music and photos when the user touches their mobile device to the HRCP (high-rate close proximity) access point which is implemented on e.g., automatic ticket gates,

・Expected performance/features when applied:

When the transmission rate is 2 Gbit/s, downloading time for a 30-minutes video file whose size is 50 MBytes will be 220 msec.

-Mitigates the wireless traffic loads in 5G mobile networks, by downloading large-volume files at these HRCP access points.

-Reduces power consumption in mobile device, because wireless communication is not required while playing video, unlike streaming usage.

・Preconditions when applied:

(1) high-rate multi-Gbit/s wireless transmission, (2) device management functions that turns on the wireless module only during the downloading, and (3) cache mechanisms for delivering the content file to HRCP access point at which the download will occur.

- Radio access technologies using unlicensed bands will be employed for (1).
- To realize (2) and (3), new management/ control functions that interoperate with 5G mobile networks are needed.

11.3.7 Information of technical works related to energy saving nature

(1). Millimeter-wave (60-GHz band) High-Speed Close Proximity Transmission Technology [66][67][68]

The Millimeter-wave High-Speed Close Proximity Transmission Technology uses wireless communications over a 60 GHz unlicensed band to enable instantaneous high-volume content transfer.

- Point-to-Point connection over millimeter-wave (60 GHz) band. Enable Point-to-Point high-speed wireless communication without interference from surrounding traffic.
- Close proximity (less than 10cm) data transfer. Prevents information leakage by

close proximity transmission.

・The technology would be useful when applied to : [mMTC]

A user can instantaneously download large-volume files such as movies, music and photos when the user touches the mobile device to the HRCP (high-rate close proximity) access point which is implemented on e.g., automatic ticket gates.

・Expected performance/features when applied:

When the transmission rate is 2 Gbit/s, downloading time for a 30-minutes video file whose size is 50 MBytes will be 220 msec.

-Mitigates the wireless traffic loads in 5G mobile networks, by downloading large-volume files at these HRCP access points.

-Reduces power consumption in mobile device, because wireless communication is not required while playing video, unlike streaming usage.

・Preconditions when applied:

(1) high-rate multi-Gbit/s wireless transmission, (2) device management functions that turns on the wireless module only during the downloading, and (3) cache mechanisms for delivering the content file to HRCP access point at which the download will occur.

- Radio access technologies using unlicensed bands will be employed for (1).
- To realize (2) and (3), new management/ control functions that interoperate with 5G mobile networks are needed.

(2). Energy Performance of 5G-NX Wireless Access Utilizing Massive Beamforming and an Ultra-lean System Design [69][70]

[The energy performance of 5GNX systems, characterized by ultra-lean design and massive beamforming, is estimated for a dense urban (major Asian city) scenario with novel power consumption models where sleep power is defined based on the maximum allowed DTX periods by each system. The results show that 5G-NX systems provide much better energy performance compared to LTE due to the ultra-lean design and the high beamforming gain, enabling longer and more efficient sleep. At expected traffic levels beyond 2020, 5G-NX is shown to decrease the energy consumption by more than 50% while providing around 10 times more capacity.

Furthermore, carrier aggregation was shown to be a promising solution that combines the benefit from higher bandwidth and beamforming capabilities at 15 GHz, and the better propagation conditions at 2.6 GHz. As a result, at expected traffic levels beyond 2020, carrier aggregation with 5G-NX provides superior performance with lower energy consumption despite the comparably energy inefficient LTE layer.

The main focus of future work will be to evaluate the energy saving potential of 5G-NX at country level considering more scenarios, alternative deployments and operating frequencies...]

・The technology would be useful when applied to : [eMBB, mMTC]

・Expected performance/ features when applied:

[Decrease the energy consumption by more than 50% while providing around 10 times more capacity compared with existing LTE.].

・Preconditions when applied:

[To be confirmed]

11.3.8 Information of technical works related to RAN virtualization

(1). RAN Virtualization [71][72][73] [74][75]

Novel heterogeneous networks can be realized by radio access network (RAN) virtualization and softwarization in the 5G mobile network. The RAN virtualization is an effective approach in a fabric of Cloud-RAN structure for example, since the mobile network needs to support flexible capabilities in terms of frequency bands, transmission schemes, antenna configuration, multiplex access attributes, etc.

Given the technical trends above, future RAN may have a capability of intelligent control on radio configurations, front/back-haul transmissions, and radio resources of a number of small cells which are organized virtually from the central controller.

A concept of RAN virtualization is illustrated in Figure 10 of Ref. [71]. In this example, several slices are configured in association with the service profile to achieve the required quality and reliability by radio network arrangement with RATs, bandwidth, antenna configuration, transmission power, latency, mobility, and so on.

・The technology would be useful when applied to : eMBB, URLLC, mMTC:

In the future, more diverse services will come out in various usage scenarios in some working environments with wide ranges of data rate, latency, connection density, data size, mobility, reliability, etc. for the associated service profiles.

Flexible programming on the softwarized virtualization can help handling those service data transmissions under the virtual control function. It will be beneficial to the network operators, service providers and end users.

・Expected performance/ features when applied:

RAN virtualization can support following radio network capabilities:

· RAN controller integrates overall network control, scheduling, and data transport control throughout the user devices, remote radio unit (RRU), radio access schemes, fronthaul, backhaul, and radio resources such as transport bandwidth, RAT attributes, signaling on BBU.

· Upper controller of cloud network can potentially cover the RAN in the end-to-end network coordination.

· Depending on the requirements for service application, the network slices are flexibly arranged and scaled up or down in a configuration set with appropriate network resources, virtual network functionalities (VNFs) in the virtual network topology in dynamic way.

· The RAN has a capability of orchestrating the VNF chain by arranging and scheduling the virtual machines, storage memories, processing units, and so on. In consequence, all the data processing functions and the transport lines become programmable in software.

· Network resources can be flexibly allocated in a scalable manner under the RAN controller. Network resources are pooled, and idle resources can be shared among some network slices.

· As a result of the expected capabilities as above, the network can provide comfortable quality of experience (QoE) for a variety of services in a reasonable cost (CAPEX and OPEX) with higher flexibility and agility.

・Preconditions when applied:

In a case of cloud-RAN model, it generally consists of RAN control platform, BBU pools, backhaul connection to core network, Fronthaul connection to a number of small cell sites for remote radio units (RRU). In a 5G novel network, the RAN is expected to have an intelligent control over functions and transport network in a virtual network structure, and a number of small cell sites with various radio network resources which can be controlled remotely from the central controller.

11.3.9 Other information of technical works related to '5G' RAN

Additional '5G' related technical works have been identified. These are microwave backhaul with multiband [76] and a white paper summarizing '5G' RAN related work as a white paper [77].

References

[1] "Mobile Communications Systems for 2020 and beyond," ARIB 2020 and Beyond Ad Hoc Group White Paper, Oct.2014.

[2] Genichiro Ohta and Takuro Sato, "An Orthogonal Frequency Multiplexed (OFDM) four-layer SSB-QAM Modulation method," IEICE RCS report, vol. 114, no. 490, RCS2014-328, pp. 159-164, March 2015.

[3] Genichiro Ohta and Takuro Sato, " A Study of Four-layered SSB-QAM Modulation Method (Part

2) ,“IEICE RCS report, vol. 114, no. 30, RCS2014-31, pp. 67-72, May 2014.

[4] Genichiro Ohta and Takuro Sato, “A Study of Multi-level modulation for SSB-QPSK method, “IEICE RCS report, vol. 113, no. 456, RCS2013-312, pp. 37-42, March 2014.

[5] Yi Jiang, Zhenyu Zhou, Masahiko Nanri, Genichiro Ohta and Takuro Sato, “Performance Evaluation of Four Orthogonal Single Sideband Elements Modulation Scheme in Multi-Carrier Transmission Systems”, Vehicular Technology Conference (VTC Fall), 2011 IEEE.

[6] G.Ohta, M. Uesugi, T. Sato and H. Tominaga, “Considerations on New Modulation Methods for Mobile-Multimedia Wireless”, IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences Vol.E87-A No.10 pp.2676-2683, Oct. 2004.

[7] G. Berardinelli, et. al. “On the potential of zero-tail DFT-spread-OFDM in 5G networks”, 2014 IEEE 80th Vehicular Technology Conference (VTC Fall), 2014.

[8] F. Hasegawa, et. al., "Static sequence assisted out-of-band power suppression for DFT-s-OFDM," in Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on , vol., no., pp.61-66, Aug. 30 2015-Sept. 2 2015.

[9] Yuya Saito, et al., “Field Experimental Trial of Filtered-OFDM for Uplink Transmission”, IEICE General Conf., Mar. 2016

[10] Huawei Whitepaper, “5G: New Air Interface and Radio Access Virtualization”, April 2015

[11] Javad Abdoli, M. Jia and J. Ma, “Filtered OFDM: A New Waveform for Future Wireless Systems”, IEEE SPAWC2015, pp. 66 - 70, 2015.

[12] Peiying Zhu, “5G Enabling Technologies – An Unified Adaptive Software Defined Air Interface”, Sept 3rd, 2014, PIMRC.

[13] Huawei Whitepaper, “5G: New Air Interface and Radio Access Virtualization”, April 2015.

[14] Huawei Technologies, “Vision on 5G Radio Access Technologies”, RWS-150006, page.14, 3GPP RAN workshop on 5G, Sept.17-18, 2015.

[15] Bin Li, Hui Shen, David Tse, "An Adaptive Successive Cancellation List Decoder for Polar Codes with Cyclic Redundancy Check", IEEE Communications Letters, Vol.16, No.12, December 2012.

[16] S. B. Korada, E. Sasoglu, and R. L. Urbanke, “Polar codes: characterization of exponent, bounds, and constructions,” IEEE Trans. Inf. Theory, vol. 56, pp. 6253–6264, 2010.

[17] E. Arikan and E. Telatar, “On the rate of channel polarization,” in Proc. 2009 IEEE Int’l Symp. Inform. Theory, pp. 1493–1495.

[18] I. Tal and A. Vardy, “How to construct polar codes,” in Proc. 2010 IEEE Inform. Theory Workshop.

[19] I. Tal and A. Vardy, “List Decoding of Polar Codes,” available as online as arXiv: 1206.0050v1.

[20] E. Arıkan, “Polar Coding: Status and Prospects”, Plenary Talk of IEEE International Symposium on Inform. Theory, Saint Petersburg, Russia, 2011.

[21] S. B. Korada, E. Sasoglu, and R. L. Urbanke, “Polar codes: Characterization of exponent, bounds, and constructions,” IEEE Trans. Inform. Theory, vol.56, pp. 6253–6264, 2010.

[22] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in Proc.IEEE Int’l Symp. Inform. Theory, Seoul, South Korea, 2009, pp.1493–1495.

[23] I. Tal and A. Vardy, “How to construct polar codes,” in Proc. IEEE Inform. Theory Workshop, Dublin, Ireland, 2010.

[24] P. Trifonov, “Efficient Design and Decoding of Polar Codes,” IEEE Trans. Commun., vol. 60, no. 11, pp. 3221-3227, Nov. 2012.

[25] B. Li, H. Shen, and D. Tse, “An Adaptive Successive Cancellation List Decoder for Polar Codes with Cyclic Redundancy Check,” IEEE Comm. Letters, vol. 16, pp. 2044–2047, Dec. 2012.

[26] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, “Hardware Architectures for Successive Cancellation Decoding of Polar Codes”, IEEE ICASSP, 2011.

[27] C. Zhang, B. Yuan, and K. K. Parhi, “Reduced-latency SC polar decoder architectures,” ICC 2012, pp. 3471 - 3475.

[28] Bin Li, Hui Shen, and David Tse, “Parallel Decoders of Polar Codes”.

[29] Liang Zhang, Zhaoyang Zhangy, Xianbin Wang, Qilian Yu, Yan Chen, "On the Puncturing Patterns for Punctured Polar Codes", 2014 IEEE International Symposium on Information Theory, pp.125-125.

[30] Huawei Whitepaper, “5G: New Air Interface and Radio Access Virtualization”, April 2015.

[31] Huawei Technologies, “Vision on 5G Radio Access Technologies”, RWS-150006, p.12, 3GPP RAN workshop on 5G, Sept. 17-18, 2015.

[32] Peiying Zhu, “5G Enabling Technologies – An Unified Adaptive Software Defined Air Interface”, p.7-10, Sept

3rd, 2014, PIMRC.

[33] Report ITU-R M.2320-0 (11/2014), “Future technology trends of terrestrial IMT systems”, 5.1.1.2 Non-orthogonal multiple access.

[34] Hosein Nikopour and Hadi Baligh, “Sparse Code Multiple Access”, 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), pp.327-331.

[35] Alireza Bayesteh, “Blind Detection of SCMA for Uplink Grant-Free Multiple-Access”, IEEE pp. 853-857, 2014.

[36] Mahmoud Taherzadeh, et al., "SCMA Codebook Design", IEEE VTC-fall, Vancouver 2014.

[37] Kelvin Au, et al., “Uplink Contention Based SCMA for 5G Radio Access”, Globecom 2014 Workshop - Emerging Technologies for 5G Wireless Cellular Networks, pp.900-905.

[38] Hosein Nikopour, et al., “SCMA for Downlink Multiple Access of 5G Wireless Networks”, IEEE pp. 3840-3844, 2014.

[39] METIS Deliverable D2.2, "Novel radio link concepts and state of the art analysis" Version 1, 31-10-13, p.42-46.

[40] METIS Deliverable D2.3, "Components of a new air interface - building blocks and performance" Version 1, METIS, 30/04/2014, p.41-42, p.90.

[41] METIS Deliverable D2.4, "Proposed solutions for new radio access", Version 1, 28/02/2015, p.152-160.

[42] METIS Deliverable D4.3, "Final Report on Network-Level Solutions", Version 1, 01/03/2015, p.13-102.

[43] METIS Deliverable D6.5, "Report on simulation results and evaluations", Version 1, 01/03/2015, p.5-.

[44] METIS Deliverable D8.4, "METIS final project report", Version 1, 30/04/2015, p.14-19.

[45] Y. Hamaguchi, et al., “Activities on technologies of massive devices and low latency communication toward 5G,” IEICE Technical Report, RCS2015-184, pp. 141-145, Oct. 2015.

[46] NGMN Final Deliverable, “NGMN 5G WHITE PAPER,” Version 1.0, 17/02/2015, p.88.

[47] W.Wang and X. Wang, “Virtual Full Duplex via Joint Selection of Transmission Point and DL/UL Configuration,” VTC spring 2015.

[48] H. Nishimoto, H. Iura, A. Taira, A. Okazaki, and A. Okamura, “Block lower multi-diagonalization for multiuser MIMO downlink,” Proc. WCNC2016 Workshop, pp. 342--347, Qatar, April 2016.

[49] H. Nishimoto, A. Taira, H. Iura, S. Uchida, A. Okazaki, and A. Okamura, “A study on Nonlinear Block Multi-diagonalization Precoding for High SHF Wide-band Massive MIMO in 5G,". IEICE Tech. Report, RCS2016-377, pp. 255--260, March 2016. (in Japanese)

[50] A. Benjebbour, et al., “Experimental Trial of TDD Downlink Massive MIMO”, IEICE General Conf., Mar. 2016.

[51] X. Wang, et al., “Experimental Trial of Downlink Massive MU-MIMO with Non-linear Precoding”, IEICE General Conf., Mar. 2016.

[52] Press release, “Huawei to bring 73GHz mmWave Mu-MIMO live demo to Deutsche Telekom”, 2016-02-18, http://www.huawei.com/en/news/2016/2/73GHzmm-Wave-Mu-MIM-live-demo.

[53] W.Roh et. al.,”Milimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results,” IEEE Commun. Mag., vol.52, no.2, Feb.2014.

[54] A.Taira et. al., ”Performance Evaluation of 44GHz Band Massive MIMO Based on Channel Measurement,” IEEE Globecom Workshops (GC Wkshps), Dec. 2015.

[55] T.Obara et. Al., “Channel Estimation for Super High Bit Rate Massive MIMO Systems Using Joint Processing of Analog Fixed Beamforming and CSI-based Precoding”, IEEE CSCN2015, Oct.2015.

[56] M.Shimizu, Y.Ohashi and M.Yoshida, “Millimeter-wave Beam Multiplexing Method Using a Hybrid Beamforming,” IEICE Technical Report, SRW2015073, Mar. 2016. (in Japanese)

[57] “5G Radio Access Technologies and Our Activity – Massive MIMO and Beam Control”, IEICE General Conference2016, BP-1-5. (In Japanese)

[58] S. Umeda, A. Okazaki, H. Nishimoto, K. Tsukamoto, K. Yamaguchi, and A. Okamura, "Cell structure for high-speed land-mobile communications," Proc. IEEE VTC2015-Fall, Sept. 2015.

[59] H. Nishimoto, A. Okazaki, Y. Kinoshita, K. Tsukamoto, S. Umeda, K. Tsuji, K. Yamaguchi, and A. Okamura, "Millimeter-wave train radio communication system based on linear cell concept," Proc. STECH2015, Nov. 2015.

[60] N. Gresset, H. Bonneville, “Fair preemption for joint delay constrained and best effort traffic scheduling in wireless networks,” Net4cars/Net4trains, 8th International Workshop on Communication Technologies for Vehicles, Sousse, Tunisia, May 2015.

[61] M.Minowa, H.Seki, Y.Okumura, S.Suyama, A.Otaka, S.Kimura, M.Nakatsuwgawa, H.Asano, Y.Ichikawa,

Y.Hirano, Y.Yamao, F.Adachi and M.Nakazawa, “[] Invited Lecture] 5G R&D Activities for High Capacity Technologies with Ultra High-Density Multi-Band and Multi-Access Layered Cells,” IEICE Technical report, RCS2015-250, Dec. 2015. (in Japanese)

[62] Riichi Kudo, B. A. Hirantha Sithira Abeysekara, Yasushi Takatori, Takeo Ichikawa, Masato Mizoguchi, Hiroto Yasuda, Akira Yamada, and Yukihiko Okumura, “Channel Access Acquisition Mechanism Coupled with Cellular Network for Unlicensed Spectrum,” in Proc., IEEE VTC2015Spring, May 2015.

[63] Riichi Kudo, B. A. Hirantha Sithira Abeysekera, Yusuke Asai, Takeo Ichikawa, Yasushi Takatori, Masato Mizoguchi, “User equipment centric downlink access in unlicensed spectrum for heterogeneous mobile network,” IEICE Trans. on commun., vol.E98-B No.10, Oct. 2015.

[64] Riichi Kudo, Yasushi Takatori, B. A. Hirantha Sithira Abeysekera, Yasuhiko Inoue, Atsushi Murase, Akira Yamada, Hiroto Yasuda, and Yukihiko Okumura, “Advanced Wi-Fi data service platform coupled with cellular network for future wireless access,” IEEE Commun. Mag., vol. 52, issue. 11, pp. 46-53, Nov. 2014.

[65] Nadia Brahmi, Osman N. C. Yilmaz, Ke Wang Helmersson, Shehzad A. Ashraf, Johan Torsner, “Deployment Strategies for Ultra-Reliable and Low-Latency Communication in Factory Automation” ,International Workshop on Ultra-Low Latency and Ultra-High Reliability in Wireless Communications, Dec., 2015

[66] -Tadao Nakagawa, Hideki Toshinaga, Toshimitsu Tsubaki, Tomohiro Seki, and Masashi Shimizu, "Millimeter-wave close proximity high-speed data transfer system", IEICE Communications Express, Vol.5, No.4, pp.114-117, 2016.

[67] Andrew Estrada, et al.,”TG3e Technical Guidance Document”, IEEE 802.15 Doc. 15/0109r7, July, 2015.

[68] Tadao Nakagawa, Hideki Toshinaga, Toshimitsu Tsubaki, Tomohiro Seki, Ken Hiraga, Masashi Shimizu, “Research on a Millimeter-Wave Close Proximity High-Speed Data Transfer System”, Thailand-Japan microwave workshop 2015 (TJMW2015), SA2-03, August, 2015.

[69] Sibel Tombaz, Pål Frenger, Fredrik Athley, Eliane Semaan, Claes Tidestav and Anders Furuskär, “Energy Performance of 5G-NX Wireless Access Utilizing Massive Beamforming and an Ultra-lean System Design”, IEEE Global Communications Conference (GLOBECOM’15), Dec., 2015

[70] “5G energy performance”, Ericsson white paper, Apr. 2015.

[71] ITU-T, FG IMT-2020, “FG IMT-2020: Report on Standards Gap Analysis”, TD208-PLEN/13, Nov.16, 2015, p.96–99.

[72] Peter Ashwood-Smith, “5G and the Wireline Network”, FOE-4, April, 2016.

[73] Report ITU-R M.2320-0 (11/2014), “Future technology trends of terrestrial IMT systems”, 5.6.4 Cloud-RAN.

[74] NGMN, “SUGGESTIONS ON POTENTIAL SOLUTIONS TO C-RAN BY NGMN ALLIANCE”, DATE: 03-JANUARY-2013, VERSION 4.0.

[75] China Mobile Research Institute, “C-RAN, The Road Towards Green RAN (Version 3.0)”, White Paper, Version 3.0 (Dec, 2013).

[76] “Microwave backhaul gets a boost with multiband,” Ericsson Technology Review, Vol.93, 2016.

[77] “5G radio access,” Ericsson white paper, Apr. 2016.

12. Network Technologies for 5G

This chapter describes network technologies for 5G. Based on the guiding concept "network softwarization", which elaborates the overall transformation trend including Network Functions Virtualisation (NFV) and Software Defined Networking (SDN), technology focus area is identified as the result of study in the network architecture group of 5GMF. The brief description of the area and the associated technical issues are described in the following sections.

12.1 Technology focus area

Fig. 12.1-1 describes technology focus area of networking technologies for 5G. It is intended to guide the research and development activities to address essential issues. The results of such activities will constitute the basis for designing 5G systems. The technology focus is divided into four areas: network softwarization, network management/orchestration, fronthaul/backhaul and mobile edge computing.

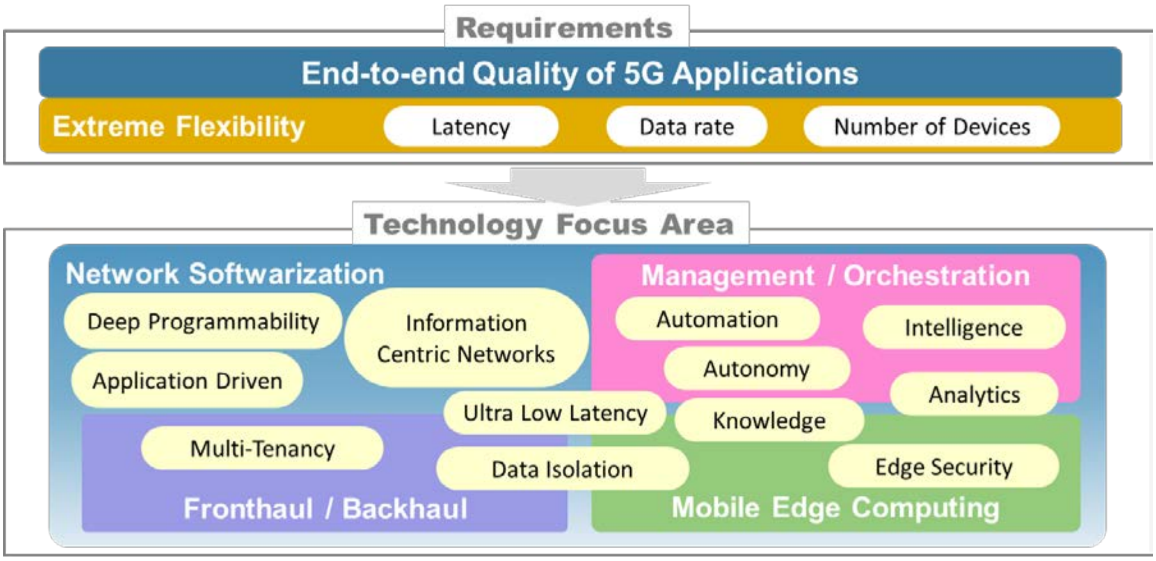


Fig. 12.1-1 Technology Focus Area

<u>Network softwarization</u>

Network softwarization is an overall transformation trend about designing, implementing, deploying, managing and maintaining network equipment and/or network components through software programming. By exploiting the natures of software such as flexibility and rapidity, the industry is working towards for a cost-optimized and value-creating telecommunications infrastructure, which enables prompt delivery of new services with lower equipment and operating expenditure. The industry effort on NFV and SDN are integral part of this transformation. The term "network softwarization" was coined by the academic community, with the aim of harmonizing a number of independent efforts in this industry. It is expected that such harmonization effort will allow operators to utilize consistent and stable foundations for realizing 5G systems.

Network management and orchestration

NFV and SDN technologies constitute the foundation for managing the life cycle of logically isolated network partitions, called "slices". When creating a slice, the management and orchestration functions, NFV-MANO, will provide primary capabilities: select functions requested, launch them on a virtualization platform, and connect them via virtual networks created on physical infrastructure. NFV-MANO is the management and orchestration function that is being defined and specified in the Industry Specification Group (ISG) on Network Functions Virtualisation (NFV) in the European Telecommunications Standards Institute (ETSI). NFV-MANO currently focuses on a single site scenario. However, it is being extended to cover end-to-end service scenarios, in which multiple sites are connected over networks of different administrative domains.

