NTT DOCOMO

Technical Journal

Vol.20 No.3 | Jan. 2019

NTT docomo

Facing the Challenges of the 5G+AI Era through Innovation

DOCOMO Beijing
Communications Laboratories
President and CEO

Lan Chen

Under the theme of co-creation, NTT DOCOMO is using 5G and AI technologies to make the lives of its customers even more convenient and enjoyable while opening up new markets and developing industry. DOCOMO Beijing Communications Laboratories (hereinafter referred to as "Beijing Labs") is contributing to this effort on a day-to-day basis as a powerful engine driving NTT DOCOMO innovation.

As in Japan, technology development and large-scale testing in China is progressing rapidly toward a 5G commercial launch by 2020 through a collaborative effort between operators and vendors under government guidance. In AI, the government is supporting collaboration among industry, government, and academia with the goal of becoming the world leader in AI by 2030. In addition, the number of highly competent, adaptive, and motivated personnel with technically advanced skills in 5G and AI is increasing yearly.

Blessed with a good location, good timing, and high-quality personnel, Beijing Labs has been ahead of the times in seizing many opportunities. In research, for example, we have made a transition from wireless communications as our single focus to wireless communications and AI as a two-pronged approach focused on "5G and beyond." In this way, we have expanded our efforts from fundamental to applied research.

In wireless communications, we have been accumulating robust wireless technologies to support the ongoing development of NTT DOCOMO while promoting the standardization of technologies essential to future networks. For 3GPP Release 15 RAN1, Beijing Labs played a leading role in formulating elemental technologies such as Massive MIMO, channel coding, and control channel design contributing 40% of the RAN1 papers submitted by NTT DOCOMO.

Furthermore, with the aim of contributing to NTT DOCOMO's "smart life" initiatives, we have been absorbing technologies quickly and producing results at breakneck speeds through joint research with universities and research institutions on cutting-edge natural language processing and computer vision.

At the same time, frequency band harmonization between China and Japan is extremely important for achieving a 5G global ecosystem. With this in mind, Beijing Labs has arranged meetings with high-level key persons of government institutions to exchange opinions, participated in the IMT-2020 promotion group led by the China Academy of Information and Communications Technology (a first for a foreign operator), and participated in 4.4 – 4.9 GHz verification tests with vendors. We have also been supporting the needs of relevant departments at NTT DOCOMO headquarters by providing them with advanced technologies from Chinese vendors.

To support these efforts, Beijing Labs puts innovation into practice in the following three ways.

(1) Results-oriented innovation

"Be trustworthy in word and resolute in deed." Taking responsibility for committing to results is essential to success in research and development. For example, in a project we undertook on the analysis of drone images for infrastructure inspection, nothing was more important than producing results that met the needs of our customer. Consequently, to compensate for a shortage of personnel in the AI department, we involved researchers from the wireless department and conducted interdisciplinary brainstorming to come up with fresh ideas. In the end, we were able to produce high-quality results satisfying the customer's order within one month.

(2) Fast-track innovation

"Good work requires good tools." At standardization meetings, it is essential that we provide facts (evaluation results) that can demonstrate the advantage of NTT DOCOMO's proposal over that of our competitors. To this end, we use a simulation evaluation platform to expand functions rapidly while maintain accuracy in results. It is precisely this asset that enables us to submit technology proposals and evaluation results in time for standardization meetings held every two months.

(3) Fearless innovation

"Pressing ahead without fear of difficulties." Prior to the standardization of uplink non-orthogonal multiple access (UL NOMA) technology, many companies made proposals resulting in more than ten technical solutions to be considered. Faced with such competition, we boldly applied radio communications theory to design optimal sequences targeting the generalized Welch-bound equality that exploited differences in transmission power thereby demonstrating the superiority of NTT DOCOMO's technology.

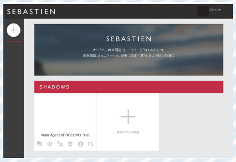
"They who know the truth are not equal to those who love it, and they who love it are not equal to those who delight in it." At Beijing Labs, the passion felt by our researchers for their work plays a big part in unending innovation.

To provide services that create a favorable response from our customers, that can induce them to say "After all, that's NTT DOCOMO!" and convince them that NTT DOCOMO is their best choice now and into the future, Beijing Labs is committed to progressing from "very good" to "excellent" and even "absolute best."

NTT DOCOMO
Technical Journal
Vol.26 No.3 Jan. 2019

[Contents]

DOCOMO Today

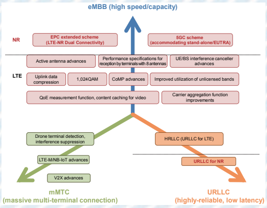


(P.10)

(P.26)

Technology Reports (Special Articles)

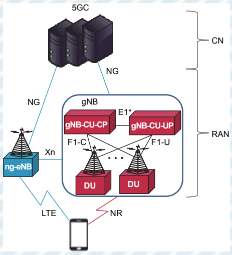
Special Articles on New AI Agent



(P.37)

Technology Reports (Special Articles)

Special Articles on Release 15 Standardization —Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

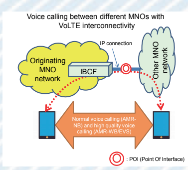


(P.62)

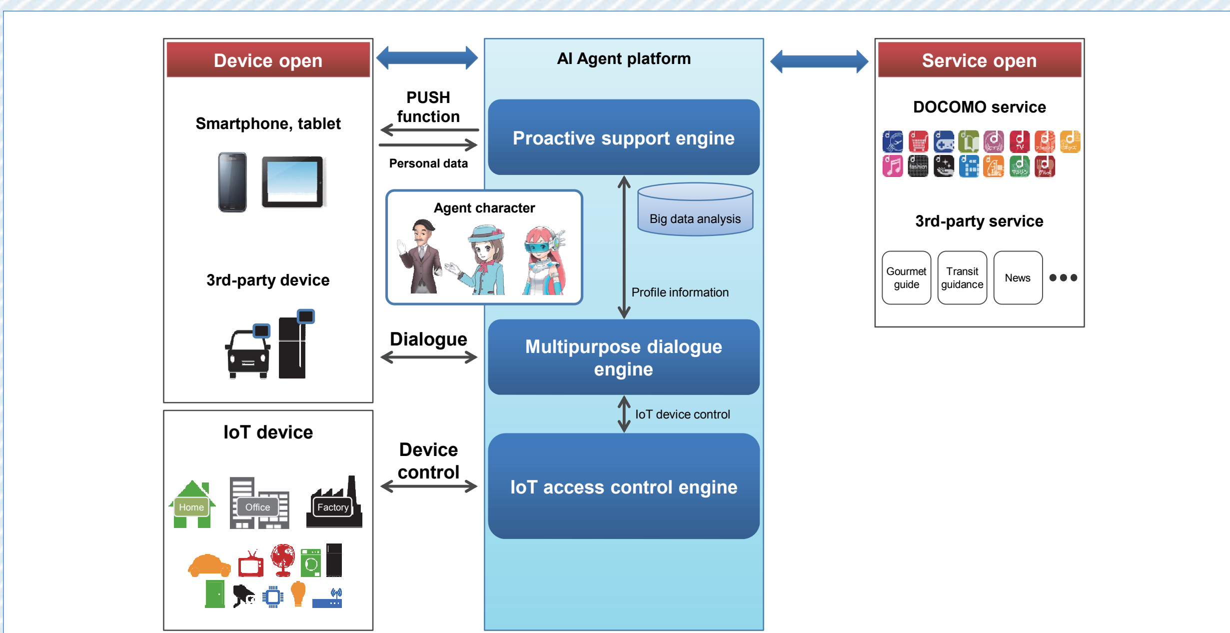
Technology Reports

(P.102)

Topics



(P.112)



Technology Reports (Special Articles)　docomo AI Agent Open Partner Initiative (P.4)

AI Agent platform structure

Technology Reports

AI Agent | Open Partner | Voice AI

Special Articles on New AI Agent

docomo AI Agent Open Partner Initiative

Service Innovation Department Takanobu Oba† Hiroki Asai
Innovation Management Department Yoshikazu Akinaga Gahi Kim

In its medium-term strategy to 2020, the "Declaration beyond", NTT DOCOMO will offer a new AI agent to revolutionize the lifestyles of its customers, and as the initial step, NTT DOCOMO announced the docomo AI Agent Open Partner Initiative, which entails co-creating voice dialogue assistance functions and hardware with partner companies. This article describes the target world of AI agent, and various related initiatives.

1. Introduction

With the spread of smartphones, and improvements to voice recognition and natural language processing technologies*1, services that assist users via voice dialogue have increased rapidly. Common examples include Apple's Siri®*2 and DOCOMO's Shabette Concier.

Additionally, the appearance of new hardware called "smart speakers" has been gaining attention in recent years. For example, Amazon's "Amazon Echo*3" enables users to access a voice assistant service by talking to the "Alexa" AI system, which responds appropriately. Google also has released its "Google Home™" smart speaker that connects to the "Google Assistant™*4" voice assist function provided by the company. LINE also released a smart speaker called "WAVE" that uses the company's "Clova*5" cloud AI platform, and plans to release a smart display called "FACE".

Thus, as its medium-term "Declaration beyond" strategy to 2020, NTT DOCOMO will offer a new AI

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently, Nippon Telegraph and Telephone Corporation Media Intelligence Laboratories

*1 Natural language processing technology: Technology to process the language ordinarily used by humans (natural language) on a computer.

agent to revolutionize the lifestyles of its customers. As the first step, NTT DOCOMO announced the docomo AI Agent Open Partner Initiative in June 2017 [1]. This initiative entails co-creating voice dialogue assistance functions and hardware with partner companies.

This article describes the features of AI agent and a platform for its achievement, and business developments and partnerships using the docomo AI Agent API.

2. AI agent World

Conventional voice assistance service design tends to include a single-personality character (hereinafter referred to as an "agent") that response to various user requests. Although such designs could be viewed as intended for agents with omniscient abilities, the functions and services that can be provided through voice assistance using the assets of only one company are limited.

Thus, initiatives to solve this issue through the provision of open platforms are ongoing.

With Alexa, it's possible to register individual functions and services called "skills". This allows developers to freely expand Alexa's functions. In other words, skills enable various companies to provide their services through Alexa.

Like Alexa, AI agent also has functions that enable other companies to provide services, but they are not exactly the same. With Alexa, other companies' services are provided through only one agent (Alexa). However, AI agent provides the freedom to individualize the characters of the agents that provide the services of each company. Thus, this enables the existence of many different agents to provide users with combined support. This is analogous to real human society in which services of companies and stores are provided by their respective staff members, and if something is unknown, then a person who knows can be asked.

AI agent is divided into two parts - main agents and expert agents.

- The main agent is positioned as the closest partner of the user, and is the first agent that responds when the user speaks to a device such as a smart speaker or smart appliance.

  The main agent is not an agent provided by NTT DOCOMO, but one configured with unique characters for the devices of individual companies. These features are called "device open".
- In contrast, expert agents appear when they are called by a main agent. These are created as experts in certain fields, and could presumably be staff member agents of companies or shops for example. This function is called "service open", and enables users to call particular agents when required, providing users with the services of each company.

This variety of agents with different characters plays an important role in bringing about new value through co-creation with partner companies. In companies, and in-service provision, brand image is of paramount importance, and as a contact point, an agent has a direct connection with a brand image. Having partner companies use DOCOMO's AI agent asset and give it their own original character will enable them to provide services with the brands that they have fostered through time, without the NTT DOCOMO brand at the forefront. In

*2 Siri®: A trademark or registered trademark of Apple Inc. in the United States and other countries.

*3 Amazon Echo: Amazon, Echo and Alexa are registered trademarks of Amazon.com, Inc. and its related companies.

*4 Google Assistant™: Google Assistant and Google Home™ are trademarks of Google LLC.

*5 Clova: Clova and WAVE are registered trademarks of LINE CORPORATION.

Japan, where the use of company mascots is particularly common, the technology will enable such companies to use their mascots as agents without having to change them.

3. docomo AI Agent API

In the docomo AI Agent Open Initiative, we are offering the docomo AI Agent API[*6] as a central system to configure the AI agent world.

The docomo AI Agent API consists of a multipurpose dialogue engine, a proactive support engine and an IoT access control engine (**Figure 1**).

(1) The multipurpose dialogue engine is a voice dialogue system that plays a central role in achieving the AI agent world. To create agents for voice dialogue, voice recognition, natural language understanding, and speech synthesis[*7] are provided as a set. The system enables both voice and text input. A variety of models are registered to enable creation of voices suitable for the various agent characters.

(2) The proactive support engine is for collecting and analyzing diverse information, and analyzing and using various user profile[*8] information. Collected information includes information related to user activities, and

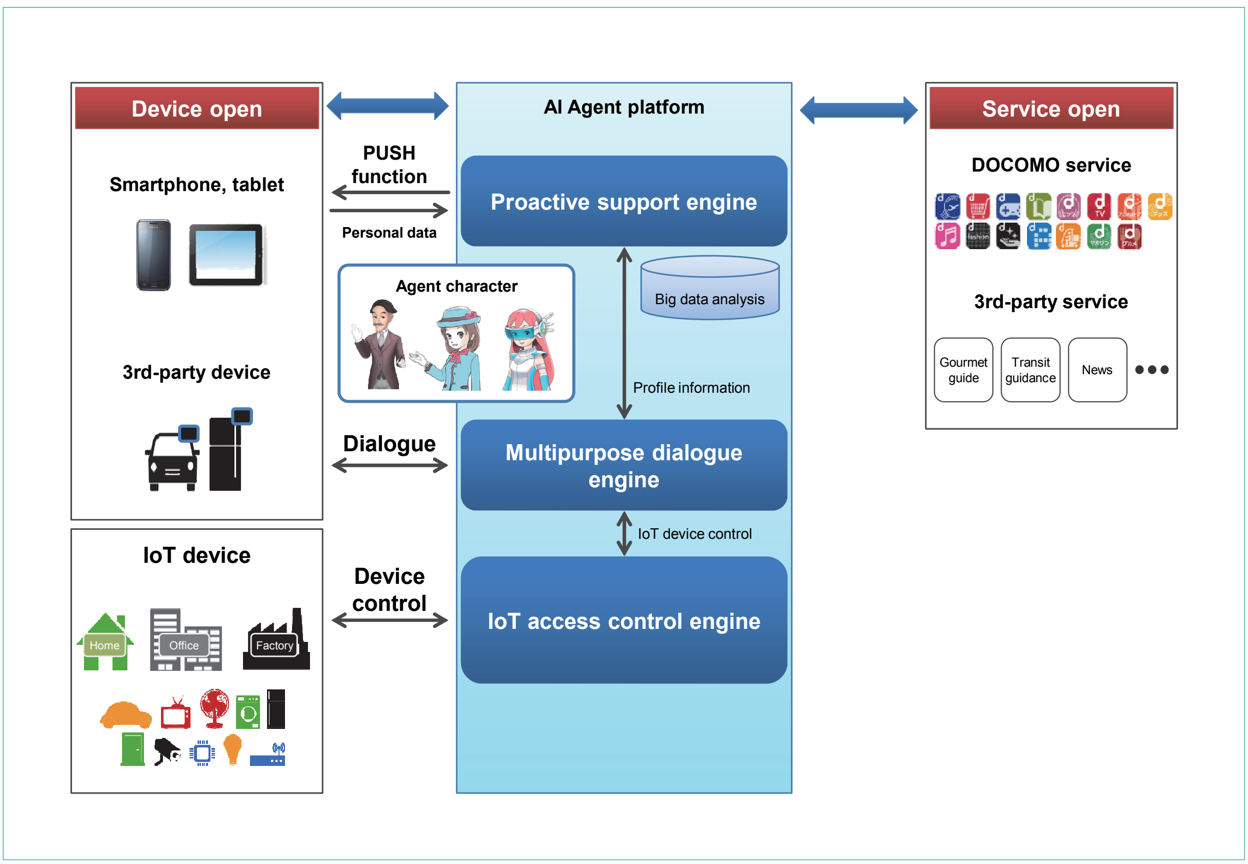


Figure 1　AI Agent platform structure

*6　API: An interface that enables software functions to be used by another program.

*7　Speech synthesis: Technology for artificially creating speech data from text and verbally reading out text.

*8　Profile: Here, this refers to information about users place of residence, tastes and interests etc. estimated from system operational history.

information of a highly public nature such as weather, public transport operation and disaster information. Analysis of this information is aimed at providing the necessary information to those who need it in a timely manner.

(3) The IoT access control engine is for using IoT devices in the home via the Internet. We adopted "DeviceConnect" software to control IoT devices with different respective specifications in a unified format [2], to enable control of different devices, with the same interface specifications.

The above three engines can be mutually interconnected, so that the multi-purpose dialogue engine can access the proactive support engine to reflect user information in conversation with an agent at the right time. Also, the multipurpose dialogue engine can access the IoT access control engine to enable operation of IoT devices with voice commands.

4. Business Developments and Partnerships Using the docomo AI Agent API.

The docomo AI Agent API is a platform to bring about new value through co-creation with partner companies, and has been released to enable anyone to freely use it. Since multiple partners can use the same platform to provide their respective services, NTT DOCOMO is engaging in mutual B2C*9 and B2B*10 business developments taking advantage of this characteristic.

4.1 my daiz - "By Your Side with Your Smartphone"

Released in May 2018, my daiz is a DOCOMO B2C service that uses the AI agent platform technology, and is a pivotal service in the docomo AI Agent Open Partner Initiative.

Aiming to build one-to-one relationships with customers, this service was developed under the concept of "by your side with your smartphone" to achieve the new AI agent to revolutionize customers' lifestyles, as stated in our medium-term strategy to 2020, the "Declaration beyond".

my daiz is characterized by the proactive support engine which is used to proactively pick up and deliver suitable information to the customer. The system displays on-screen information based on profile information estimated by the proactive support engine and updates it as necessary, and analyzes customer situations from registered schedules etc. to deliver suitable information with suitable timing.

Also, my daiz includes the concept of "members". "Members" are various services using the multipurpose dialogue engine, which can be added free-of-charge, and include many services offered by partner companies. NTT DOCOMO plans to expand these "members" going forward. When using members, agents created by other companies appear as contacts, and the multipurpose dialogue engine is used for this dialogue function. Also, my daiz enables IoT infrared remote control via the IoT access control engine.

4.2 Collaboration with Other Companies Using the docomo AI Agent API.

The docomo AI Agent API can be used to build

*9 B2C: Business-to-consumer transactions.
*10 B2B: Business-to-business transactions.

services or products of other companies. In the lead up to full commercial offering planned for spring 2019, discussions are ongoing towards co-creation and collaboration with a range of companies. Here, we introduce a few of these initiatives.

YKK AP Inc. announced its "UPDATE GATE" door of the future with a built-in AI and facial authentication system. The company has used the docomo AI Agent API for the door's AI dialogue system. The door has been developed with the concept of "Update with each daily pass" and aims to enrich the daily lives of residents by providing them with required information such as weather and transport information, etc., as shown in **Photo 1**.

Photo 2 shows NTT Resonant Inc.'s "Oshieru: romance consultation bot robot". Oshieru was originally a text input-type Web service, but the company held discussions with NTT DOCOMO and used the docomo AI Agent API to turn the system into a robot for exhibitions, etc.

In this way, as well as providing voice recognition and speech synthesis as a set, the multipurpose dialogue engine can cooperate with other companies' servers to add voice input/output functions to existing text input/output services.

Like the above examples, NTT DOCOMO has engaged in a range of consultations with various companies who plan release of diverse devices that should gain attention. Above, we introduced some examples of utilization of the multipurpose dialogue engine. Similarly, we had been taking inquiries on utilization of the proactive support engine and IoT access control engine, and we expect further usage of the docomo AI Agent API in various scenes in the future.

Photo 1 Usage image of the "UPDATE GATE" door of the future

Photo 2 "Oshieru" AI romance consultation robot

4.3 Fostering a Developer Community

For docomo AI Agent API to become a platform supporting a wide range of services, it's necessary for both partner companies providing services and developers building the systems for those services to understand the advantages of docomo AI Agent API and how to use it. For this reason, NTT DOCOMO is putting efforts into fostering a developer community.

First of all, NTT DOCOMO has opened a trial site for developers [3]. This site provides information necessary for development such as instructions on how to use docomo AI Agent API and sample code information, and enables developers to actually try out the docomo AI Agent API.

NTT DOCOMO also has been holding study meetings and hackathons*11 around Japan that have been attended by more than 1,000 people to date.

Going forward, NTT DOCOMO plans to proactively disseminate information to developers around Japan as well as provide them with rich content including a wide range of development examples and sample code.

5. Conclusion

This article has described the features of AI agent and a platform for its achievement. The article has also introduced services provided by NTT DOCOMO using the docomo AI Agent API, products achieved through collaboration and co-creation with other companies, and initiatives to foster a developer community.

NTT DOCOMO has begun studies with partner companies on specific products using this API to strengthen solutions. However, many issues with voice user interfaces (VUI) remain, such as the fact that design methods have not been established. Going forward, NTT DOCOMO will take initiatives to solve these issues to promote usage of the docomo AI Agent API.

Lastly, NTT DOCOMO planned the concept of this AI agent initiative around June 2016. At the time, this in-house project was called “Sebastien” and was developed by a mere seven people. However, having gained cooperation from many others, the project was included in the medium-term strategy and became a company-wide initiative in less than a year. Just as the name “Partner Initiative” suggests, in the future we expect these initiatives to bring about a wide range of solutions through cooperation with the people in a wide range of our partner companies.