A number of technical challenges are necessary in this area, so as to make the best use of the foundations available in the industry. It includes how to efficiently manage individual functions that constitute end-to-end service context and how to define management models to establish service level agreement when those functions are deployed in different administrative domains. Other challenges include automation and autonomy capabilities which provide easy-to-use workflow procedures for prompt delivery of services and analytics capabilities that will guide optimum placement of functions.

Fronthaul/backhaul

In order to support increasing traffic, mobile operators will need to introduce a number of small cells through the addition of base stations or remote radio heads (RRHs) operated with baseband units (BBUs). Mobile fronthaul (MFH) is a transport network connecting RRHs to BBUs and mobile backhaul (MBH) is a transport network connecting BBUs with core network functions, such as MME, S-GW/P-GW and so forth.

The current MFH is realized by a high speed digital link technology called common public radio interface (CPRI). The wireless signal received and transmitted by RRHs is digitized and coded with CPRI and transferred through optical fibers. For 5G and beyond, the capability of CPRI needs to be advanced so as to match the data transfer requirements, by using techniques such as, high-speed signal processing and precise clock skewing. In addition, new signal processing method and redesign of functional components among RRHs and BBUs will be required.

Considering the economics of building MFH and MBH, it is essential for mobile operators to make the best use of existing physical infrastructure. In Japan, optical fiber networks are available in most of the urban and suburban areas, while other types of networks are utilized in other counties and regions. The international standardization organization is expected to take the leadership role to establish industry-wide standards by incorporating various regional requirements on existing physical infrastructure.

Mobile edge computing

Mobile edge computing (MEC) will play a central role in order to support end-to-end quality of applications and services. In December, 2014, ETSI established an ISG on MEC. Telecom operators, vendors and service providers have been studying techniques and methodologies to distribute functions with the aim of creating open standards. MEC is expected to provide the means to address the support of latency sensitive or high bandwidth applications. Technical challenges include how to decompose functions, where to place the functions to sustain the quality and how to design edge computing platform in an economically viable manner.

12.2 Network softwarization

12.2.1 General definition

Network softwarization is an overall transformation trend about designing, implementing, deploying, managing and maintaining network equipment and/or network components. It exploits the nature of software such as flexibility and rapidity the lifecycle of network functions and services. It will enable re-design of network and service architectures, in order to optimize processes and expenditure, enable self-management and bring added values in an infrastructure.

The term “network softwarization” was first introduced at the academic conference, NetSoft 2015, the first IEEE Conference on Network Softwarization. It encompasses broader ideas in the industry including Network Virtualization, NFV, SDN, MEC, Cloud/IoT technologies and so forth.

12.2.2 Network softwarization in 5G

12.2.2.1 Network softwarization view of 5G systems

The term "network softwarization" is introduced to describe the view of 5G systems with the notion of programmable software defined infrastructure.

The basic capability provide by "network softwarization" is "Slicing" as defined in [ITU-T Y.3011], [ITU-T Y.3012]. Slicing allows logically isolated network partitions (LINP) to exist in an infrastructure. Considering the wide variety of application domains to be supported by 5G systems, it is necessary to extend the concept of slicing to cover a wider range of use cases than those targeted by NFV/SDN technologies, and a number of issues are to be addressed on how to compose and manage slices created on top of the infrastructure.

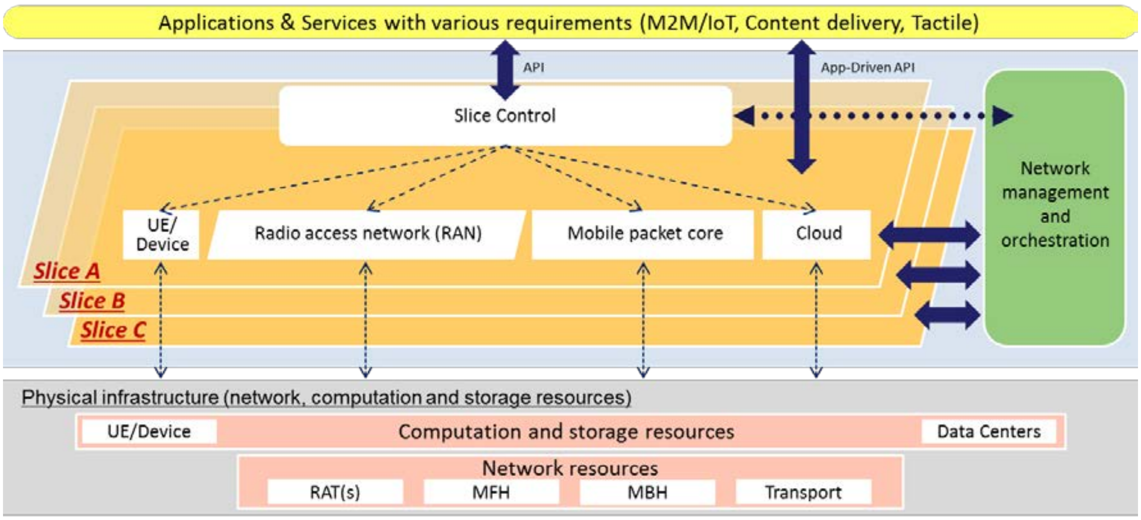


Fig. 12.2-1 Network softwarization view of 5G systems

Fig.12.2-1 illustrates the network softwarization view of 5G systems, which consists of a couple of slices created on a physical infrastructure and a "network management and orchestration" box. A slice is a collection of virtualized or physical network functions connected by links, and it constitutes a networked system. In this figure, the slice A consists of a radio access network (RAN), a mobile packet core, an UE (User Equipment)/device and a cloud, each of which are a collection of virtualized or physical network functions. Note that the entities in Fig.12.2-1 are described symbolically: links are not described for simplicity. The box “network management and orchestration” manages the life cycle of slices: creation, update and deletion. It also manages the physical infrastructure and virtual resources, abstraction of physical ones. The physical infrastructure consists of computation and storage resources that include UEs/devices (e.g. sensors) and data centers, and network resources that include RATs, MFH, MBH and Transport. It should be noted that both computation/storage resources and network resources are distributed and are available for virtualized network functions wherever required.

In addition, virtualized network functions and other functions assigned to a slice are controlled by the “slice control”. It oversees the overall networked system by configuring its entities appropriately. It may include network layer control, and service/application layer control. In some cases, it makes a part of infrastructure being service-aware. It depends on the requirements presented for the networked system, for example, a slice to provide the support of information centric networks (ICN).

Orchestration is defined as the sequencing of management operations. For example, a customer may send a request to the “network management and orchestration” box with their own requirements of an end-to-end service and other attributes related. The request is handled in the box and network programmability functions, if they exist, in the fronthaul/backhaul, core networks, software-defined clouds and mobile edge computing. This involves,

- support for on demand composition of network functions and capabilities and
- enforcement of required capability, capacity, security, elasticity, adaptability and flexibility where and when needed.

<u>Step 1: Creating a slice</u>

Based on a request, the “network management and orchestration” creates virtualized or physical network functions and connects them as appropriate and instantiate all the network functions.

<u>Step 2: Configuring the slice</u>

The slice control takes over the control of all the network functions and network programmability functions if they exists, and configure them as appropriate to start an end-to-end service.

12.2.2.2 Horizontal extension of slicing

In 5G, to satisfy end-to-end quality is an important requirement. Especially as wireless technologies are expected to advance, networking technologies should support as appropriate to sustain end-to-end quality of communications. Therefore, it is natural to consider extending the slicing concept to cover end-to-end context, i.e., from UE to Cloud. Issues in extending slices have to then be addressed, not only the software defined infrastructure in a limited part of a network, but also the entire end-to-end path.

The scope of the current SDN technology primarily focuses on the portions of the network such as within data-centers or transport networks. In 5G, it is necessary to consider end-to-end quality. Therefore, there exists a gap between the current projection of SDN technology development and the requirement for end-to-end quality. It is desired that an infrastructure for 5G will support end-to-end control and management of slices and the composition of multiple slices, especially with consideration of slicing over wireless and wireline parts of end-to-end paths.

Fig. 12.2-2 shows the breakdown of the end-to-end latency in the current mobile network. This figure implies that the network architecture needs to allow latency-aware deployment of network functions and services in order to satisfy end-to-end latency requirements.

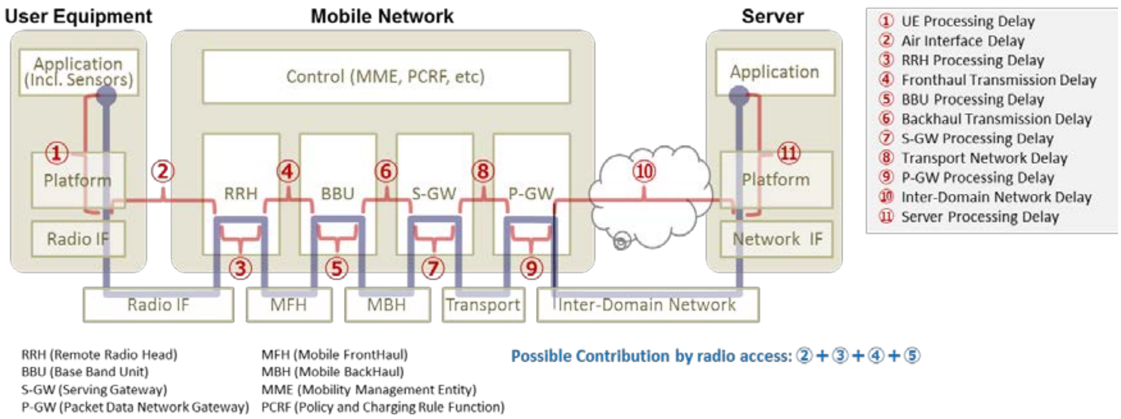


Fig. 12.2-2 Breakdown of end-to-end latency of the current mobile network

3GPP carried out latency studies for 3G, which are documented in specifications TR 25.912, TR25.913, TR36.912 and TR36.913. 3GPP has carried out studies for future network service requirements, which are documented in TR22.891. Operators are building LTE networks to meet the latency budget provided in the 3GPP specification.

Latency studies carried out on many LTE deployed networks demonstrate that the 3GPP specifications provide adequate guidelines. Actual LTE network performances varied, however, due to a variety of variables as well as adjacent ecosystems.

For 5G, an extensive latency study should be carried out in order to provide guidelines for a number of latency-critical services. In order to structure the latency study framework, it is suggested to use breakdown of latency according to Fig. 12.2-2.

12.2.2.3 Vertical extension of slicing (Data plane enhancement)

5G systems may support various communication protocols, even those that have not yet been invented, for services such as Internet of Things (IoT) and content delivery provided by information centric networking (ICN) and content centric networking (CCN). Advanced infrastructure may need the capability of data-plane programmability and associated programming interfaces, which we could call the vertical extension of slicing. The current SDN technology primarily focuses on the programmability of control-plane, and only recently the extension of programmability to data-plane is being discussed in the research community and in ITU-T SG13 without well-defined use cases. For 5G, there are several use cases for driving invention and introduction of new protocols and architectures especially at the edge of networks. For instance, the need for redundancy elimination and low latency access to contents in content distribution drives ICN at mobile backhaul networks. Protocol agnostic forwarding methods such as protocol oblivious forwarding (POF) discuss the extension to SDN addressing forwarding with new protocols. In addition, protocols requiring large cache storage such as ICN needs new enhancement. A few academic research projects such as P4[1] and FLARE[2] discuss the possibility of deeply programmable data-plane that could implement new protocols such as ICN, but there is no standardization activity to cover such new protocols to sufficient extent. Therefore, there exists a gap between the

---

[1] Pat Bosshart, Dan Daly, Martin Izzard, Nick McKeown, Jennifer Rexford, Cole Schlesinger, Dan Talayco, Amin Vahdat, George Varghese, David Walker, “Programming Protocol-Independent Packet Processors”, http://arxiv.org/abs/1312.1719

[2] Nakao, Akihiro. "Software-defined data plane enhancing SDN and NFV." *IEICE Transactions on Communications* 98.1 (2015): 12-19.

current projection of SDN technology development and the requirements for deep data-plane programmability. The infrastructure for 5G is desired to support deeper data-plane programmability for defining new protocols and mechanisms.

12.2.2.4 Considerations for applicability of softwarization

In general, not every component of infrastructure may be defined by software and made programmable, considering the trade-off between programmability and performance. Therefore, it is necessary to clearly define the role of hardware and software according to the potential use cases when softwarizing infrastructure.

SDN is primarily motivated by reduction of operating and capital expenditure and flexible and logically centralized control of network operations. Operators might be motivated to softwarize everything everywhere possible to meet various network management and service objectives. In addition, traffic classification is often per flow basis.

In 5G, some applications have stringent performance requirements such as ultra-low latency and high data rate, while others may require cost-effective solutions. A range of solutions exists from application driven software-based solutions executed on virtualization platform with hypervisor, container or bare metals, to complete hardware-assisted solutions. The former may need performance enhancement enabled by hardware-assisted solutions, while the latter may be facilitated by software-based solutions. The infrastructure for 5G may need to support traffic classification performed not only by flow-basis but also by other metrics and bundles such as per-device and per-application basis so as to apply software/hardware based solutions appropriately for individual use cases. Therefore, there exists a gap between the current projection of SDN technology development and the requirements for applicability of softwarization.

12.2.2.5 End-to-end reference model for scalable operation

Softwarized systems should have sufficient levels of scalability in various aspects of functions, capabilities and components. Firstly, the target range of the number of instances should be considered, e.g. service slices to be configured and to be in operation concurrently. The number of clients and service providers accommodated by each service slice is also an important metric for the practical deployment of the systems. The main constraints for scalability would be the dynamic behavior of each slice and control granularity of physical resources. The communication session established by mobile packet core, however, would be challenging, because it requires a dedicated system for such an extraordinary multiple-state and real-time control, especially for mobility handling. The coordination and isolation between these systems should be clearly defined. Nevertheless, scalability for other types of sessions would also be an issue concerning architectural modelling, including application services, system operation or advanced network services.

In addition to the dimensions and dynamics of the systems, further research is required from the perspective of resiliency and inter-system coordination. For resiliency, some new aspects might be considered other than traditional mean-time-between-failure (MTBF) type faulty conditions. In case of disaster, for example, fault localization, analysis and recovery of softwarized systems could be more complicated. Traditional operation architecture also finds it difficult to cope with misbehaviors caused by human factors because of the indirectness arisen when operating softwarized systems.

The inter-system coordination architecture should be clearly structured and modelled for efficient standardization and for scalability evaluation of softwarized systems. There might be two categories of the coordination, namely horizontal and vertical. The horizontal coordination is for between slice, cloud systems, and UE; in other words, the end-to-end system coordination. Vertical coordination can be distinguished in two ways. One way is for slice and service provider through APIs and the other way is for virtual and physical resource coordination aimed to efficient resource handling through policy and analytics.

In summary, softwarized systems should have sufficient levels of scalability as follows:

- The number of instances/service slices to be supported
- Series of capabilities provided by service slices
- The number of service sessions to be handled concurrently
- Dynamic behavior of instances and slices
- Granularity of resource management, especially for policy control and/or analytics
- Resiliency for various faulty conditions
- Intra-slice coordination among end-to-end resources
- Inter-slice coordination, specifically with various external systems.

Intensive studies are required on both the dimension and the dynamic behavior of softwarized systems, since such systems will have an enormous number of instances and their reactions are not easy to extrapolate from the current physical systems.

Virtual resource handling must be an essential part of the scalable and novel operation architecture, which potentially improves conventional network operations and possibly even up to the level of supporting disaster recovery by using network resiliency and recovery of/with the systems both in a single domain and in multiple domains.

The end-to-end quality management is a key capability required for 5G. However, this capability will be established on the complex interaction among softwarized systems including UEs, cloud systems, applications and networks. An appropriate end-to-end reference model and architecture should be intensively investigated for such complex systems.

12.2.2.6 Coordinated APIs

It may be useful to define APIs so that applications and services can program network functions directly bypassing control and management to optimize the performance, e.g., to achieve ultra-low latency applications.

Discussions on the capabilities of the programmable interface should be objective-based: for example, accommodating a variety of application services easily, enabling higher velocity of service deployment and operation and efficient physical resource utilization. Users or developers who utilize the APIs can be categorized according to their roles. Application service providers will enable value added services over the end-to-end connectivity through the APIs. Advanced network service providers will add some sophisticated functions to communications sessions, such as security and reliability, in order to facilitate faster application service deployment by the aforementioned application service providers. Network management operators will also utilize the APIs for more efficient and agile resource handling.

Information modelling should be the most significant issues for API definitions. It should include virtual resource characteristics, relationships between various resources, operational models, and so on. Levels of abstraction should be carefully investigated, so that the model and APIs should be human-readable and machine/system-implementable at higher performance simultaneously. Since considerations on software development methodologies will have an impact on the development model, the choice of the proper methodology for each capability will be important.

The system control and coordination architecture is another issue that will affect the achievement of scalable and agile APIs. Not only the traditional provisioning/configuration or distributed control of networking systems, automatic and autonomic system control should be the main target. The closed loop control architecture might be the most innovative enhancement from the traditional networking systems even for the APIs.

The robustness and fault tolerance are absolutely necessary for open systems controlled through the APIs by various providers. Isolation over virtual resources should be carefully structured with the APIs' functionalities and constraints.

In summary, discussions on the programmable interface capabilities should embrace:

- Level of abstraction sufficient both for system operations and for customization of the capability provided by the interfaces;
- Modelling for virtual/abstracted resources in a multiple-technology environment;
- Ease of programming for service and operation velocity;
- Technologies for automatic and/or autonomic operations;
- Provisioning of classified functional elements suitable for a range of system developers such as supplication service providers, network service providers, and network management operator.

12.2.3 Information Centric Network (ICN) enabled by network softwarization

12.2.3.1 General Characteristics

a. Overview

One of the aims of 5G is the provision of the emerging network paradigm which fits social requirements. ICN is a promising candidate, with a variety of R&D activities ongoing worldwide. ICN has several merits, including:

1. Server location independent access by contents name
2. Traffic reduction by in-network caching
3. Easy provisioning of in-network data processing
4. Contents security
5. Robustness to network failures by multi path routing

Details of these aspects are described below.

This paradigm, however, adopts a new data forwarding mechanism different from the current Internet. Therefore, it is necessary to have data-plane programmability.

b. Contents/service delivery by its name

The prime difference between ICN and the current internet is how content is accessed. Content is accessed on the Internet through knowing where the content server is located on the network. ICN, on the other hand, content is accessed by submitting a request of the name of the content is on the network. The network will then route the request to the appropriate network node which is storing or caching the named content. The capability to access content by finding "named content" is the basis of ICN, by which the point where named content is stored dynamically moves to the node where the content is most frequently requested and therefore is more efficiently served to the end-user. This can also apply to in-network data processing services. Accessing named content also makes it easier to support consumer mobility by making the ability to serve content more efficient, as well as improving human readability of content requests.

c. Traffic reduction by in-network caching

Another feature of ICN is in-network caching. ICN network nodes are equipped with a content cache server which caches content going through a particular node. The server will then autonomously select which content to cache based on the need of the users accessing the node. Generally, despite different use-cases, content will generally move towards the network edge node where the specific named content is frequently requested. Once the most popular content is cached at the network edge node, subsequent content requests will be served at the particular network edge node, with future communication being terminated at this edge, resulting in a total reduction of the network traffic and lessening the overall server load.

d. In-network data processing

In-network data processing will provide network nodes to do network wide data processing and provide application services on network nodes. The current configuration will need a basic structural change to handle the increase of video traffic and the expansion of IoT as well as to provide shorter response times. Currently data processing is done at a remote data center and the network functions only as a data pipe. In 5G, data processing for application services will be provided with the aim of reducing network congestion as well as shortening response time when necessary. Two typical examples of in-network data processing are ICN, which reduces traffic congestion and response time through the use of a network cache, and edge computing, which provides data processing and service provisioning at the network edge. In-network processing can be considered generally an expanded form of edge computing, where data processing and service provisioning will be provided dynamically any place on a network that is appropriate. Due to the dynamic nature of service and data processing points, ICN's basic mechanism of accessing requested content by name rather than location is especially suitable to provide in-network data processing. Edge computing also is efficient in terms of shortening response times and reducing network congestion when the target data for computing is close to an edge node area. Some IoT use-cases will, however, have target data needed for processing across many edge node areas, therefore the inner node of a network will be more appropriate for processing. Another example is on-path data processing, which data processing is applied in tandem on a transmission path. This is frequently used in big data processing. There are also some use-cases in which the inner network node is better suited to perform data processing, for example when users for a particular service are few in number and yet distributed across several edge nodes.

e. Content security

In some ICN architecture such as CCN and NDN, content security is provided as a basic function. Since security is a key concern in several systems like content delivery and IoT, having a built-in security mechanism is very attractive point of ICN.

f. Robust to network failure by multi-path routing

To enable the content access by name, ICN routing/forwarding is capable of multi-path routing, because the contents once cached in certain node will not be available at the next chance. In ICN multi-path routing, when the response does not come back from the direction the interest is sent out, the node will automatically issue the same request to another direction. This mechanism is very helpful when the part of the network failed down such as the disaster case, and makes the network robust to the failure.

12.2.3.2 Applications of ICN

a. Networking in a disaster area

This service scenario describes ICN as a communication architecture which provides an efficient and resilient data dissemination in a disaster area.

A provider using ICN will be able to directly disseminate emergency data to specific individuals or groups. In this use case, the consumers in advance will express their interest in a specific type of emergency data and information, which ICN will deliver when available. Providers will also be able to directly disseminate emergency data to its users in case of an emergency regardless of any prior requests for this service, as well.

A provider can push emergency data to the cache or storage of ICN nodes, and then the ICN nodes can indirectly deliver the emergency data from their cache or storage to specific individuals and groups as well as to a larger population using the network on a case-by-case basis.

ICN nodes have sufficient storage capacity and so they can hold emergency data for a long time. Both providers and consumers can use the ICN storage system as intermediate devices to share any emergency data with others during a disaster period.

Providers will be able to efficiently and resiliently disseminate emergency data in a disaster area due to the forwarding and caching functions of ICN, in which emergency data is forwarded to an intermediate ICN node where the data is kept (cached) first and then sent to the final destination or to another intermediate ICN node. The caching data can be served for other consumers (efficient data dissemination) even when the original provider is not available due to a temporal network partition (resilient data dissemination).

A consumer can retrieve emergency data even from an intermittently connected network during a disruption or disaster period. ICN follows a receiver (consumer)-driven communication model where receivers can regulate if and when they wish to receive segments of data, and so continuous data retrieval from multiple ICN caching points is possible without regard to end-to-end session.

Operators will benefit from the reductions in system construction costs related to protecting their networks in case of a disaster. Due to name based communication, there is no clear, functional boundary between the network and end devices in ICN. This means ICN nodes can act on behalf of end devices by recognizing and responding to user requests. For example, all ICN nodes will be able to respond to all consumers using its storage capability to share information in a disaster area. This will be a particularly useful feature at times when it is impossible to predict which parts of a network will not be accessible during a disaster.

b. Advanced metering infrastructure (AMI) on a smart grid

This service scenario involves smart meters, communications networks and data management systems that provide two-way communication between utility companies and their customers. Customers will be given assistance through devices such as in-home displays and power management tools. On the communication network, ICN nodes can be installed in order to keep a copy of this data in its cache, which can then be used to present the data in a desired format for the convenience of both the consumer as well as the utility company. The ability to quickly retrieve use pattern data of a particular service is very important in order to efficiently plan and consume services provided by utilities to consumers.

By using the ICN nodes in the network, efficient resources usage and effective load control is possible. Besides an ICN approach for AMI systems in smart grid, they can efficiently control network congestion, support mobility and ensure security.

Since with ICN it is possible to secure data itself, customer will feel more comfortable with the smart grid infrastructure based on the ICN. Furthermore, the operator (utility company) can manage the data more cost-effectively. It also adds value in the scalability issue.

c. Proactive Caching

This service scenario involves people who will access the internet by through their portable device, such as a smartphone or laptop, while passengers of a moving vehicle, such as trains, cars and buses. A certain passenger wants to watch a video-on-demand on her smartphone. If this passenger is on a commuter train, the desired video will be proactively cached in every train station's ICN node according to the scheduler, which decides how much video content should be proactively cached according to video and transportation information. If the user is a vehicle passenger, such as a car, the vehicle mobility information, accessed from the navigation system, can be used to choose an ICN node where content/video will be cached proactively.

The quality of video delivery can be significantly improved by using proactive caching integrated with ICN nodes. Since an ICN node fetches a data object in advance, data objects requested by the mobile user will be immediately available after changing the Point of Attachment. The delay will be minimized due to the reduction of number of hops taken during data transmission. In addition, since the ICN node will maintain a cache of a particular data object, all subsequent requesters of the same data object will reuse the data already cached by the ICN node.

Network operators will benefit, as well. First, bandwidth consumption will decrease due to caching and data-reuse. Second, energy consumption will be reduced since data objects accessed from ICN nodes through Wi-Fi, reducing traffic on 3G/4G networks. The reduction of transmission delays will also allow providers to offer enhanced user experiences for their customers.