REFERENCES

[1] NTT DOCOMO Press Release: “DOCOMO Announces docomo AI Agent Open Partner Initiative,” Jun. 2017. https://www.nttdocomo.co.jp/english/information/media_center/pr/2017/0623_00.html

[2] Device WebAPI Consortium Web site. https://en.device-webapi.org/

[3] SEBASTIEN Web site (In Japanese). https://docs.sebastien.ai/

Technology Reports | AI Agent | Natural Language Processing Technology | Voice Recognition/Synthesis

Special Articles on New AI Agent

Multipurpose Dialogue Engine for Converting Natural Language into Instructions to Enable Services

Innovation Management Department Tetsuo Sumiya Koji Yamazaki
Service Innovation Department Noriyoshi Kamado Go Tanaka
Service Design Department Kosuke Kadono

Spoken words regularly used by people in daily life often entail ambiguous expressions. The Multipurpose dialogue engine has functions to analyze and interpret such diverse and unclear natural language and achieve services. It consists of a natural dialogue platform, a voice recognition function, a speech synthesis function, a service platform front-end, a dashboard for users and a dashboard for developers. This article provides a general overview of the multipurpose dialogue engine and describes its components.

1. Introduction

As well as having voice recognition/synthesis and natural language processing capabilities, multipurpose dialogue engine (**Figure 1**) is characterized by its ability to link external services through the Application Program Interface (API)*1. Thus, it enables users to have a wide range of new experiences through dialogue with users on various devices, content provision, and easy mounting of agents that operate devices.

Since launching voice agent services such as the "Shabette Concier" in 2012, the "natural dialogue engine" in 2015, and "Oshaberi Robot for biz" in 2016, NTT DOCOMO has built a log containing more than three billion user utterances from more than eight million people to improve the dialogue performance of its natural language processing

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 API: An interface that enables software functions to be used by another program.

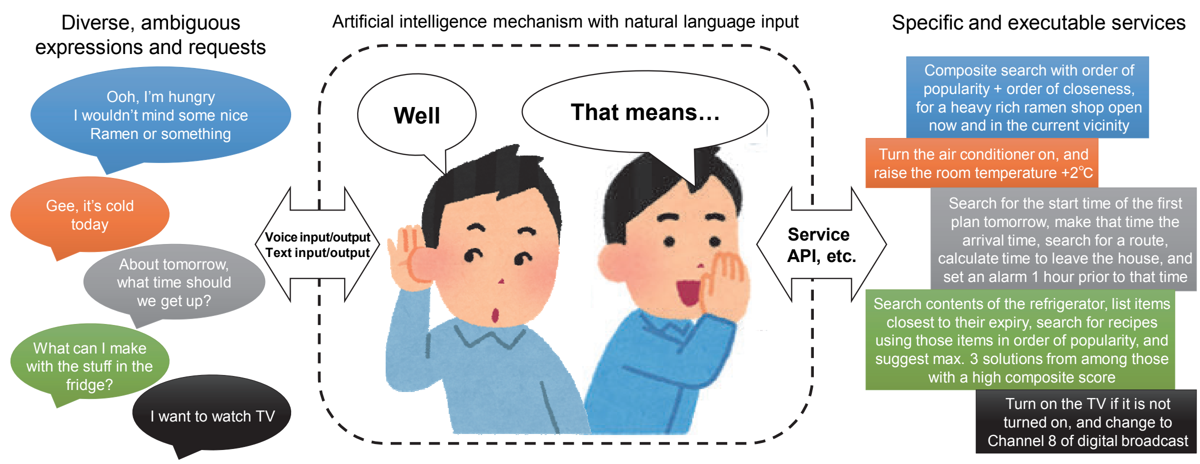


Figure 1　Multipurpose dialogue engine overview

technology. The multipurpose dialogue engine is the result of this six-year accumulation of know-how.

2. Multipurpose Dialogue Engine Technology

2.1 Multipurpose Dialogue Engine System Architecture

Multipurpose dialogue engine consists of the Service Platform Front-end (SPF), Automatic Speech Recognition (ASR), Natural Language Understanding (NLU) platform, a Text-To-Speech (TTS) speech synthesis[*2], a User DashBoard (UDS), and a Developer Dashboard (DDS). **Figure 2** shows the overall system flow of dialogue.

- Voice recognition from user utterance

Normally with voice recognition, devices include a Software Development Kit (SDK)[*3]. According to user utterances, the device sends authentication tokens acquired in advance and either voice or text to the engine. At first, the SPF receives all user requests and verifies the authentication token. After the authentication token is verified, voice data is sent to ASR, and the recognition result response is given in text format.

- Linking external services from natural language processing

NLU processes the recognition result with natural language processing and then recognizes tasks. According to the recognized tasks, NLU links to external services, and retrieves the information required by the user. For example, for a recognition result

*2　**Speech synthesis:** Technology for artificially creating speech data from text and verbally reading out text.

*3　**SDK:** A set of documents, tools, libraries, sample programs, etc. needed to create applications.

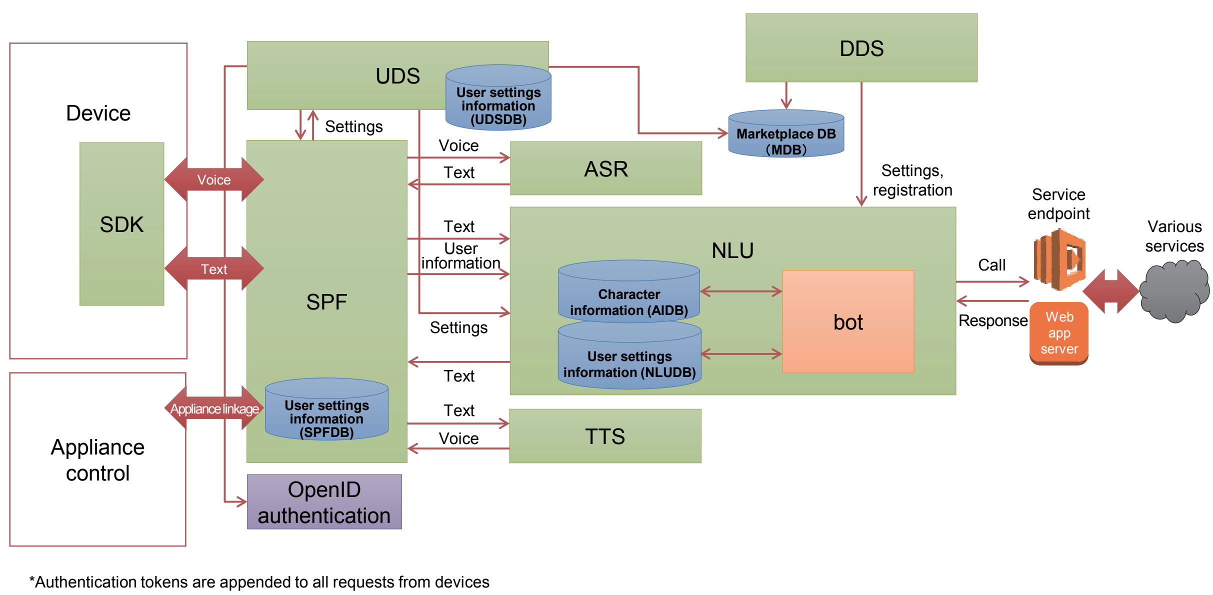


Figure 2　Multipurpose dialogue engine system configuration

for "what's today's weather?", NLU determines that it should link to a weather forecasting service, makes a request using an API for acquiring weather forecast information, and based on the response to the request, creates a text response for the user.

- Responding to users with speech synthesis

  The response created for the user is sent to TTS, which converts the text into voice data using a voice model specified in advance. The created voice data is then sent to the device through SPF to provide a voice response to the user's inquiry. The device SDK and SPF maintain a WebSocket connection from the user's initial utterance through to the end of the conversation, which enables high-speed responses regardless of the fact that voice response data is larger than text and takes longer to load.

2.2 SPF

As mentioned, for continuous dialogue requests including voice and text from SDKs, the SPF has functions to generate appropriate responses in real time for each request by combining responses from the backend engine consisting of ASR, NLU and TTS connected at the latter stage to respond to SDKs. This software architecture is shown in **Figure 3**.

The SPF consists of two types of processing units called "Block" and "Edge". Block performs the processing required to generate a response to

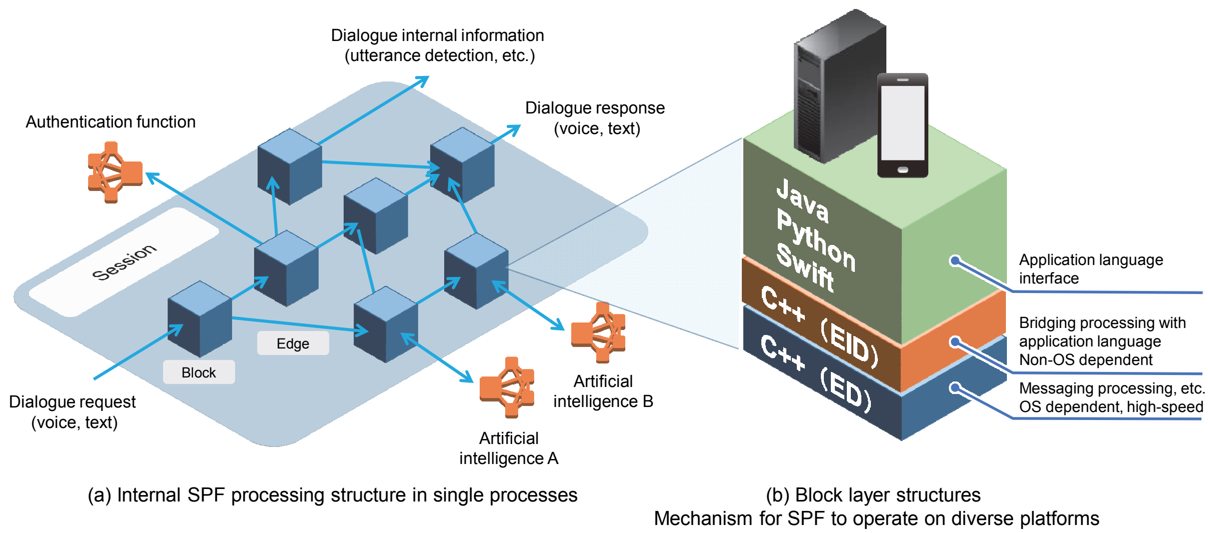


Figure 3 SPF software architecture

a request sent from a SDK, while Edge passes data between Blocks. By combining the results of multiple Blocks with Edge, the appropriate response can be returned to the SDK. This architecture increases the flexibility to change the processing within the SPF and allows an ever-evolving spoken dialogue technology to be quickly incorporated into the SPF.

Also, by separating processes as Block and Edge, it's possible to aggregate data shared between the Blocks on the Edge and raises the level of parallel processing. This enables processing of not only text but also enables real-time streaming*4, which requires parallel processing for large amounts of data such as voice.

The software layer structure of the Blocks is described in Fig. 3 (b). In SPF, the speed of Environment Dependent (ED) processing that requires high-speed and OS optimization is increased with C++ (and other languages which are dependent on the environment).

Also in SPF, through an environment and programming language non-dependence layer (the Environment-InDependent (EID) layer) wrapping the layer described above, it's possible to describe block processing in diverse programming languages. In this way, using SPF as the medium between programs written in different languages enables short service development time.

Using device SDKs with this architecture also achieves high speed and stable voice dialogue processing not only in server environments but also

*4 Streaming: A communication method for sending and receiving audio and video data over the network, whereby data is received and played back simultaneously.

in environments such as smartphones and embedded Linux devices*5.

2.3 NLU

NLU is part of the dialogue system that achieves the agents using a description language called xAIML. xAIML is a description language for creating an artificial intelligent dialogue agent on the NLU, and is based on AIML1.1 but with functionality expanded by utilizing NTT DOCOMO's natural language processing technologies*6.

With xAIML, when creating a single dialogue agent, it's possible to develop efficient systems by designing the respective functions of the dialogue agent as bots which can be linked to achieve various functions in the main agents*7 and expert agents*8 in the multipurpose dialogue engine.

1) Design Patterns for Creating Main Agents

NLU is a platform that can be freely customized and incorporated into services and products by developers to develop dialogue services and products able to converse with people, and thus enables quick and easy development of ideal dialogue agents by free combination of the components useful to dialogue system development. The multipurpose dialogue engine is able to create and link its two types of main and expert dialogue agents using this NLU.

A main agent is a main character in the dialogue system, and is the agent that controls dialogue with the user. When the user speaks to an agent (hereinafter referred to as "user utterance"), the main agent to connect is determined on identification of the user. Then, the user's intentions are interpreted*9 with the main agent from the content of utterances, and if tasks are executable in the main agent, they are executed. On the other hand, if the user makes an utterance such as "d Gourmet please" to call a particular expert agent, the user utterance is passed to the expert agent, and the subsequent processing and dialogue with the user are handled by that expert agent. In this way, the roles of main and expert agents are clearly defined, yet can be freely customized.

It's also possible to give priority to group scenarios such dialogue for tasks, calling expert agents, making commands or chatting so that scenarios can be prioritized for individual services. Partner companies can design and develop main agents their preferences. They cannot only freely combine the above scenarios but also create original main agents.

Expert agents are specialized for particular fields, and are called from the main agent. Anyone can easily design these using DDS, and release them onto the designated AI agent API marketplace once the developed expert agents have been screened. Then, users can freely add or delete released expert agents that they want to try, which enables users to individually manage their own expert agents with the UDS.

2) Achieving Expert Agents through a Connected bot Structure

Expert agents in the multipurpose dialogue engine are generated as independent bots using DDS. Users can freely select and enable expert agents through UDS to use the following expert agent-related functions.

- Input a specific word into the main agent to call an expert agent so that it can take over.
- User input is transferred to the called expert agent until dialogue with that expert

*5 Embedded Linux device: A kind of special-purpose device that is incorporated into mobile data terminals, digital appliances and other such products, and that has a CPU and software that runs on the Linux OS.

*6 Natural language processing technology: Technology to process the language ordinarily used by humans (natural language) on a computer.

*7 Main agent: The agent at the forefront of dialogue with users. Service providers can create main agents with the characters they like, and use them to connect services with devices.

agent finishes.

- Outputs from expert agents that are against public order or morals can be filtered, and the attempted output sent to the administrative module of DDS.
- System administrators or expert agent developers can suspend the functions of an expert agent.

In the multipurpose dialogue engine, the above capabilities are achieved by designing functions to manage expert agents of individual users, pass user input to a bot in a dialogue with the user, filter inappropriate language and manage expert agent status as independent bots (**Figure 4**).

Also, the main agents created using the multipurpose dialogue engine can easily implement the functions to offer expert agents by using these bots.

2.4 UDS

UDS enables settings for user device authentication and expert agents. Users can use both GUI*10 and REST API*11 to authenticate devices.

Figure 5 shows an image of UDS in the trial environment. UDS offers the following functions.

- Authentication: The dashboard supports both d Account authentication with OpenID Connect (OIDC)*12 and Google*13 authentication.
- Device list/new registration: Users can register device IDs mapped to devices and map

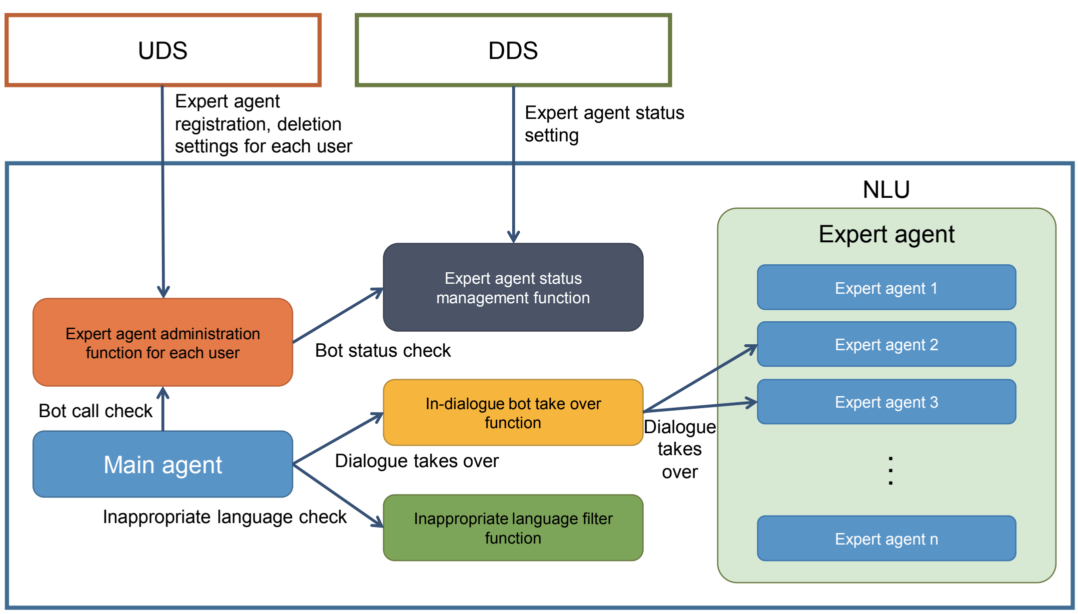


Figure 4 Bot structure to achieve expert agents

*8 **Expert agent:** An agent specialized for a service and called by a main agent. Service providers can freely provide existing main agents (robots, bots, etc.) with services.

*9 **Intention interpretation:** Technology that uses machine learning and so forth to determine the user's intention from the user's utterances (natural language). User intentions are called "tasks." For example, all the utterances "What's tomorrow's weather?," "I wonder if tomorrow will be fine?," and "Is it going to rain tomorrow?" are judged as weather tasks.

*10 **GUI:** A superior type of interface that offers visibility and intuitive operability by expressing operations and objects visually on a screen.

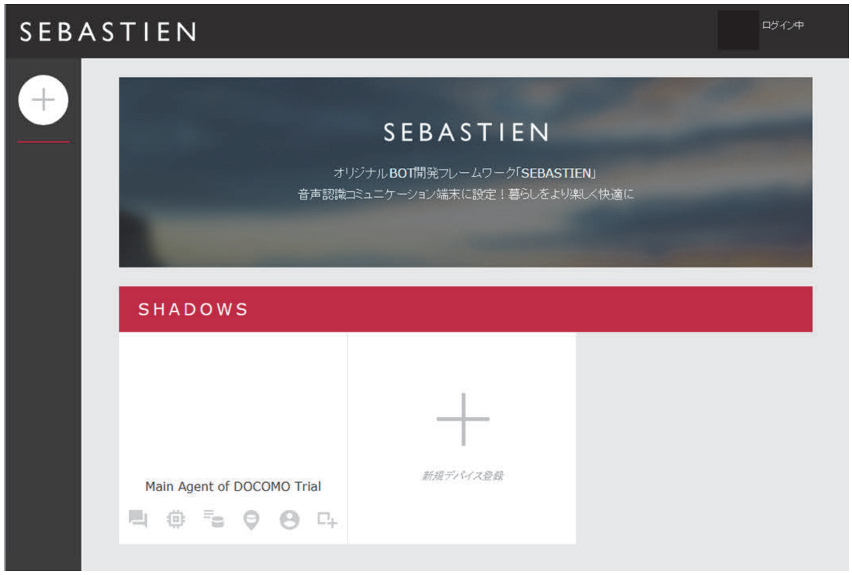


Figure 5　User dashboard image

them to users. Users can display a list of currently registered devices and delete devices if required.

- Dialogue history: Users can display a history of dialogue with main and expert agents.
- Home device linkage: By linking to IoT access control engine, users can make settings to perform administrative tasks such as registering or tagging linked appliances from UDS or operate appliances from expert agents.
- Expert agent configuration: This enables registration and listing of expert agents with DDS. It's necessary to register expert agents with DDS in advance to use them. If an expert agent needs to link an account, OAuth 2.0 authorization is required when registering the expert agent.

2.5 DDS

DDS offers expert agent creation functions [2].

1) Design Concept Behind the Expert Agent Creation Function Provided with DDS

The final objective of an expert agent is to provide a specific service to the user. Thus, dialogue design is for that purpose.

Also, dialogue design does not require any specialist knowledge about natural language processing or dialogue design - tools have been designed to enable the process to be completed just by entering setting values on a screen, which enables developers who have never created a dialogue service to create agents simply and quickly.

2) Expert Agent Dialogue Design

Expert agents created with DDS adopt task-oriented dialogue design with slot filling*14 suitable for service provision. The specific dialogue sequence

---

*11　REST API: An API conforming to REST. REST is a style of software architecture developed based on design principles proposed by Roy Fielding in 2000.

*12　OIDC: A mechanism to link IDs so that authentication can be done with one ID for using various Web sites on the Internet and mobile applications, etc.

*13　Google: A trademark or registered trademark of Google LLC.

*14　Slot filling: A function to extract required parameters from text when executing tasks. For example, to execute a weather search task, the "time" and "place" parameters of "today" and "Tokyo" are extracted from the text "What's the weather in Tokyo today?".

is as follows.

(1) The expert agent extracts Intent*15 and parameters (slots) from user utterances.

(2) If the user makes an ambiguous utterance, the expert agent asks the user again to confirm the user's intent before executing a task.

(3) Finally, the specific intent information (intent, slot) is sent to a set external program (API) with a POST request*16.

(4) Based on the received intent information, the external program executes a service.

(5) The external program returns a response (dialogue wording) for the user according to the results of executing the service.

(6) The expert agent gives the response from the program to the user.

3. Conclusion

This article has described the multipurpose dialogue engine which interprets natural language including diverse and ambiguous expressions uttered by users, and converts them to specific executable services. With the concept of main and expert agents, this system offers users with new experiences of dialogue with various devices, content provision and device operation. Going forward, we would like to engage in research and development such as common dialogue technologies required across multiple platforms and advanced intent interpretation technologies in various fields, and take initiatives to develop agent generation technologies to enable high-level dialogue and advance APIs.

REFERENCES

[1] SEBASTIEN Web site (In Japanese).
https://users.sebastien.ai/

[2] Expert Agent Developer Dashboard Web site (In Japanese).
https://developers.sebastien.ai/

*15 Intent: Refers to tasks defined in expert agents.

*16 POST request: A type of request sent to a Web server from a client (such as a Web browser) with HTTP communications, for sending data from a client to a program etc. specified with a URL. While requests such as GED or HEAD are only headers, POST requests have body sections in which the data desired for sending is described. This is used for sending large amounts of data or files to servers.

Technology Reports

AI Agent | Prediction | Personalization

Special Articles on New AI Agent

Proactive Support Engine to Actively Support User Activities

Service Innovation Department Naoharu Yamada Masato Yamada Norihiro Katsumaru Kyo Hattori

Human Resources Management Department Masahiro Inoue

R&D Strategy Department Yukinobu Takehana

Accurately understanding user activities and characteristics is of paramount importance in achieving agent services to actively support users in their daily lives. Smartphones contain a lot of personal data such as location information, schedules, email and images, and it's possible to accurately understand user activities by analyzing this data. For this reason, NTT DOCOMO has developed a proactive support engine as a platform for integrated analysis of a wide range of personal data. This article describes an overview of the engine and its various functions.

1. Introduction

In recent years many companies have been providing assistance services, including services such as Google Assistant™*1, Apple's Siri®*2, DOCOMO's Shabette Concierge and Microsoft Cortana*3.

Currently, these assistance services are evolving from passive types that respond to user inquiries to active types that provide support by proactively interacting with users. Conventional passive assistance services cannot provide suitable support if the user does not make explicit inquiries. In contrast, the active types can sense trouble that the user doesn't know about such as incoming rain squalls or train delays, and use push notifications and so forth to convey them to the user, hence enabling more active support.

It's crucial to predict the user's profile and subsequent activities using the personal data on their smartphone to achieve these proactive assistance

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 Google Assistant™: A trademark or registered trademark of Google LLC.

*2 Siri®: A trademark or registered trademark of Apple Inc. in the United States and other countries.

*3 Cortana: A trademark or registered trademark of Microsoft Corporation in the United States and other countries.

services. These days nearly everybody carries around a smartphone for a wide range of purposes such as managing their schedules, communicating with family and friends, shopping online, taking photographs and videos and searching for travel routes. Therefore, accurately understanding user activities and characteristics based on the location, schedule, mail and image data acquired and accumulated in those processes will enable suitable support provision.

To date, to understand user activities and characteristics, there has been much research on estimating the user's residence and place of work and predicting subsequent destinations from smartphone location information. However, using location information to predict destinations is based on history of past movements, and only enables prediction of places that are frequently visited - places like beauty clinics that are not often visited or places which are being visited for the first time are fundamentally impossible to predict.

For this reason, NTT DOCOMO developed a proactive support engine that accurately estimates user activities from multiple types of personal data such as schedules, email and images in addition to location information. This has enabled expansion of support by expanding what can be estimated and increasing variety. For example, although predicting destination using location information alone only enables prediction of places that the user has already been, it's possible to also determine and predict places not regularly visited or places visited for the first time, such as those involved with travel, by analyzing schedule and email data.

This article describes an overview of the proactive support engine, and functions comprising the engine.

2. Proactive Support Engine

The proactive support engine analyzes and uses the personal data on a user's smartphone to provide individualized information and services to the user with optimal timing.

Personal data includes location information, schedule information, email and images. Using this information, the proactive support engine can estimate current activities, predict future activities and estimate a profile of the user (static characteristics). In addition, the personalized push function enables extraction of optimal contents based on the results of these estimations and push delivery of such content (**Figure 1**).

Following describes the location analysis, schedule analysis, email analysis, image analysis and personalized push functions.

2.1 Location Analysis Function

It's possible to understand when and where users have been staying and where they have gone by periodically obtaining location information from their smartphone, since they always carry it around. From this periodically acquired location information, the proactive support engine is able to estimate current activity, predict future activity and estimate a user profile.

1) Current Activity Estimation

Based on periodically acquired location information, the system can judge whether the user is stationary or on the move.

If the user is stationary, the system determines the facility they are in from their stationary location.

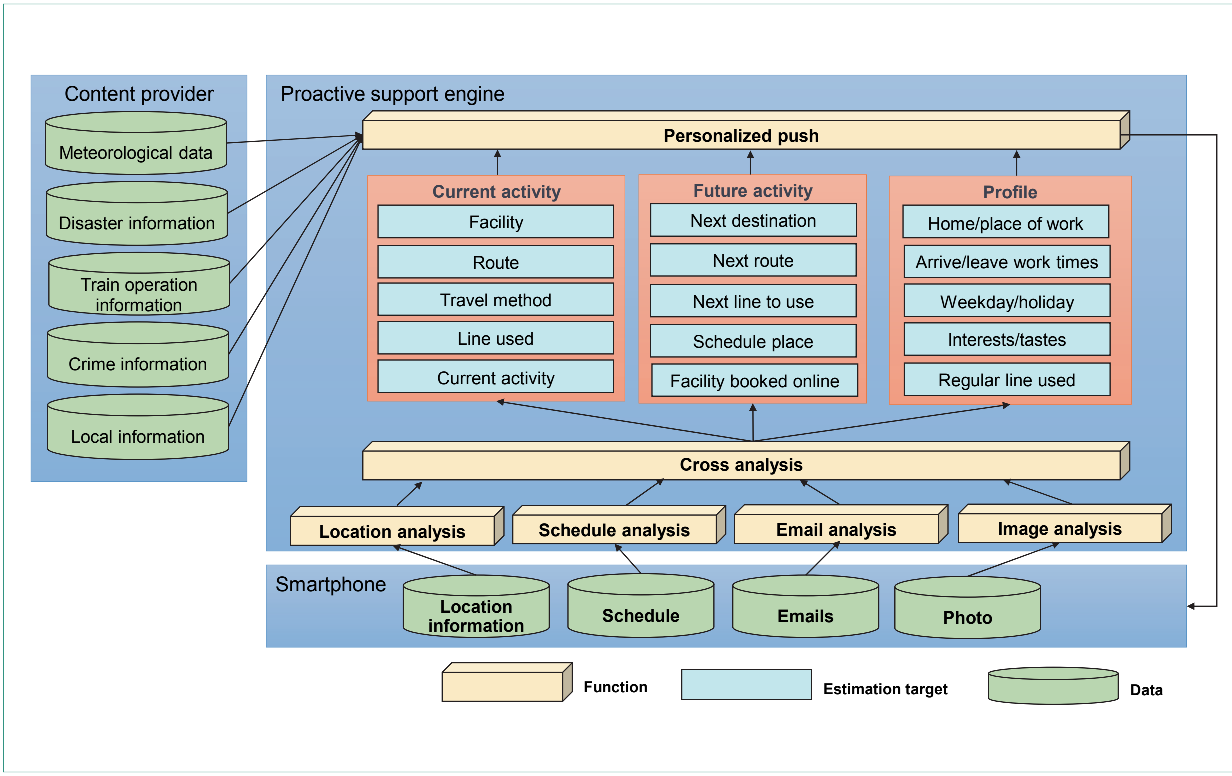


Figure 1　Proactive support engine function overview

However, in facilities such as golf courses that are spatially large, it's not possible to correctly extract the facility by simply extracting the center of the facility closest to the user's stationary location. To counter this issue, this system uses machine learning*4 models that use the positional relationship between the user and the boundary (shape) of the facility grounds or building, etc. as feature values*5 to estimate the facility that the user is visiting. **Figure 2** describes an image of estimating the facility a user is visiting. In this example, although the facility central position of Facility D is the closest to the stationary location of the user, it's possible to estimate that the facility being visited is Facility A when considering the facility boundary.

If the user is moving, the system determines the route they are traveling from the acquired serial location information. However, to reduce smartphone power consumption, the location information acquisition time interval is often lengthened, which makes the location information during movement sparser, making it more difficult to accurately determine the user's route. This system improves the accuracy of route estimation by using location information acquired from the user's repeated passing of the same route. Specifically, the route is first estimated from location information acquired during movement. This is then compared to past route estimations, and a judgment is made about whether the route is the same as one from the past. Thus, accuracy is improved by integrating the user's route with the location information for a past route if

*4　Machine learning: A mechanism allowing a computer to learn the relationship between inputs and outputs, through statistical processing of example data.

*5　Feature value: Values extracted from data, and given to that data to give it features.

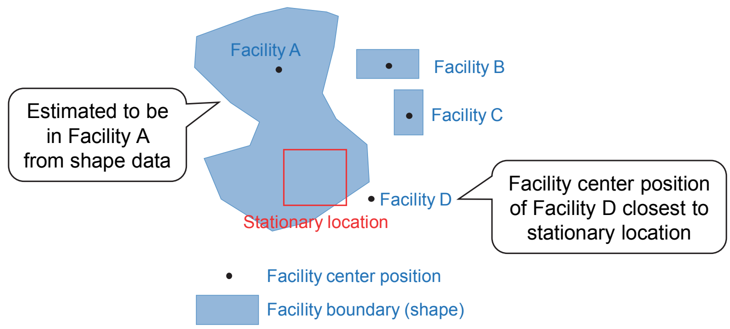


Figure 2 User's visited facility estimation image

it exists. Then, the system can judge whether the user is on a train, in a car or walking from the estimated route, location information for train stations and the user's speed of movement, etc. If the user is traveling on a train, the name of the line traveled, the station name where the user boards, the station name where the user alights and station names where the user changes trains can be determined.

2) Future Activity Prediction

As the user's past movement history, the proactive support engine retains routes and the number of times the user took them, on what days and in what time slots. Then, based on the current time and current location, the system predicts subsequent visit destinations, the route to get there, whether by train, car or on foot, and which train line to use, etc., from movement history information for the same time slot and same locations.

3) Profile Estimation

The system calculates the number of days and the length of time the user was in locations based on daily location information, and from machine learning that uses those feature values, estimates the position of the user's residence and place of work. Then, based on the time that the user leaves for work every day, the system estimates the days that the user goes to work, the times that the user leaves home by the day, the time that they arrive at work, the time that they leave work, and the time that they arrive home. The system also determines places that the user often stays, and train lines that are often used from stationary history and route history. From the history of facilities that they have visited, the system can estimate the user's interests and tastes. For example, if the user is often in art galleries, then it's most likely the user has an interest in art.

2.2 Schedule Analysis Function

Users often register irregular event information such as travel plans, beauty salon appointments and hospital visits in a schedule application to avoid forgetting them. Data for these scheduled events often includes text indicating their location, which can be determined to enable prediction of the user's activities.

The text that represents these scheduled locations includes entries for addresses and entries for Point Of Interest (POI)*6 information such as "Tokyo Tower." While address expressions can be extracted using regular expressions*7, the exact POI name cannot be obtained from POI information with simple morphological analysis*8, for example with "Tokyo Tower," only the location name "Tokyo" is extracted.

Therefore, to accurately obtain POI names, the proactive support engine extracts them using a sequence labeling*9 method called Conditional Random Fields (CRF)*10. This entails attaching labels to morphemes indicating location names in the schedule data. **Table 1** describes this labeling. For a schedule entry called "lunch at Tokyo Tower," location labels "B-LOC" and "I-LOC" are attached manually to "Tokyo Tower" which indicates the location. Manual labeling of large amounts of scheduled data and learning with CRF enables high accuracy extraction of location names. Evaluating this technology with actual data shows that POI names can be searched with 90% accuracy.

2.3 Email Analysis Function

Recently, online opportunities to book hotels, high-speed rail tickets, restaurants, lessons and so forth and purchase a wide range of products have been on the rise. In such cases, email is sent to the user to confirm a booking or purchase. By analyzing these, it's possible to determine when and where users are going, and what users have purchased and at what price.

Mail analysis is performed as a combination of a rule base and machine learning.

1) Information Extracted with the Rule Base

Confirmation emails for purchases and bookings use the unique formats of individual companies and are sent to users automatically by machines. Here, the system extracts booking and purchase information from email for each sender mail address based on predetermined rules. The rules describe which character strings are to be extracted from the character strings in the emails. For example, for a booking confirmation email from a certain company assumed to be in the format "Reservation number: ABC123," a predefined rule to extract "ABC123," the character string after "Reservation number:" for this email enables the proactive support engine to accurately extract the reservation number.

2) Information Extracted with Machine Learning

An issue with extracting information using a rule base is that rules must be created manually,

Table 1 Image of labeling location names

Morpheme	Part of speech	Label
lunch	Common noun	Other
at	Preposition	Other
Tokyo	Proper noun, location name	B-LOC
Tower	Common noun	I-LOC

*6 POI: Refers to a shop or facility.

*7 Regular expression: A method of expressing a collection of character strings as a single character string.

*8 Morphological analysis: The process of dividing a sentence into a sequence of morphemes, which are the smallest units of meaning.

*9 Sequence labeling: A method for automatically applying appropriate labels to members of sequences such as character strings, based on decision criteria obtained through statistical processing of examples.

*10 CRF: A method of assigning pre-defined labels to a sequence of input entities based on the feature values of the entities.

which means it's practically impossible to create them for all Internet Web sites. To solve this, this system applies a machine learning approach using learning data from reservation mail collected based on the rule base.

Specifically, this first entails categorizing emails. There are ten types of these categories, including net shopping purchase confirmations, movie reservations, air ticket reservations, and coupons, etc. Morphemes in the email body are analyzed to create a vocabulary, and categories are created through multi-class logistic regression*11.

Next, required items in categories are defined (for example the title, movie date and time, name of cinema, and reserved seat number, etc. in a movie reservation confirmation email) and information related to each item is extracted using CRF. Specifically, "K" is attached to items to extract, "V" is attached to values to extract, and "S" is attached to the symbol separating these for the character string undergoing the morpheme analysis. As an example, **Table 2** describes an image of CRF labeling of the character string "Reservation Reception Date:12/17/2014 (Tuesday)." In this example the label "K" is attached to "Reservation Reception Date," "S" is attached to ":" and "V" is attached to "12/17/2014 (Tuesday)." Labeling large amounts of email data in this way and learning with CRF enables extraction of information from email. Evaluating this technology with actual data shows that information can be extracted with 90% accuracy with *F* values*12.

Table 2 Image of labeling information extracted with machine learning

Morpheme	Part of speech	Label
·	Symbol	O
Reservation	Noun	K
Reception	Noun	K
Date	Noun	K
:	Symbol	S
2014	Noun	V
–	Symbol	V
12	Noun	V
–	Symbol	V
17	Noun	V
(	Symbol	V
Tuesday	Noun	V
)	Symbol	V

2.4 Image Recognition Function

Images held by users often give clues to their interests and tastes. For example, there is a high possibility that dog owners will take a lot of pictures of their dog, and keen travelers will take a lot of pictures on their overseas trips. Thus, estimating the scenes and events depicted in user images is directly connected to estimating the user's interests and tastes.

The proactive support engine uses two approaches to estimate scenes and events from images. One of these entails estimation using a single real image, while the other entails estimation using multiple images.

1) Estimating Scenes and Events Using Single Real Image

To estimates scenes and events using a single real image, a convolution neural network*13, which is a deep learning method, is used. This method requires processing to learn data, and it's accuracy

---

*11 **Multi-class logistic regression:** Logistic regression is a type of statistical regression model for variables that follows a Bernoulli distribution. While logistic regression entails binary classification, multi-class logistic regression is a method which expands logistic regression for multi-level classification.

*12 ***F*** **value:** A scale used for comprehensive evaluation of accuracy and exhaustiveness, and it is calculated as the harmonic mean of precision and recall.

is heavily dependent on whether enough good quality learning data has been prepared. However, much of the image data available on the Web was captured by professional camerapersons, and the image trends in these photographs often differ from photographs captured by ordinary users. Therefore, NTT DOCOMO has improved accuracy by using images captured by users as learning data.

Figure 3 describes accuracy with only images available on the Web and user data as learning data. In the graph, both a high precision*14 and recall*15 can be seen with user data as learning data.

2) Estimating Scenes and Events Using Multiple Real Images

Estimating scenes and events which are difficult to understand from a single real image can be estimated using multiple real images. For example, let's assume that multiple images were captured during a visit to Disneyland*16. While it's only possible to attach the scene "dining" to a captured image of a dining scene, it's possible to attach the event "Disneyland" by collectively analyzing multiple images captured at Disneyland.

The following procedures are performed to analyze multiple real images. First, using image metadata such as capture date and time and capture location, the images are grouped. Then, scenes and events are estimated for the groups of images.

(1) Grouping entails making collections of photographs with generally close capture locations and capture times as the photographs of the same event, from among a series of user photographs. Estimation by grouping enables event estimation from clues from other photos in the group even if it's not possible to estimate the event from one photo.

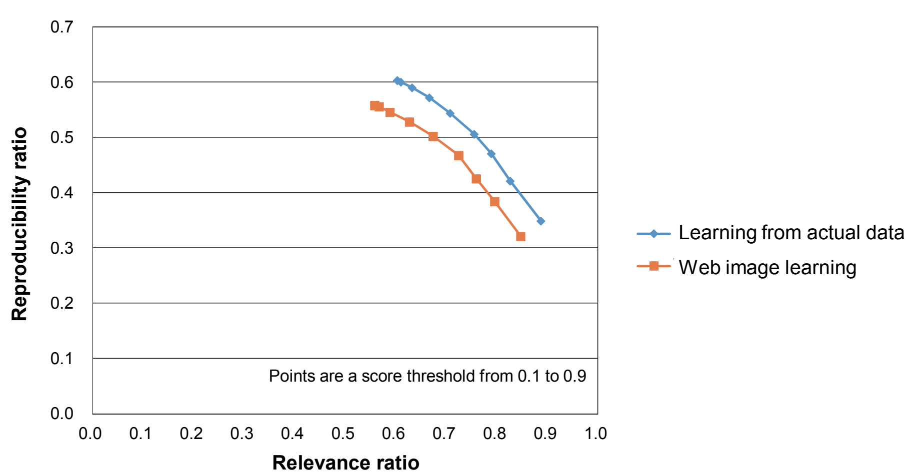


Figure 3 Accuracy using only Web images and user data (real data)

---

*13 Convolution neural network: A type of deep learning method that demonstrates superior performance particularly in the field of image recognition, and entails the inclusion of layers with characteristic functions such as convolution layers and pooling layers in a neural network consisting of several deep layers.

*14 Precision: An index that can express the accuracy of estimation results, but cannot express the comprehensiveness of estimation results.

*15 Recall: Expresses comprehensiveness as a lack of leakage with estimation results, but cannot express precision of estimation results.

(2) Event estimation entails estimation of possible events for the entire group using the results of image recognition, capture location and date and time information for the grouped multiple photographs. For example, if there are many photographs of animals that have capture locations in the vicinity of a zoo, then it can be estimated that the photography took place at a zoo.

2.5 Personalized Push Function

The personalized push function delivers content optimized for users with optimal timing by combining the estimation results described above with content information such as weather, transport operations and events.

To achieve personalized push, first of all ECA rules*17 are created to describe push delivery conditions. ECA stands for “Event Condition Action.” Event describes the time for push delivery, Condition describes who the delivery is for, and Action describes what is to be delivered. For example, the case of alerting users of impending rain will be Event: probability of rain exceeds 50%, Condition: people in the area where the probability of rain exceeds 50%, and Action: “It looks like it’s going to rain.” This system enables new push delivery simply by adding ECA rules.

The ECA rule Event is periodically checked for matching conditions. Thus, if conditions match, users to which Condition applies are extracted, and the message described in Action is delivered with push.

3. Conclusion

This article has described NTT DOCOMO’s proactive support engine - a platform for analysis of personal data. Understanding user activities and profiles from personal data is indispensable in achieving assistance services that actively support users.

The proactive support engine analyzes smartphone data including location information, schedule, email, and image data to enable estimation of the user’s current activities, future activities and profile.

Smartphones also contain a wealth of personal data other than that described above, such as application usage logs and terminal sensor data. By expanding personal data targeted for analysis, we plan to improve the accuracy of estimation of user activities and profile, and take initiatives to estimate new profile aspects such as life events.

*16 Disneyland: A trademark or registered trademark of DISNEY ENTERPRISES, INC.

*17 ECA rules: A prescription consisting of a combination of events, conditions and activities for a processing method.

Technology Reports

AI Agent | Device WebAPI | IoT

Special Articles on New AI Agent

IoT Access Control Engine Enabling Control of Various IoT Devices from the Cloud

Service Innovation Department Takashi Yoshikawa Takafumi Yamazoe
Ken Yamashita Yuki Tanaka Shoichi Horiguchi

Although many IoT devices have been released onto the market, the interconnectivity between them is still not very good. NTT DOCOMO developed a "Device WebAPI" as technology for easily connecting such devices and an "IoT access control engine" to handle them through the cloud. This article describes the circumstances leading to these developments, characteristics of the technologies and service examples.