12.2.3.3 Migration scenario

5G will co-exist with legacy network equipment and be compatible with existing network technologies. In other words, it should work in a hybrid manner: it may be composed of classical physical network appliances and softwarized appliances during the intermediate phase towards full deployment. Therefore, migration from the starting network to the target one will gradually be accomplished by using a hybrid deployment model, as shown in the following three-steps-migration path:

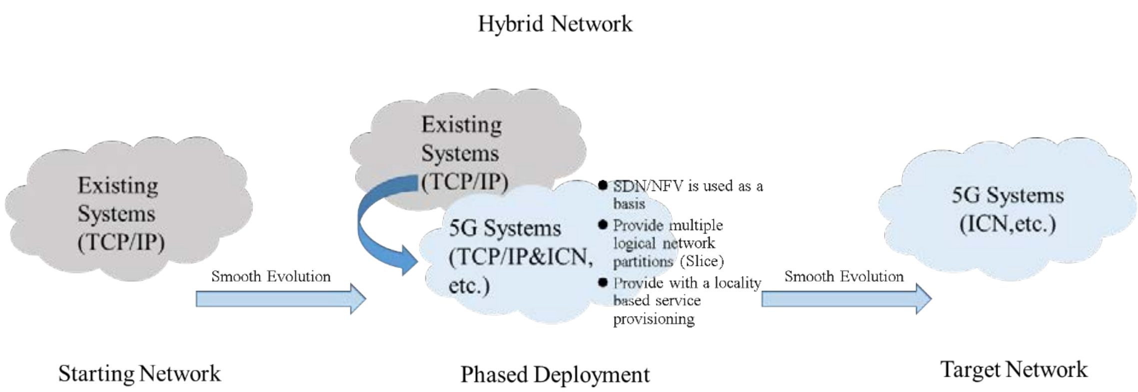


Fig. 12.2-3 Phased Migration

Starting network:

The starting network phase utilizes current and state-of-the-art network technologies (existing technologies), including LTE and IP-based networks.

Phased deployment (intermediate phase):

The benefit of this model of deployment during the migration intermediate phase is all end-to-end resources can still be maintained through conventional communication means in order to communicate with each other. As a result, this mechanism enables migrated end-to-end resources that have been deployed in conjunction with existing devices. It enhances the migration process feasibility by enabling both the gradual deployment of 5G while maintaining current communication models simultaneously during intermediate period.

The requirements for 5G migration are as follows:

- 5G is a foundation of future services and having a mechanism to smoothly evolve to the one which is under discussion in ITU-T SG13/Q15;
- Migration scenarios from the early stage of 5G;
- Locality based service provisioning mechanisms and architecture: mobile edge

computing, a major topic of interest in 5G discussions, and local area computing are examples;

- Possibilities of the in-network data processing/service provisioning capability, where each network node carries out some data processing and service provisioning, a feature especially useful for the efficient management of IoT devices and big data.
- Adoption of emerging network technology.
- Possible technological directions include:
- Application of network softwarization as a core technology of 5G, such as SDN and NFV;
- Adoption of multiple logical networks (slice), each having different architecture that fits to the services provided on the slice. Candidates include: IP , ICN, IoT, and low latency;
- Having a clear API to provide for the development and distribution of a variety of applications and services.

Moreover, 5G will need to provide in-network data processing capabilities, whereby each network node carries out some data processing and service provisioning. This feature will allow 5G to handle IoT devices and big data efficiently.

Target network:

This will also benefit network operators. First, the bandwidth consumption will be low due to caching and data-reuse. Second, energy consumption will be reduced by accessing data objects from the ICN node through Wi-Fi, reducing the 3G/4G traffic. Since the transmission delays will be minimized, network operators will be able to provide an enhanced user experience, as well.

12.3 Management and Orchestration

12.3.1 Overview

12.3.1.1 Management and orchestration technologies

Network management will have a more important role than today in order bring about the full capabilities and services that 5G can provide. In this context, the scope of management and orchestration should cover mechanisms of providing application-driven flexible network as well as managing FCAPS (Fault, Configuration, Accounting, Performance, and Security). In addition, the aforementioned mechanisms will need to be able to provide non-continuous service based upon user needs.

In the following section (i.e., 12.3.2), approaches to these mechanisms are discussed.

12.3.1.2 Challenges and requirements

To discuss challenges and requirements, we need to clarify the current state of the art regarding the future of network management and orchestration. In the area of NFV modeling, a reference architecture of network management and orchestration has been established by ETSI NFV ISG as shown in Fig. 12.3-1. Based on this architecture, relevant interfaces and functional requirements are defined as not only in ETSI but also in other standard organization (SDO) such as TMForum and 3GPP. Technical challenges for the time being are coherent harmonization among those SDOs to enhance inter-operability of interface protocol, data model and so forth. Those challenges should be resolved within the coming few years in order to become fundamental enablers more extensively on network management for 5G systems. Not only SDOs but also emerging open source development efforts will accelerate the resolutions at the implementation level.

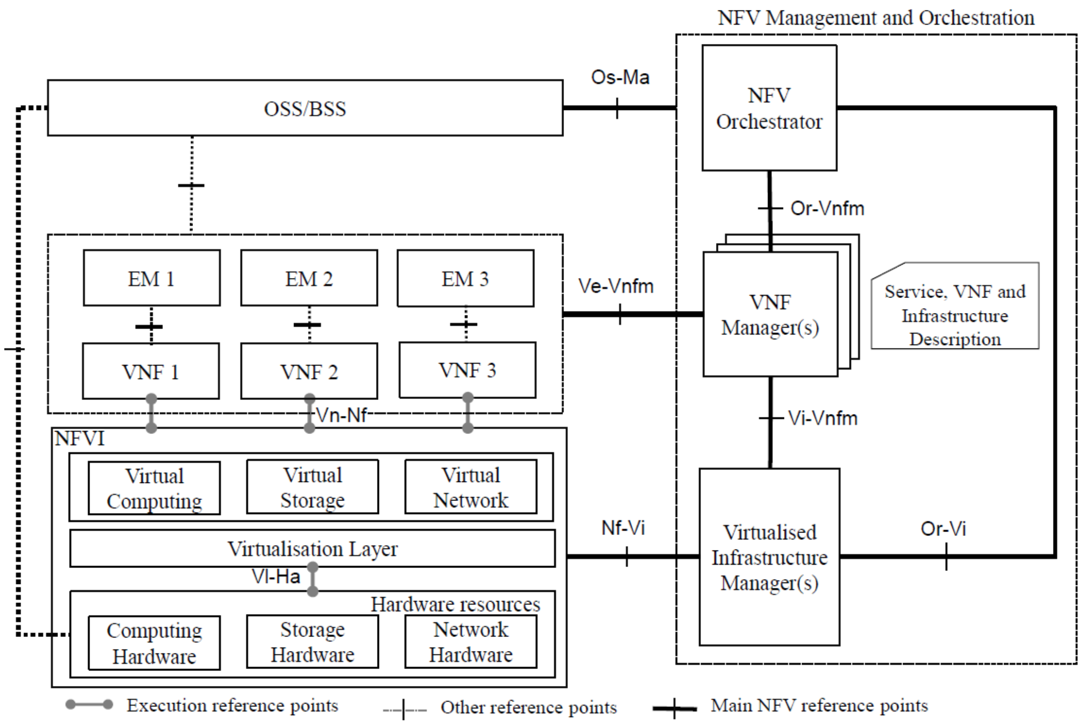


Fig 12.3-1: NFV reference architectural framework [ETSI ISG NFV]

Beyond those challenges, network management and orchestration need to be further application and user driven to tailor slices for each purpose. In addition, as a basic requirement, network management is required to simplify the management of complex network and reduce burden of network operators. As discussed in [FMN-AH1 WP], analyzing the large amount of management data (e.g., statistics, syslogs, events, alarms) for preventing serious event and introducing more distributed way of processing management data will be an additional requirement toward 5G.

12.3.2 Approaches for 5G network management

12.3.2.1 Flexible network for optimal performance and resources

Background and Motivation

Future mobile networks are expected to provide connectivity with a vast variety of applications and services requiring a wide range of levels of quality in terms of respective performance.

For example, some types of unique variations will be required in following use cases:

- Super high data rate services (e.g. future video applications)
- Ultra-low latency services (e.g. tactile and quick response interactive applications)
- Massive number of connections (e.g. M2M/IoT sensors and actuators)
- Super high quality of mobile services (equivalent quality to fixed line services)
- Super reliable data communications (e.g. autonomous driving, life-line tele-communication)

Data traffic varies across a wide range in the use-cases, depending on time (e.g. daytime vs. midnight), location (e.g. indoor vs. outdoor), and the usage environment.

Scenes of dynamic traffic change can be found in situations such as the dynamic hotspot inside a stadium during a sporting event, a concert hall, a station platform, an ongoing festival, and emergency calls in disaster scene and so forth.

The following chart shows the actual traffic volume of broadband Internet data (e.g. DSL, FTTH) measured in Japan by the MIC from 2009 to 2014. While data traffic is increasing every year, the data amount varies in a range of four times or more depending on the time of day and the day of the week as observed in statistics.

Such a behavior of traffic variation is also the case in the mobile application data as illustrated in Fig.12.3-2 below.

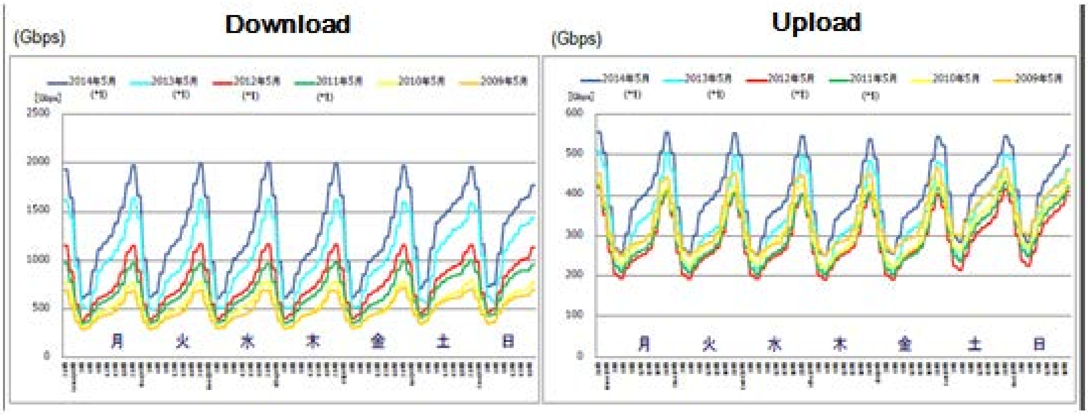
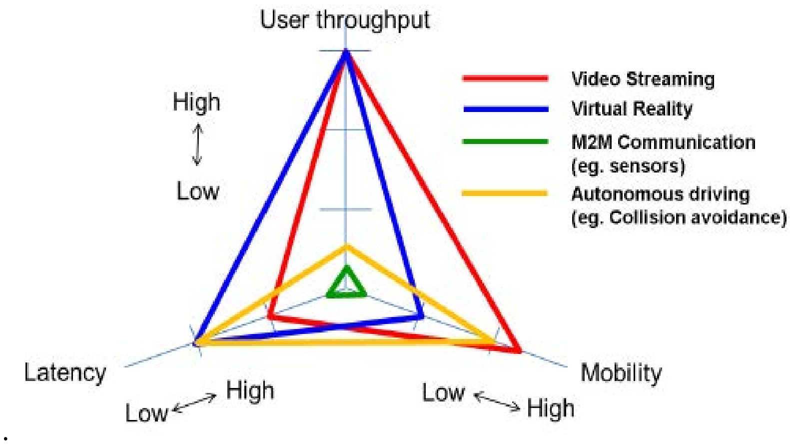
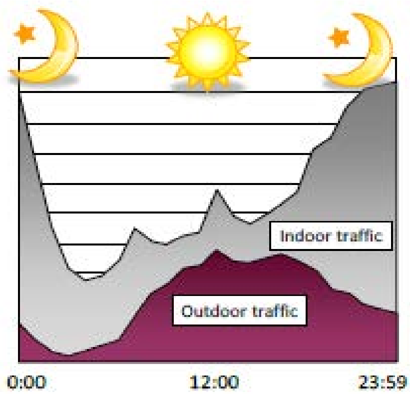


Fig. 12.3-2 Traffic fluctuation of Internet user data in Japan [MIC]



(a)



(b)

Fig. 12.3-3 Diverse capabilities depending on applications, and on the time/location domain [ARIB]

This variation depends on the service application such as video streaming, virtual reality, M2M, and autonomous driving, as shown in Fig. 12.3-3 (a). And it should be noted that the user service does not always require a higher level of performance as presented in Fig. 12.3-3 (b).

Similar views of the application dependency can be found in the Rec. ITU-R M.2083-0, where enhancement of key capabilities is described as the targets for IMT-2020.

Table 12.3-1 Key capabilities and the extreme target
in IMT.VISION, ITU-R Rec.2083-0 (09/2015)

Key Capabilities	**Extreme Target**
Peak data rates	20 Gbps
Latency (air interface)	1 ms
Connection density	10^6 /km^2
Mobility	500 km/h

This table provides the future visions of key capabilities of IMT-2020 from a radio network perspective. These numbers envisage 5G encompassing a wide range of network performance capabilities. In other words, maximum performance capabilities will not always necessary for serving applications to meet the user needs. In fact, the Rec. ITU-R M.2083-0 also provides a picture of the key capabilities variation for different usage scenarios as presented in Fig. 12.3-4 below.

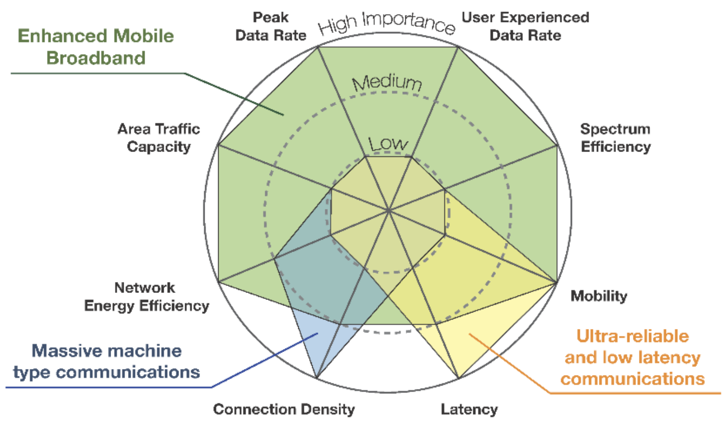


Fig. 12.3-4 The importance of key capabilities in different usage scenarios [ITU-R 2083]

The above chart presents three types of usage scenarios below.

- Enhanced Mobile Broadband
- Ultra-reliable and low latency communications
- Massive machine type communications

Depending on the service type, the required level of capability varies in a scale by several magnitudes to the 10^{th} index power for each capability.

Because of these aspects of future network services, 5G should offer flexible virtual network capabilities to meet specific service demands using the network resources obtained from the infrastructure facilities and physical resources. Key requirements for the virtual network may be identified as:

- guarantee quality and performance in accordance with the service level requirements;
- provide specialised handling of traffic flows in the network segments;
- open and programmable configuration for specialised traffic processing;
- efficient sharing of network resources pooled in the infrastructures and the physical domains;

Given these network capabilities, the network structure has to be scalable enough to be able to cope with flexibility and agility with the changes of traffic loading in order to save operational costs, for example power, link usage, and at hardware facilities.

Consequently, in order to realize the service-oriented optimized network, a virtual network and functional nodes on the associated topology, protocols, and data transport mapped on a specific slice need to be configured flexibly depending on the application type, service profile, operation environment and service quality by means of programmable controllers organized by the management entity. The operation of controllers together with the network resources management are to be activated, coordinated, and organized comprehensively in an intelligent manner by the network orchestrator.

Research and Future Challenges concerning the introduction of Flexible Networks

The following research has been identified as necessary for the introduction of flexible networks to be able to achieve optimal performance and resources utilization;

Study 1: Virtual network structuring with programmable control under the management and orchestration

5G should be designed considering the factors discussed above. In addition, the associated control/management software needs to be developed to organize user data transportation and processing, in the distributed functional nodes on the network slices. This technology should also include developing mechanisms to virtualize network functions and relocate as appropriate for flexible use.

In order to introduce the service-oriented network, a virtual networking, with optimal topology, functional nodes, protocols, and data transport paths need to be configured flexibly in a suitable way to the application type, service profile, device environment and the service demand under programmable controllers coordinated by the management entity. Those operations along with the network resources are to be activated, managed, and organized comprehensively in an intelligent manner by a unified orchestrator.

Smart network concept with the virtual network slices and the associated management and orchestration are illustrated in a sketch below to achieve optimal performance with efficient use of network resources.

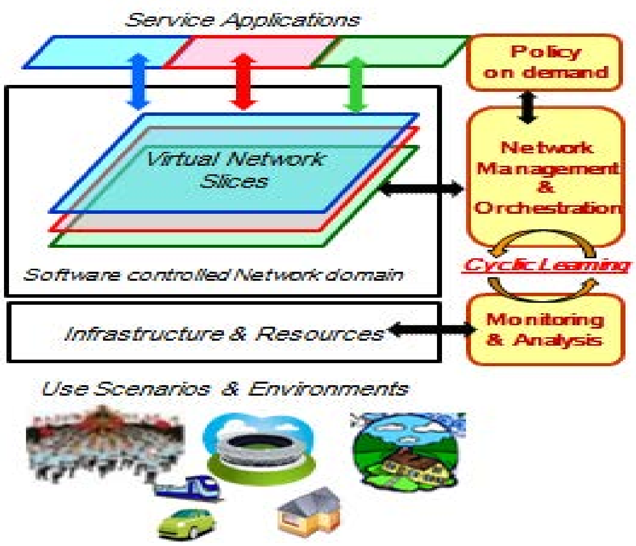


Fig. 12.3-5 Conceptual view of flexible smart network

5G should be designed by considering the factors below.

Challenges:

- Flexible, scalable and dynamic network building
- Capability and suitable QoE provision for diverse service requirements

• Autonomous network organization with intelligence

Approaches:

• Organization and optimization of the virtual network slices and network resources
• Capability of demand based policy execution
• Deep learning with autonomous analysis

For these purposes, intelligent control/management software needs to be developed to organize user data transportation and processing, in the distributed functional nodes on the network slices. This technology should also include developing mechanisms to virtualize network functions and to relocate network resources as appropriate for flexible use.

Study 2: Resource Management for the service profiles using pooled resources

Mobile network resource management in the flexible service-oriented architecture may be driven by having with three aspects below:

Software defined topology: Determination of the logical data plane topology for a given service consisting of the selected physical network nodes. Different services may need different functions as defined by service function chain and the physical nodes where functions need to be instantiated in this logical topology.

Software defined transport and resource allocation: This is the step of determining physical transport paths and the required resources in these paths for the data flows on the data plane, once the logical topology is determined. This would require traffic engineering to establish a reasonable link loading balance and node resourcing (e.g. processing, energy).

Software defined protocol: This is the step of determining the end-to-end (e.g. including RAN, Fronthaul/Backhaul, Core network, for example) data plane transport protocols under a software based management plane and control plane. This includes the establishment of the protocol stack and adjustment of the logical functional units depending on the application type, the expected QoE, and the physical recourse mapping.

Fig. 12.3-6 represents an example of logical structure of flexible mobile network.

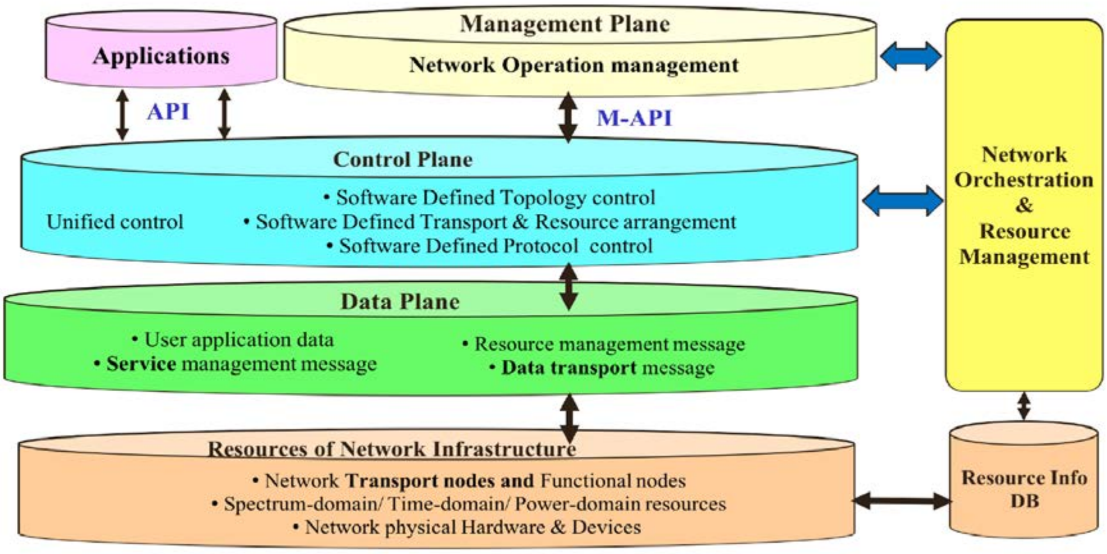


Fig. 12.3-6 Flexible mobile network – a logical structure

With the structure of Fig. 12.3-6, some capabilities should become available for intelligent and elastic network realization as follows:

- Scalable network control for Dynamic Hot-spot with Time-variant/Location-variant data calls and the traffic.
- Service-oriented QoE with optimal set of Throughput, Latency, Connectivity, etc. for diverse applications.
- On demand based network functional nodes application for different type of network services.
- Contingency networking by the flexible routing path against unpredictable network failures.
- Energy saving with the optimal set of resources by the resource management and orchestration.
- CAPEX/OPEX reduction for the network operators, due to efficient utilization of the minimal set of hardware.

<u>Functional view of Flexible Networking</u>

Following note further describes how the key entities (i.e. Management &

Orchestration, and Control management) can work in a flexible mobile network as illustrated in Fig. 12.3-7 below.

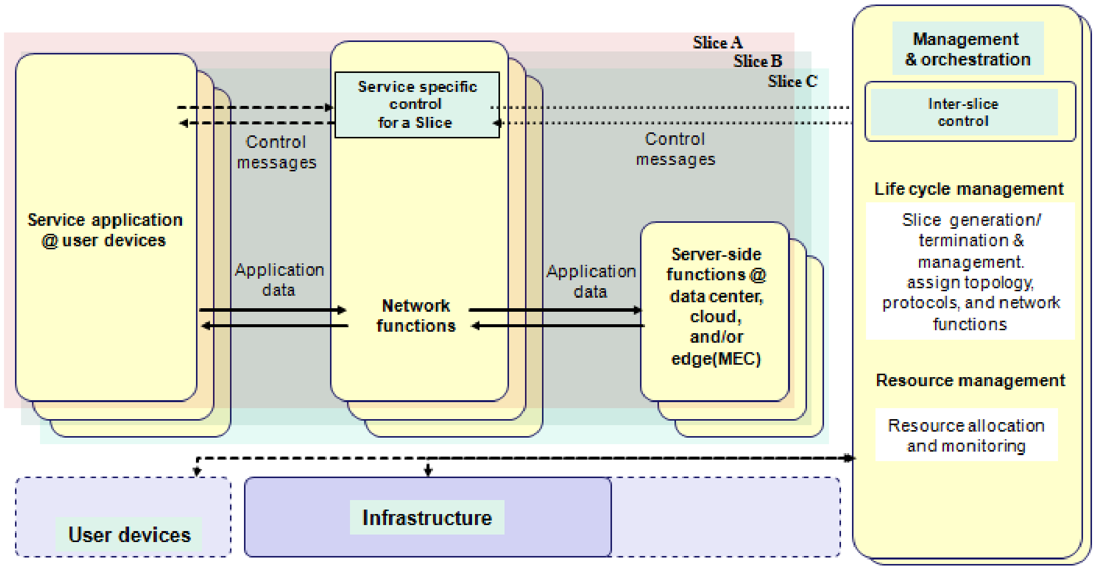


Fig. 12.3-7 Flexible network functional view

<u>Management & Orchestration for Virtual Network Design</u>

The management and orchestration (M&O) block is responsible for life cycle management of network slices. It performs placement and instantiation of network functions. Furthermore, it performs association to the function on user devices and server-side functions.

At the time when a service-specific slice is about to be created, requests may be generated by the service-specific controller indicating what transport network and the functions are needed (e.g. any MTC service, CDN service, public safety) and what type of devices & applications (e.g. video, device data /real-time or not) are used in their locations.

The M&O block is responsible for resource management of infrastructure, which manages the allocation of network functions and virtual networks which are used by the slices. It examines the requests and determine the resources to be allocated, then it instantiates the network functions and virtual networks on the slice on associated physical infrastructure.

The main task of the M&O box is to decide the placement of the VNFs and instantiate them, and to manage life cycle of all the virtual resources and virtual network functions, which are used by these slices.

During the service specific slice creation process, some requests may be generated by the specific service provider indicating what service functions are needed (e.g. any MTC service, CDN service, public safety) and what type of devices & applications (e.g. video, device data /real-time or not) are used in their locations.

Having with those formations, a decision of performance optimization need to be taken based on the network analysis as for where to place these functions in the virtualized infra-structure for providing best performance for the service. As a result of this designing process, the software-defined Topology, Protocol, Resource allocation, and Data processing are configured on each of slice.

Once it is decided, then the management entities in the O&M instantiates the virtual functions on the slice in the associated physical nodes of infra-structure.

In addition, given the result of network analysis for performance optimization, the software-defined protocols, resource allocations, and data processing are configured on each slice.