1. Introduction

In recent years, the Internet of Things (IoT) is a word that has begun appearing in a range of situations.

It is about 30 years since Mark Weiser first advocated ubiquitous computing [1]. In that time, differing from the so-called pervasive computing [2], wearable computing [3], mobile computing, and Zen computing, the concept of "ubiquitous computers blending into kinds of places to enrich the lives of people" has been successively continued as a fundamental aim of researchers.

Although various technical issues surrounding IoT have been raised such as data amounts, communications networks, security, data analysis technologies, and costs [4], NTT DOCOMO has taken a particular focus on interconnectivity, and is proceeding with research and development in that area.

To build applications and services using IoT devices, developments have to be done according to

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

each device. For this reason, device manufacturers respond to these needs by distributing software libraries and Software Development Kits (SDK)*1 in an effort to facilitate easier development. However, since the need to learn about the unique implementation and individual device technologies remains, there has not been much reduction of the workload on developers. Another issue is the difficulty of both libraries and SDKs to support all operating systems and development environments. Furthermore, implementation methods differ for each device and manufacturer, which makes it difficult to support applications and services with different devices once they have been developed - there have already been many cases of redevelopment.

These affect the interconnectivity of IoT devices, which is one issue holding back the rapid spread of IoT applications and services. To solve this issue, NTT DOCOMO abstracted a wide range of IoT devices at the functional level, and developed a "Device WebAPI" to enable access with common RESTful*2 WebAPI*3, and developed an "IoT access control engine" as a cloud-based platform enabling unified handling of IoT devices.

This article describes the technical characteristics of the Device WebAPI and the IoT access control engine, and describes examples of their use with the AI agent platform.

2. Device WebAPI and IoT Access Control Engine

2.1 Overview

One of the values of IoT is that it enables visualization of never-before-seen data, which can lead to the discovery of new value through analysis and hence new solutions. In other words, the real value with IoT comes from interacting with various types of devices and data, which means unified handling of these devices and data is indispensable in spreading IoT services.

However currently, a wide range of IoT device standards exist both in Japan and around the world, and many manufacturers use proprietary specifications for their IoT devices. Therefore, IoT service developers have to understand the specifications and build source code for various devices, which is also a huge impediment to the spread of IoT services. To address this issue, NTT DOCOMO developed its Device WebAPI and IoT access control engine.

The Device WebAPI is an interface abstraction technology that achieves (1) a common device access method using RESTful WebAPI, (2) device abstractions at the functional level, and (3) high versatility and expandability based on plug-in architecture. These three characteristics can help to solve the issue of IoT interconnectivity [5]. This technology has been standardized as a GotAPI by the Open Mobile Alliance (OMA)*4 [6].

Also, the IoT access control engine is a cloud platform that enables remote access to the Device WebAPI, so that in addition to the Device WebAPI technical characteristics, it also provides (4) unified IoT device remote management and (5) multiple permissions*5 management functions.

The IoT access control engine is also one of the engines in the AI agent platform, and can interact with other engines to enable remote control of IoT devices from multipurpose dialogue engine and

*1 SDK: A set of tools and technical documentation required for developing software.

*2 RESTful: The idea of obtaining/providing information by directly pointing to the information to be provided in a stateless manner.

*3 WebAPI: An HTTP-based API.

*4 OMA: An industry standardization organization that aims to standardize service and application technology and achieve interoperability in mobile communications.

*5 Permission: The right of access to a system. In this article, this refers to access rights set in an API to access IoT devices.

proactive support engine and sensor information collection.

Figure 1 shows an overview of the system architecture using the IoT access control engine. An IoT service application accesses the IoT access control engine using a prescribed Application Programming Interface (API) to uniformly authenticate users, control various types of IoT devices and reference accumulated data, etc.

2.2 IoT Access Control Engine API

There are currently three types of API available for the IoT access control engine, as shown below.

(1) Management API: The API for authenticating and authorizing users. It also enables device registration and creation of common users. It is OAuth 2.0*6 compliant.

(2) Device API: The API for controlling IoT devices. Achieves the same actions with a common interface.

(3) Archive API: The API for acquiring data accumulated from IoT devices.

2.3 Home Gateway

To control IoT devices using short-range radio systems such as Bluetooth®*7 and Wi-Fi®*8, a home gateway is required to mediate between the IoT access control engine and IoT devices (Fig. 1).

Agent software using the Device WebAPI operates in the home gateway. Agent software currently supports the Android™*9 and Node.js*10 platforms, which means an Android smartphone can be used as the home gateway.

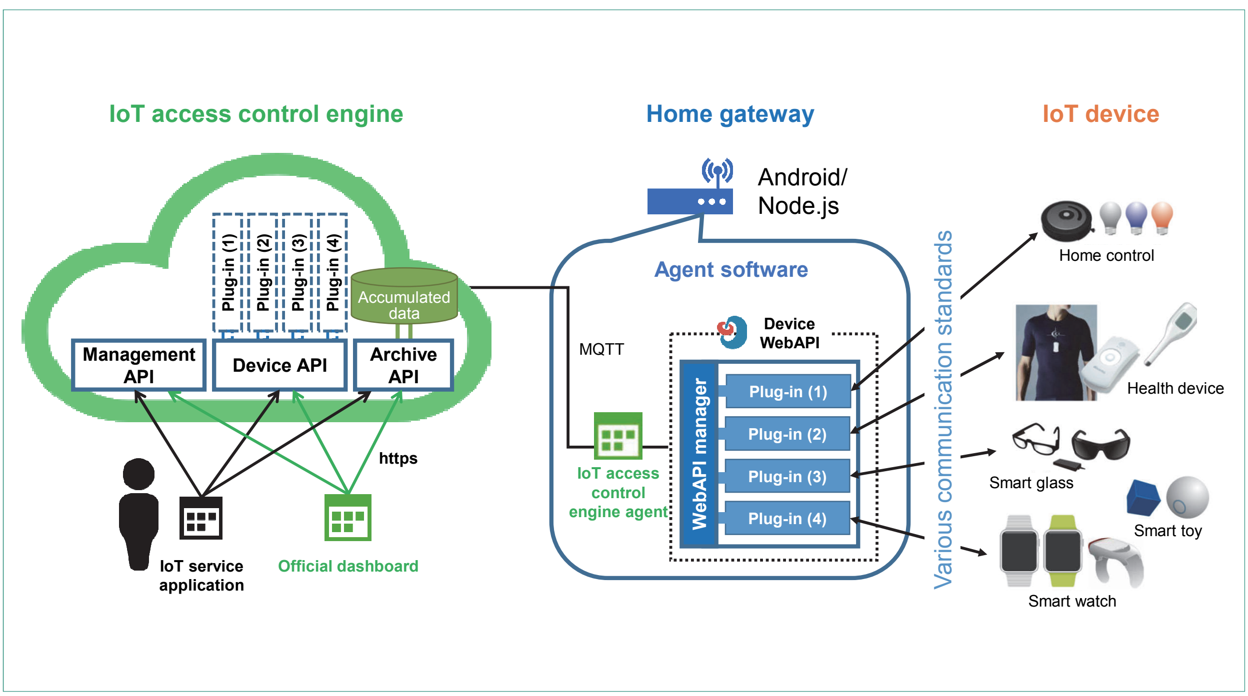


Figure 1 Schematic of the system using IoT access control engine

*6 OAuth 2.0: A mechanism to authorize system operations for a legitimate client. RFC6749.

*7 Bluetooth®: A short-range wireless communication standard for interconnecting mobile terminals such as mobile phones and notebook computers. A registered trademark of Bluetooth SIG Inc. in the United States.

*8 Wi-Fi®: The name used for devices that interconnect on a wireless LAN using the IEEE802.11 standard specifications, as recognized by the Wi-Fi Alliance. A registered trademark of the Wi-Fi Alliance.

*9 Android™: A software platform for smartphones and tablets consisting of an operating system, middleware and major applications. A trademark of Google LLC.

2.4 Agent Software

Agent software is software that is installed in the home gateway and performs intercommunications between the IoT access control engine and IoT devices. It consists of the IoT access control engine agent, a Device WebAPI manager (virtual server) and plug-ins. The IoT access control engine agent connects using Message Queuing Telemetry Transport (MQTT)*11 protocol between the IoT access control engine and the home gateway, and relays API signals received from the cloud environment*12 to the Device WebAPI.

Handling existing IoT devices with the IoT access control engine is done by including software called plug-ins in the home gateway - no hardware modifications are required. Plug-ins absorb the various differences in IoT device specifications, and the API is defined using these plug-ins. Some of the devices already supported are shown in **Table 1**.

2.5 Official Dashboard

An official dashboard (Web site) is available to access the IoT access control engine API from a Web browser. This dashboard is created using the

Table 1 IoT access engine supported devices

Device type	Manufacturer	Product name
Medical thermometer	A&D	UT-201BLE
Body weight scale	A&D	UC-352BLE
Sphygmomanometer	A&D	UA-651BLE
Activity meter	Fitbit	charge2
Open/close sensor	Ermine	STM250J
Human sensor	Simics	HM92-01WHC
Human sensor	Optex	CPI-J
Human sensor	OMRON	HVC-F
Smart light	Philips	Hue single lamp
Smart light	Philips	Hue go
Smart lock	SESAME	Sesame smart lock
Infrared learning remote control	RATOC Systems	REX-WFIREX1
Infrared learning remote control	RATOC Systems	REX-WFIREX2
Infrared learning remote control	RATOC Systems	REX-WFIREX3
Dust sensor	RATOC Systems	REX-BTPM25
Dust sensor	RATOC Systems	REX-BTPM25V
Environment sensor	Pressac Sensing	CO_2, Temperature and Humidity Sensor

* Products appearing in the table are trademarks or registered trademarks of their respective manufacturers.

*10 Node.js: A software execution platform that enables JavaScript to run on various platforms. Node.js and the Node.js logo are trademarks or registered trademarks of Joyent, Inc. Oracle and Java are registered trademarks of Oracle Corporation and its subsidiaries and related companies in the United States and other countries.

*11 MQTT: A Pub/Sub-type light weight message queue protocol. Used for exchanging messages between various IoT devices and servers.

*12 Cloud environment: A virtual computing environment created on a network for use at the required time and in the required amount. Examples include AWS, etc.

IoT access control engine API, and is used for checking operations of developed IoT service applications using the IoT access control engine, etc.

3. Technologies Making Up IoT Access Control Engine

As characteristics of the IoT access control engine, the dynamic API generation, multiple device control and permission management, highly convenient data collection are enabled by the following mechanisms.

3.1 Dynamic API Generation for Individual Functions

The IoT access control engine can upload design information of functions in local environments such as gateway devices to which IoT devices are connected and smartphones to the cloud environment, manage functions in local environments from the cloud environment, and externally open these as APIs. Because no function design information is required in advance at the cloud environment side, it's possible to dynamically expand functions even if they are unknown.

The Device WebAPI, a technology that achieves commonality of IoT device control, is used as design information of functions in local environments.

By preparing virtual service in local environments, the Device WebAPI achieves function access without dependence on:

(1) Communications protocols such as wireless LAN or Bluetooth

(2) Operating system or execution environments

(3) Development language or development environments (building SDK environments for individual devices, dependency resolution, etc.)

Also, secure design and free functionality expansion with plug-ins for the virtual server in the local environment used by the Device WebAPI is standardized by OMA [6], and by including API design with the Swagger*13 (OpenAPI Specification) [7] WebAPI standardized specification in plug-ins as function design information, connectivity can be ensured even if the function is unknown.

As an implementation based on the Device WebAPI, NTT DOCOMO has released "DeviceConnect®*14 WebAPI" as MIT licensed*15 open source software on GitHub®*16 [8].

DeviceConnect WebAPI supports the Android, iOS*17 and Node.js environments. To develop plug-ins, source code generation tools are provided for plug-in output for each environment at a time just with API design based on Swagger. For this reason, Device WebAPI can be used in each environment with the least amount of development to correlate APIs and functions. For API design, function granularity, guidelines for describing APIs and function design patterns are prescribed, which makes abstracted design easy and independent of device structure or specifications.

With the IoT access control engine, functions are available from the cloud environment in a similar way from the local environment by generating a configuration for dynamic function access by designing APIs with the Device WebAPI and uploading them to the cloud through MQTT. Also, in development with APIs, functions can be used

*13 Swagger: A framework for building RESTful API or a standard format for describing an interface. Lead by Open API Initiative, also referred to as Open API Specification.

*14 DeviceConnect®: Software for interconnecting various devices through a WebAPI. A registered trademark of NTT DOCOMO, INC.

*15 MIT license: A software license whose license notation enables free and unlimited use, although usage is not covered by guarantee.

*16 GitHub®: A software development platform to promote development through the exchange of source code among multiple developers. A registered trademark of GitHub Inc.

*17 iOS: A trademark or registered trademark of Cisco in the United States and other countries and is used under license.

without any awareness of the local or cloud environments just by changing the API reference destination, if IoT access control engine security authentication is done. This means developers and service users do not need any awareness of the messages exchanged with MQTT.

3.2 Controlling Multiple Devices and Managing Their Permissions

Because the IoT access control engine includes API design for function access as the aforementioned mechanism, it's possible to express function access instructions and results as an API design. Handling these functions structured based on that API design with the IoT access control engine dashboard enables not only operation of various functions and data acquisition, but also settings for the scope of operations and data acquisition for individual functions, and release of those scopes to third parties (particular people or external services) (**Figure 2**).

Although it's possible with the API to operate the IoT access control engine environment to open functions that don't require user settings for individual functional scopes, or achieve mechanisms to grant permissions for function usage requests from external services, currently, for security reasons, these are restricted.

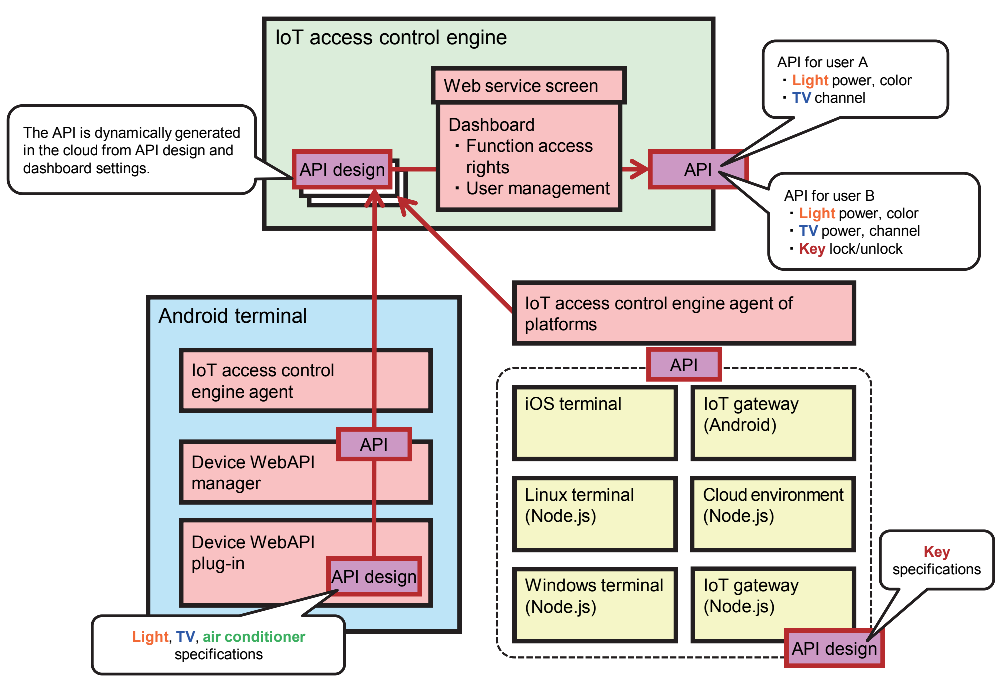


Figure 2　Dynamic API generation for individual functions

3.3 Data Collection in a Highly Convenient Form

The IoT access control engine has achieved abstracted function access with unified API design. Therefore, it's possible to reference data collected for specific purposes or check operations logs even with different devices or environments.

With the IoT access control engine, it's possible to easily achieve AI usage cases that would normally require substantial system design knowledge, because the architecture is consistent from the local through to the cloud environments and data collection regardless of differences in device specifications.

Specifically, there are prospects for usage methods, for example, using cloud services for machine learning[*18] in the IoT access control engine as cloud plug-ins that don't depend on a particular platform, or inputting data accumulated in the IoT access control engine into machine learning services as learning data and correct data for machine learning and then executing the generated learning model as rules in the local environment without dependence on a particular inference engine.

4. Interaction in AI Agent Platform

This chapter describes usage methods in the AI agent platform of the "IoT access control engine" discussed above. Also, as an example, this chapter briefly describes the "Kaden-kun" appliance control service that enables appliance operation with voice by interacting with "multipurpose dialogue engine (the engine for interpreting users' natural language)."

4.1 Remote Control

One usage of the IoT access control engine is IoT device remote control (downlink). **Figure 3** describes the most popular example, controlling appliances with dialogue, using "Kaden-kun." When the user makes utterances such as "turn on the TV" or "turn off the lights" for dialogue-enabled devices/applications, the multipurpose dialogue engine executes voice recognition and natural dialogue processing, and generates a request to an external service. Then, through the action of the appliance operation dialogue scenario, an API request processed in the IoT access control engine is generated.

For example, this will be "DELETE/device/tv" for the utterance "turn off the TV," or "PUT/device/light?color="FF0000"" for "change the light color to red."

When a request is sent to the IoT access control engine endpoint using the REST API[*19], routing is done to the relevant manager and plug-in in the engine, appliance control is executed and a response is returned. Finally, the appliance control dialogue scenario generates a system utterance in the multipurpose dialogue engine, the utterances "I've turned off the TV" or "I've made the light turn red" are made to the user with speech synthesis[*20], and processing finishes.

4.2 Information Collection

Another use of the IoT access control engine is remote information collection from various sensors and IoT devices (uplink). This is also done through interaction with the multipurpose dialogue engine. For example, when the user asks "What is the

---

*18 **Machine learning:** A mechanism allowing a computer to learn the relationship between inputs and outputs, through statistical processing of example data.

*19 **REST API:** A style of software architecture used on the Web.

*20 **Speech synthesis:** Technology for artificially creating speech data from text and verbally reading out text.

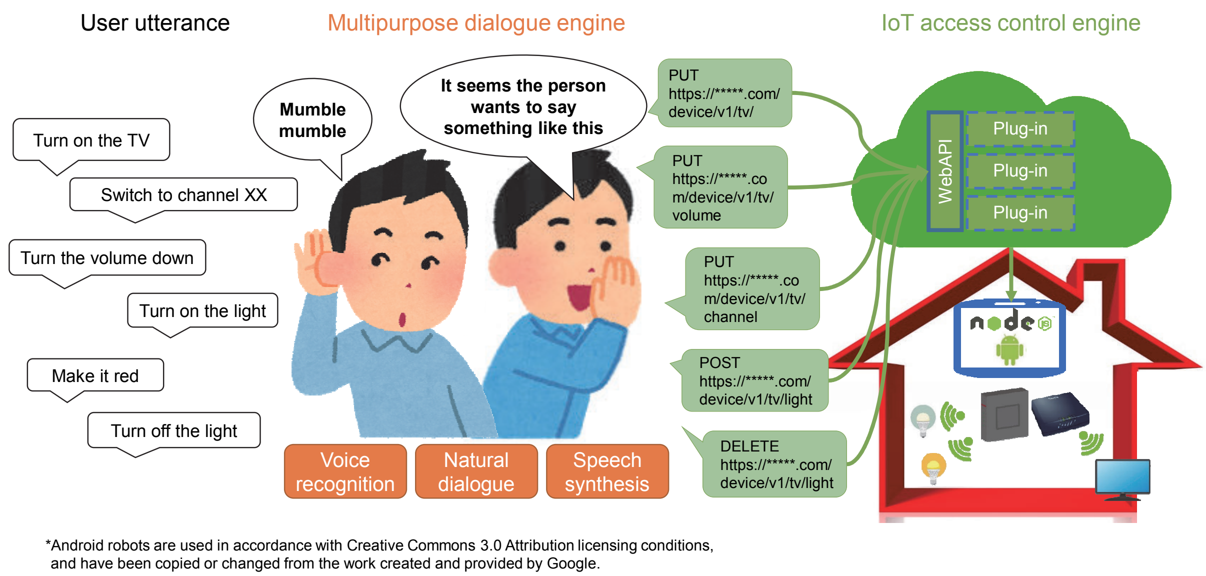


Figure 3　An example of operating appliances with dialogue

temperature of the room?" or "Is the house locked?", the IoT access control engine makes "GET/device/temperature/" and "GET/device/lock/" requests respectively, and calls the endpoint*21 in HyperText Transfer Protocol (HTTP). Then the system returns the utterances of "the current temperature is 26 ℃,"or "the house is unlocked."

4.3 Interacting with Multipurpose Dialogue Engine

Although it's possible to control appliances with natural dialogue by interacting with the multipurpose dialogue engine as discussed above, there are issues with combining natural dialogue processing and IoT. The following describes "Kaden-kun" solutions to these issues.

The first issue is control target determination. Since usage cases can entail users wanting to control more than one appliance with voice, or wanting to operate multiple devices at once (for example turning multiple lights on or off), determining targets for control is an issue.

We attempted to solve this issue by setting nicknames or group names by users and having users utter them. **Figure 4** describes an image of the "Kaden-kun" settings site. In this way, we

*21 **Endpoint**: URI to access to API.