Scalable Management for Network Resource Control and Service Quality

Management of mobile network resources (e.g. functional node, anchoring node, access network, MFH/MBH elements, transport lines, spectrum-/time-/power-domain resources) for providing a wide range of connectivity services is a task of the control plane which enables the optimal virtual network operation. The control plane should interface with data plane via a control interface to negotiation requirements per the service/application/virtual operations, and the interface with data plane also provides w instructions for resources to be allocated for a particular service.

A Common Control Manager & Service specific Control Managers

Service-specific controller for each application is allocated on each slice. Different application services may have different requirements which request different types of functions and resources (physical and virtual) and topologies to be instantiated and different configurations to be maintained during their life time.

Inter-slice manager coordinates service-specific controllers for slices and manages a common control functions in the Management and Orchestration block. It interfaces with service-specific controls to perform life cycle management and resource management of slices.

While a service specific controller may track authentication of its service application, a physical device may be tracked by the management and orchestration block in some way as a particular device may be connected to multiple slices simultaneously.

Control and data plane functions specific to each application are allocated on each slice as different parts of the network because different services may have different control functions. These functions may be instantiated at different physical nodes and the virtual topologies might be quite different.

There should be an entity which co-ordinates those individual control functions, while managing a common control function, on management plane. That entry may contain a common Connectivity Manager (CM) and a common customer Service Management.

A common CM may also perform certain functions, even if a user is attached to only one slice. Examples include:

• When a user first sends an attach request – it has to first go to Common CM, then forward to specific slice CM;

When some request messages are made where the user is located (e.g. paging), the requests first come to the Common CM, since other entities may not know to which slice the UE belongs.

A service specific CM may track UE's relative location and authentication. A device may be connected to multiple slices simultaneously. These may be tracked by a common CM on the management plane. Subscription management of the devices may be conducted by the common CM, and the session request for devices may send from the common CM to individual CMs.

12.3.2.2 Application-driven network configuration management

Scope

Current mobile network mainly deals with the Internet access from smart phones and feature phones. However, it is presumed that services provided over 5G, including IoT/M2M, will have different requirements for the network. These requirements could include latency, bandwidth, communication frequency, communication topology, and security needs. Therefore, network management on 5G systems will be needed to manage physical and virtual networks accommodating services that will have various requirements.

Challenge

Challenges are

-To improve QoE for each service with minimum network infrastructure

-To provide very low latency services

Approach

- In-network application processing

Each service provided over 5G will have different characteristics. Some real-time services such as Augmented Reality (AR) and ITS require that networks provide low-latency communication between IoT devices and application servers. Because latency between them largely depends on the distance between them, it is efficient to locate application servers near devices. Mobile edge computing (MEC) is one of such solution. Other services deal with a large amount of data raised from sensors causing high traffic to the core network. For such services, a part of applications can be located on MEC and it executes some pre-processing function to reduce the traffic between MEC and application servers through the core network.

- Dynamic application allocation in the network based on service requirements

Each service consists of one or more applications. To improve QoE for each service with a minimum network infrastructure, applications related to the service should be located in 5G systems appropriately. When a new service is installed, dynamic application-allocation function locates each application on appropriate computing resources such as base stations, network nodes, servers based on the requirements from the service.

- Dynamic network resource allocation based on service requirements

Each service will require network functions and resources such as mobility, security and transport. To improve QoE for services with minimum network infrastructure needs, the appropriate network functions and resources should be allocated for those services. When a new service is installed over 5G, dynamic network-resource-allocation function creates a new virtual network, a "slice", for the service and allocates appropriate network functions on the slice based on the requirements from the service. Dynamic network-resource-allocation function may also create a new slice by combining existing slices if they can be reused. Dynamic network-resource-allocation reallocates the network functions and resources on the slice when requirements from services are changed by the environmental reasons for example the increase of users and traffic.

- Interworking between network function allocation and application allocation

The location of applications on the network affects the allocation of network functions. When a user application runs on a base station for instance, 3GPP Evolved Packet Core (EPC) Gateway function should be allocated at the same base station. The EPC gateway function terminates the tunneling protocol against user equipment, so that applications allocated in the base station can handle the data from the user equipment. Therefore, the service management and the network management need to be coordinated each other.

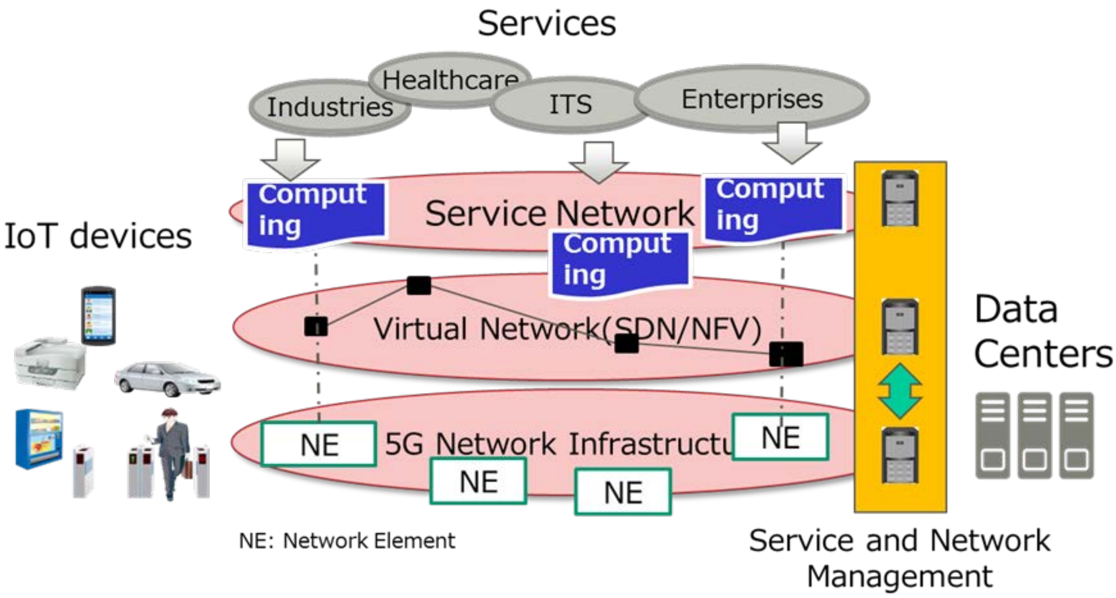


Fig. 12.3-8 The overview of Application-driven network configuration management.

Various services like enterprises, ITS, healthcare run on 5G. Since each service has different requirements, the network management (shown on the right side of the figure) sets up virtual network and the service management allocates applications realizing the service on a virtual network based on its requirements as necessary. The service management and the network management collaborate each other to provide the service appropriately and efficiently.

12.3.2.3 Forward to providing service function in network from data-transmission network

The next generation network needs to accommodate diversified application services and meet various requirements from them. For example, one application service would demand large bandwidth dynamically. Another application service would be sensitive about end-to-end data transmission time. In this sense, customized resources are needed for each application. While resource management becomes more complex as resource usage is customized depending on the needs of each application, reducing communication data produced by MTC/IoT object could conserve a large amount of resources, as well. In order to meet these requirements, the next generation network needs to create various service functions. The following are the challenges for realizing this.

<u>Challenges</u>

- On-demand application-driven configuration

Application services provided by the network have become varied, and conditions of network resources requested by them have also diversified. In addition, new application services are dynamically created and provided using a virtual machine. Therefore, network resources are needed to be configured dynamically by the application services.

- Data processing network for MTC/IoT

In order to accommodate vast amounts of Machine Type Communication (MTC) or Internet of Things (IoT) devices, the network needs to handle a large number of varied data flows. However, MTC/IoT devices can generate a large amount of data and may cause degradation of data transmission quality due to network congestion, etc. Therefore, the amount of data needs to be reduced or transformed into statistics information by data processing inside the network.

- Complex and virtual network management

Conventionally, one application service is provided to many users in the same quality. However, the preferences and environments of users are typically different. Providing a customized service environment to each user using virtually separated network resources is important. The management of multiple virtualized networks is complicated, however, so management of virtual and complex networks is needed.

- End-to-end experience quality management

Lately, quality of service is evaluated based on user experience since the quality felt by people is not the same as the data transmission quality. In addition, end-to-end data transmission is done through multiple networks including wireless and wired networks, and evaluation scheme for heterogeneous networks is also an issue. Therefore, management to guarantee end-to-end experience quality is needed.

<u>Approach</u>

In order to address above challenges, an establishment of a framework to provide customized service functions in the network is important. Fig. x shows an overview of the framework that is composed of three resource-management layers and two resource-management components.

The three resource-management layers consist of a physical infrastructure layer, a virtual networking layer, and a network service layer. The physical infrastructure layer consists of various resources such as a radio access network, fronthaul/backhaul network, backbone network resources and so forth. The virtual networking layer consists of logically integrated resources and their management functions to create multiple slices that are composed of multiple resources that are isolated between them. The network service layer consists of multiple network service slices that are composed of MTC/IoT resources, mobile-edge computing functions, mobility control functions, and so forth.

On the other hand, two resource-management components that are the keys to providing service functions in the network consist of service interfaces and management functions. These services interface and management functions are used to control and manage the three resource management layers. Details of the resource-management components are as follows:

- Service interfaces

Service interfaces that are used by multiple applications can be divided into three categories: are the end-user service interface, the network service interface, and the network management interface. The end-user service interface is used by various customers to confirm that the utilities, transportation, and other services they use are receiving an adequate amount of resources in order to function properly. The network service interface is used by a network service provider to add advanced network functions. For example, highly reliable and low latency network functions could be provided. The network management interface is used by a network operator to control and manage resources according to various management aspects such as energy efficiency, autonomic network configuration, orchestration of resources, and so forth. In order to provide suitable operation environment for each application service, defining above three interfaces is important.

- Slice management for multi-layer/multi-domain virtual resource

In order to provide a customized operation environment for each application or customer, slicing logically integrated resources that are composed of multi-layer and multi-domain resources are indispensable. In addition, the resources for each customer must be isolated from each other. Otherwise service quality for each customer is not guaranteed. Besides, in order to support short-term life-cycle applications, it is crucial that the configuration of resources and functions should be promptly executed. Therefore, dynamic slice network management for multi-layer and multi-domain network resources and functions.

- Programmable/scalable network management

In a conventional network, the management function itself is basically stable and is not enhanced very often. However, data plane functions are dynamically enhanced in a customized slice network. Therefore, network management functions should be programmable to be enhanced and customized dynamically along with the changing of the data plane. On the other hand, in order to accommodate a large number of sliced networks, the management function should be scalable to withstand increasing number of resources and devices.

- Real-time end-to-end QoE monitoring and management

Quality of services should be monitored and managed in real time for end-to-end communication. However, if communication data are transmitted through heterogeneous networks, monitoring end-to-end quality is not easy. Therefore, a scheme to monitor the quality for end-to-end communication is valuable. In addition, schemes to visualize, analyze, and evaluate monitored data are also needed to control application service quality since its quality felt by people is not the same to the quality of a numerical evaluation. Besides, it is needed to create a scheme to manage and orchestrate resources in order to guarantee the quality demanded by an application service.

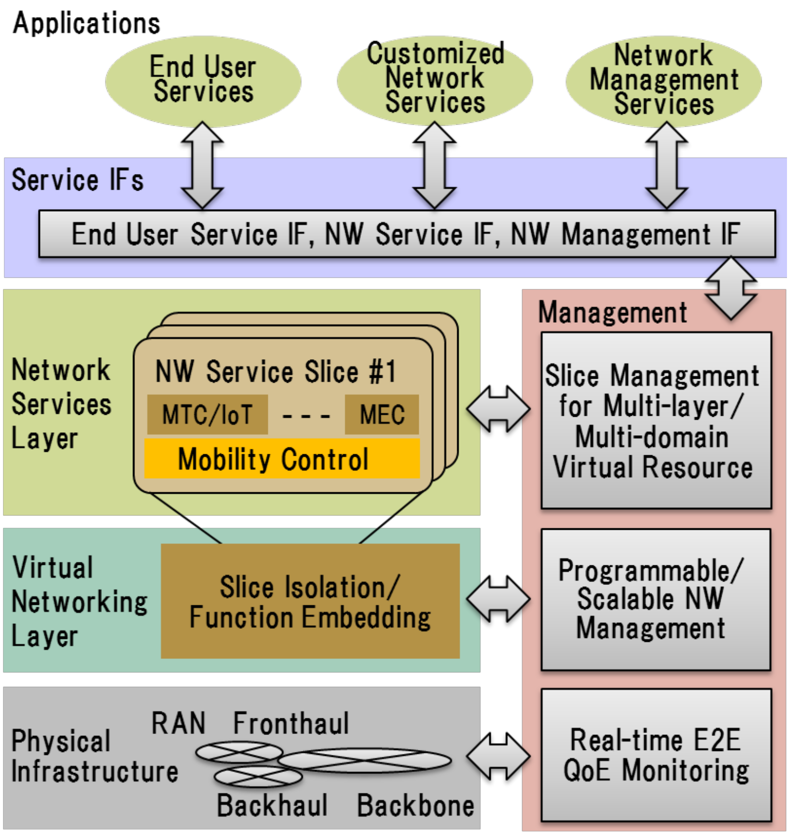


Fig. 12.3-9 Approach to providing service function in the network

12.3.2.4 Service aware device management architecture

<u>Objectives</u>

In 5G era, 50 billion or more devices are expected to be connected to the network in geographically distributed locations. Even if a 1 billion terminal management system provides fast-enough response in accessing the management system and identifying the terminals for a specific service, if the 1-billion-terminal system is extended to support 50 billion devices, the extended management system will not be able to provide same quick level of response for access and the identification. In order to provide the quickest response, we should choose a separate set of devices to be used for each service and build a service-aware device management architecture where the set of devices and corresponding information are managed service-independently. For example, consider an automatic driving support network and a disabled-person's wheel-chair mobility support network. These network systems require very low latency to access and identify devices to avoid impending unsafe conditions and to control the devices safely.

Therefore, to support mission-critical applications and services, one of the aspects of 5G network management is to intelligently enhance handling of new services and applications, especially for 2020 and beyond.

<u>Challenges</u>

Our challenge is to build a network with an intelligent device management architecture where the 50 billion devices are maintained and operated in accordance with service profiles of use and locations. The network needs to provide a diverse set of secure, short response-time required of IoT services, where the response time is defined as the time from the instance of the data is generated from a device to the instance when a target device is actuated by the corresponding control command. Therefore, 5G systems should have the device management capability to identify and operate the IoT devices immediately. Device identification, access, and data transfer should be secured and isolated among different types of services.

<u>Approaches</u>

Addressing the challenges of meeting low-latency responses for device identification, access, and data transfer, we would design the architecture by taking into account the following four features inspired by a notion of service-aware network management and network softwarization:

- Design a device management architecture where sensing devices that are information sources and destination actuator devices are maintained efficiently;

- Design a device management model where mobile operators distribute and maintain separately information about individual devices (e.g., IDs, usages, locations, users);
- Design a service-isolated device management model where a billion of devices are maintained in a distributed system for each service-usage toward low response time to devices and services;
- Design a service model by combining the necessary devices, processing resources, and device management system on the virtualized resources in a slice.

Fig. 12.3-10 shows a complete device management system (green) and service-specific device management system (red and white). The complete system maintains the registration of 50 billion devices in one trillion records within an infrastructure provider's network. The records are stored in a cloud. If the system directly provides something for an individual service (e.g., automatic driving support and an impaired-person's wheel-chair mobility support), the latency requirement may not meet the requirements for that individual service. Thus, for each service, necessary information of devices is retrieved and the necessary required records are formed as a service-independent registry system and located close to the user in order to provide low-latency access to the record as well as allowing for faster updating of the records. The updated information is then synchronized with the information of the complete management system for consistency.

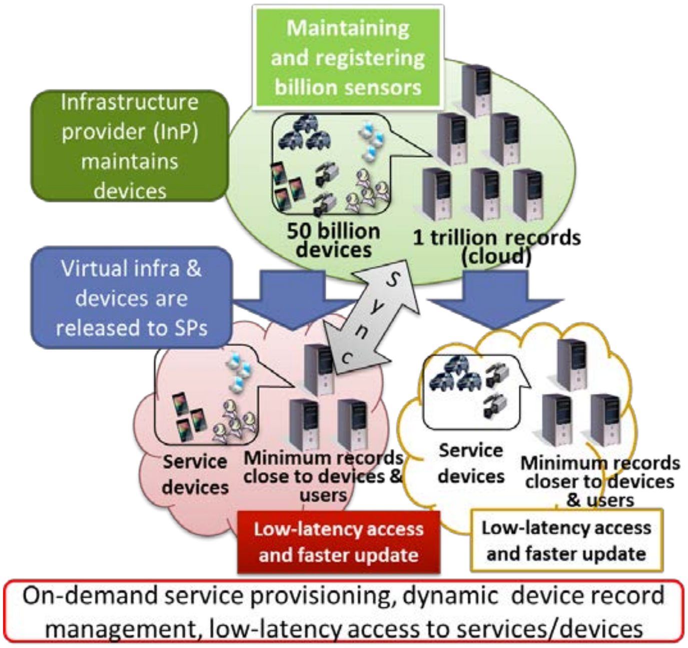


Fig. 12.3-10. A whole device management system (green) and service-specific device management systems (red and white)

12.3.2.5 Personal identification and flexible accounting in 5G

Scope and Challenges

To handle 2020 period applications in 5G, the following intelligence management schemes will be required.

Personal identification as network function

A *Personal*, which is defined in this section as an individual unit on networks such as user, organization and device, cannot be identified on current networks, and a user or an organization uses many IDs (e.g. account and address) to use network services. This situation creates problems as follows (also shown in Fig. 12.3-11):

- Since a user or an organization cannot remember many IDs, a user sets easy IDs and password. It creates opportunities for criminals to steal user's information though spoofing or phishing scams. Many users and organizations today have become victims of such scams, having their bank accounts or credit card information stolen and used by criminals.
- A user cannot identify definitely who sent information by e-mail, web and etc. on networks. This causes a lot of unknown and unrecognized information to cross networks.

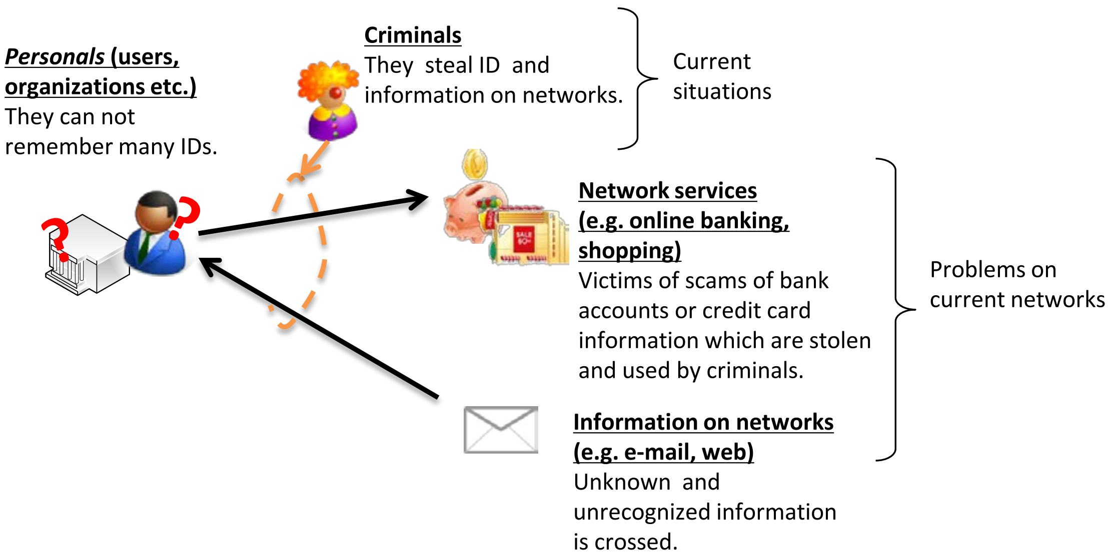


Fig. 12.3-11 Identification problems on current networks

If a *Personal* can be identified on networks as one of the network function, *Personal* who use a network service or send information can be easily and definitely identified as in Fig.12.3-12. If a network gives a unique identification to network users as a part of the network functions, it will allow for a safer and less stressful experience for network users Therefore, *Personal* identification as one network function is an important scheme for 5G.

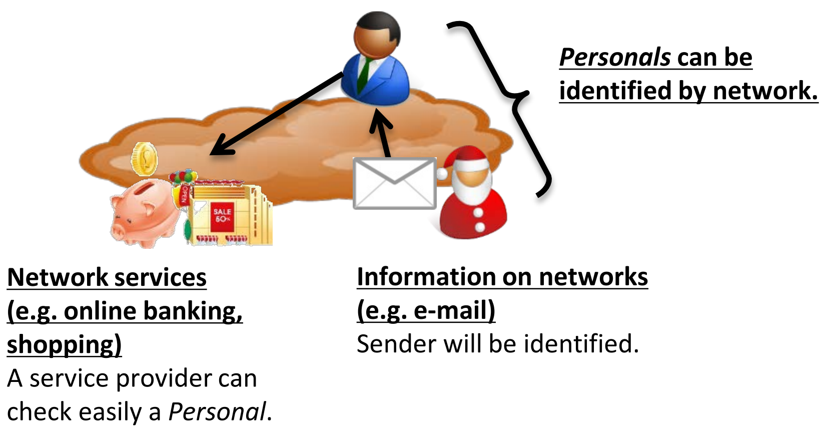


Fig. 12.3-12 Effect of *Personal* identification as a network function

Flexible accounting

Currently mobile network users must contract with an individual operator for accounting and authorization purposes. In this situation, when an operator A network is congestion and an operator B network is empty in a certain place, a user which has the contract only of operator A network cannot use operator B network (Fig. 12.3-13). In addition, since complex networks will be built into 5G from current discussion on 5G, this form of contract will not provide users with the best experience. For example, complex networks, many operators such as mobile virtual network operator (MVNO) offer network resources, many mobile communication systems with many frequency bands (including high frequency bands) and network virtualization.

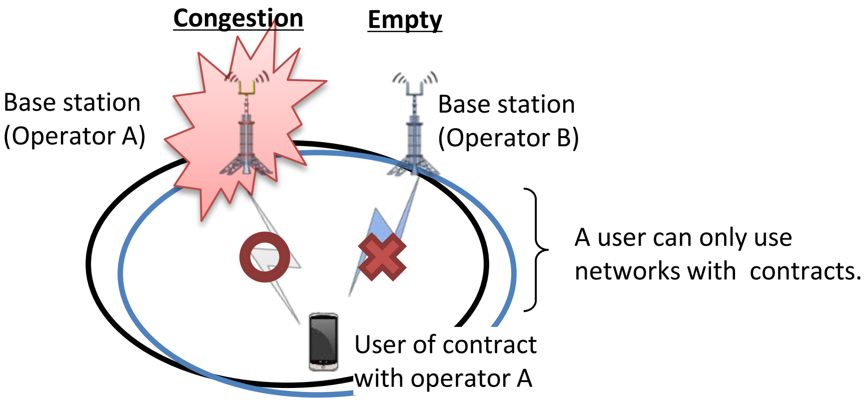


Fig. 12.3-13 A problem with contract based network

Flexible accounting will solve the above situation. Fig. 12.3-14 shows an overview of flexible accounting; a user pays the cost of each network used each time without a contract from each individual operator. When flexible accounting is used, users can access freely available across many networks in a certain place as required by the services they want to use.

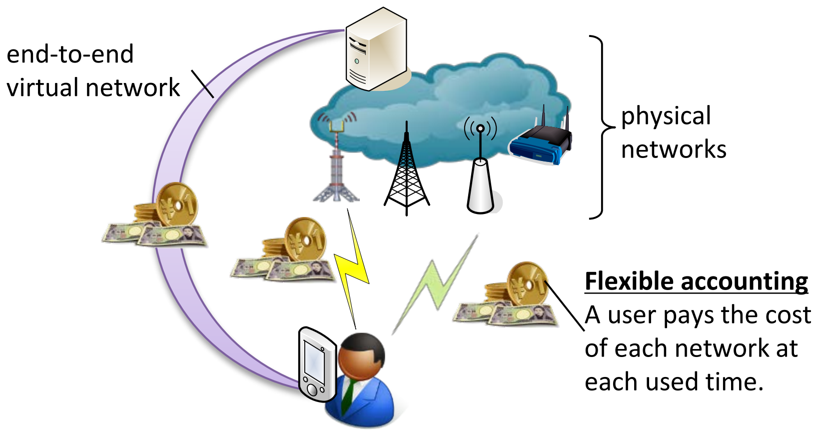


Fig. 12.3-14 Overview of flexible accounting image

Approaches

In order to implement *Personal* identification as a network function and flexible accounting in the future, evolved management schemes will be studied, such as:

- required nodes and arrangement for management on 5G systems
- large scale *Personal* and accounting information management
- secure control message exchange without any changes

These approaches will be considered based on 5G characteristics, such as virtualization and softwarization which are discussed on other sections of this whitepaper and [IMT]. In addition, *Personal* identification as network function is required to realize flexible accounting, because *Personal* identification is needed for accounting.

References:

[ETSI NFV] ETSI GS NFV 002, ETSI, Oct. 2013.

[FMN-AH1 WP] TTC Ad Hoc Group on Future Mobile Networking, White Paper, TTC, Mar. 2015.

[MIC] Ministry of Internal Affairs and Communications: "Data volume of internet traffic in Japan" (2014.10.07)
http://www.soumu.go.jp/menu_news/s-news/01kiban04_02000086.html

[ARIB] ARIB 20B-AH Whitepaper, Section 7.3 and 4.2.