Figure 4　The "Kaden-kun" settings site

implemented processing in which users set nicknames or group names, and then nicknames or group names with longest matching are extracted by linking the dialogue scenario with the IoT access control engine. When the user utters "turn on (nickname)" or "turn off (group name)," it's possible to correctly determine the target for control and control it.

Secondly, there is the issue of the desire to operate a target without specifying it, instead of specifying a target every time. The subject of sentences is characteristically omitted from Japanese language, but with the IoT access control engine, the subject had to be uttered to make a request using the API by determining the target for operation as a service ID*22.

We solved this issue by implementing cache*23 processing for the control targets in dialogue scenarios. Specifically, we solved the issue by inserting processing to continue caching the control target slot until the control target is uttered again or cancel is uttered, once the control target is input into the slot*24 for the control target (**Figure 5**).

5. Conclusion

This article has described a Device WebAPI and IoT access control engine, technologies to solve issues with IoT device interconnectivity.

The Device WebAPI abstracts IoT devices at

*22 Service ID: A unique identifier in the IoT access control engine for identifying particular functions in particular devices.

*23 Cache: Temporarily stored data to be distributed.

*24 Slot: A data model that complements information required to launch actions as a result of speech dialogue. For example, the slots required to operate an appliance are "control target" and "operation details," or for a weather forecast are "location" and "time."

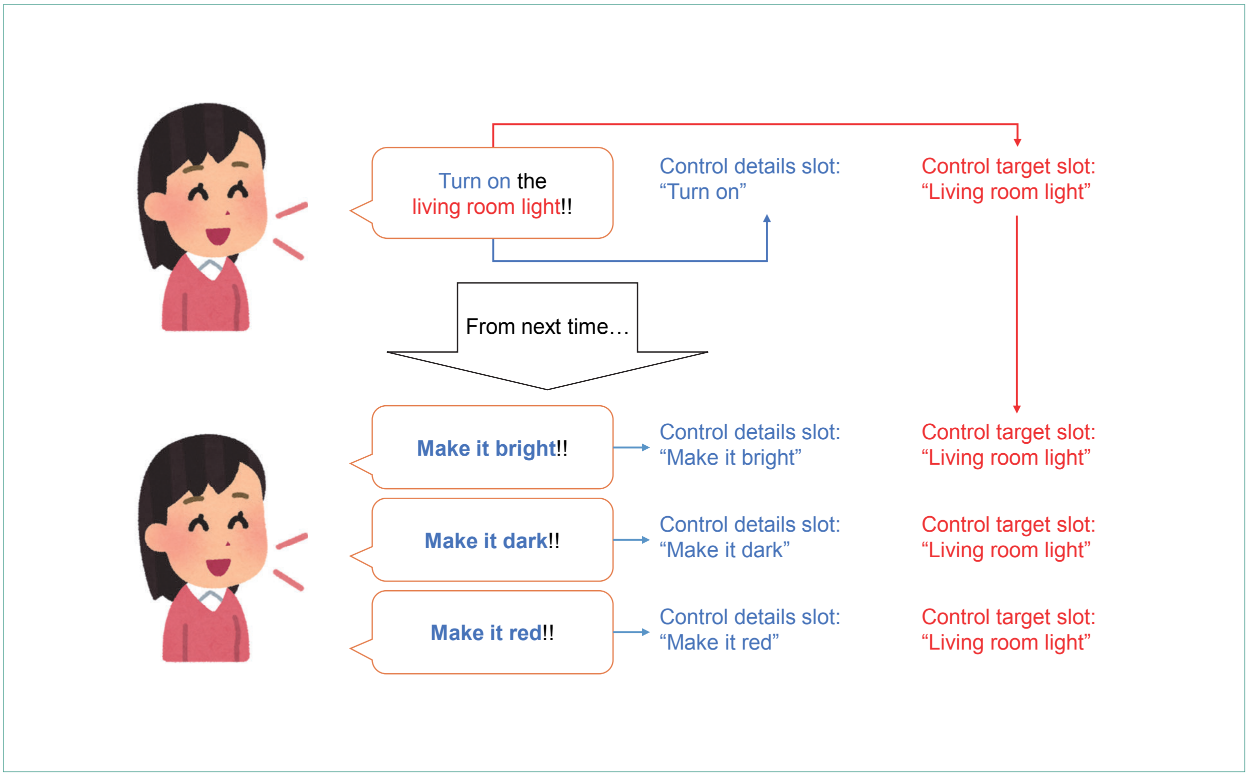


Figure 5 Control target slot caching

the functional level, and enables access with the common RESTful WebAPI. The IoT access control engine provides remote control and diverse permissions management functions. Moreover, as one of the engines on the AI agent platform, the IoT access control engine enables both remote control of IoT devices and information collection from them by interacting with other engines.

In future, we plan to provide cloud plug-in functions and a rule engine to expand functionality.

With cloud plug-in functions, the Device WebAPI is not held in the local environment, but in the cloud environment. By controlling it, it's possible to directly use devices that require interaction between cloud services.

Also, having a similar design to function access using the Device WebAPI, the rule engine is achieved with plug-ins to generate and control rules, which enables remote handling of local environment rules from the IoT access control engine. In addition, we plan to offer technologies to make rule description easy across multiple gateway devices and cloud services with the IoT access control engine.

To further advance the AI agent platform, we will continue research and development into more efficient protocols specialized for IoT, support for various communications networks such as Low Power, Wide Area (LPWA)*25, and log data analysis and machine learning specialized for IoT.

*25 LPWA: Wireless communications technology that can support a wide communications area at the kilometer level with low power consumption.

REFERENCES

[1] M. Weiser: "The Computer for the 21st Century," Scientific American 265, No. 3, pp. 94-104, Sep. 1991.

[2] M. Satyanarayanan: "Pervasive computing: Vision and challenges," IEEE Personal communications, Vol.8, No.4, pp.10-17, Aug. 2001.

[3] T. Starner: "Human-powered wearable computing," IBM Systems Journal, Vol.35, Issue 3.4, pp.618-629, 1996.

[4] G. D. Abowd: "Software engineering issues for ubiquitous computing," Proc. of the 21st international conference on Software engineering, ACM, 1999.

[5] T. Yamazoe et al.: "Device Connect WebAPI – Web Interface for Variety of Smartphone-linked Devices –," NTT DOCOMO Technical Journal, Vol.17, No.1, pp.4-9, Jul. 2015.

[6] OMA SpecWorks: "GotAPI." https://www.omaspecworks.org/what-is-omaspecworks/iot/gotapi/

[7] Swagger: "OpenAPI Specification." https://swagger.io/specification/

[8] GitHub: "DeviceConnect." https://github.com/DeviceConnect

Technology Reports 5G NR LTE

Special Articles on Release 15 Standardization
—Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

Technical Overview of the 3GPP Release 15 Standard

5G Laboratories Satoshi Nagata Kazuaki Takeda
Radio Access Network Development Department Hiromasa Umeda Hideaki Takahashi
Core Network Development Department Kenichiro Aoyagi

3GPP Work items on detailed specifications for RAN and CN have progressed toward commercialization of 5G, and non-standalone specifications for providing services combining LTE and NR areas were completed in December 2017. Then, in June 2018, the Release 15 standards were completed, including standalone specifications for providing NR-only areas and for advances to LTE/LTE-Advanced. This article gives an overview of the NR and LTE specifications from the 3GPP, completed in Release 15.

1. Introduction

The 3rd Generation Partnership Project (3GPP), which is an international standardization organization for mobile communications systems, completed its Release 15 specifications in June 2018. This included the first complete specification for 5th Generation (5G) mobile communications systems. The specifications set regulations for the New Radio (NR) communication system, satisfying 5G use cases

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

and requirements and not backward-compatible with earlier LTE/LTE-Advanced systems, and for the new 5G Core Network (5GC). At the same time, it also sets new specifications for LTE/LTE-Advanced, continuing to extend functionality and increase performance. This article gives an overview of the main radio and core network functionality specified in Release 15, and discusses background considerations for the release.

2. Release 15 Specification Functional Overview

3GPP Release 15 contains new NR and 5GC specifications, and extended specifications for LTE/LTE-Advanced. These include enhanced Mobile BroadBand (eMBB) technologies for achieving higher speed and capacity as with earlier mobile communications technologies, but also actively studies technologies for implementing Ultra-Reliable and Low Latency Communications (URLLC). Technical extensions to LTE/LTE-Advanced for massive Machine-Type Communications (mMTC), to accommodate large numbers of Internet of Things (IoT) terminals are also included. **Figure 1** shows functionalities specified for the three usage scenarios: NR, 5GC and LTE/LTE-Advanced.

3. New Radio Access Functionality Defined in the Release 15 Specification

3.1 Functionality for NR

For NR, technologies to provide for non-standalone*1 and standalone*2 operation of eMBB are specified [1]. The main technology for providing non-standalone operation is Dual Connectivity*3 for LTE and NR. Technology to provide URLLC on NR is also specified.

1) Functionality for eMBB

The main NR features for realizing eMBB are (a) high-frequency/ultra-wideband transmission, (b) Massive Multiple Input Multiple Output (Massive MIMO)*4 transmission, and (c) flexible frame*5 structure and physical channel*6 structure. These features are described below.

(a) High frequency/ultra-wideband transmission

NR in Release 15 anticipates high-frequency bands up to 52.6 GHz and radio performance is specified for frequency range 1 (FR1) from 450 to 6,000 MHz and frequency range 2 (FR2) from 24,250 to 52,600 MHz. As with LTE/LTE-Advanced, FR1 is specified assuming conducted testing, but FR2 is assumed to only be used Over The Air (OTA)*7 [2]. For ultra-wideband transmission, up to 100 MHz per Component Carrier (CC)*8 is specified for channel bandwidth in FR1, while up to 400 MHz per CC is specified in FR2. The physical layer*9 specifications also support Carrier Aggregation (CA)*10 and dual connectivity, to realize ultra-high-data rate transmission by bundling up to 16 NR CCs. This support for large channel bandwidth and extremely high frequency bands is a major difference between NR and LTE/LTE-Advanced.

NR also supports multiple Orthogonal

---

*1 Non-Standalone: An operation format in which terminals connect to a mobile communications network via multiple radio technologies.

*2 Standalone: An operation format in which terminals connect to a mobile communications network via a single radio technology.

*3 Dual Connectivity: A technology that achieves wider bandwidths by connecting two base stations in a master/secondary relationship and performing transmission and reception using multiple component carriers supported by those base stations.

*4 Massive MIMO: MIMO systems transmit radio signals overlapping in space by using multiple antenna elements for transmission and reception. Massive MIMO systems aim to achieve high-speed data communications with greater numbers of simultaneous streaming transmissions while securing service areas. They achieve that aim by using antenna elements consisting of super multi-element arrays to create sharply formed radio beams to compensate for the radio propagation losses that accompany high-frequency band usage.

*5 Frame: The period in which an encoder/decoder operates or a data signal of length corresponding to that period.

*6 Physical channel: A generic term for channels that are mapped onto physical resources such as frequency or time, and transmit control information and other higher layer data.

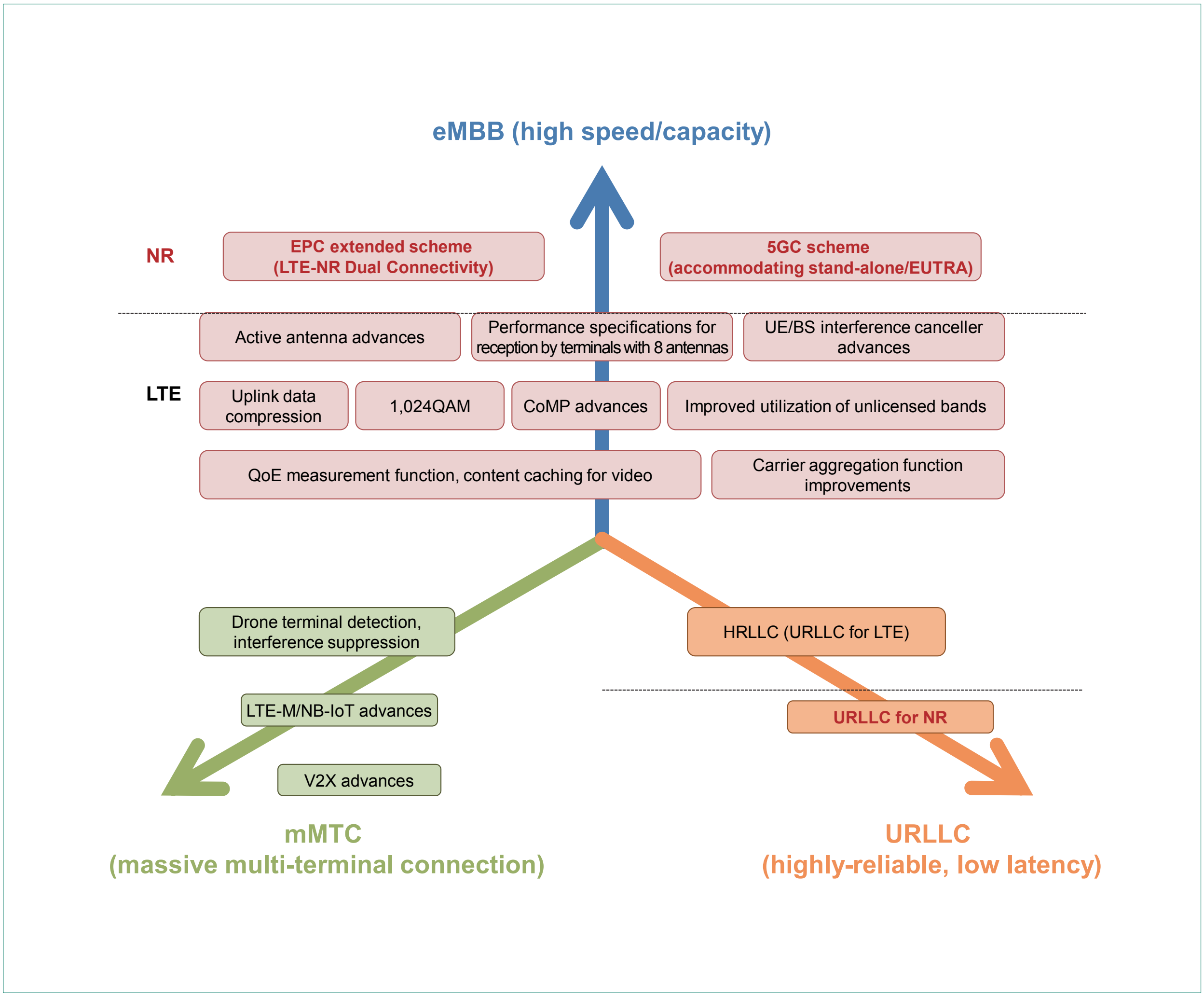


Figure 1 Main functions specified in Release 15 for NR, 5GC, and LTE/LTE-Advanced

Frequency Division Multiplexing (OFDM)*11 sub-carrier*12 spacings, of 15, 30, 60, and 120 kHz, for data transmission. This is because generally if OFDM sub-carrier spacings are the same at all frequencies, high frequency bands have significantly degraded resistance to multipath fading*13 and phase noise*14, compared with lower frequency bands.

OFDM based technology was also adopted as the multiple access*15 system because it has good affinity with MIMO technology and strong resistance to frequency-selective fading*16. OFDM is supported on both the uplink and the downlink. To ensure coverage

*7 OTA: A method for setting specified points and measurement points in a radiowave-propagation space, specifying wireless performance (including antenna emission and reception characteristics), and measuring those parameters.

*8 CC: Term denoting each of the carriers used in Carrier Aggregation.

*9 Physical layer: First layer of the OSI reference model; for example, "physical-layer specification" expresses the wireless interface specification concerning bit propagation.

*10 CA: A technology that achieves high-speed communication while maintaining backward compatibility with existing LTE, by using multiple component carriers simultaneously to expand bandwidth for transmission and reception.

*11 OFDM: A digital modulation method where the information is transmitted over multiple orthogonal carriers and sent in parallel. It allows transmission at high data rates.

*12 Sub-carrier: Individual carrier for transmitting signals with multi-carrier transmission such as OFDM.

*13 Multi-path fading: A phenomenon whereby a radio wave is subjected to repeated reflection and diffraction due to geographical features and buildings, and thus reaches to a receiver as multiple radio waves.

on the uplink, in addition to OFDM, Discrete Fourier Transform-spread (DFT-spread) OFDM*17, which has low Peak-to-Average Power Ratio (PAPR)*18, is also supported for single stream transmission.

(b) Massive MIMO transmission

The Release 15 physical layer specification specifies component technologies such as the reference signal structure and beam management, which assume Massive MIMO transmission in high frequency bands with up to 256 antenna elements on base stations and up to 32 antenna elements on terminals. To realize high data rate and capacity, the downlink supports single user MIMO*19 transmission with up to 8 layers and multi-user MIMO*20 transmission with up to 12 layers, and the uplink supports single-user MIMO transmission with up to four layers.

With high frequency bands, beam forming*21 is usually an important technology to compensate for reduced coverage due to propagation losses*22. Since LTE/LTE-Advanced did not anticipate use of high frequency bands, it assumed a digital beam forming implementation, with beam forming generated in the digital domain. However, it is difficult to implement high-frequency-band Massive MIMO in this way, so NR assumes a hybrid implementation that combines both digital and analog beam forming. As a result, major NR functions in the standard specification, such as initial access, scheduling, and Hybrid Automatic Repeat reQuest (HARQ)*23 retransmission, give consideration to the new high-frequency-band hybrid type of beam forming as well as the current, low-frequency-band digital beam forming [2].

(c) Flexible frame structure/physical channel structure

As mentioned earlier, NR supports multiple sub-carrier spacings, and with a wide subcarrier spacing in the frequency direction, OFDM symbols*24 are shorter in the time direction. For example, with a 120 kHz subcarrier spacing, the OFDM symbols are 1/8 the length of with a 15 kHz sub-carrier interval, as with LTE/LTE-Advanced. Using wide sub-carrier spacings in this way makes low-latency transmission possible. The number of OFDM symbols in an allocation unit for control and data channels can be changed flexibly [2], and the uplink and downlink slot ratio in the frame structure can be changed flexibly, according to the traffic ratios on the uplink and downlink.

2) Functionality for URLLC

URLLC supports and assists usage scenarios for mission critical services*25 that require near-real-time and high reliability. These are a type of IoT that includes self-driving vehicles, industrial robots, and remote medicine. As mentioned earlier, low latency is implemented by using a wide sub-carrier spacing, and reducing the number of OFDM symbols used for data assignment. On the other hand, to implement high reliability, new Channel

---

*14 Phase noise: Phase fluctuation that occurs due to frequency components other than those of the carrier frequency in a local oscillator signal.

*15 Multiple access: Indicates methods in a radio system in which channels are assigned from among multiple vacant radio channels for communication, when multiple UE are communicating within the system.

*16 Frequency selective fading: A phenomenon in which the received level is not uniform along the frequency axis of the received signal because signals (frequencies) arrive through various paths due to reflection from buildings, etc.

*17 DFT-spread-OFDM: A digital modulation format. It is able to reduce PAPR by multiplying the signal from a user by a DFT precoder before performing OFDM modulation.

*18 PAPR: The ratio of the maximum power to the average power. If this value is large, the amplifier power back-off has to be large to avoid signal distortion, which is particularly problematic for mobile terminals.

*19 Single-user MIMO: Technology that uses MIMO transmission at identical temporal frequencies for a single user.

*20 Multi-user MIMO: A technology that uses MIMO to transmit signals to multiple users at the same time using the same frequency.

Quality Indicator (CQI)*26 and Modulation and Coding Scheme (MCS)*27 tables for URLLC are specified to support lower signal ratios than eMBB.

3.2 Functionality for LTE/LTE-Advanced

Release 15, also extends earlier specifications, with technologies that extend LTE/LTE-Advanced to implement eMBB, including support for 1024 Quadrature Amplitude Modulation (QAM)*28 to further improve the frequency utilization, enhanced OTA regulation for the active antenna system*29, and technical extensions to use unlicensed bands.

1) eMBB Technology for Increasing Bandwidth Utilization

Technologies to extend LTE/LTE-Advanced and increase frequency utilization are described below.

(1) 1024QAM support

To increase the peak data rate further, 1024QAM and DeModulation Reference Signal (DM-RS)*30 overhead*31 reduction are specified.

(2) Coordinated Multi-Point transmission/reception (CoMP)*32 advances

Non-coherent joint transmission, in which two base stations transmit different data sequences without knowing Channel State Information (CSI)*33 of the other base station is supported. With this, extensions were made to QCL, control information, and CSI feedback functions.

(3) Performance specifications for reception by terminals with eight antennas

Radio performance is specified for realizing expanded cell coverage with terminals having eight receiver antennas and maximum transmission speed using eight layers on the downlink is increased.

(4) Expansion of OTA based requirements for active antenna systems

Only two items for OTA based requirements were specified in Release 13 (Equivalent Isotropic Radiated Power (EIRP)*34 and Equivalent Isotropic Sensitivity (EIS)*35) for active antenna system radio performance, and these basically assumed conducted testing. In Release 15, all the conducted based requirements other than the two in Release 13 are replaced with those based on OTA, introducing specifications that enable end-to-end performance to be guaranteed, including the antennas. With these requirements, it is now possible to require performance to be guaranteed based on 3GPP specifications, for active antenna systems that do not have antenna connectors. For equipment with this type of structure, power losses within the equipment are reduced, and multiple transceivers and antennas can be implemented with high density, forming tighter beams (high gain), and compensating for the higher radio wave propagation losses in high frequency bands.

(5) Technologies to reduce various types of interference

Several functions to extend LTE/LTE-Advanced have been specified, to suppress

---

*21 Beamforming: Technology for generating a directional pattern for transmission and/or reception by using multiple antennas (by means of controlling amplitude and phase of each of multiple antennas) and increasing or decreasing antenna gain in regard to specific directions.

*22 Propagation losses: The amount of attenuation in the power of the signal emitted from the transmitting station till it arrives at the reception point.

*23 HARQ: A technique that compensates for errors in received signals through a combination of error-correcting codes and retransmission.

*24 OFDM symbol: A unit of transmission data consisting of multiple subcarriers. A Cyclic Prefix (CP) is inserted at the front of each symbol.

*25 Mission critical service: A type of communications service defined for 3GPP mobile communications networks, provided for public safety; mainly police and fire prevention.

*26 CQI: An index of reception quality measured at the mobile station expressing propagation conditions on the downlink.

*27 MCS: Combinations of modulation scheme and coding rate decided on beforehand when performing AMC.

*28 QAM: A modulation method that uses patterns of both amplitude and phase for modulation, with varieties according to the number of such patterns such as 16QAM and 64QAM.

interference from other cells and increase user throughput. A function has been specified to reduce the amount of Cell-specific Reference Signal (CRS)*36 transmitted by base stations during times and in areas where the volume of data traffic is low, reducing both interference due to CRS and base station power consumption. An interference canceller*37 function for uplink reception on base stations has also been specified. Performance of interference suppressing technologies for low cost terminals with only one receiver antenna, such as IoT terminals, has also been specified.

2) Other Technologies for eMBB

Besides technologies to improve frequency utilization, other improvements have been made based on experience operating LTE/LTE-Advanced networks. The main such technologies are described below.

(1) Improving utilization of unlicensed bands

The uplink function for Licensed Assisted Access (LAA)*38 that was specified in Releases 13 and 14 has been extended. Specifically, it is possible to change the position of the first and last symbols more flexibly when transmitting uplink data. Autonomous UL transmission is also supported, enabling terminals to begin uplink transmission autonomously.

(2) CA functionality improvements

With earlier CA, delay for the following required terminal processing occurs before the Secondary Cell (SCell)*39 can be configured.

- The time required to measure the quality of candidate carriers increases, depending on the number of SCell candidate-measurement carriers.
- When communication with the SCell begins, several tens of milliseconds are needed to start up the Radio Frequency (RF) channel.

To solve these issues, a mechanism has been specified to measure radio signal quality of SCell candidates while a terminal is in the Idle state, along with the new dormantSCell state, in which an SCell is configured with an initialized RF circuit before the SCell starts communication.

(3) Uplink data compression

For Time Division Duplex (TDD)*40, the ratio of uplink to downlink usually emphasizes the downlink, so the radio resources available for transmission on the uplink are limited. To transmit uplink user data using the radio resources efficiently, a mechanism for Uplink Data Compression (UDC) is specified, which compresses headers for packets at the IP level and above (IP, User Datagram Protocol (UDP)*41, Real-time Transport Protocol (RTP)*42, etc.). Headers are compressed at the Packet Data Convergence Protocol (PDCP)*43 layer, using the DEFLATE*44 algorithm standardized by the Internet Engineering

*29 Active antenna system: A system that integrates antenna elements and RF circuits that have traditionally been separated thereby providing a more efficient system.

*30 DM-RS: A user-specific reference (pilot) signal known by the base station and mobile station for estimating the fading channel used for data demodulation.

*31 Overhead: Control information needed for transmitting/receiving user data, plus radio resources used for other than transmitting user data such as reference signals for measuring received quality.

*32 CoMP: Technology which sends and receives signals from multiple sectors or cells to a given UE. By coordinating transmission among multiple cells, interference from other cells can be reduced and the power of the desired signal can be increased.

*33 CSI: Information describing the state of the radio channel traversed by the received signal.

*34 EIRP: The transmission power at the reference point in radio radiation space.

*35 EIS: The received power at the radiated requirement reference point in radio reception space.

*36 CRS: A reference signal specific to each cell for measuring received quality in the downlink.

Task Force (IETF)*45.

(4) QoE measurement functionality/content caching for video

With the recent rise of services for viewing video on for smartphones, increasing the image quality and Quality of Experience (QoE) for video in mobile communication environments has become an issue for operators. To measure QoE on real networks, a mechanism enabling the network to collect QoE measurements from terminals, called Minimization of Drive Test (MDT)*46, has been specified. A mechanism which places a video content cache servers near base stations has also been studied, to reduce delay when downloading video. With this mechanism, terminals download data from the content server by connecting and communicating through the base station directly with the content server, rather than communicating through the Evolved Packet Core (EPC)*47.

3) Technologies for mMTC

In response to recent increasing anticipation in the market for IoT, the 3GPP has also specified technologies for machine communication and inter-vehicle communication. Advancements in these functionalities have been implemented in Release 15, specifying enhancements for drones and URLLC.

(1) Drone terminal detection/interference suppression

With the spread of services using drones, demand has also increased for mobile communication systems to provide wide-area communication for drone terminals. To meet this anticipation in the market, the 3GPP has studied whether providing communication for drone terminals using LTE/LTE-Advanced is feasible, and found that even existing specifications have the necessary and sufficient functionality to provide communication for drone terminals.

On the other hand, a problem of increasing uplink interference from overhead drone terminals to the base station has been identified. Some countries and regions also require a license for drone terminals to fly, so the issue of how the network can determine if a drone is licensed or not was also studied.

For the issue of uplink interference, a mechanism has been specified to adjust the open-loop*48 transmit-power control parameters for individual terminals, setting them to control the target Signal to Interference Ratio (SIR)*49 even when propagation losses from the base station are small.

To determine whether a drone terminal is licensed, a mechanism is specified to link with the Home Subscriber Service (HSS)*50 within the network and determine the license status from subscriber information.

(2) LTE-M*51/NarrowBand (NB)-IoT*52 advances

Release 15 also adds extensions to TDD support and low-power features for IoT terminals through LTE networks.

---

*37 Interference canceller: A method for separating multiple combined signals received at the same time, by successively detecting and then cancelling each signal from the received signal. It generally yields better performance than Minimum Mean Square Error (MMSE) detection.

*38 LAA: A generic name for radio access methods in which terminals obtain configuration information from an LTE carrier using a licensed band, and then use an unlicensed band for radio communication.

*39 SCell: Carriers other than the PCell with multiple carriers in CA. Also referred to as the secondary cell.

*40 TDD: A format in which downlink and uplink communication is segmented in time, with transmission and reception alternating.

*41 UDP: A standard Internet protocol above the IP layer. In contrast to TCP it does not provide functions to establish a connection between server and terminal or to retransmit data that does not reach the destination.

*42 RTP: A real-time multimedia transport protocol used on IP networks. Defined by the Internet Engineering Task Force (IETF, See *43).

*43 PDCP: A sublayer of layer 2. A protocol for ciphering, validation, ordering and header compression, etc.

*44 DEFLATE: A data compression algorithm standardized by the IETF.

(a) Idle mode power conservation technology (Wake-up signal)

To reduce power consumption in Idle mode, a new Wake-up signal has been specified. Normally, terminals in Idle mode attempt to decode the downlink control periodically to obtain paging*53 information. They do not know whether there is paging information to obtain until the channel has been decoded, so it must be done periodically, increasing power consumption. Release 15 introduces a simple process for detecting a Wake-up signal. This provides a way to determine whether there is paging information and should reduce terminal power consumption.

(b) Reduced latency for small-packet communication

In use cases such as smart meters, data transmissions are expected to use relatively small packets. For such cases, a procedure for starting transmission of small amounts of data has been added within the random access*54 procedure, which normally cannot be used to start data transmission. This promises to reduce latency for small-packet communication.

(c) TDD support

In Releases 13 and 14, LTE-M and NB-IoT were specified, targeting operation in Frequency Division Duplex (FDD)*55 bands. Release 15 also enables operation in TDD bands.

(3) Vehicle to Everything (V2X) communication advances

Release 15 adds extensions to the V2X communication functionality specified in Release 14. Specifically, to increase the data rate and bandwidth of V2X communication, CA has been introduced to Mode 4*56, which enables terminals to select transmission resources autonomously from a resource pool*57. Support for the 64QAM modulation scheme has also been added for V2X. New terminal performance specifications have also been added to satisfy low latency requirements.

4) Highly-Reliable Low-Latency Communication (HRLLC, URLLC for LTE)

There has been increasing demand recently for services requiring real-time performance, such as self-driving vehicles, industrial robots, remote medicine and virtual reality (VR), and mission-critical services requiring high reliability. As such, functionality implementing low latency and highly reliable communication on LTE/LTE-Advanced has been specified.

(a) Technologies for highly-reliable communication

For highly-reliable communication, functionality has been introduced to improve transmission quality on the downlink control channel, and uplink and downlink data channels.

For the downlink control channel, in conventional LTE/LTE-Advanced, the Physical Control Format Indicator Channel (PCFICH)*58

*45 IETF: A standardization organization that develops and promotes standards for Internet technology.

*46 MDT: A technology standardized by the 3GPP for gathering QoE information. Terminals send information to the network regarding incidents such as interruption of communication or failed handover as they occur, such as location and cause of the incident.

*47 EPC: The core network on 3GPP mobile communication networks, mainly accommodating E-UTRA.

*48 Open loop: A type of control that operates on the input, without using the output for feedback.

*49 SIR: The ratio of desired-signal power to interference power.

*50 HSS: A subscriber information database that manages authentication and current location information in a 3GPP mobile network.

*51 LTE-M: An LTE communication specification for terminals that communicate at low speed using narrow bandwidth, for IoT devices (sensors, etc.).

*52 NB-IoT: An LTE communication specification for terminals that communicate at even lower speed and narrow bandwidth than LTE-M, for IoT devices (sensors, etc.).

*53 Paging: A method and signal for calling a visiting UE that is on standby when a call is received.

was detected and the number of PDCCH OFDM symbols was identified. However, in this case, the quality of the entire downlink control channel is constrained by PCFICH detection errors and error detection. As such, a method to improve the quality of the downlink control channel has been specified, avoiding earlier effects of PCFICH detection errors and error detection by notifying of the number of PDCCH OFDM symbols using higher layer signaling.

To improve the quality of transmission on the uplink and downlink data channels, new functionality was introduced to repeat transmission of the same data over multiple Physical Downlink Shared Channels (PDSCH), Short Physical Downlink Shared Channels (SPDSCH)*59, Semi-Persistent Scheduling Physical Uplink Shared Channels (SPS-PUSCH)*60 or Semi-Persistent Scheduling Short Physical Uplink Shared Channels (SPS-SPUSCH)*61.

(b) Technologies for low-latency communication

For low-latency communication, the signal processing times for the 1 ms Time-To-Interval (TTI)*62 was reduced from earlier values, and a new short TTI was specified.

Functionality has been specified to reduce the signal processing times from when data is received on the downlink till when HARQ feedback is sent, and from receiving on the downlink control channel, which schedules uplink data, till data is sent on the uplink, by 25%, from the prior minimum of 4 ms to 3 ms.

To also reduce the size of the 1 ms TTI itself, which is composed of 14 OFDM symbols, slots composed of 7 OFDM symbols and sub-slots composed of 2 or 3 symbols have been specified as short TTIs. Accordingly, functionality has also been specified for scheduling resources using these units of data on the downlink SPDSCH and uplink SPUSCH. To implement Short TTI data allocation, uplink and downlink control channels, Short Physical Downlink Control Channel (SPDCCH) and Short Physical Uplink Control Channel (SPUCC) are also specified.

4. Core Network Specified for Release 15 NR

Figure 2 shows the two methods for providing NR as specified in Release 15 from the core network perspective. They are the extended EPC scheme, which extends the existing EPC to provide NR in non-standalone operation, and the 5GC scheme, which introduces the newly specified 5G core network (5GC) to provide NR in standalone operation [1].

4.1 Features of Extended EPC Scheme

1) EPC Dual Connectivity

Cases of NR operation, such as deployment in localized areas or limited coverage, are expected because NR has features such as use of high-frequency bands. On the other hand, the quality of existing services such as Voice over LTE (VoLTE) and IoT

*54 Random access: A procedure executed by mobile terminals and base stations for connecting uplink signals and synchronizing their transmission timing.

*55 FDD: A method for implementing simultaneous transmission and reception with radio communications etc, in which transmission and reception are done using different frequencies.

*56 Mode 4: A type of resource allocation method used with LTE V2X.

*57 Resource pool: A set of resources achieved by bundling together many units of hardware each possessing certain types of resources (CPU, memory, HDD, etc.). Various types of virtual machines can be created from a resource pool.

*58 PCFICH: A physical channel used to notify of the number of symbols for PDCCH transmission on LTE.

*59 SPDSCH: PDSCH with a short TTI.

*60 SPS-PUSCH: A method for periodically sending different uplink data based on demodulation of downlink control information.

*61 SPS-SPUSCH: PUSCH with a short TTI.

*62 TTI: Transmission time per data item transmitted via a transport channel.

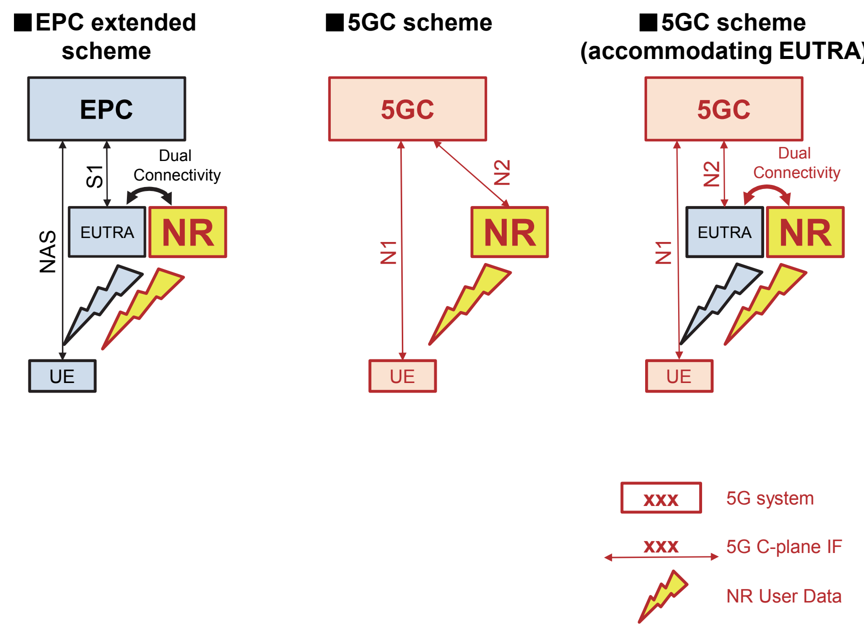


Figure 2 Provision of NR from the core-network perspective

must be maintained and require adequate area coverage, so operation will need to vary according to the service and environment. Existing EPC facilities and coverage provided by LTE/LTE-Advanced will be used for these services, and NR will be used for specialized services such as viewing high-definition video. The main feature of EPC Dual Connectivity is that various settings on existing EPC equipment, the S1 Interface (S1-IF)*63 with evolved NodeB (eNB)*64, and the Non-Access Stratum interface (NAS-IF)*65 are used with the terminal to minimize any effect on core network equipment. This enables NR to be introduced relatively quickly.

2) 5G NR Service Identification (Control) Functionality

In addition to the Dual Connectivity function described above, various other functions have been specified for providing NR much more appropriately and flexibly, considering roaming and various other types of contract and forms of operation.

- 5G area notification function: When the core network receives from a terminal, as during the attach*66 procedure, the core network sends 5G subscriber information to the terminal. The terminal then notifies the user

*63 S1-IF: The functional layer between eNB and EPC.
*64 eNB: A base station for the LTE radio access system.
*65 NAS-IF: The functional layer between the mobile terminal and core network through the Access Stratum (AS).
*66 Attach: The process of registering a mobile terminal to a network when the terminal's power is turned on, etc.

whether it is in an NR area and whether a 5G NR connection is possible using an icon or other means, based on the subscriber information and cell configuration information from the base station.

- 5G NR connection decision function: As described above, this notifies the terminal and the base station regarding the subscriber's 5G contract. This allows control of network operation such as enabling or disabling 5G and making NR connections while roaming, or only allowing specific terminals used for testing to make NR connections.
- 5G Gateway (GW*67) selection function: When providing GW equipment optimized for throughput and capacity, which are 5G features, this function gives priority to terminals capable of using NR when connecting to the GW equipment.
- 5G-Data Reporting function: With Dual Connectivity, the base station allocates transmission of user data to either Evolved Universal Terrestrial Radio Access (E-UTRA)*68 or NR, according to factors such as the radio environment. This function counts and reports the amount of user data transmitted by NR and reports it from the base station so that the core network can know how much user data was actually transmitted by NR.

4.2 New 5G Core Network Features

5GC is the new 5G core network equipment specified in Release 15. See reference [1] regarding the main features of 5GC.

One of the main features of 5GC is that it is able to accommodate E-UTRA together with NR in a Next Generation Radio Access Network (NG-RAN)*69. eNB that are able to connect to 5GC are called ng-eNB, and use the same NG-AP interface when connecting to 5GC as they do when connecting to NR. A fallback technology is also provided, so that when a terminal is in a 5GC area and starts IP Multimedia Subsystem (IMS)*70 voice services, the terminal will be connected to EPC, and these services will be provided through EPC. This enables radio equipment built with E-UTRAN, and its settings and coverage to be used on 5GC, preserving the quality and area of existing services while providing new 5G services and implementing migration smoothly.

5. Conclusion

This article has given background and described the main functionality in the Release 15 specifications for NR, 5GC and LTE/LTE-Advanced. For more details regarding NR and related network management specifications introduced here, please see other articles of this special feature, references [2] to [5]. The 3GPP began work creating the Release 16 specifications in October 2018. With respect to NR, the intention is to create specifications that will advance multi-beam/MIMO for millimeter wave and expand application areas for URLLC and IoT. Regarding the core network, the intention is to study improvements to various platform functions

---

*67 GW: A node having functions such as protocol conversion and data relaying.

*68 E-UTRA: An air interface used for advanced wireless access schemes in 3GPP mobile communication networks.

*69 NG-RAN: A RAN connecting to the 5G core network using NR or E-UTRA as radio access technology.

*70 IMS: A subsystem that provides IP multimedia services (e.g., VoIP, messaging, presence) on a 3GPP mobile communications network.

for commercial introduction of 5GC, and new services such as URLLC, V2X and IoT, enhancing and applying them on 5GC.

REFERENCES

[1] A. Minokuchi, et al.: "5G Standardization Trends at 3GPP," NTT DOCOMO Technical Journal, Vol.19, No.3, pp.5–12, Jan. 2018.

[2] K. Takeda, et al.: "NR Physical Layer Specifications in 5G," NTT DOCOMO Technical Journal, Vol.20, No.3, pp.49–61, Jan. 2019.

[3] T. Uchino, et al.: "Specifications of NR Higher Layer in 5G." NTT DOCOMO Technical Journal, Vol.20, No.3, pp.62–78, Jan. 2019.

[4] Y. Sano, et al.: "5G Radio Performance and Radio Resource-Control Specifications," NTT DOCOMO Technical Journal, Vol.20, No.3, pp.79–95, Jan. 2019.

[5] K. Tsubouchi, et al.: "Network Management Specifications for 5G Era," NTT DOCOMO Technical Journal, Vol.20, No.3, pp.96–101, Jan. 2019.

Technology Reports | 5G | Design of Physical Signals and Channels | NR Physical Layer

Special Articles on Release 15 Standardization
—Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

NR Physical Layer Specifications in 5G

5G Laboratories Kazuki Takeda Hiroki Harada Ryosuke Osawa
Communication Device Development Department Yuichi Kakishima
DOCOMO Beijing Communications Laboratories Co., Ltd. Lihui Wang Runxin Wang

A Work Item of the detailed specifications of 5G RAN for 5G commercial services is being conducted in 3GPP. The non-standalone NR specifications, which enable early deployment of NR carriers by exploiting LTE/LTE-Advanced carriers as the master node carriers, were completed in December 2017. In June 2018, the standalone NR specifications, which require full NR functionality and do not rely on LTE/LTE-Advanced carriers, were also completed. In this article, we present the NR physical layer specifications.

1. Introduction

With the spread of smartphones and tablet terminals, it has become easier for people to collect information and enjoy rich contents such as videos and music at any time and in any location. The quality of the content available to these terminals has also been improved dramatically. Furthermore, it is expected that services made possible by the "Internet of Things" (IoT) era, where everything is connected to networks, will become increasingly important in the future.

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

Under such circumstances, the 3rd Generation Partnership Project (3GPP) formulated the New Radio (NR) standard for wireless communication in 5th Generation mobile communication systems (5G), which doesn't have backward compatibility with LTE and LTE-Advanced. NR supports various use-cases such as enhanced Mobile BroadBand (eMBB), Ultra-Reliable and Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC), over a very wide frequency range from below 1 GHz to 52.6 GHz. This article describes the underlying technologies from the physical layer perspective of NR that are specified to realize these requirements from the various use-cases.