[ITU-R 2083] Recommendation, ITU-R 2083-0 (09/2015)

[IMT] ITU-T FG IMT-2020, "Report on Standards Gap Analysis" Section 7.2, Focus Group on IMT-2020 IMT-O-016, Oct 2015.

12.4 Fronthaul and Backhaul

12.4.1 Overview

12.4.1.1 Terminology Definitions

Fronthaul

The intra-base station transport, in which a part of the base station function is moved to the remote antenna site. (Note that this definition is equivalent to the definition given in MEF 22.1.1 for the current 4G technology.)

Backhaul

The network path connecting the base station site and the network controller or gateway site.

12.4.1.2 Motivation

1) Large capacity

According to [12-4-1], the traffic in mobile communication networks is increasing at an annual rate of 61% and projected to grow 1000 times in the future. Therefore, it is required to discuss as to whether future requirements can be supported by the current network architecture for mobile communications.

Fig.12.4-1 provides a VAN diagram outlining the requirements for future mobile communications. Compared with 4G, the future mobile communication requires larger capacity in extreme areas, faster communication in areas such as rural, urban, dense, etc. and expanded coverage in isolated areas.

Regarding the capacity increase it is assumed that applications like AR (Augmented Reality) will have real-time cloud access with data rate requirements of 100 to 1000Mbps at any given time and around 10Gpbs at peak.

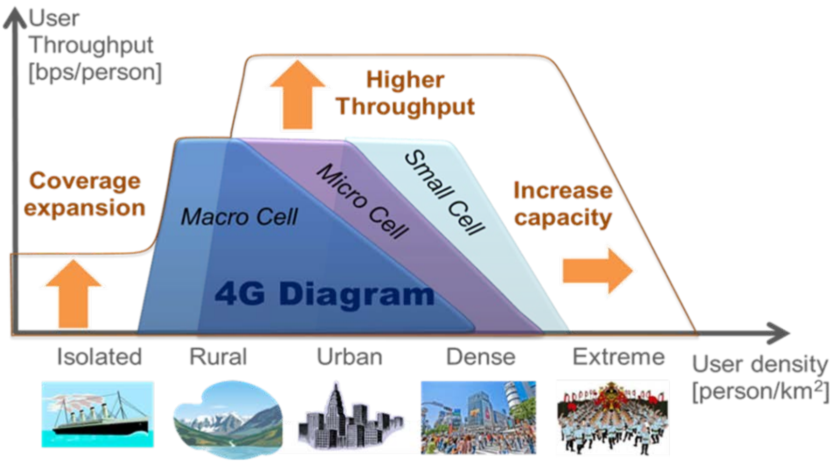


Fig.12.4-1 Requirements for future mobile communications

2) Large number of small cells

Fig.12.4-2 shows the configuration of the Mobile Fronthaul. Due to high-speed data rate of mobile terminals (each cell having a large capacity), the capacity of the line used for the mobile fronthaul needs to be increased. For example, transmission capacity of about 160Gbps (about 16 times) is required to support 10Gbps terminals in the current CPRI-based mobile fronthaul.

Furthermore, widespread deployment of small-size cells is expected to support high-speed and large-capacity mobile communications. In addition to macro cells with a radius of several kilometers, small cells with a radius of a few dozen meters to more than several hundred meters are being considered to be deployed together. For instance, assuming that a macro cell of 2km radius is replaced with small cells of 200m radius, the number of cells calculated based on the area above would increase 100 times. This brings up a concern about sharp increase of network cost due to increases in the number of links in the P2P configuration used for the current fronthaul.

Fig.12.4-3 and Fig.12.4-4 provide the number of links in the macro/small cell. If a macro cell (2km radius) is replaced with small cells (200m radius), the following is expected:

- The number of small cells increases 100 times.
- Required fibers and MFH optical transmission equipment also increase 100 times due to the increase in the number of small cells.
- The cost increase due to large capacity of MFH optical transmission equipment needs to be taken into account.

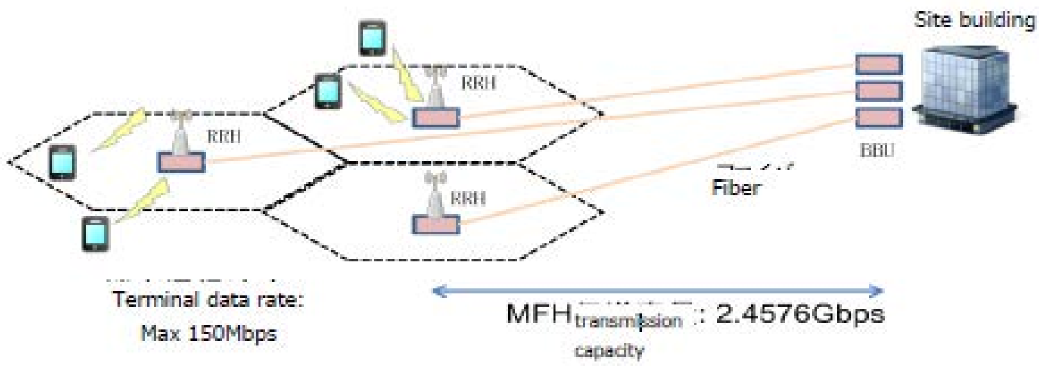


Fig. 12.4.-2 Configuration of Mobile Fronthaul

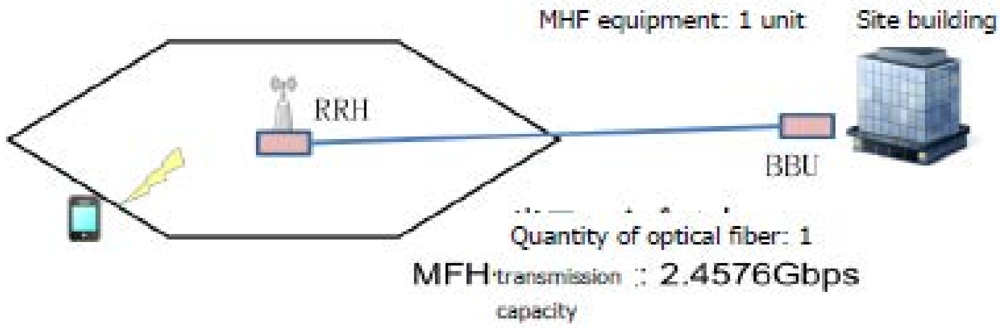


Fig. 12.4-3 Number of links at macro cell

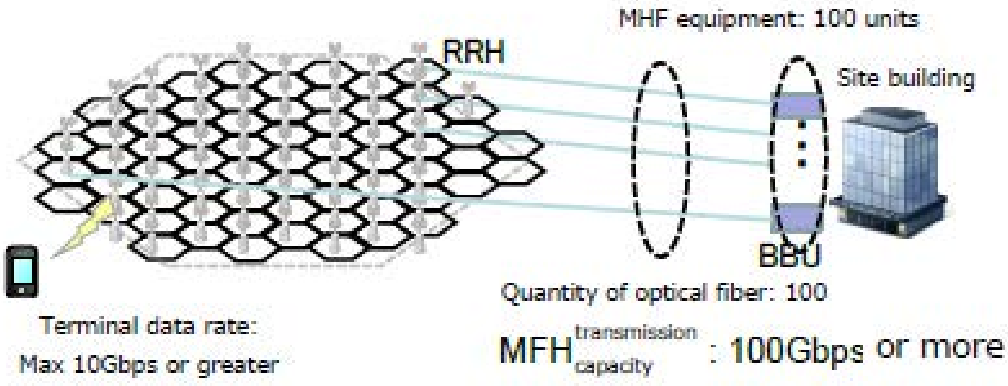


Fig. 12.4-4 Number of links at small cell

3) Low latency

Future mobile network will be required to provide new services requiring real-time performance and requirements of 1ms or less latency is being considered for E2E. Latency due to physical transmission distance cannot be ignored, so it is required to establish technologies such as (1) minimized routing path with optimized layout for each transmission equipment, (2) reduction of processing latency for modulation/demodulation processing time, protocol conversion processing time, etc. and (3) study of overall network architecture that incorporates these technologies.

It is expected that some new mobile services with very low latency requirements will appear, which could not be provided with 4G. Specifically, the E2E latency requirement of 1ms is being considered for such extreme applications as tactile communication, AR and autonomous vehicles.

4) Low Power

There are concerns about how future mobile networks will cope with an increase in power consumption as a result of increased transmission rates and the number of devices in the MFH/MBF. In light of the growing importance of energy issues such as global warming, development of new technologies is expected to achieve at least the level of 4G, with an efficiency target at one-tenth of the current rate. To do that, it is required to study technologies including equipment with high energy efficiency, active system control according to traffic fluctuation and a new MFH transmission method in building a system or network.

5) Low Cost

Regarding the increase in the number of links, the number of fibers and equipment is expected to increase as long as the current P2P configuration is used, causing an increase in costs. Costs will also increase due to the large capacity of MFH/MBF optical transceivers. Optimization of the cost is necessary for the rapid and smooth deployment of 5G.

6) Large-scale disaster/congestion/failure resilience

Future mobile networks will expect to accommodate an increase in traffic as well as the expansion of connected terminals, including those for IoT. This will mean the importance of mobile networks as part of society's infrastructure will be greater than ever. Therefore, the network needs to be more robust than ever against congestion and failure in the event of a disaster. The existing network can only cope with congested traffic during a disaster by temporarily managing network resources and therefore does not ensure sufficient network resources necessary during an emergency. It is necessary to make fundamental changes to future networks, such as allowing for prompt enhancement.

Disaster resilience can be considered from congestion and failure resilience perspectives.

The traffic during the Great East Japan earthquake needs to be considered when thinking about congestion resilience. Traffic in 2011 was 50 to 60 times higher than normal with regard to voice communication via cellar phones. Concentrated service requests from base stations that cover a wide area caused resource shortage and congestion. Telecommunication carriers then implemented 80 to 95% traffic control [12.4-2]. It was extremely difficult for users to establish a voice connection. According to the survey results, people made a call about 12 times on average until they succeeded and about 14 times on average until they gave up in disaster-stricken areas [12.4-3].

For failure resilience, with regard to unexpected communication process disruption due to damage of network functions, the earthquake and tsunami caused collapse, flooding and washout of building facility, split and damage of undergrad cables, duct lines, etc., damage of utility poles, damage of aerial cables and collapse and washout of mobile base stations, which resulted in severe damage [12.4-2].

Although no specific numerical target levels are shared as a future scenario in terms of disaster resilience, the government and users both demand further enhancement of telecommunication networks based on these lessons learned from the Great East Japan earthquake.

7) Diversified types of terminal/traffic/operator

Future mobile networks are expected to permeate further into society, even more than the conventional mobile network has. It will not only be utilized by people using conventional terminals like feature phones and smartphones, but also a by a number of terminals assumed to be embedded in devices are expected to emerge, creating a variety of equipment. As a result, traffic patterns may also be different. The end point of communication will be machines instead of people, and the number of terminals for M2M communication is expected to increase exponentially. Furthermore, M2M information exchange is expected to have a traffic pattern that differs significantly from the server-client data exchange in conventional IP networks. In addition, a variety of operators are expected to operate mobile networks. Thus, new challenges concerning the network are generated by this diversification of terminal requirements, traffic patterns and mobile network operators.

Traffic has already been increasing with conventional terminals with large screens because of an increase in video services delivered by OTT content providers. Furthermore, as M2M devices become more popular, M2M traffic is expected to increase sharply, as well.

In general, a connection topology like sensor network is assumed for M2M devices, with possible use cases such as management, monitoring and remote control of production facilities, lifelines, building and housing, vending machines and heavy equipment. Device mobility will be relatively low and both the occurrence frequency and data volume of each traffic tend to be small, but the number of terminal connections per unit area becomes very large. From 2020 and onward, along with advances toward IoT and IoE incorporating M2M, the devices and applications to be accommodated will further diversify. It is also expected that there will be many new players in the mobile service industry as MVNO.

8) End-to-End QoS

Most discussions on QoS requirements of the traditional mobile networks have focused on those in the RAN, which is defined as a section between user equipment and the gateways located in mobile core networks. However, the service quality that end user experience depends on not only QoS in the RAN but also End-to-End QoS including data forwarding quality in fixed networks, the gateways, and processing in the servers and the user equipment. In particular, QoS will strongly depend on backhaul that provides data transmission between base stations (BBU) and mobile core networks, as well as the Core IP network that provides data transmission between the gateways.

Fig 12.4-5 shows an example model of End-to-End components of mobile network. MBH is deployed as a usually metro area network, although it can also be a part of a nationwide network. Transmission latency due to the signal propagation delay on the communication lines, such as fiber, copper and microwave, and delay and jitter due to the store-and-forward packet forwarding, queuing and congestion in network nodes affect end-to-end QoS significantly. In the study in 5GMF, there is a consensus to consider end-to-end QoS by means of whole network architecture redesigning. Moreover, the required QoS is expected to vary from application to application. For example, a target end-to-end delay of less than 10 ms or 100ms would be acceptable depending on the service characteristics. Some applications may require high bandwidth and low latency, while others will require low bandwidth and extremely low jitter. Therefore, MBH for 5G systems should provide transmission links with a variety of QoS based on application requirements and flexible network resource management in addition to new technologies for higher bandwidth and lower latency.

Moreover, an aspect of mobile services, UE moves around in the geographically distributed user plane, should be considered to guarantee End-to-End QoS. When UE moves out from an area of its serving base station, MBH has to provide dynamic connection/bandwidth management for a new serving base station. Therefore, flexible network connection management is an inevitable feature of MBH for 5G systems.

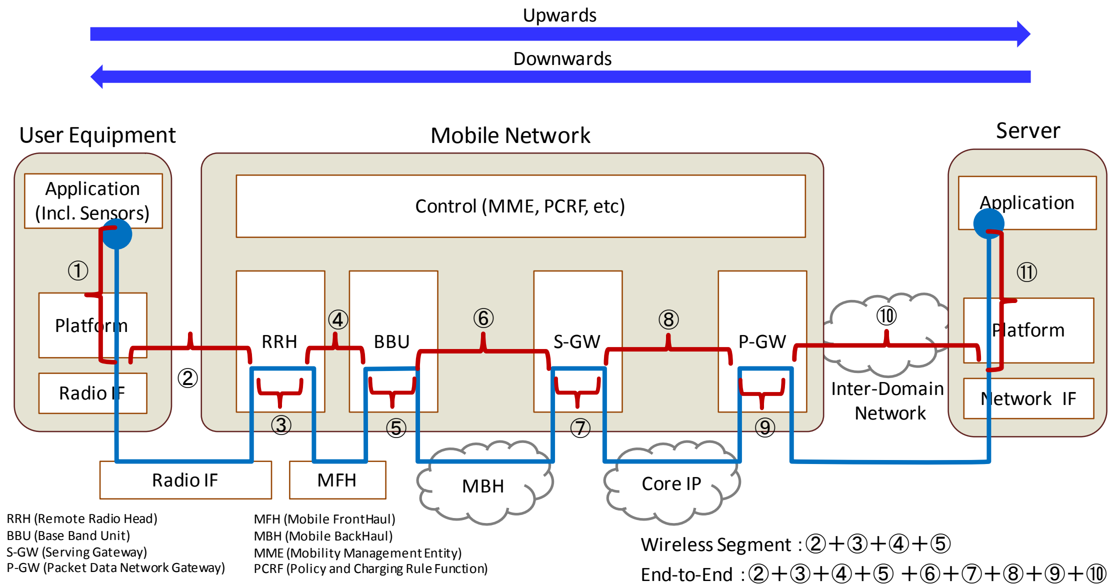


Fig. 12.4-5 End-to-End Network Component Model

12.4.1.3 Technical Challenges

1) Transport bandwidth

Unprecedented explosion in bandwidth demand is expected in mobile fronthaul and Backhaul networks for the 5G era. The bandwidth explosion in fronthaul comes from two major factors. First is the increase of available frequency bandwidth would require more than 10 times transport bandwidth for each sector. Second is the introduction of higher order MIMO, such as 64x64 MIMO, requiring 32 times bandwidth compares to typical 2x2 MIMO of today. Considering these factors, several hundred times of transport bandwidth would be required for fronthaul, the section between RRH site and centralized BBU site, in the 5G era. This means that several hundred Gbps transport bandwidth may be required. IMT-2020 FG says "a single 200 MHz sector of 5G would need about 400 Gb/s of capacity" in IMT-O-016, for example.

The bandwidth explosion in backhaul also comes from two major factors. First, the peak data rate is expected to reach 10 Gbit/ in 5G. This is 10 times higher rate than that of IMT-advanced 1Gbit/s peak rate is anticipated. Second, widespread deployment of small cells is expected to support larger capacity wireless communications in addition to macro cells with cover a radius of several kilometers. Assuming that small cells of 200m radius are deployed in an area that could be covered by a macro cell with a 2km radius, the number of cells deployed in the same area would increase a hundred times. Even with the expected statistical multiplexing effect of packet communication, it can easily be assumed that ultra large capacity circuits will be required to support Backhaul between the base stations and the core networks. For example, a minimum 10Gbps access link will be needed for a cell site which has multiple sectors with peak rate of 10Gbps. A N x 100Gbps uplink to a Metro network will be required at an aggregation site which accommodate several hundred cell sites even if 1/10 statistical multiplexing is assumed. Moreover, the required link rate at a Metro Core network that accommodates dozens of Metro Networks would reach to N x Tbps. Table 12.4-1 shows a comparison of the required bandwidth for existing 4G Backhaul networks and 5G Backhaul networks.

Table 12.4-1 Typical Line Rate [bps]

Generation	Backhaul			Mobile Core
	Access	Metro	Metro Core	Core IP
4G	100M to 1Gbps	10G	10G to 100Gbps	100G
5G	10G	N x 100G	N x T	N x 10T?

2) Functional split

The number of mobile networks employing centralized radio access network (C-RAN) architecture, which consists of base stations (BSs) and remote antenna sites (RASs), is now increasing because of its flexibility to deploy RASs and easiness to realize coordinated multi point (CoMP) transmission/reception. The current C-RAN uses a common purpose radio interface (CPRI) as the de facto standard interface. This CPRI requires a large overhead for sampling analog wireless signals. In detail, the optical bandwidth required by CPRI is more than 10 times as large as the wireless transport rate. If CPRI is still applied to 5G, the required optical bandwidth will be tens of Gbps or more, and it must be a serious problem from the view point of the transceiver cost and its power consumption. Therefore, new interface should be studied to provide effective C-RAN transport.

Bandwidth compression technique is one solution to reduce the optical bandwidth. It is reported that the bandwidth can be reduced to almost half of CPRI, while it degrades the wireless signal.

Another approach is to re-allocate the functions between the BSs and the RASs, which means that some functionality of the BSs is moved to the RASs. This would enable to transport digital data between BSs and RASs so that the large overhead required in CPRI can be suppressed. On the other hand, it also would prevent large part of CoMP gain. So, the challenge is to reduce the optical bandwidth while obtaining the CoMP gain at the same time.

3) Efficiency of fronthaul

There is a case which any UEs do not exist in a small cell, since the area of cells is reduced. Therefore, the system with all base stations are working all the time wastes power consumption. Thus, the movement prediction of UEs and the sleep control of base stations are required.

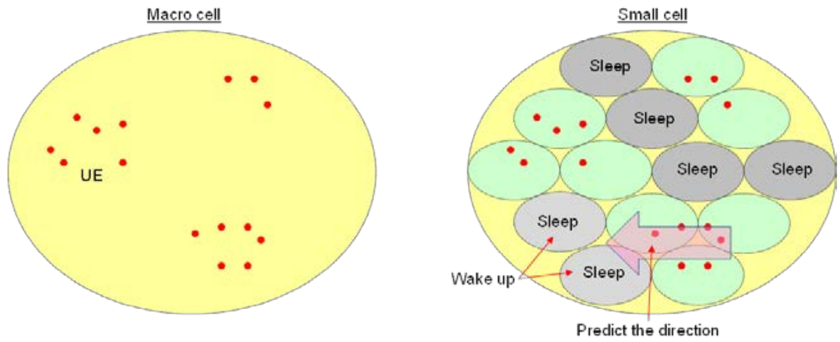


Fig. 12.4-6 Sleep control by the movement prediction of UEs

4) Reliability and resilience

Mobile networks have become a primary source of information and assistance and many services necessary for daily life now depend on it. 5G is inherently expected to be a part of the social infrastructure, supporting not only voice and Internet services but also providing every day and emergency services essential for people's daily lives, such as sensor networks and autonomous driving. Therefore, the network has to be more robust than ever against any events, such as network node failure and link failure due to equipment errors, human errors and natural disasters. A variety of protection and restoration methods are available to make fronthaul/backhaul networks robust. In a traditional fronthaul/backhaul system, the network systems have redundant components and network topology is designed to have multiple redundant routes. Protection and restoration protocols across multiple network layers, such as synergy of optical network and packet network, provide automatic recovery from communication failures. In addition to these existing features for reliability, other aspects need to be considered for 5G:

- Reliability of fronthaul

Fronthaul links require a stable low latency and small jitter transport links. The traditional protection protocol doesn't work well to recover links without influences of mobile protocols working on it and to user communications. Therefore, new protection features customized for fronthaul is required.

- Efficient Multi-Layer Protection and Restoration

Network slicing, meaning introducing separated logical network systems for specific requirements from applications, is one of the key architectural concepts for 5G. From the view point of fronthaul/backhaul networking, this means providing optimized transport lines for each logical network system by utilizing the capabilities of each network layer. For example, an optical layer can provide an ultra-broadband and low latency "hard pipe", and packet layer can provide a flexible packet multiplexing "soft pipe". Fronthaul/Backhaul is constructed by a combination of multiple layer network technologies. Therefore, studies on multi-layer orchestration to provide the best protection and restoration with comprehensive viewpoint are expected.

- Disaster Resilience

Fronthaul/backhaul for 5G will be expected to provide a more reliable transport between base stations and mobile core networks than ever before. Although network operators have made substantial capital investment in physical facilities, such as hardening of buildings and underground cabling, more research is required to improve resilience against large scale disasters such as earthquakes and floods. Multiple backup routes and restoration with consideration of geographical distribution is one of ideas.

5) Diversified types of terminal/traffic/operator/FH&BH

It is expected that in 5G many different types of base stations/devices are likely to be deployed with different transportation requirements and targets.

Current transport between wireless base stations are mostly optical fiber and microwave transmission, and the transport may be inefficient and costly to provide the transport in the future dense deployment using small cell. Wireless transport would be introduced for its inherent flexibility, low cost, and ease of deployment.

The capability of flexible topology and the capability of flexible resource assignment or sharing are necessary for new MFH/MBH to effectively utilize radio resources. Such flexibility also needs to improve other issues, such as reliability, co-existence with other solutions, fast deployment, support of multiple applications with different QoS, network level energy efficiency, etc. For the reliability of shared MFH/MBH, reserved resources have to prevent resource wastefulness or creating obstacles that affect other resources.

6) Support of network slicing / management with FH&BH

One of the major architectural themes in 5G is network slicing. The goal of this concept is to provide dynamic resource allocation and configuration management of the underlying network to the upper layer applications. The underlying network consists of RAT, MFH, MBH, and Transport as Fig. 12.4-7 shows.

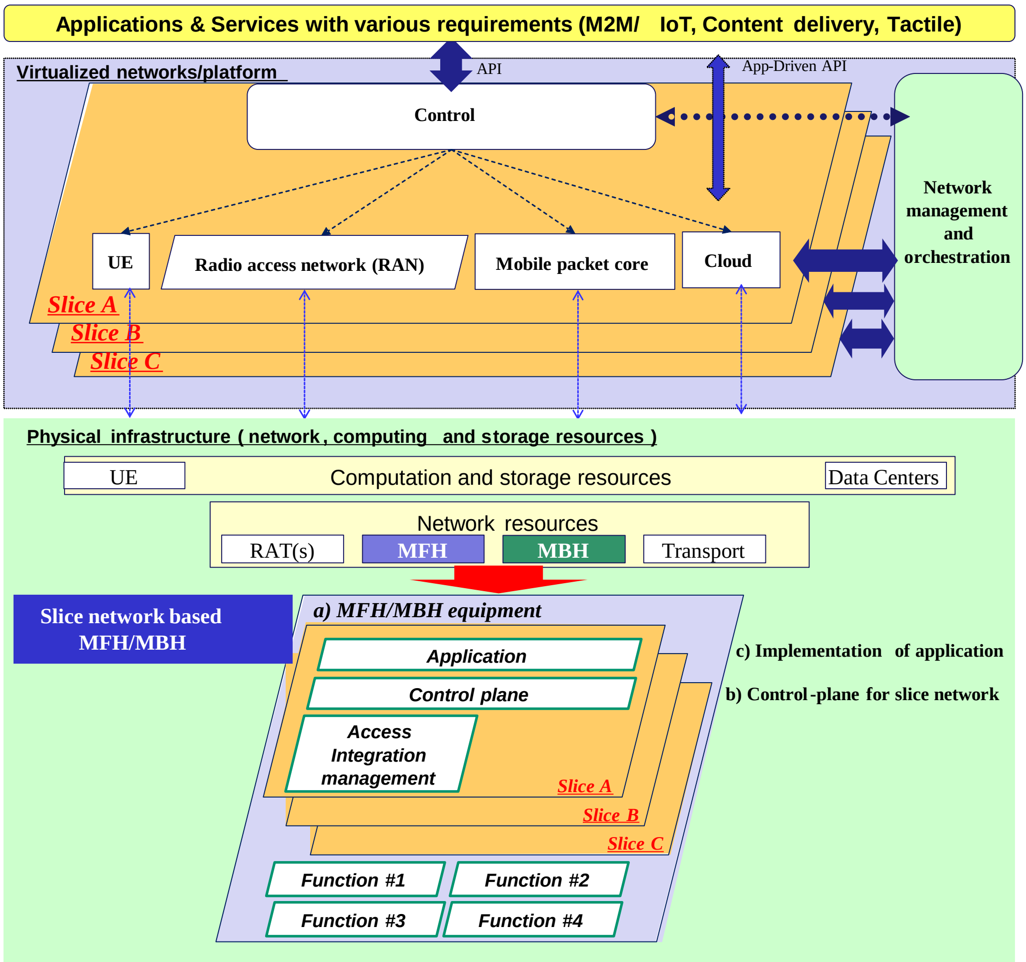


Fig. 12.4-7 The sliced network and its relationship to MFH and MBH

a) MFH/MBH equipment

MFH/MBH equipment supporting network slicing, as Fig.12.4-7 shows, has a control plane, data plane and application in every logical slice. Network slices can be a physical or virtual network, and will be necessary for slices to support a characteristic (bandwidth, latency, priority, topology, etc.) of the application. In addition, the transport system in the core networks and MBH/MFH should keep interconnectivity with the existing IP network technology where possible.

b) Control-planes for slice networks in MFH and MBH

For flexible control of mobile network, a control-plane for the slice network is required in MFH and MBH. The assignment of control functions can be configurable for the requirements of each slice network.

c) Implementation of application in MFH and MBH

Implementation of application at the aggregation part of MFH and MBH realizes flexible control according to use-cases and requirements of applications by the appropriate use of the API, where possible. However, it must be noted that typical MFH networks transport very low-level unresolved wireless signals, and so the control would be at an aggregate level.