2. Frame Structure/Duplex Mode

2.1 Support for New Subcarrier Spacings

NR adopts a radio access scheme called Orthogonal Frequency Division Multiplexing (OFDM)*1, which is the same scheme used in LTE. To adapt to services requiring low latency and to enable the use of higher frequencies (including millimeter-wave*2 frequencies), NR supports higher subcarrier*3 spacings of 30, 60, 120, and 240 kHz based on the LTE subcarrier spacing of 15 kHz. Note that a subcarrier spacing of 240 kHz is only used for Synchronization Signal (SS)*4/Physical Broadcast CHannel (PBCH)*5 blocks, as described later (In addition, the SS/PBCH block does not support a subcarrier spacing of 60 kHz).

When user equipment initially accesses the network, it detects an SS/PBCH block by assuming a subcarrier spacing that can be set in this carrier, and based on the PBCH detected in this block, it identifies the subcarrier spacings of the control/data channels, etc.

2.2 Frame Structure

In NR, multiple OFDM symbols*6 are used to construct slots*7, subframes*8, and frames*9. A slot consists of 14 OFDM symbols for the given subcarrier spacing, a subframe is defined as a 1 ms interval, and a frame is defined as 10 subframes. These relationships are shown in **Figure 1**.

In the frequency domain, a resource block consists of 12 consecutive subcarriers for the given subcarrier spacing.

In NR, unlike LTE, the frame structure is not dependent on the duplex mode*10. In other words, it uses a common frame structure regardless of the duplex mode.

NR provides much greater flexibility than LTE in the uplink and downlink patterns of Time Division Duplex (TDD)*11 communication. It is possible to semi-statically or dynamically set various uplink/downlink patterns by the system information and/or by user-specific higher layer signaling*12 and/or L1 signaling*13.

It is also possible to perform TDD communication without using higher layer/L1 signaling to indicate the uplink/downlink patterns. The user equipment can recognize the direction of uplink/downlink communication based on periodic transmission and reception that are configured by higher layer signaling and/or indicated by dynamic signaling in the physical layer.

*1 OFDM: A multi-carrier modulation format where information signals are modulated with orthogonal subcarriers. A type of digital modulation scheme where information is split across multiple orthogonal carriers and transmitted in parallel. It can transmit data with high spectral efficiency.

*2 Millimeter waves: Radio frequency band with wavelengths in the range of 1 to 10 mm.

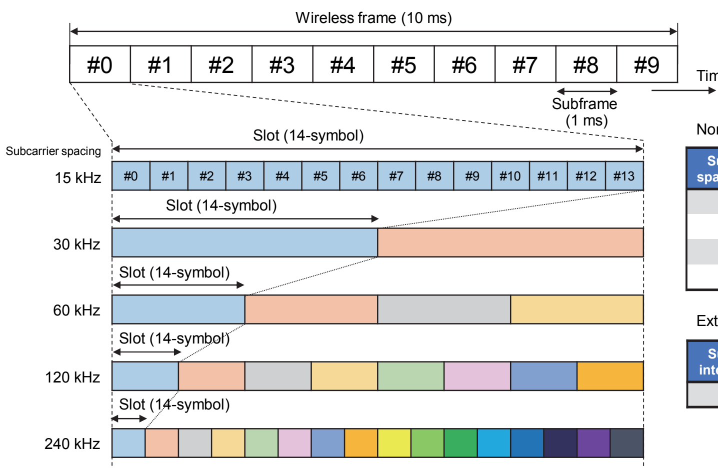
*3 Subcarrier: Each carrier in a multi-carrier modulation system that transmits bits of information in parallel over multiple carriers.

*4 SS: A physical signal enabling the mobile station to detect cell frequency, reception timing, and cell ID in order to begin communications when powering up.

*5 PBCH: A channel for broadcasting essential wireless parameters for receiving control channel and corresponding shared channel (such as system frame number, control channel configuration including sub-carrier spacing and so on).

*6 OFDM symbol: A unit of transmission data consisting of multiple subcarriers. A Cyclic Prefix (CP) is inserted at the front of each symbol.

*7 Slot: A unit for scheduling data consisting of multiple OFDM symbols.



Normal cyclic prefix

Subcarrier spacing [kHz]	No. of symbols per slot	No. of slots per subframe	No. of subframes per frame
15	14	1	10
30	14	2	10
60	14	4	10
120	14	8	10

Extended cyclic prefix

Subcarrier interval [kHz]	No. of symbols per slot	No. of slots per subframe	No. of subframes per frame
60	12	4	10

Figure 1 The relationship between slots, subframes and frames

3. Initial Access and Mobility

A user equipment's initial access in NR is performed according to a procedure that involves the steps of detecting a Synchronization Signal (SS), acquiring broadcast system information[*14], and establishing connection with the network by a random access[*15] procedure.

3.1 SS/PBCH Block

As in LTE, the SS in NR consists of two signals: a Primary Synchronization Signal (PSS)[*16], and a Secondary Synchronization Signal (SSS)[*17]. The SS together with the PBCH and the DeModulation Reference Signals for PBCH (DMRS for PBCH)[*18] forms an SS/PBCH block as shown in **Figure 2**.

The base station uses this SS/PBCH block to provide information that is essential for initial access and mobility, including system parameters whereby user equipment can discover NR cells[*19], establish frame synchronization, measure the downlink reception quality, and carry out other actions necessary for the reception of system information. The base station can set the transmission timing and transmission period of the SS/PBCH block for each carrier, and indicates this information to the user equipment.

In NR, multiple resources for SS/PBCH block transmission are defined within a half frame of 5 ms duration, where the maximum number of SS/PBCH block transmissions per carrier depends on the frequency band. As shown in **Figure 3**, the number of SS/PBCH blocks to be transmitted can be set according to factors such as the base station

*8 Subframe: A unit of radio resources in the time domain, consisting of multiple slots.

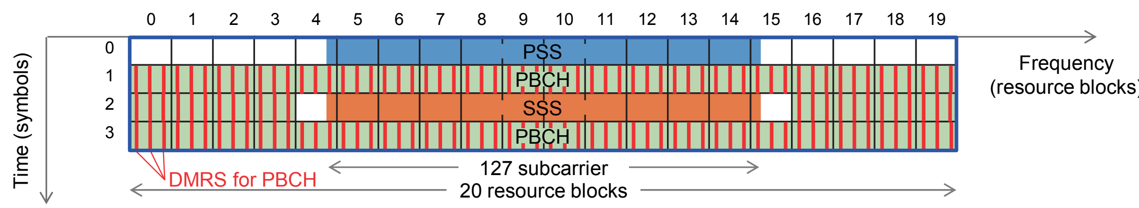
*9 Frame: The period in which an encoder/decoder operates or a data signal of length corresponding to that period.

*10 Duplex mode: A communication scheme where transmission can be performed in the uplink and downlink simultaneously. Generally implemented as Frequency Division Duplex (FDD) or Time Division Duplex (TDD) (see *11).

*11 TDD: A bidirectional transmit/receive system. This system achieves bidirectional communication by allocating different time slots to uplink and downlink transmissions on the same frequency.

*12 Higher layer signaling: In this article, upper layer signaling refers to messages that are transmitted and received in order to control terminals in the Medium Access Control (MAC) layer and higher layers. Examples include Radio Resource Control (RRC) messages and MAC control elements.

*13 L1 signaling: In this paper, L1 signaling refers to messages that are transmitted and received in order to control terminals at layers above the MAC layer. Examples include Downlink Control Information (DCI) and Uplink Control Information (UCI).



Subcarrier spacing of SS/PBCH block	Transmission bandwidth of SS/PBCH block
15 kHz	3.6 MHz
30 kHz	7.2 MHz
120 kHz	28.8 MHz
240 kHz	57.6 MHz

Figure 2 SS/PBCH block configuration

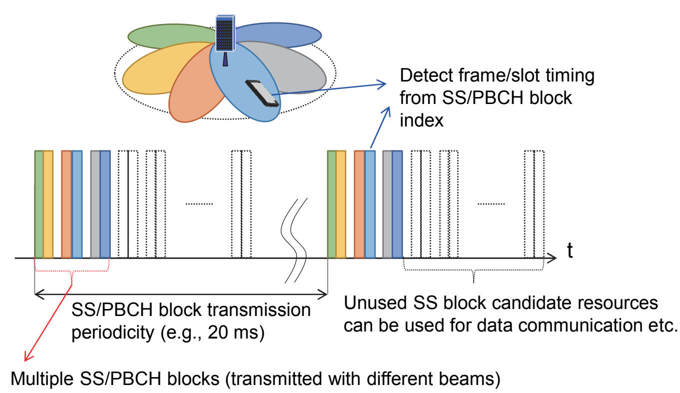


Figure 3 SS/PBCH block transmission configuration example

antenna configuration. With multiple SS/PBCH blocks, different beamforming*20 can be applied to each SS/PBCH block in order to increase the communication range and expand the area of coverage.

3.2 System Information Notification

Broadcast information in NR can be classified into three types: broadcast information transmitted by the PBCH, system information necessary for initial access, and other system information.

The PBCH includes a System Frame Number (SFN)*21 and information that user equipment needs to establish frame synchronization with a NR cell after detecting an SS/PBCH block, such as an index

*14 Broadcast system information: Essential system information (including cell access information required for executing the procedure for connecting mobile terminals to cells, random access channel information and so on) to be broadcast within a cell.

*15 Random access: A procedure executed by mobile terminals and base stations for connecting uplink signals and synchronizing their transmission timing.

*16 PSS: A known signal that the user equipment first searches for in the cell search procedure.

*17 SSS: A known signal transmitted to enable detection of the physical cell ID in the cell search procedure.

*18 DMRS for PBCH: A known signal transmitted to measure the state of a radio channel for PBCH demodulation.

*19 Cell: The smallest unit of area in which transmission and reception of wireless signals is done between a cellular mobile communications network and mobile terminals.

for identifying the symbol position of the detected SS/PBCH block in a half frame. The PBCH also carries system parameters that are needed for the reception of System Information Block[*22] type 1 (SIB 1), which is described below.

To perform random access, it is necessary to have information such as the uplink carrier information and random access signal configuration information. This is included as part of the information necessary for initial access, which is broadcast to the user equipment in an NR cell as SIB1.

3.3 Random Access

Random access in NR is performed in four steps in the same way as in LTE.

At the first step, the user equipment transmits a Physical Random Access CHannel (PRACH)[*23] to the base station. As shown in **Table 1**, NR defines 13 PRACH formats in total, including some formats with fixed subcarrier spacings that follow the design of LTE PRACH, and other formats with variable subcarrier spacings that can fit into the duration of an integer number of symbols in an NR slot.

When a base station is operated with beamforming where different beams are applied to multiple SS/PBCH block transmissions, different PRACH transmission resources are associated with different SS/PBCH blocks. User equipment transmits a PRACH to initiate random access by using the PRACH resource associated with the selected

Table 1　PRACH format

Format number	Sequence length	No. of OFDM symbol repetitions	Subcarrier spacing	Time duration	Bandwidth
0	839	1	1.25 kHz	1 subframe	1.05 MHz
1		2		3 subframe	
2		4		3.5 subframe	
3		4	5 kHz	1 subframe	4.20 MHz
A1	139	2	{15, 30, 60, 120} kHz	2 symbols	{2.09, 4.17, 8.34, 16.68} MHz
A2		4		4 symbols	
A3		6		6 symbols	
B1		2		2 symbols	
B2		4		4 symbols	
B3		6		6 symbols	
B4		12		12 symbols	
C0		1		2 symbols	
C2		4		6 symbols	

*20 Beamforming: A technique for increasing or decreasing the gain of antennas in a specific direction by controlling the amplitude and phase of multiple antennas to form a directional pattern with the antennas.

*21 SFN: A number allocated to each radio frame of 10 ms.

*22 SIB: A specific block of system information broadcast from an base station to mobile terminals. There are multiple types of SIBs.

*23 PRACH: A physical channel used by mobile terminals as an initial transmitted signal in the random-access procedure.

SS/PBCH block. In this way, the base station can use the received PRACH resources to figure out which SS/PBCH block (and thus which beamforming direction) was received by the user equipment transmitting a PRACH. Therefore, in the subsequent random access procedure consisting of random access response reception, connection request message transmission and contention resolution message reception, the base station can use transmission/reception beamforming directed specifically to the user equipment.

3.4 Mobility

In NR, as in LTE, the base station performs tasks such as selecting serving cells, performing handovers and adding/deleting secondary cells based on the report from corresponding user equipment regarding measurement results on downlink reference signals. The SSS included in the SS/PBCH block transmitted by the base station is a basic cell-specific reference signal in NR. The user equipment uses it to measure and report the Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) in each cell according to settings from the base station.

4. MIMO/Beamforming

4.1 MIMO Transmission Method

In the high frequency band, it is very important to obtain a high beamforming gain*24 by using many antennas to compensate for the effects of radio wave attenuation. The high frequency band of NR requires the use of Multiple Input Multiple Output (MIMO)*25 technology that assumes the availability of hybrid beamforming using up to 256 antenna elements at the base station and up to 32 antenna elements at the user equipment.

Hybrid beamforming is a technique that can reduce the cost of implementing circuits for beamforming in high-frequency bands. It consists of digital beamforming to perform signal control in the baseband, and analog beamforming to perform signal control in the RF (Radio Frequency). Since analog beamforming cannot be controlled in sub-band units due to implementation constraints, wideband beamforming is generally used.

To improve the spectral efficiency*26, it is essential to use spatial multiplexing in addition to beamforming. For the downlink, codebook*27 based closed-loop*28 precoding*29 is specified to accommodate up to eight layers of single-user MIMO*30 and up to twelve layers of multi-user MIMO*31. For the uplink, two transmission methods are supported, namely codebook-based and non-codebook-based transmission. It is possible to transmit up to four layers with single-user uplink MIMO. The non-codebook-based transmission method is designed by assuming beam reciprocity*32 is supported in the user equipment. The user equipment receives a downlink reference signal (RS) such as SS/PBCH blocks or Channel State Information RS (CSI-RS)*33, and determines the uplink precoder based on the received downlink RS.

4.2 Reference Signal Structure

The RS structure of NR basically follows that of LTE while achieving the flexibility to adapt to

*24 Gain: One of the radiation characteristics of an antenna. An indicator of how many times larger the radiation strength in the antenna's direction of peak radiation is relative to a standard antenna.

*25 MIMO: A signal transmission technology that improves communications quality and spectral efficiency by using multiple transmitter and receiver antennas to transmit signals at the same time and same frequency.

*26 Spectral efficiency: The number of data bits that can be transmitted per unit time and unit frequency band.

*27 Codebook: A set of predetermined precoding weight matrix candidates.

*28 Closed-loop: A method of using feedback information from receivers.

*29 Precoding: A process for improving the quality of signal reception by multiplying signals before transmission with weights according to the current radio propagation channel.

*30 Single-user MIMO: Technology that uses MIMO transmission at identical temporal frequencies for a single user.

operation in various different frequency bands and scenarios. An example of a downlink RS structure is shown in **Figure 4**.

If an RS is transmitted constantly, it may cause design constraints for possible RS and channel designs in future releases. Therefore, NR tries to avoid specifying an RS that is constantly transmitted (e.g., Cell-specific Reference Signal (CRS) in LTE), and implements function of CRS by using multiple RSs instead.

Specifically, the CSI-RS, DeModulation RS (DM-RS) and Tracking RS (TRS) are specified for channel state information estimation, data demodulation and time/frequency tracking, respectively.

DM-RS enables suitable tracking of a wide range of channel fluctuation speeds and is specified with a structure having a Front-Load (FL) DM-RS mapped in the front part of the data channel as well as additional mapping of 0–3 symbols of Additional (AD) DM-RS in order to suppress overheads[*34].

TRS has the same signal sequence generation with CSI-RS. However, TRS is transmitted in the intervals of 4 subcarriers and 4 OFDM symbols so that sufficient tracking accuracy is achievable with a reasonable overhead.

In high-frequency bands, phase noise[*35] would be a serious issue. Therefore, in NR, Phase-Tracking Reference Signal (PT-RS) is newly specified as a UE-specific reference signal.

4.3 Beam Control Techniques

Beam control in L1/L2[*36] can be divided into beam management and CSI acquisition.

Beam management is a particularly effective technique at high frequencies and is generally aimed at establishing and maintaining transmitting/

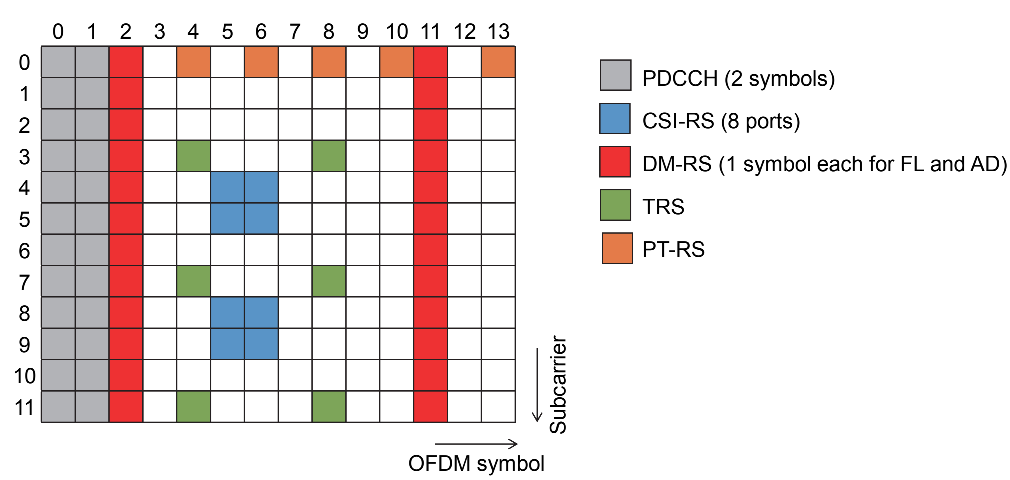


Figure 4 Example of downlink reference signal configuration

*31 Multi-user MIMO: Technology that uses MIMO transmission at identical temporal frequencies for multiple users.

*32 Beam reciprocity: A technique whereby radio equipment determines which transmitting beam to use based on information derived from the received beam.

*33 CSI-RS: A downlink reference signal used by mobile terminals to measure the state of the radio channel.

*34 Overhead: Control information needed for transmitting/receiving user data, plus radio resources used for other than transmitting user data such as reference signals for measuring received quality.

*35 Phase noise: Phase fluctuation that occurs due to frequency components other than those of the carrier frequency in a local oscillator signal.

*36 L2: The second layer of the OSI reference model (data link layer).

receiving analog beam pairs between the base station and user equipment. For example, the user equipment compares the L1-Reference Signal Received Power (RSRP)*37 of multiple SS/PBCH blocks and CSI-RS to which different beams have been applied by the base station, and selects a suitable transmit beam to be reported to the base station. The base station reports the beam information applied to the downlink channel, so that the user equipment can select the corresponding reception beam to receive the downlink channel. A beam failure recovery technique is also specified, whereby user equipment that detects deterioration in the characteristics of a base station beam can request a switch to a different beam.

On the other hand, CSI acquisition is used for purposes such as determining the choice of transmission rank*38, digital beams and Modulation and Coding Scheme (MCS)*39. The codebook used for digital beam control is specified as Type I and Type II, which have relatively low and relatively high quantization granularity*40, respectively. In Type II, information about two beams and their linear combination*41 information is reported to the base station, enabling beam control with higher spatial granularity.

5. Scheduling/HARQ

In NR, as in LTE, the downlink data channels and uplink data channels are scheduled based on the Downlink Control Information (DCI)*42. The DCI is transmitted and received via the Physical Downlink Control CHannel (PDCCH)*43. As in LTE, it is possible to use the frequency domain resource assignment field included in DCI to allocate frequency-domain resources in resource block units. In addition, in NR, the DCI is also able to indicate the time-domain resources for data channel scheduling.

The base station allocates a Physical Downlink Shared CHannel (PDSCH) to the user equipment and uses PDCCH to transmit downlink control information for the PDSCH. The user equipment receives and demodulates the PDCCH, and receives the PDSCH based on this downlink control information in the PDCCH. After receiving the PDSCH, the user equipment sends a Hybrid Automatic Repeat reQuest ACKnowledgment (HARQ-ACK)*44 to feed back the results of decoding. HARQ-ACK is transmitted via the Physical Uplink Control CHannel (PUCCH)*45, as are Scheduling Requests (SR)*46 and CSI. The PUCCH transmission timing and resources can also be indicated by the DCI in the same way as the data channel. **Figure 5** shows an example of data and HARQ-ACK resource allocation.

In HARQ retransmissions, in addition to the method whereby the entire transport block as initially transmitted and received is retransmitted, another method called code-block-group-based retransmission is specified. In code-block-group-based retransmission, when a transport block consists of multiple code blocks, only the code block groups that contain errors are retransmitted instead of retransmitting the entire transport block.

In general, uplink data scheduling is followed by scheduling request procedure, in which the user equipment transmits a SR using configured SR resources (**Figure 6** (a)). In this scheme, however,

*37 **RSRP**: Received power of a signal measured at a receiver. RSRP is used as an indicator of receiver sensitivity of a mobile terminal.

*38 **Transmission rank**: The number of layers (spatial streams) transmitted simultaneously in MIMO.

*39 **MCS**: Combinations of modulation scheme and coding rate decided on beforehand when performing Adaptive Modulation and Coding.

*40 **Quantization granularity**: The spatial granularity of beams that are capable of being formed.

*41 **Linear combination**: The linear sum of vectors. The vectors are multiplied by constant factors and added together.

*42 **DCI**: Control information transmitted on the downlink that includes scheduling information needed by each user to demodulate data and information on data modulation and channel coding rate.