12.4.2 Fronthaul technologies

12.4.2.1 Economization using PON technology

Since it is expected that a large number of cells will be deployed, economical structure and operation of fronthaul are key issues. Using a PON topology network to solve these issues is one solution, for two reasons:

1) Reuse of existing access networks, and
2) Economic aspects of PON itself.

At present, broadband access with capabilities over 1 Gb/s are widely deployed in several countries. The major relevant systems are G-PON, GEPON, XG-PON, and 10GEPON. These operate using a TDM/TDMA scheme that shares a single optical wavelength channel, and the systems provide generic packet transport. An actual TDMA PON transmission system would be beneficial by applying radio-over-packet style of interworking technology. Technology based on the interworking between radio base station and fronthaul is needed to realize low latency using TDM/TDMA scheme. For a large capacity transmission system, using WDM technology is one reasonable solution to provide more transmission bandwidth, without network restructuring. As basic PON technologies which support these systems/new technologies, bandwidth allocation/multiple access control technologies are used.

12.4.2.2 Dynamic control of NW resources and path optimization

To realize the high efficient utilization of network resources (bandwidth and power), the virtualization of MFH with WDM technologies is considered. Fig. 12.4-8 shows the configuration of the virtualized MFH. In this system, when additional bandwidth is required, network capacity increases with adding wavelengths, or if too much excessive bandwidth is used, network devices will go to sleep with decreasing wavelengths to reduce power consumption. Furthermore, a virtualized MFH can achieve diverse QoS requirements using wavelength groups (e.g. low latency service).

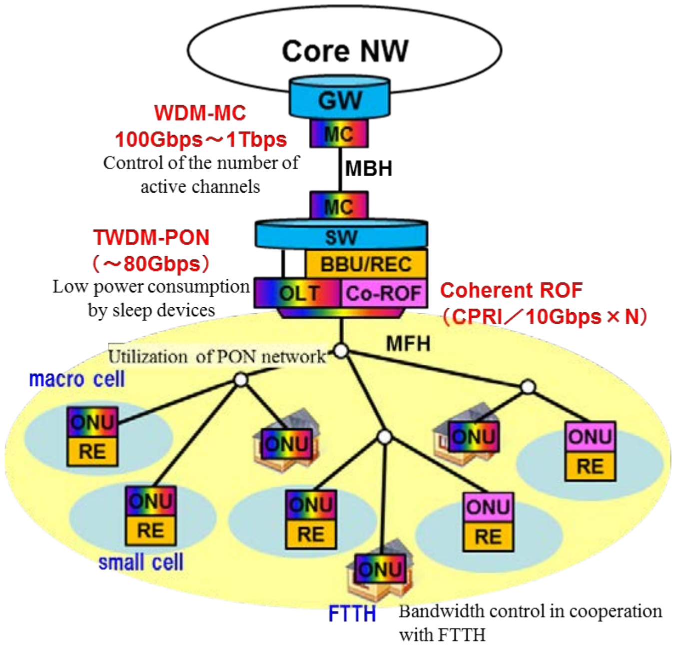


Fig. 12.4-8 Configuration of virtualized MFH/MBH using WDM technologies

12.4.2.3 Function Splitting

Re-allocation of the functions between the base station and the remote antenna site can reduce the capacity required in MFH. Several function split points are under consideration as shown in Fig. 12.4-9. When the function split point is defined in a higher layer (at a more left point in the figure), the required capacity becomes smaller, but it becomes difficult to realize Coordinated Multi-Point (CoMP) transmission/reception.

The following options are possible split points:

(a) CPRI (conventional)

(b) Split PHY

(c) MAC-PHY
(d) Split MAC
(e) RLC-MAC
(f) PDCP-RLC
(g) Service

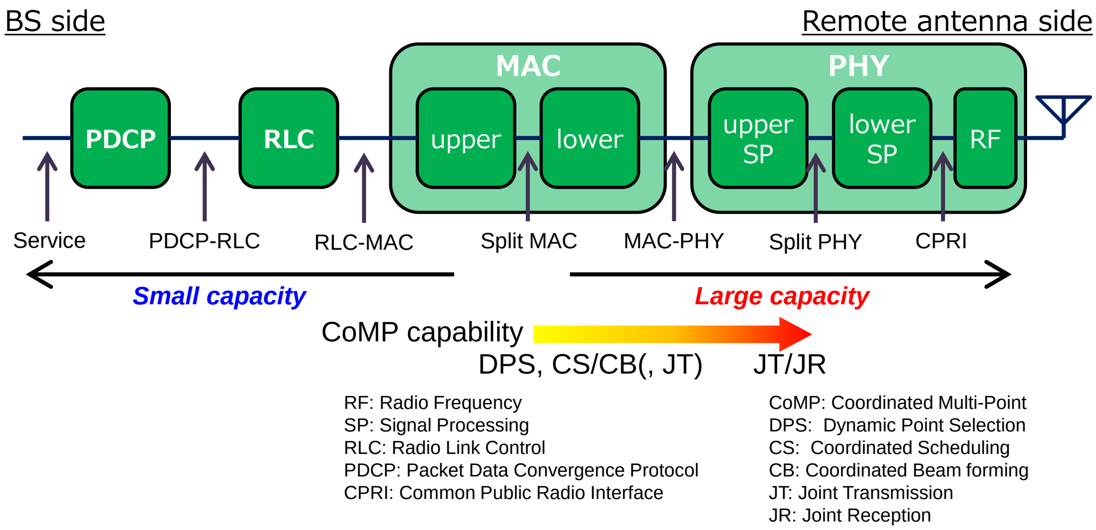


Fig. 12.4-9 Options of function split and required capacity

To realize the future MFH with a new function split discussed above, we need a new signal format, i.e. a frame. That can not only reduce the capacity in MFH compared with CPRI but also allows various wire-line networks to be used as the base for the MFH. For example, Ethernet frame is one of the candidates.

12.4.3 Backhaul technologies

12.4.3.1 QoS classify/slicing using virtualization

QoS classify

QoS is essential for network slicing because QoS defines network requirements: guaranteed bit rate, latency, and so on. Especially in terms of E2E latency, MBH has more influence than other network segments, since MBH has long-distance and multi-hop network. Therefore, QoS management on MBH is one of key themes for 5G.

QoS doesn't define granularity of network slice because the same QoS can be applied to multiple network slices. Granularity of network slice can be defined in the similar way of MEC. The following is the MEC's recommendations for identification of mobile application;

- E-RAB policy: Subscriber Profile ID (SPID), Quality Class Indicator (QCI), Allocation Retention Priority (ARP)
- Packet: 3-tuple (UE IP address, network IP address, IP protocol)

Among these parameters, QCI is the most important for QoS on MBH, because QCI defines latency and error rates as in Table. 12.4-2. Therefore, network slices on MBH should meet QoS defined by QCI. However, QCI isn't attached to mobile user-plane packets, therefore network equipment need to be able to recognize QCI indirectly from them, for example by associating QCI with TEID (Tunnel Endpoint ID) in GTP header or 3 tuple as shown above.

Table 12.4-2 QCI definition in 3GPP TS 23.203

QCI	Resource Type	Priority Level	Packet Delay Budget	Packet Error Loss Rate	Example Services
1	GBR (Guaranteed Bit Rate)	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
65		0.7	75 ms	10^{-2}	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66		2	100 ms	10^{-2}	Non-Mission-Critical user plane Push To Talk voice
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9			
69		0.5	60 ms	10^{-6}	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)
70		5.5	200 ms	10^{-6}	Mission Critical Data (e.g. example services are the same as QCI 6/8/9)

Slicing using virtualization

It is certain that both eNB and EPC will be fully virtualized in future mobile networks. Since the main role of MBH is providing IP reachability between eNB and EPC, network slicing on MBH should adjust itself to influence from virtualization of eNB and EPC. This requires future MBH to provide multipoint VPNs for multi cloud environments as in Fig. 12.4-10. The influence of both virtualizations of eNB and EPC should be examined before considering future MBH.

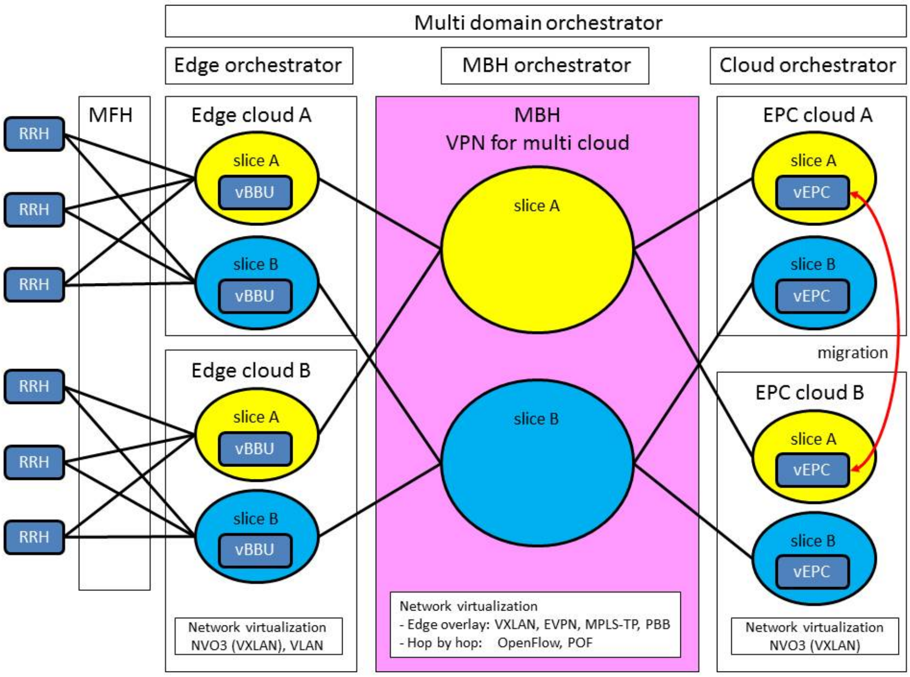


Fig. 12.4-10 MBH connections between Edge clouds and EPC clouds

With the help of NFV, virtualization of EPC has been evolving in regards to resiliency and load balancing of EPC. Future MBH should allow virtualized EPC to migrate within and among clouds. One possible method to migrate virtualized EPC among clouds is edge overlay technology, for example VXLAN standardized by IETF NVO3. Edge overlay technology has potential to create flexible network slice by decoupling IP address of EPC from underlay network management. Therefore, VXLAN is a current leading technology to realize network slicing within an EPC cloud.

In addition, eNB is also expected to be virtualized in future because of not only NFV and also CRAN evolution. This means that edge clouds will emerge between MBH and MFH. Moreover, MEC will be supposed to be deployed in these edge clouds to provide additional network services, especially for ultra-low-latency application. Through evolution of these technologies, virtualized BBU will be deployed for each network slice. However, virtualized BBU doesn't need to migrate among edge clouds. This allows use of VLAN for network slice within edge cloud, in addition to VXLAN.

MBH will need to combine seamlessly both network slices of an EPC cloud and an edge cloud. The methods of network slicing on MBH are categorized to the following two types;

- Edge overlay: VXLAN, EVPN, MPLS-TP, PBB
- Hop by hop: OpenFlow, POF

One advantage of the edge overlay model is the decoupling of the virtualized overlay network from physical underlay network, because edge overlay is an encapsulation technology. This allows operators to enhance total network systems by just updating edge network equipment without updating core network equipment. And also, VXLAN and EVPN can use an existing IP/MPLS network as its underlying network. Protection for network failure can be delegated to this underlying network function.

On the other hand, an advantage of hop-by-hop technology is full control of MBH, because the central controller can manage all SDN network equipment. That allows an operator to manage their network as they like, especially regarding latency.

Latency of network slice within MBH

In both of edge overlay and hop-by-hop technology, network latency stems from the physical network. Therefore, monitoring the latency of MBH will be more important in 5Gboth before and after the creation of network slices, no matter if MBH uses edge overlay or hop-by-hop virtualization.

This requires MBH orchestrator to gather network performance information from the physical network and compare it to required QoS as Fig. 12.4-11 shows. After slice control receives network requirements from upper API for application and services, it needs to propagate QoS requirements to not only to the core network orchestration but also MBH orchestration. At this point, MBH orchestration should refer to monitored performance of physical network, and then create network slice with appropriate QoS. After creation of network slice, MBH orchestration should monitor network performance regularly to assure SLA for each network slice.

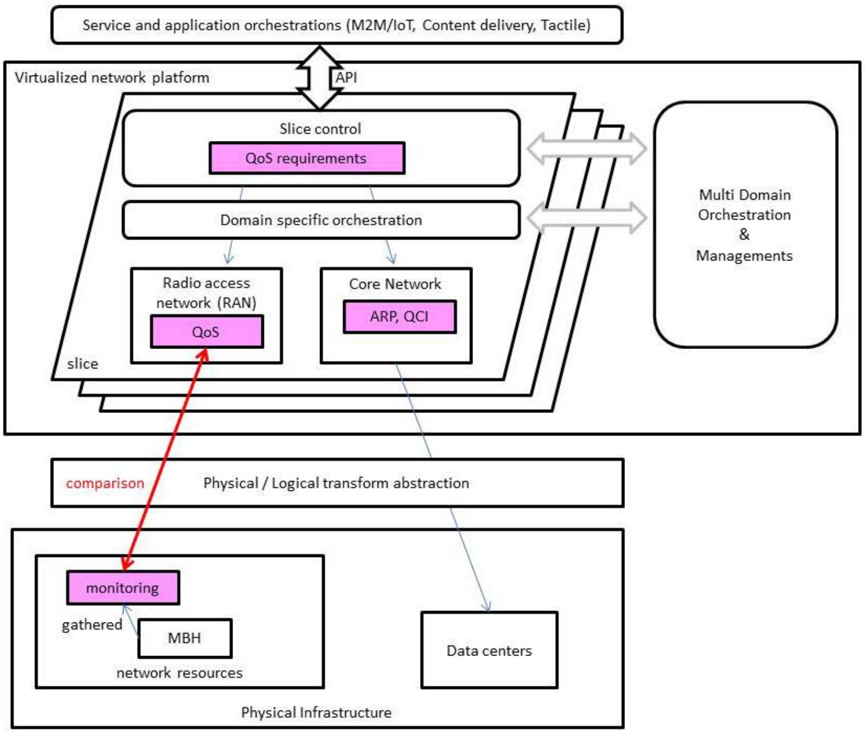


Fig. 12.4-11 Handling QoS with MBH orchestrations

12.4.3.2 Dynamic control of NW resources and path optimization

Backhaul/fronthaul provides transport links between base stations and mobile core networks. In 5G, mobile core functions and application computing capability would be built in a cloud computing environment, and distributed from the network core to the network edge to handle massive traffic or realize the ultra-low latency required by applications.

Fronthaul/backhaul should have dynamic control feature of network resources, such as optical wavelength, transmission bandwidth and priority control. A dynamic path route control with consideration of global resource status is required to achieve resource usage optimization. Fig. 12.4-12 shows an example of resource controls to provide appropriate transport path for each network slice. In Network Slice #1, the direct optical path allows ultra-broadband and low latency communication between the BBU and the Edge/Metro cloud where mobile core features and application servers are enabled. In Network Slice #2, the hop-by-hop packet network allows economical communications with statistical multiplexing between the BBU and the Core cloud where the traditional mobile core and application servers are located.

The resources for these slices should be quickly reserved to guarantee service quality when the new network services are started. Moreover, it should be dynamically controlled when the requirements are changed.

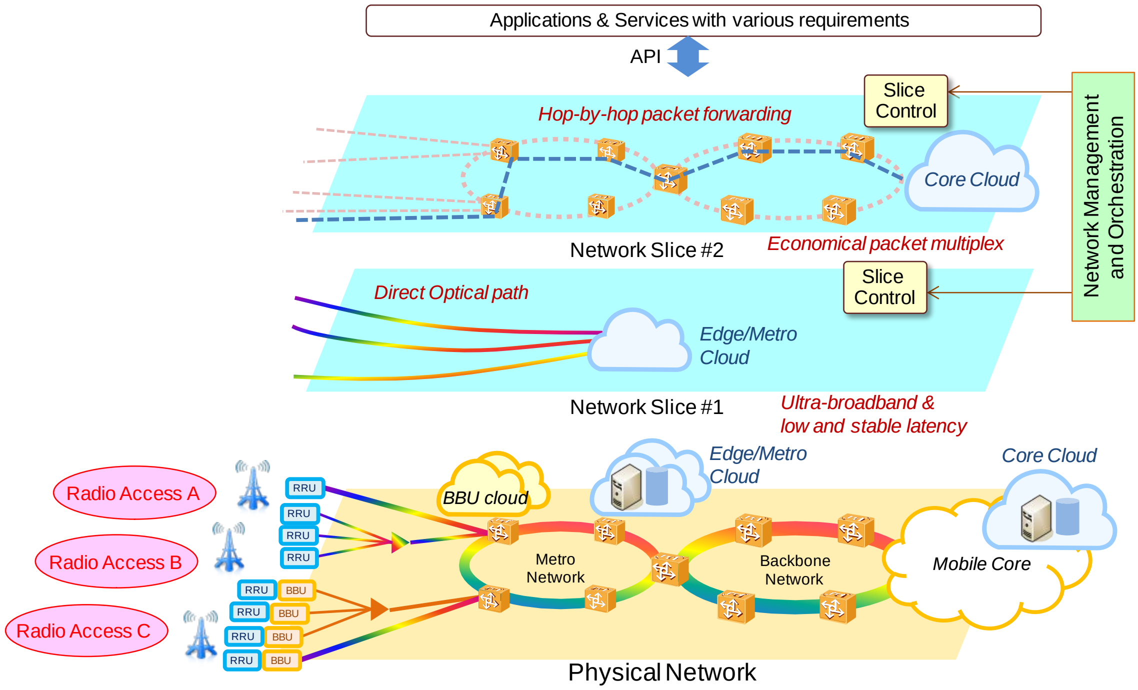


Fig. 12.4-12 Network resource Control and Path optimization for the Slicing

12.4.3.3 Energy saving methods

In MBH, more efficient power saving methods are required, since a large power consumption will occur by higher line rate than current MFHs. A virtualized MBH using WDM technologies is considered to be an energy saving method (Fig. 12.4-8). In the same way as a virtualized MFH, optimal power consumption is achieved by controlling the number of wavelengths according to required traffic amount. Furthermore, the power consumption can be further reduced by the line rate control of an optical transceiver according to required traffic amount.

References

[12.4-1] Cisco Visual Networking Index (VNI) "Global Mobile Data Traffic Forecast Update" (http://www.gsma.com/spectrum/wp-content/uploads/2013/03/Cisco_VNI-global-mobile-data-traffic-forecast-update.pdf)

[12.4-2] Ministry of Internal Affairs and Communications: "2011 WHITE PAPER on

Information and Communications in Japan”
http://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2011/2011-index.html
(http://www.soumu.go.jp/johotsusintokei/whitepaper/h23.html)"

[12.4-3] Mobile Society Research Institute, NTT DOCOMO, INC. “Disaster resistant information society”, NTT Publishing Co., Ltd., 2013.

12.5 Mobile Edge Computing (MEC)

12.5.1 Overview of MEC

12.5.1.1 General description

As we are approaching year 2020, new network service applications are emerging endlessly. While they may bring amazing experiences to the end user, they also require a more efficient, personalized, intelligent, reliable and flexible network.

Many OTT application providers have identified the demand of managing data at the mobile edge, which has significant advantages. OTT application providers will be able to access to the real time network context information so that they can adjust traffic transmission in a timely fashion. It will also benefit some OTT applications running in the cloud with locally processing huge amounts of data at the mobile edge. This data will only be used for a few seconds and doesn’t have to be sent to the cloud. Mobile users will be able to enjoy the personalized service with ultra-low latency and higher bandwidth.

Recently operator’s key role is to maintain efficient bearing networks, including core networks, radio networks, fronthaul/backhaul networks and backbone networks. The investment and maintenance of them, especially radio access nodes (e.g. base stations and eNBs) and mobile backhaul, is quite costly. Handling data traffic at the mobile edge while providing network context to OTT applications will not only help operators explore new business opportunities but also can reduce radio and mobile backhaul resource consumption.

With the demands of all stakeholders, the concept of mobile edge computing is being seriously considered in the industry. Mobile edge computing is an open IT service environment at a location considered to be the most lucrative point in the mobile network, the radio access network (RAN) edge, characterized by proximity, ultra-low latency and high bandwidth. This environment will offer cloud computing capabilities as well as exposure to real-time radio network and context information. Users of interactive and delay-sensitive applications will benefit from the increased responsiveness of the edge as well as from maximized speed and interactivity.

IT economies of scale can be leveraged in a way that will allow proximity, context, agility and speed to be used for wider innovation that can be translated into unique value and revenue generation. All players in this new value-chain will benefit from closer cooperation, while assuming complementary and profitable roles within their respective business models.

12.5.1.2 Features

Mobile Edge Computing technology enables a lot of new features in the mobile network.

- **Consumer-oriented services**: these are innovative services that generally benefit directly the end-user, i.e. the user using the UE, which includes gaming, remote desktop applications, augmented and assisted reality, cognitive assistance, etc.
- **Operator and third party services**: these are innovative services that take advantage of computing and storage facilities close to the edge of the operator's network. They are usually not directly benefiting the end-user, but can be operated in conjunction with third-party service companies, for example: active device location tracking, big data, security, safety, enterprise services, and etc.
- **Network performance and QoE improvements**: these services are generally aimed at improving performance of the network, either via application-specific or generic improvements. The user experience is generally improved, but these are not new services provided to the end-user. These include content/DNS caching, performance optimization, video optimization, etc.

<u>Augmented reality</u>

Augmented reality allows users to have additional information from their environment by performing an analysis of their surroundings, deriving the semantics of the scene, augment it with additional knowledge provided by databases, and feed it back to the user within a very short time. Therefore, it requires low latency and computing/storage either at the mobile edge or on the device.

In augmented reality services, UE can choose to offload part of the device computational load to a mobile edge application running on a mobile edge platform. UE needs to be connected to an instance of a specific application running on the mobile edge computing platform which can fulfil latency requirements of the application, and the interaction between the user and the application needs to be personalized, and continuity of the service needs to be maintained as the user moves around.

Data analytics

Some data analytic services need gathering of huge amounts of data (e.g. video, sensor information, etc.) from devices analyzed through a certain amount of processing to extract meaningful information before being sent towards central servers.

In order to support the constraints of the operator or the third party requesting the service, the applications might have to be run on all requested locations, such as mobile edge servers which are very close to the radio nodes. The application running on mobile edge server processes the information and extracts the valuable metadata, which it sends to a central server. A subset of the data might be stored locally for a certain period for later cross-check verification.

Mobile video delivery optimization using throughput guidance for TCP

Media delivery is nowadays usually done via HTTP streaming which in turn is based on the Transmission Control Protocol (TCP). The behavior of TCP, which assumes that network congestion, is the primary cause for packet loss and high delay, can lead to the inefficient use of a cellular network's resources and degrade application performance and user experience. The root cause for this inefficiency lies in the fact that TCP has difficulty adapting to rapidly varying network conditions. In cellular networks, the bandwidth available for a TCP flow can vary by an order of magnitude within a few seconds due to changes in the underlying radio channel conditions, caused by the movement of devices, as well as changes in system load when other devices enter and leave the network.

In this feature, a radio analytics Mobile edge application, which uses services of Mobile Edge Computing, provides a suitably equipped backend video server with a near real-time indication on the throughput estimated to be available at the radio downlink interface in the next time instant. The video server can use this information to assist TCP congestion control decisions. With this additional information, TCP does not need to overload the network when probing for available resources, nor does it need to rely on heuristics to reduce its sending rate after a congestion episode.

12.5.1.3 Key challenges

Mobile Edge Computing uses a virtualisation platform for applications running at the mobile network edge. The Mobile edge platform provides a framework for providing services to applications it hosts, with a basic set of middleware services already defined, allowing these applications to have a rich interaction with the underlying network environment, especially to be aware of the radio network status so that appropriate handlings will be made to adapt to the underlying network environment. In addition, radio analytic is exposed to applications through standardized API. See Fig. 12.5-1, below for an overview of MEC framework.

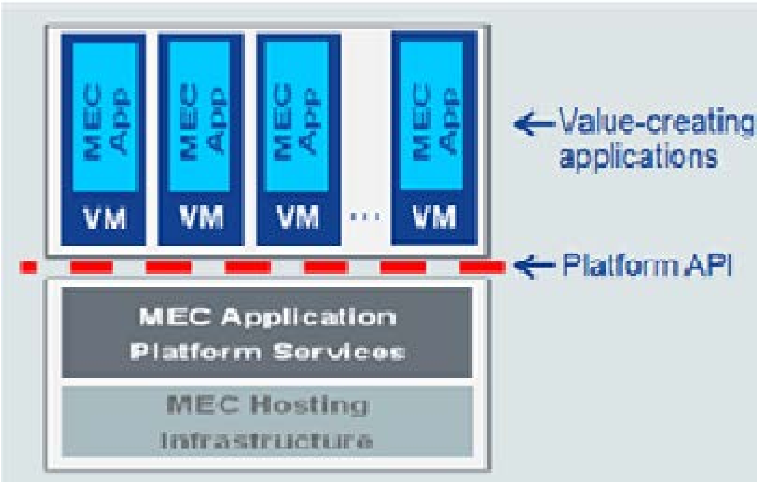


Fig. 12.5-1 Overview of MEC framework

To achieve that, there are some key challenges to be considered:

__Virtualization__

Mobile Edge Computing uses a virtualisation platform for applications running at the mobile network edge. Network Functions Virtualisation (NFV) provides a virtualisation platform to network functions. The infrastructure that hosts their respective applications or network functions is quite similar.

In order to allow operators to benefit to as much as possible from their investment, it would be beneficial to reuse the infrastructure and infrastructure management of NFV to the largest extent possible, by hosting both VNFs (Virtual Network Functions) and Mobile edge applications on the same or similar infrastructure.

__Mobility__

Mobility is an essential component of mobile networks. Most devices connected to a mobile network are moving around within the mobile network, especially when located at cell edge, but also when changing RATs, etc., or during exceptional events.

Some mobile edge applications, notably in the category "consumer-oriented services", are specifically related to the user activity. These applications need to maintain some application-specific user-related information which is synchronized with the instance of that application running on another mobile edge server. Therefore, service continuity should be maintained while the user is moving to an area served by another mobile edge platform which hosts the application.

Simple and controllable APIs

In order to enable the development of a strong ecosystem for Mobile Edge Computing, it is very important to develop APIs that are as simple as possible and are directly answering the needs of applications. To the extent this is possible, Mobile Edge Computing specifications need to reuse existing APIs that fulfil the requirements.

Application lifecycle management

The Mobile edge platform shall be available for the hosting of Mobile edge applications. The MEC management functionality shall support the instantiation and termination of an application on a Mobile edge server within the Mobile edge system when required by the operator or in response to a request by an authorized third-party.

Platform service management

The Mobile edge platform provides services that can be consumed by authorized applications. Applications should be authenticated and authorized to access the services. The services announce their availability when they are ready to use, and mobile edge applications can discover the available services.

Traffic routing

The mobile edge platform routes selected uplink and/or downlink user plane traffic between the network and authorized applications and between authorized applications. One or more applications might be selected for the user plane traffic to route through with a predefined order. The selection and routing during traffic redirection are based on re-direction rules defined by the operator per application flow. The selected authorized applications can modify and shape user plane traffic.

Data forwarding to edge or conventional computing server

User data needs to be placed into one of two different categories, depending on the service nature. One category would be data which are processed in application server of data center (DC) or the cloud. The other category is service data which should be processed near the edge. For example, delay critical application data or localized proximity service data should be processed in the edge network, while some other application data are addressed to the conventional servers in DC or cloud. In order to conduct that way systematically, an identifier presenting data types and the control entity will be required in order to address the application data to edge network or to the conventional network.

Control signal transfer management

Because some types of user application data should be processed in the edge network, service specific control signals may be needed to be combined with the edge local operation in order for data to be transferred to the edge network efficiently. Hence, a management capability will be required so that the control signals are combined or transferred to the local edge control entity for processing MEC application data.

Inter-edge mobility

Mobile edge service areas may consist of contiguous spots or isolated spots. The question arises about how those proximity services can be seamlessly transmitting data even when the devices are moving around local areas across multiple edge networks. One solution is requiring transferring cached service data from a source edge to a destination edge server. In addition, sharing device positioning information among neighbor edge sites will be useful for tracking the mobile device, especially in the case that pin-point serving spots are distributed. That capability may be realized by means of some positioning systems or any type of spot marking assistance technologies.

Gap analysis

Support enhanced MEC management of virtualization

Mobile Edge Computing uses a virtualisation platform for applications running at the mobile network edge. Although Mobile edge server lifecycle management supported by existing NFV-MANO, while MEC management should support some enhancements in following aspects:

1) Mobile edge application lifecycle management: The MEC management functionality should support the instantiation and termination of an application on a Mobile edge server within the Mobile edge system when required by the operator or in response to a request by an authorized third-party.

Mobile edge application service management: The Mobile edge platform provides services that can be consumed by authorized applications. Applications should be authenticated and authorized to access the services. The services announce their availability when they are ready to use, and mobile edge applications can discover the available services.

Support inter-edge mobility

Mobility, of course, is an essential component of mobile networks. Considering some mobile edge applications are specifically related to the user activity, it needs to maintain some application-specific user-related information that needs to be provided to the instance of that application running on another mobile edge server. Therefore, service continuity should be maintained while the user is moving to an area served by another mobile edge platform which hosts the application. So MEC system should to support inter-edge mobility mechanism for service continuity.

Support more simple and controllable APIs

In order to enable the development of a strong ecosystem for mobile edge computing, it is important to develop APIs that are as simple as possible and are directly meeting the needs of applications. In addition, radio analytics/radio network information is provided through a standardized API and if there are enhancements required. MEC system should optimized existing APIs to make it more simple and controllable.

Support traffic routing among multiple applications

The mobile edge platform routes selected uplink and/or downlink user plane traffic between the network and authorized applications and between authorized applications. More than one application might be selected for the user plane traffic to route through properly (e.g. video optimization, augmented reality). The MEC system should support traffic routing mechanism among multiple applications: selection and routing during traffic redirection based on re-direction rules which is defined by the operator per application flow, and selected authorized applications can modify and shape user plane traffic.

12.5.2 Application of MEC

12.5.2.1 Ultra-low latency networking

In the 5G era, non-perceptional latency is expected for realizing zero-distance user experience. That will be necessary in some features of delay critical interactive services or systems, since sub-1ms response time is required to realize quick recognition, reaction, and control.

In fact, we experience interaction with any system as intuitive and natural, only if the feedback of the system is adapted to our human reaction time. The required response time for interactive systems enabling real-time reactions depends on perceptual human senses.

Following description texts are extraction from ITU-T Technology Watch Report “The Tactile Internet” (August 2014).

==

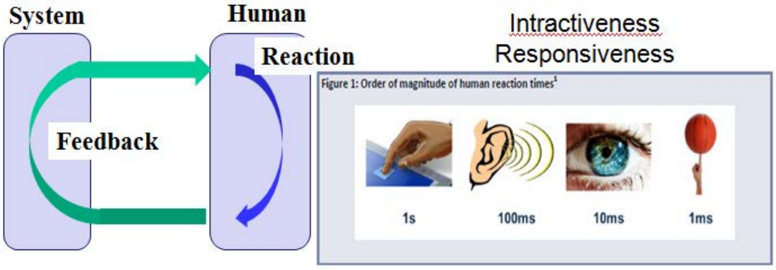


Fig. 12.5-2 Order of magnitude of human reaction times

Source: The Tactile Internet, ITU-T Technology Watch Report, Aug. 2014

An intuitive example is interactive **web browsing.** To experience immediacy, the page build-up after clicking on a link should be a fraction of the human unprepared reaction time. Real-time experience for browsing interaction is achieved only if a new web page can be built-up **within a few hundred milliseconds** of a user clicking on a hyperlink. If a human is prepared for a situation, it is clear that a faster reaction time is needed.

The human auditory reaction time is about 100 milliseconds. To enable natural conversation, modern telephony is designed to ensure that voice is transmitted within 100 milliseconds. Higher latencies would disturb us.

A typical human visual reaction time is in the range of 10 milliseconds. To allow for a seamless video experience, modern TV sets have a minimum picture-refresh rate of 100 Hertz, translating into a maximum inter-picture latency of 10 milliseconds.

But if a human is expecting speed*, such as when manually controlling a visual scene and issuing commands that anticipate rapid response,* ***1-millisecond reaction time is required.*** Examples are moving a mouse pointer over a screen and viewing a smooth path of the pointer over the screen, or moving our heads while wearing Virtual Reality (VR) goggles and expecting an immediate response from the visual display.

In principle, all of our human senses can interact with machines, and technology's potential in this respect is growing.

==

It should be noted that these levels of quick response with low latency are required not only for services that augment human perception, but also some delay-critical applications for M2M/IOT systems as well.

In addition, quick connections and quick responses from the network are also desired for the control signal processing on the control plane as well.

Requirement and motivation:

As noted in previous sections, low latency is a crucially important capability for some delay-critical service applications that must be supported by 5G Fig. 12.5-3 is a chart mapping some envisaged 5G use cases on the plane of Quality (Reliability, Low Latency) and Quantity (Peak data rate, Number of devices).

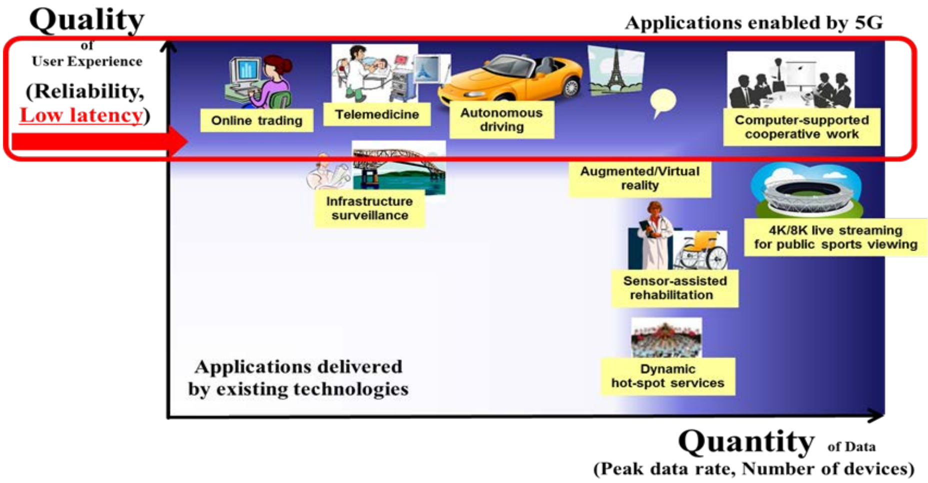


Fig. 12.5-3: Low latency in 5G Quantity by Quality mapping

Ultra-low latency use cases are shown in the upper portion on that plane, including

- On-line trading;
- Telemedicine (Tactile remote manipulations, Medical surgery);
- Autonomous driving, Vehicle Telematics;
- Augmented/Virtual reality (AR/VR);
- Computer-supported cooperative work;

Additional use-cases include:

- Automatic Speech Recognition, Text to Speech, Real-time Translation
- Delay-critical IoT services by M2M data communication
- Remote manufacturing machines, Remote driving machines

In order to provide those delay critical tactile application services, the processing time and transmission delay need to be minimized in network elements all the way from user devices to the application server.

However, today's typical network structure of mobile network shown below consists of a functional chain on the end-to-end transport path.

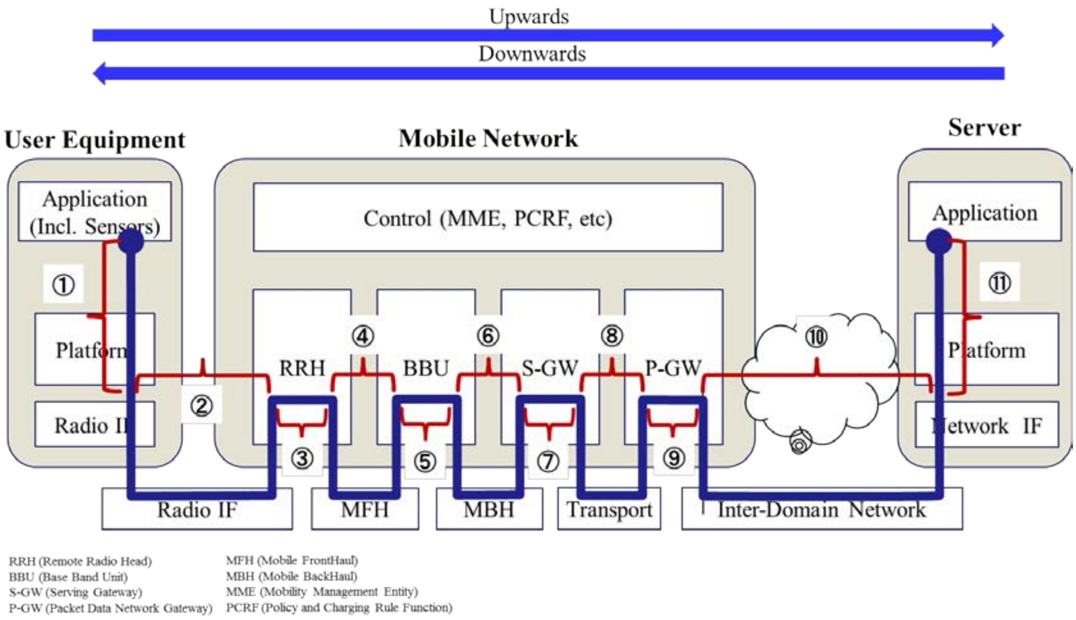


Fig. 12.5-4; Today's typical network model representing end-to-end functional chain

In this network structure, each delay of network components is added up, and the total results in a slow response on the line from end-to-end. In the particular situation in which heavy traffic is loaded on network transport lines and functional processors, the data queuing and the processing time in each functional block will be added up, creating a much longer delay, ultimately causing traffic congestion in the network.

Approach with Edge Computing for Ultra-Low Latency

Because of the issues described above, it is important to envisage functional chain overall in the data path between the user device and the application server, in order to achieve 1-millisecond order latency for the tactile services. For this purpose, an innovative approach is necessary from a network architecture perspective at the system level consideration. One expected solution is placing data computing and content caching servers near the edge of network to achieve fast access and quick response times, instead of placing those at the far end of a cloud network in a data center. This can be achieved with the architectural approach of mobile edge computing (MEC). The delay presentation diagram below is excerpted from the ITU-T Technology Watch Report. This diagram shows an example of an IoT application with the combination of a sensor and an actuator as data originator and the action driver respectively. In the mobile edge cloud shown on the right-hand side, the appropriate control and steering are processed and send back from the associated server, which will be able to achieve a low delay action; e.g. 1m second.

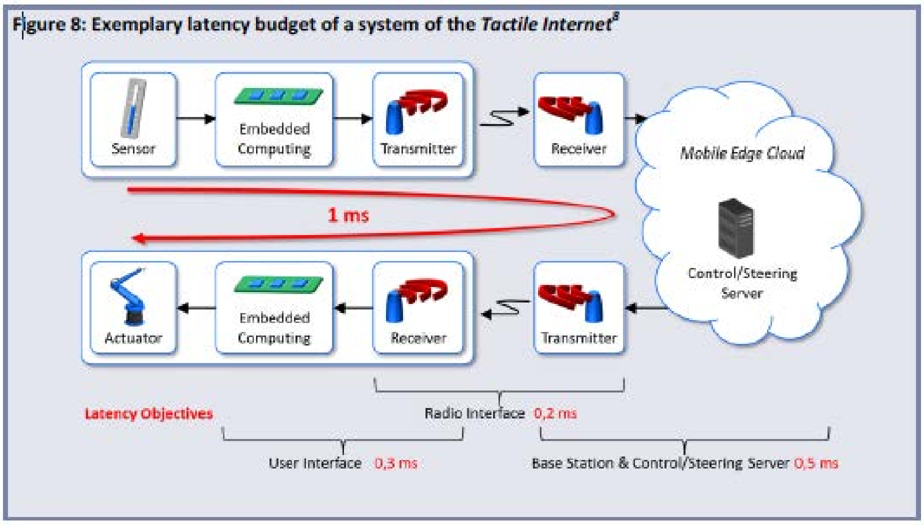


Fig. 12.5-5: Exemplary latency budget of a system of the Tactile Internet
Source: The Tactile Internet, ITU-T Technology Watch Report, Aug. 2014

This diagram represents the mobile edge computing model that can be implemented to introduce the concept of a functional chain executed on the edge side of the network. In this model, application data is processed in a server that combines computation and storage which is then placed in an edge cloud network near to the user rather than a data center in the service cloud that is located far from the user.

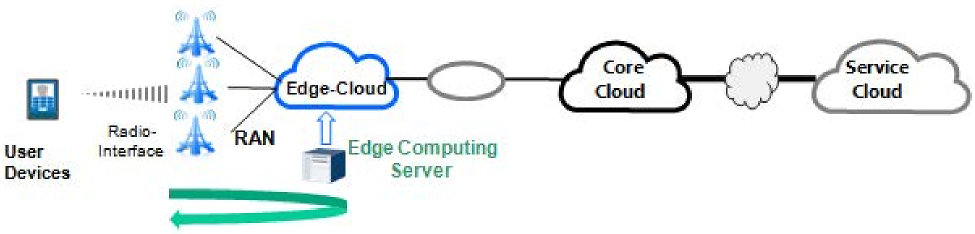


Fig. 12.5-6 Edge computing model for ultra-low latency networking

In Fig. 12.5-6, user data transactions are processed in a local edge network with the computing and cache server. Those functions are moved "closer" to the edge depending on the service, considering all requirements in regards to the application, mobility, traffic volume, and/or latency in order to optimize data transmission.

In addition, the mobile-edge computing architecture potentially owns some more network capabilities as follows:

- Location-/Service-awareness proximity services
- Big-data collection and processing on a real-time basis
- Disaster relief emergency services provision from local edge servers

Furthermore, it should be noted that the edge computing will be able to work not only on the user application data but also on some sequential signaling message processing as well. By introducing the control plane processing conduction in the edge network for some application signaling, the terminal devices will have a benefit of low delay control of quick attaching to network by reducing the connection time.

The figure below shows an image of mobile-edge computing network attached to the conventional mobile network. It shows that some application data are going into edge networks, while other ordinary data are going into the conventional core network and data center.

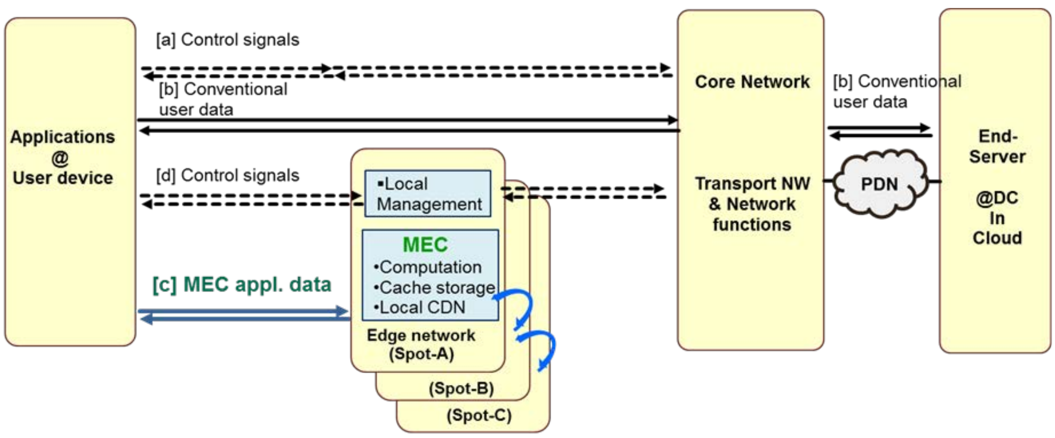


Fig. 12.5-7 Data flow image in the mobile edge network and the conventional network

However, it should be noted that ultra-low latency performance should be facilitated for a fraction of traffic compared to total traffic of network. Mobile-edge computing will effectively work for the delay critical service data, but it should not be processed for other data which do not need for delay sensitive services. It is necessary that some suitable application data are addressed to the mobile edge computing network to obtain the benefits of ultra-low latency and so on, while other ordinary data are transported to conventional core network or further into the data center to get the appropriate performance of services.

This situation creates some trade-offs and so a good balance of edge computing and conventional network operation is required. In the case that application data computing is processed in the edge side of network, then the response time will become quite small. In addition, the data traffic loaded on the backhaul network and the core network functional nodes will be relaxed with mitigation to some extent. However, on the other hand, the data processing workload together with the required storage capacity become relatively much heavy and larger in the edge network side, and more network facility and higher level of performance of data computing capability will be required in the edge network.

In addition, because the edge-computing local clouds are placed in some distributed locations for some proximity service processing, the user device mobility across those edge networks needs to be considered in order to realize seamless handover of inter-edge networks for some concerned mobile applications.

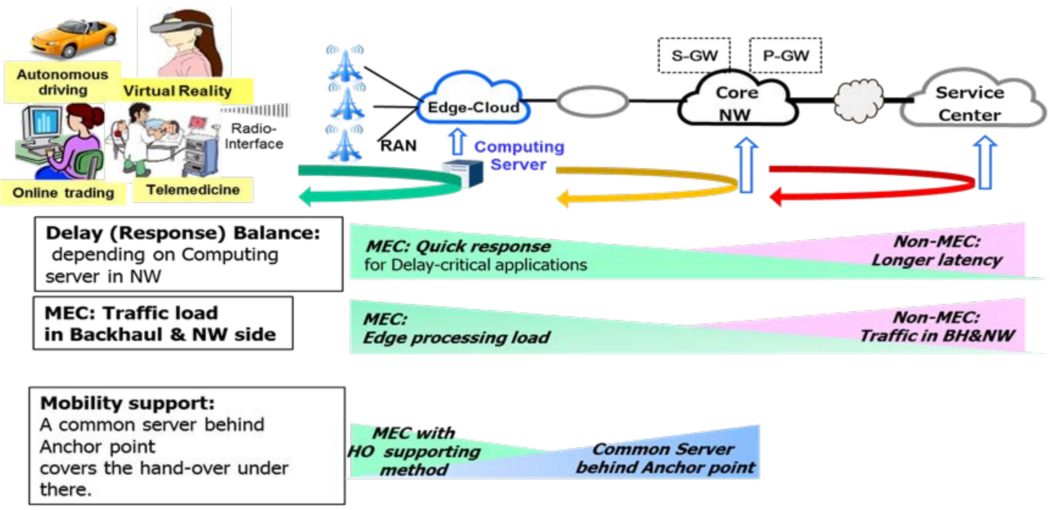


Fig. 12.5-8 Application of MEC

Challenge for better mobile-edge computing

The Following items need to be considered as a potential challenge to be resolved for a better introduction and operation of mobile edge computing network in order to provide ultra-low latency.

(1) Data forwarding to edge or conventional computing server

User data can be placed into two categories depending on the nature of the service.

One category is data which are processed in application server of data center (DC) or the cloud. Another category is the service data which should be processed near the edge. For example, delay critical application data or localized proximity service data should be processed in the edge network, while some other application data are addressed to conventional servers in DC or cloud. In order to direct traffic systematically, an identifier presenting data types and the control entity will be required in order to address the application data to edge network or to the conventional network.

(2) Control signal transfer management

Because some types of user application data should be processed in the edge network, the service specific control signals may be needed to be combined with the edge local operation, or need to be transferred to the edge network for processing the control signals efficiently. Hence, a management capability will be required so that the control signals are combined or transferred to the local edge control entity for processing MEC application data. This mechanism of control signal manipulation on the edge network side would also help in making a benefit of short connection time to network with terminal devices.

(3) Inter-edge mobility

Mobile edge service areas may consist of contiguous spots or isolated spots. A question arises how those proximity services can be seamlessly provided even when the devices move around the local areas across multiple edge networks. One solution to this issue is to transfer data that has been cached from a source edge server to a destination edge server. If device positioning information can be shared among destination edge severs, this information can also be used to more efficiently distribute data across edge servers. This capability may be realized through use of positioning systems or any other spot marking assistance technology.

12.5.2.2 Control and Management for low latency and resilient networks

Introduction of the massive number of UEs in 5G and their frequent mobility would impose challenges to providing low latency communication in robust infrastructure. MEC can play an important role in addressing this issue because the current technologies of IP mobility management such as Mobile IP or Proxy Mobile IP are not enough. They require a single common anchor point to sit in both the control and data planes for maintaining reachability (i.e. location management) information of UEs, performing handover signaling (i.e. location update), and tunneling data packets. The requirement of having a single anchor point for mobility management would have negative consequences of suboptimal communication with longer end-to-end communication path (or delay) as well as vulnerability to a single point of failure in the system. The problems are clarified by depicting the operation of Mobile IP and Proxy Mobile IP in Fig. 12.5-9. In Mobile IP, in which the UE participates in the signaling for mobility management (also known as host-centric mobility management) has the home agent as the anchor point through which all control and data packets have to pass. In Proxy Mobile IP, in which the UE is not required to participate in mobility signaling (also known as network-based mobility management) as the UE's mobility is traced by an access network node, called Mobility Access Gateway (MAG), has the Local Mobility Anchor (LMA) as the anchor point. Thus, the mobility management by employing the single anchor point is counterproductive to achieving low latency communication and making robust network infrastructure.

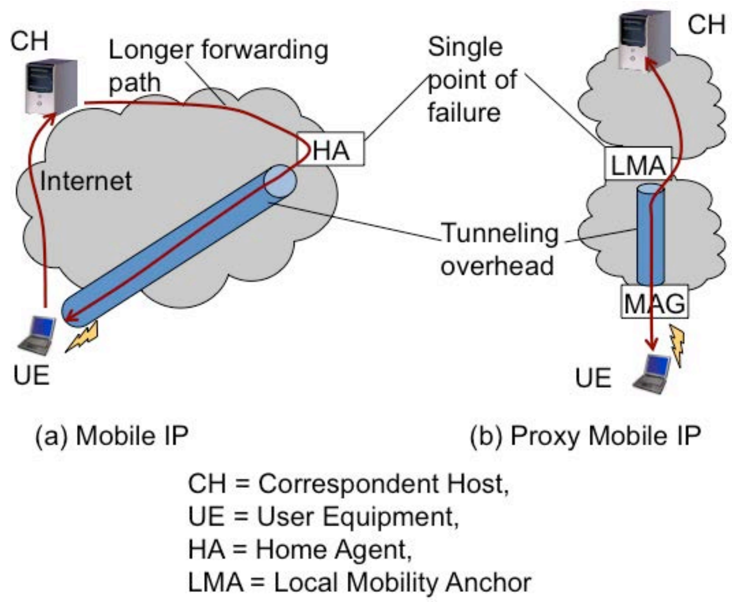


Fig. 12.5-9 – Limitations of IP mobility management

CH = Correspondent Host, UE = User Equipement, HA = Home Agent, LMA = Local Mobility Anchor

4G network now use the S-GW as the single anchor point and is likely to impose longer communication path or delay because all the communication, no matter if the correspondent node and the UE are in the same access network or different access networks, has to pass through the S-GW which is located in the core network. Therefore, in order to meet the low latency, high throughput communications to massive number of devices in 5G, we must develop distributed mobility management architecture where there exist multiple anchor points located closer to the UE in the access network. MEC concept can be extended for this purpose as described below.

Fig. 12.5-10 shows a possible structure of 5G systems with distributed mobility anchor (MA) points collocated with the BBU. The BBU are connected to multiple SGWs, which are further connected to multiple PGW. This multihoming configuration makes the network resilient to the node and link failure (e.g., due to overload or natural disasters). Namely, the network remains functioning even when a PGW, SGW, BBU or link between them gets damaged. In this configuration the BBUs get different IP address prefix blocks from different SGWs. Consequently, a UE located in a cell can be assigned with multiple IP addresses anchored with different SGWs. In this case, when the UE moves from one cell to another within the domain of a BBU, it is not required to change any of its addresses and continue communication using them. Even if the UE moves from a domain of one BBU to another BBU it is not required to change its addresses because the addresses assigned to the UE in the previous BBU are still valid in the new BBU as they belong under the same sets of SGWs.

The data being sent from the external data server will be passing through the PGW, SGW, BBU and this path remains the same when the UE moves from one cell to another belonging to the same BBU. When the UE moves to a new cell belonging to a different BBU, the addresses are still valid and the communication from the external data server can continue via the same SGW. However, the path may not be optimal when the distance between the new BBU and the SGW has become longer than the distance of the previous BBU from the SGW. The suboptimal path would be detected by the MA collocated in the new BBU and it would instruct the UE to switch the IP address to the other one which is anchored with the shortest distant SGW. For example, in Fig. 12.5-10, when the UE moves from position A to B, the MA of the new BBU instructs the UE to use IP address with the prefix assigned from the right side SGW for the optimal shortest path communication.

To allow the UE dynamically change the IP addresses used for a communication session without interrupting the application, the application should not use the IP address for the identification of the service or the communication endpoints. The application should use location-independent static IDs and these IDs should be able to be mapped to different IP addresses in the underlying layers of the communication protocols stack. The communication by using IDs is known as the ID-based communication.

In ID-based communication, it is also necessary to store the mapping records between the IDs and addresses (also known as locators). The ID registry (IDR) system, collocated with some other component, would store the ID/address mapping records and provide the record to a correspondent node that wishes to communicate with the UE.

Communication between two UEs located in cells belonging to the same BBU will take place through the BBU, and the communication between UEs located in different BBUs will also take place through the SGW. It means the end-to-end latency of communication between UEs would be at most one round trip time between the UE and the SGW. This latency remains the same even when the UEs move from one cell to another.

Moreover, to reduce the latency of data downloading services, such as popular events video or news, provisioning caching facility collocated with the MA would be helpful. In this case, when the video or news data is downloaded from the external data server for the first time, the data is cached in the BBU so that whenever a new UE requests for the same data service, the request will be served immediately with the data cached in the BBU. For the purpose, the information centric networking approach is useful. The BBU would have both the cache storage facility as well as computing or in-networking processing facility (provided by MEC) so that it would be able to serve user requests not only for the cached data but also for additional intelligence derived from the data. This offloading of computation tasks from the mobile UE would help in reduce service latency because the UE is not required to download the related huge data and perform heavy computation. This would greatly help in improving quality of experience of mobile communication services and applications.

Thus, we can conclude that the adoption of distributed mobility management, ID-based communication, and information centric networking in 5G network architecture from the design phase would be helpful to achieve low latency communication and make the failure resilient system.

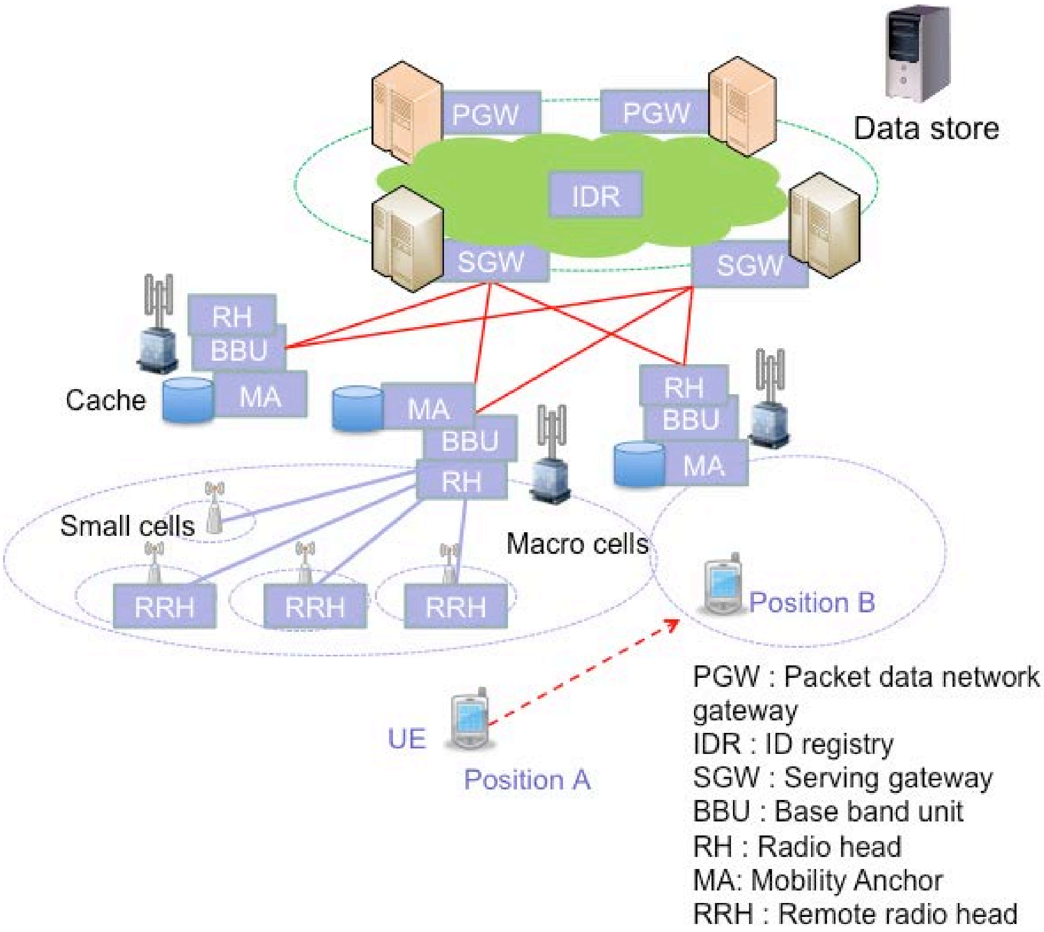


Fig. 12.5-10 – Distributed mobility anchor points and caching provisioning in 5G systems

13. Conclusion

This white paper provided information on research into everyday uses of mobile applications, including use scenes in industry, transportation, education, logistics, medical, health and welfare services, safety, emergency, and disaster relief. The research presented looks at these mobile applications from many different viewpoints, clarifies the technical requirements for the mobile communication systems as a fundamental part of society. Additionally, it reports on high quality and cutting-edge services demanded by consumers, and results of research and analysis into the trends of society and markets. It also predicts on the use scene of the 2020s and applications that will be needed in that time frame. Then, based on these expected use scenes, key concepts of 5G, requirements, capabilities, architecture, and key technologies for 5G, and the desirable radio frequencies for 5G are discussed. Below is a description of some of the results about the main features of 5G that came out of this research.

5GMF proposed in this white paper two key concepts for 5G: “Satisfaction of End-to-End (E2E) quality” and “Extreme Flexibility.” "Satisfaction of E2E quality" means providing every user access to any application, anytime, anywhere, and under any circumstance. “Extreme Flexibility” is the communications system which will allow 5G networks to always achieve E2E quality.

This white paper identified two key technologies necessary to support the wide range of use cases expected in the 5G era through “Extreme Flexibility”. The first is an “Advanced Heterogeneous Network”. The second is “Network Softwarization and Slicing”.

It is hoped that 5G standards will allow for wireless and wired networks to have the ability continue to handle growing demand for larger capacity and higher speeds as previous mobile communication systems have. It is expected that data traffic in the 2020s will be 1,000 times larger than that of 2010, meaning 5G standards will need to able to support this high level of data traffic. In addition, in order for users to be able to comfortably access rich, data-intensive content, 5G standards will need to support high speeds of more than 10Gbps.

Although transmission latency of 10s of milliseconds have already been realized by LTE/LTE-Advanced, new use cases in the 2020s such as haptic communication, robot control systems, and other control systems, will require lower latency in addition to other possible use cases that will require both low latency and high reliability. Based on these use cases, E2E latency will need to be on the order of milliseconds. Transmission latency over wireless sections of the network will especially need to be kept at less than 1 millisecond while maintaining 99.999% reliability.

Previous generations of mobile communication systems did not design for handling massive number of devices with simultaneous connections. 5G will need to meet this requirement due in part to the expected dramatic increase of IoT devices in the near future, for which 5G will be expected to support 100 times or more simultaneous connections than that currently supported.

Key technologies corresponding to these requirements will be utilized up to its maximum potential and will enable us to support new use scenes for the 2020s and beyond. Examples of these use scenes include: an air ambulance that can support surgery en route to a hospital, which will require a high capacity, low latency, and a disaster resilient network; micro robots for use in next generation agriculture, requiring high capacity, massive number of devices with low power consumption; streaming HD video while moving at ultra-high speeds, which will require ultra-high speed mobility and high capacity; and experiencing sports events in the viewpoint of players through an HD 3D live feed, which will require high capacity, support of massive number of devices simultaneously connected, and low latency. 5G will provide the opportunity to provide these revolutionary services to everyone.

Using the above research as a base, the 5GMF has contributed to ITU and 3GPP in frequency coordination, standardization, and other related activities, built collaborative relationships with 5G related organizations internationally, and disseminated 5G related information to the relevant industry sector. Along with continuing to carry out these activities going forward, in order to support the successful implementation of 5G, 5GMF plans work with partners from Japan and abroad to hold 5G verification trials under the actual condition to attract the relevant industry to utilize 5G, to give demonstrations of 5G characteristics, to consider a platform where service providers will be able to easily offer 5G related services to their customers, and to acquire the necessary frequencies bands for 5G both domestically and internationally. These activities by 5GMF will accelerate the pace of actions needed to successfully implement 5G by the year 2020.

In addition, 5GMF expects that the results of the research reported in this white paper will support ongoing and new research and development, standards activities, and radio frequency allocation coordination, as well as strengthening and extending international partnerships, and will promote to build collaborative relationship with a variety of industries in order to make the best use of 5G in user scenarios for entertainment, transportation, industry/verticals and emergency and disaster relief.

Annex Future Business and Services

Following sections, using market trends and future capabilities, introduce the perspectives of future business and services, with illustrations, eight business models and services.

With 5G technology will bring about the creation of many new mobile services. However, 5G by itself will not determine whether a service is used. Other factors include the national policies where an operator is located, individual operator business models, the cost of using 5G devices and infrastructure. In addition, industries like education and health care that will utilize 5G will also be limited by laws and regulations on their use. In this chapter, however, we will not take into account these uncertainties. Instead, we will discuss the special features and key abilities of 5G architecture and how they can be used in business and services.

The defining capabilities of 5G capabilities include a peak data rate of more than10 Gbps, mobility of 500 km/h, and latency of less than 1ms, an individual cell that can connect to 10,000 devices, 1000 times the capacity of 4G. Additionally, 5G mobile will see a drastic reduction in the use of electrical power. Before discussing the specific services and business these capabilities will bring about, we will first describe the typical use scenes.

① Surgery done in medical helicopters -- with a high peak data rate and an ultra-low latency, even in disaster users will be able to connect to 5G networks.

② Watching high definition films while moving at high speeds -- even while moving at high speeds, people will be able to download high definition films

③ A new age in agriculture with micro robots -- rather than using large devices that need to consume a lot of power, these devices will use a limited amount of power which can be used over a long period of time.

④ Athletes practicing using live hi-definition 3D systems – with a massive number of connections set up on a race track will allow uploads from the stadium interior, which, combined with the ultra-low latency of the network, means athletes can participate in a virtual race on the track.

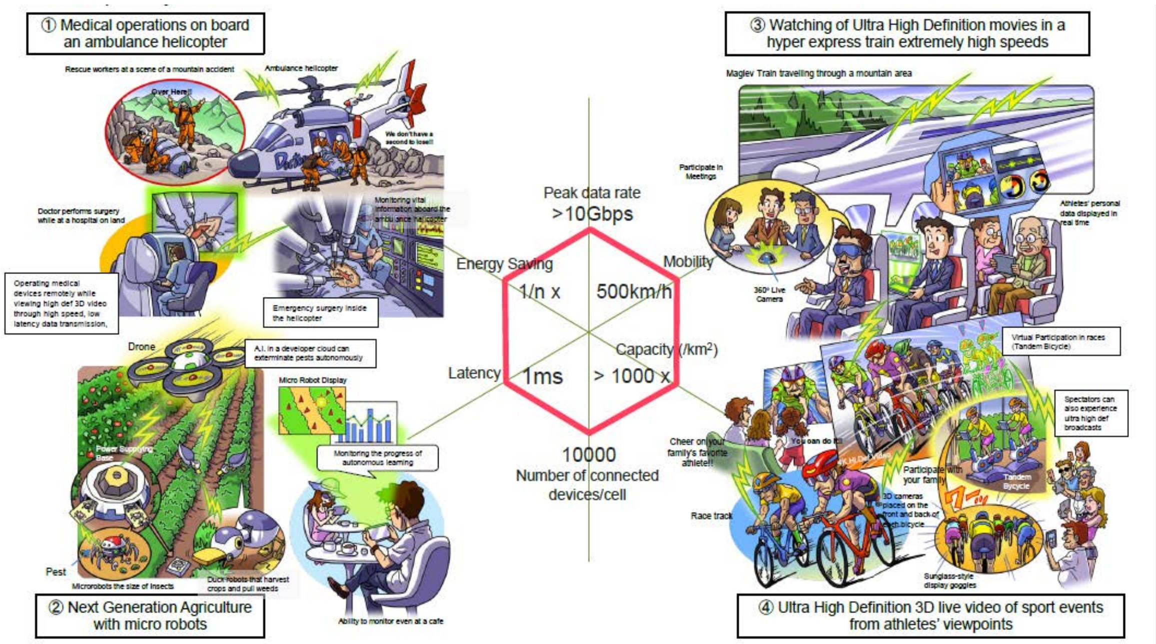


Fig.13.1 5G capability

[Further expectations to '5G' services]

1. Location Based-Services using Small Cell Technology

Current 4G smartphones have spread with the use of location-based services using GPS, such as Google Maps. One issue with these applications is the inability to access GPS while inside or around a large building. Although people are not able to access GPS inside a shopping mall, an airport, a stadium, office buildings, with the use of 5G small cell technology deployed in areas where there is a large gathering of people, devices will be able to recognize users within range of the closest small cell. That small cell base station will be used as function of location-based services. Current smartphones continual use of GPS also means the phone is wasting a large amount of power. 5G will able to utilize small cells, providing a good location service even without the use of GPS.

5G, unlike its predecessor 4G with its macro cells, will be able to deliver fast broadband with ultra-low latency utilizing multiple small cells, bringing with it the possibility of many new business opportunities. Users will be able to opt in to use small cells of mobile operator utilizing these capabilities. Mobile operators will be able to remove personal information for those users who opt in to provide location-based services. Small cells will also be able to participate in digital beam forming. With this feature, users will be able to receive information about their location near the closet small cell. Utilizing these capabilities, service providers will be able to offer users timely services at the user's specific location.

2. Services Connecting to Local Wireless Networks

5G smartphones, which are always on close at hand, will have many use scenes, but not only with the network provided by the mobile operator's frequencies, but also with local wireless networks using Wi-Fi and Bluetooth.

White goods such as vacuum cleaners and washing machines can be connected to networks. For example, now when people purchase these goods the software cannot be upgraded during the objects life. However, this may be possible in the future. Beyond this, many household functions done by internal CPUs can be transferred to the cloud on networks. For example, in the future robot vacuum machines can sense the state of the house and transfer that information into the cloud, and can download the proper algorithm to do the necessary cleaning. Washing machines and refrigerators already have many sensors, but being able to connect them to a household network and then send that data to the maker, with the power of big data in real time, can help determine whether the object is nearing a breakdown or needs necessary maintenance. In addition, using the cloud can bring down the need for using the goods own CPU, decreasing the cost of the actual product.

Consumer electronics, in order to connect to these kinds of networks, need to be able to connect to the internet using household Wi-Fi as well as through 5G smartphones in order to easily connect to the internet. It might be difficult for people, such as senior citizens, to be able to accomplish this easily. With 5G, however, users will not need to purchase separate routers to connect to the internet. For example, when they buy a smart phone and make a contract with the operator, the smart phone will register a family name for electrical appliances with the mobile operator. Household ~~goods,~~ devices, when purchased, will then be registered with the family name. When brought back to one's household, the internet connection will look for the smartphone, and the smartphone will automatically connect to other devices that are already in the household. In addition, the connected goods announce their state for other goods to connect to the smart phone through the connected goods.

3. Eco-mode Services

All resources, including frequencies, are finite. All 5G users cannot use the network at the same time with a high level of efficiency. Like users who put their phone in manner mode so they don't bother people around them by having their phone vibrate when they get a telephone call, people can only use the bandwidth they need, keeping care to think about the other users around them, with their 5G smartphone's eco-mode. The eco-mode, a low rate connection service whose periodic connection interval by 5G base stations is long, needs to be prepared. When the consumer needs more bandwidth, a pay as you go packet system can be used with 5G's broadband and low latency capacities. During a large scale emergency all users can be required to use the eco-mode to conserve bandwidth. While in eco mode, people can be connected to 5G location-based services and those in disaster areas can be provided with information immediately all at the same time. In addition, victims fined their loved ones as well through being provided information services by government and other aid agencies with priority.

4. Mobile Operator Business Model Revolution

Traffic flowing on networks has rapidly increased while the average revenue per user has declined. As mobile operators deploy nationwide 5G networks, the high cost of the initial equipment will require a high level of capital investment. User will expect that they will be able to access more content at lower service fees. 5G, to be able to cope with huge equipment costs, in addition to sharing base stations, there are various business models to work that can be implemented.

① Private small base stations. -- One model is that local stores can increase users by establishing Wi-Fi access points. Similar to 5G, they will want to build small base stations. Mobile operators in areas where 5G will be established later can build small base stations like local community antenna for television. In addition, individual users who are fiber optic subscribers can also install small base stations in order to resell their system to mobile operators.

② 5G CDN. -- In 5G, base stations will connect to edge cloud servers. End users will access the edge cloud. OTT contents will be far from the edge cloud, so they will need to access Content Delivery Networks. Mobile operators with 5G base stations and edge cloud will also have the opportunity to get into the content delivery network business.

③ Use of Personal Information in Big Data -- Information Age culture, beginning with SNS, has many merits, one of which is the sharing of personal information. Mobile operators may provide free 5G service for end users who share their personal information, for example locations they visit. This information can be collected and analyzed with big data. The results can be used to provid~~ing~~ new business opportunities for everyone.

5. Automatic Mobile operator Selection

Until today, people have been able to recognize the strength of nearby radio waves by a display on their phone. For the end user, the ideal situation is to be able to automatically choose a signal from the best operator. In the past, it was expected that a similar system to Least Cost Routing（LCR）, which was introduced with land line phones, would be introduced with mobile phones. LCR chooses the cheapest connection with an operator, but now what is most desirable is that the user chooses the best connection based on the user's throughput. For example, a heavy blog creator would choose the best throughput for uploading data but rather downloading data. Parameters that users would be able to use to choose from would include cost, throughput, and upload and download speeds.

6. Capability on Demand

The special technology of 5G includes capabilities of broadband, high density of connections, high connection speeds, low latency, and low usage of power. There are many opportunities for new business opportunities but they don't all come from these capabilities. For example, lower speed applications get lower power consumption as result. It is expected that 5G capability will be controlled by the end user on demand. The end user, with their smart phone control panel, will have a 6 pointed radar chart, from which the user can change and configure the capabilities of their phone. If the user wants a high security level, the user's phone will not use a local Wi-Fi network. 5G's communication infrastructure functions as software defined networking (SDN), so not only the telecommunication operator but the user can control their own network, allowing for the end user to have "Capability on Demand" resulting in extreme flexibility when compared to current mobile phone networks.

The end user will not have to choose their optimum capability every time for themselves, instead application developers using the device and 5G SDK/APIs can create different applications. For example, a stock purchasing financial application, when the application is active, can decide to utilize a very low level latency capability. On the other hand, to support someone running a full marathon, a sports app may only turn on the GPS and not use other power consuming capabilities in order to save energy.

These functions can also be used for IoT service industries. Businesses using 5G to connect electrical appliances or other household devices can ensure minimum bandwidth to save costs, only changing capabilities when an update is needed. This is a jump in sophistication from the current 4G MVNO business oriented interface.

7. Other Application Services

5G mobile's technological specifications include capabilities of broadband, high density, multiple connections, low latency, and low energy output. While many services can be offered to consumers using these capabilities, not all of these capabilities need to be used for every service. A service menu that restricts some of these capabilities is needed.

Looking from the perspective of the regular user the most important issue is costs. The same types of services that 3G provides will cost only 1/2 or 1/3. 5G's high efficiency will be a selling point and early adapters will be satisfied, but for the regular users, the late majority, will want to be able to choose from an inexpensive list of services. It is expected that 5G will provide inexpensive services, including the current prices of 3G but with ~~the~~ currently unavailable features that 5G broadband will provide.

Agricultural field sensors as well as sensors calculating the inventory of vending machines do not need to access high speed broadband networks. Narrowband connections will be sufficient. Pay as your go fees will be inexpensive and transmission modules will use the smallest amount of electricity that is necessary for such service. Communication intervals for reporting can be allowed to take more time than other situations, especially considering cases such as rainy weather or high temperatures. Field sensors will recognize the situation and startup for only for the certain amount of time needed to report the information and then will shut down.

Real estate customers will choose the service that provides various security and anti-theft measures. These services would potentially include location and time reporting functions, end-to-end encryption and above-mentioned field sensor functions.

For self-driving cars, automated driving levels 2 to 4 can be processed in the cloud, so the latency of the connected network is very important. But that means not only the latency of the connection needs to be guaranteed but the latency within the network needs to be guaranteed. Latency within 4G networks can't be guaranteed, but this service will need to be provided in order to provide automated driving services.

Another service that can be provided is high definition video connections to 5G small cells using beam forming tracking for users who watch a lot of rich video content.

This chapter describes some changes to the communication operator's business model, they may provide services that include edge cloud data processing functions and 5G. All IoT use scenes require low spec CPUs in devices for cost reduction. In these use scenes, not only will data transmission need to be guaranteed, but also the 5G network's ability to support a specific IoT device's data processing power will be needed, as well.