*43 **PDCCH**: Control channel for the physical layer in the downlink.

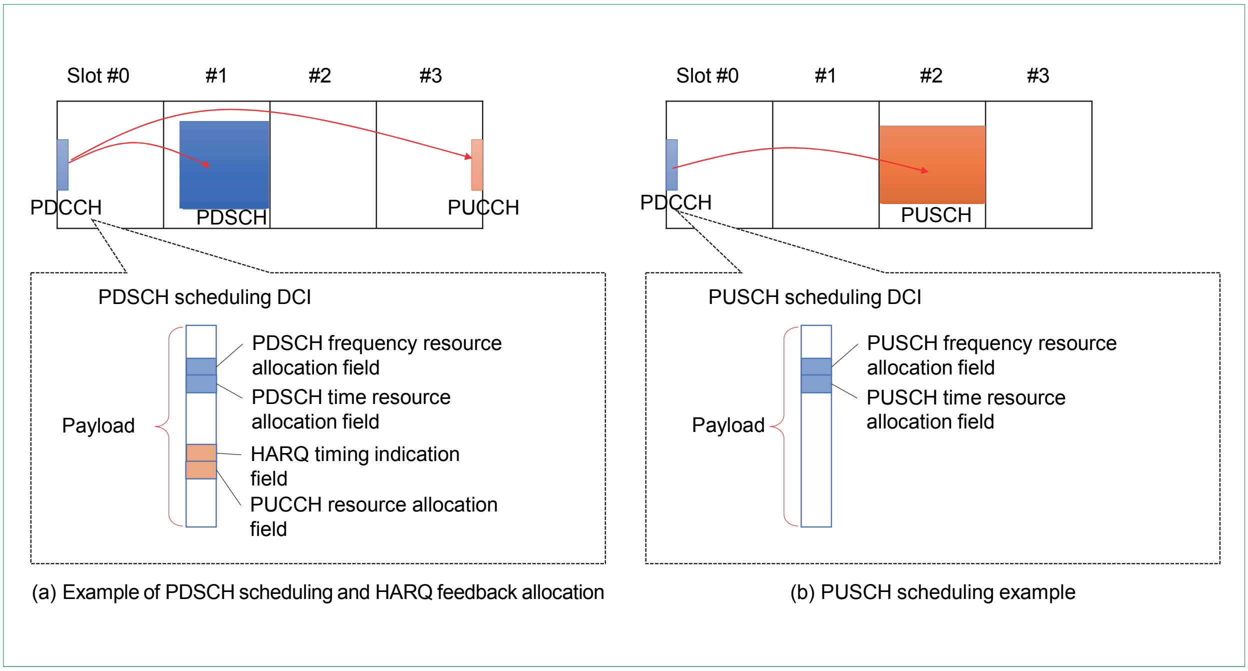


Figure 5　Example of resource allocation for data channel and HARQ-ACK

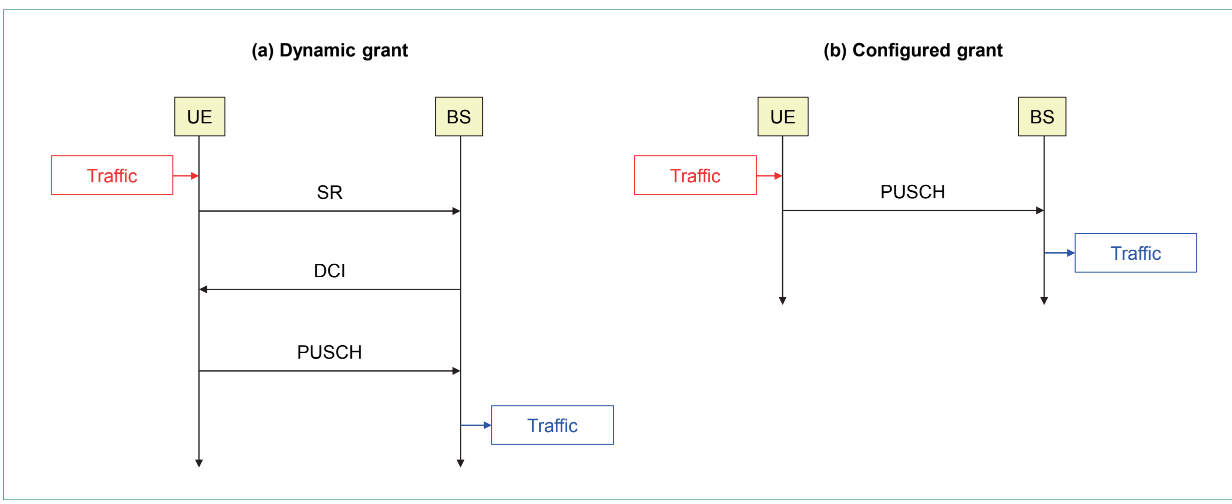


Figure 6　Dynamic grant and configured grant

there is an unavoidable delay due to the scheduling request step. Therefore, NR supports configured grants, whereby a Physical Uplink Shared CHannel (PUSCH)*47 resource is configured to a user equipment, and the user equipment can transmit a PUSCH with this allocated resource when uplink

*44　HARQ-ACK: A receive acknowledgment signal whereby a receiving node can tell the sending node whether or not the data was successfully received (decoded).

*45　PUCCH: Physical channel used for sending and receiving control signals in the uplink.

*46　SR: A signal from the user to the base station requesting radio resource allocation for uplink.

*47　PUSCH: Physical channel used for sending and receiving data packets in the uplink.

data traffic arrives without performing a scheduling request procedure (Fig. 6 (b)).

6. Modulation and Channel Coding Schemes

Tables 2 and **3** show the modulation and channel coding schemes that can be used for NR downlink and uplink communication, respectively.

1) Modulation Schemes

As the primary modulation scheme*48, in addition to the schemes specified since Release 8 LTE, 256QAM and π/2-BPSK that are available in the latest LTE are also supported in NR.

For the secondary modulation scheme*49, the same scheme as in LTE — Orthogonal Frequency Division Multiplexing with a Cyclic Prefix (CP-OFDM*50) — is applied to the downlink channels. For the uplink channels, it is possible to apply

Table 2 Modulation schemes

Secondary modulation scheme	Primary modulation scheme	Downlink	Uplink
CP-OFDM	π/2-BPSK	—	—
	BPSK	—	PUCCH format 1
	QPSK	PBCH, PDCCH, PDSCH	PUCCH format 1/2, PUSCH
	16QAM	PDSCH	PUSCH
	64QAM	PDSCH	PUSCH
	256QAM	PDSCH	PUSCH
DFTS-OFDM	π/2-BPSK	—	PUCCH format 3/4, PUSCH
	QPSK	—	PUCCH format 3/4, PUSCH
	16QAM	—	PUSCH
	64QAM	—	PUSCH
	256QAM	—	PUSCH

Table 3 Channel coding

	Downlink	Uplink
LDPC	DL-SCH, PCH	UL-SCH
Polar coding	BCH, DCI	UCI (payload size ≧ 12)
Block coding	—	UCI (payload size < 12)

*48 Primary modulation scheme: The method used to embed a digital data stream on a radio carrier wave.

*49 Secondary modulation scheme: The method for further manipulation of the primary modulated data for purposes such as frequency spreading.

*50 CP-OFDM: An OFDM method that adds a guard time (CP) between symbols to suppress inter-symbol interference caused by multipath propagation or the like.

both Discrete Fourier Transform Spreading OFDM (DFTS-OFDM), which suppresses the Peak-to-Average Power Ratio (PAPR)*51 to allow broader coverage, and the same CP-OFDM scheme as used in the downlink. This enables the use of unified access schemes in the downlink and uplink, resulting in simpler systems.

2) Channel Coding Schemes

Low Density Parity Check (LDPC) coding and polar coding are specified as the channel coding schemes in NR, in addition to the block codes that have already been used in LTE.

The LDPC coding scheme used in the data channel (DL-SCH: DownLink Shared CHannel, UL-SCH: UpLink Shared CHannel, PCH: Paging CHannel) is able to perform parallel processing to minimize decoding latency, and can thus operate close to the Shannon limit*52. The polar coding scheme used in the control channel (DCI, UCI/uplink control information with a payload size of 12 or more) and BCH (Broadcast CHannel) has less decoding computation complexity compared with the convolutional code used in LTE (TBCC: Tail-Biting Convolutional Coding*53), while exhibiting superior characteristics that approach the Shannon limit. The same block codes as used in LTE are applied to UCIs with a small payload size.

7. Transmit Power Control

7.1 Transmit Power Control Taking Beam Control into Consideration

In NR, since transmitting/receiving beams can be formed in both the base station and the user equipment, transmission power control*54 is specified by taking into consideration the beamforming operation, which causes beamforming gain variation when switching the transmitting/receiving beam.

The user equipment can use reference signals to which the transmitting/receiving beams are applied in order to estimate the path loss*55 including the beam gain. Specifically, the path loss is estimated by comparing the transmit power before applying the transmit beam with the received power after applying the receiving beam. By applying different transmitting/receiving beams to multiple reference signals, the reference signal to be used can be dynamically switched to adapt to path loss fluctuations caused by switching of the transmitting/receiving beams.

Also, as in LTE, fractional transmission power control is specified as a way of reducing the transmit power at cell edges where the path loss is large. For each transmitting/receiving beam, the power control related parameters can be set differently so that the path loss compensation and power offset can be optimized.

7.2 Distribution of Transmission Power between LTE and NR

Release 15 specifies an upper limit of transmit power based on guidelines for the protection of humans. For the case of simultaneous transmissions of LTE and NR in the same frequency range, the total power between LTE and NR shall be within a specified range of values [1].

In the Release 12 LTE DC, a user equipment can be configured with a minimum guaranteed power

---

*51 PAPR: As the ratio of maximum power to average power, an index expressing the peak magnitude of the transmit waveform. If this value is large, the amplifier power back-off has to be large to avoid nonlinear distortion, which is particularly problematic for mobile terminals.

*52 Shannon limit (also known as "Shannon communication-channel capacity"): Theoretically derived from bandwidth and Signal-to-Noise (SN) ratio, the maximum amount of information that can be transmitted.

*53 TBCC: A type of error-correcting code; namely, a coding scheme that generates codewords by using convolution calculation.

*54 Transmission power control: A technique of controlling transmission power such that the Signal-to-Noise Ratio (SNR) and Signal-to-Interference and Noise Ratio (SINR) at the receiver exceed the required values.

*55 Path loss: Propagation path loss estimated from the difference between the transmitted power and received power.

for each CG (cell group), and the user equipment dynamically distributes the transmit power between CGs such that each CG can at least use the minimum guaranteed power. However, it is difficult for early commercial user equipment to support this, since it requires dynamic power sharing between LTE and NR transmitters. Therefore, Release 15 specifies a scheme whereby the base station semi-statically configures the maximum transmission power for each CG. The behavior of user equipment for power control could be different depending on whether the user equipment supports dynamic power sharing between LTE and NR.

When the user equipment has a dynamic power sharing capability, it adjusts the NR transmission power so that the instantaneous total power does not exceed the specified value. An example of dynamic power distribution is shown in **Figure 7**. When the maximum transmission power of LTE and NR (P_{LTE}, P_{NR}) and the specified total power

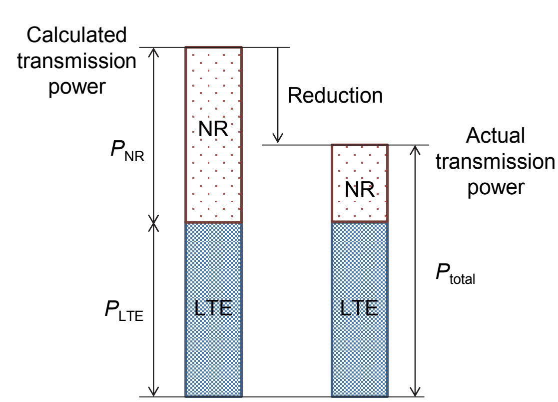


Figure 7 Distribution of transmission power between LTE and NR

(P_{total}) have been set and the instantaneous calculated total transmission power exceeds P_{total}, the NR transmission power is reduced so that the actual transmission power of the user equipment does not exceed P_{total}.

Note that the base station is able to configure the LTE and NR maximum powers so that their total exceeds the specified value. However, if user equipment does not have the ability to perform dynamic power sharing, its uplink transmissions will be performed using time division multiplexing to switch between LTE and NR.

8. BWP/CA

In NR, the maximum bandwidth per carrier is much larger than in LTE: 100 MHz at frequencies below 6 GHz and 400 MHz at higher frequencies. For carriers operated with such a large bandwidth, NR supports the BandWidth Part (BWP) concept whereby user equipment can use smaller bandwidths than the carrier bandwidth used by the base station.

The base station uses higher-layer signaling to set up a BWP configuration (bandwidth, frequency position, subcarrier spacing) for the user equipment to use during communication. Different user equipment can be configured with different BWP configurations (**Figure 8**). The BWP configuration can be changed by higher layer signaling or L1 signaling. Therefore, even when the user equipment supports a sufficiently wide bandwidth, it is possible to set a narrower BWP configuration when there is no data traffic, thereby reducing the communication

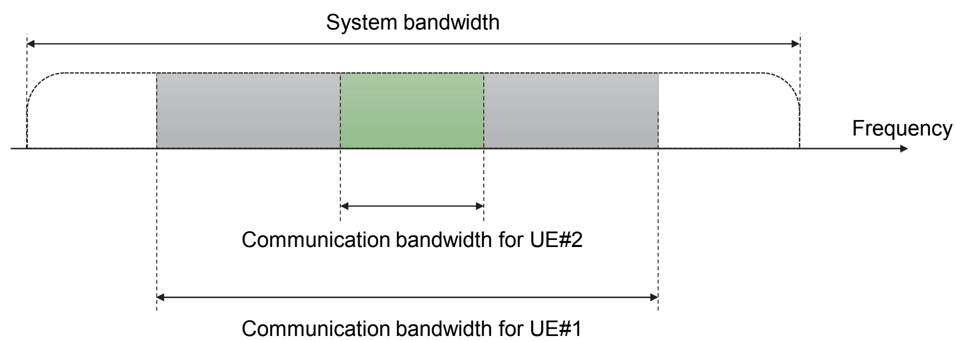


Figure 8 Bandwidth of an individual UE

bandwidth and power consumption.

In NR, Carrier Aggregation (CA)*56 is defined in the same way as in LTE. When CA is performed, BWP is configured for each Component Carrier (CC)*57. CA using CCs with different subcarrier spacings is also supported, making it possible to efficiently aggregate a wide range of frequencies ranging from existing cellular frequencies to millimeter waves.

9. Conclusion

This article has presented the NR Release 15 specifications, and the newly introduced functions in the physical layer. These functions make it possible to deliver the high speed, large capacity and low latency services that are expected for 5G. NTT DOCOMO will continue to promote the standardization activities of 3GPP to make further advances for 5G technologies to provide even lower latency and higher reliability for IoT applications, reduce power consumption, and improve direct communication between terminals.

REFERENCE

[1] 3GPP TS38.101-1 V15.3.0: "NR; User Equipment (UE) radio transmission and reception," Sep. 2018.

*56 CA: A technology for increasing bandwidth and data rate by simultaneously transmitting and receiving signals for one user using multiple carriers.

*57 CC: Term denoting each of the carriers used in CA.

Technology Reports | 5G | NR | NR Higher Layer

Special Articles on Release 15 Standardization
—Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

Specifications of NR Higher Layer in 5G

Radio Access Network Development Department Toru Uchino Kenji Kai
Teruaki Toeda Hideaki Takahashi

In June 2018, the 3GPP finalized the NR specifications of wireless communication for 5G networks. This paper describes the specifications of the NR higher layer that are required for non-standalone operation (by NR in combination with existing LTE/LTE-Advanced systems) and standalone operation (by NR alone). First, we describe a standardized architecture and a bearer type defined for non-standalone operation involving the simultaneous use of LTE and NR radio links. Then, among the main functions of the higher layer, we describe the layer 2/3 protocol functions not mentioned in the special issue of October 2017, and also explain the network interface used in the base stations providing NR.

1. Introduction

Since March 2016, the 3rd Generation Partnership Project (3GPP) has been studying 5th Generation (5G) mobile communication systems. A feature article in 2017 outlined the discussions of the Study Item (SI)*1 on feasibility studies [1] [2]. In March 2017, 3GPP shifted to the Work Item (WI)*2 phase based on the issues considered and agreed upon at the SI stage, and work began on the actual standard specification process.

The Release 15 standard specifications include

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 SI: Work on a feasibility study and an identification of high-level features to be specified.

*2 WI: Work that involves determining all functions to be specified and formulating detailed specifications for those functions.

one for non-standalone operation, in which New Radio (NR) and existing LTE/LTE-Advanced networks are combined, and one for standalone operation, in which NR can be operated alone. They were released in December 2017 and June 2018, respectively. This article describes NR's higher layer specifications as standardized by 3GPP.

2. Standardized Architecture and Bearer Types

2.1 Architecture

As stated above, 3GPP Release 15 provides specifications for both non-standalone operation and standalone operation for NR. Non-standalone operation is where services are provided in combination with existing LTE/LTE-Advanced networks. In contrast, stand-alone operation is where NR services can be operated alone. **Figure 1** shows the configuration of a Radio Access Network (RAN)*3 for these operations. In the Core Network (CN)*4, the Evolved Packet Core (EPC)*5, which was conventionally used for LTE, is used for non-standalone operation, while a new 5G Core (5GC) is used for standalone operation.

To address the demands from operators wishing to introduce 5G as early as possible, 3GPP released the 5G standard specifications in several

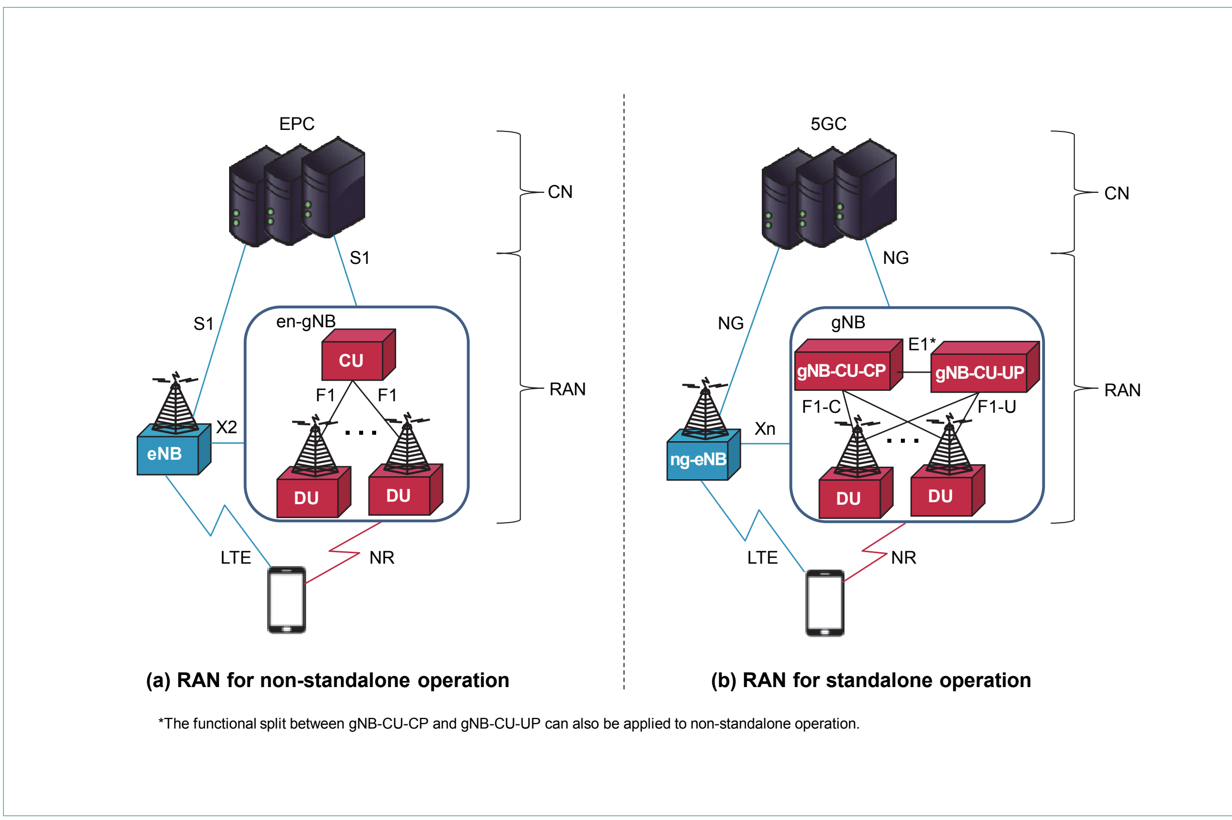


Figure 1 5G RAN configurations

*3 RAN: A network consisting of radio base stations and radio-circuit control equipment situated between the CN (see*4) and mobile terminals.

*4 CN: A network comprising switching equipment, subscriber information management equipment, etc. A mobile terminal communicates with the core network via a radio access network.

*5 EPC: A core network that can accommodate diverse radio access systems including LTE.

stages. First, for non-standalone operation, the specifications for higher layer*6 functions and the network interface*7 between LTE base stations (evolved Node B (eNB)) and NR base stations (gNB) were released in December 2017. Then in June 2018, for standalone operation, specifications for the additional functions and network interfaces were released, together with an LTE extension for connecting to 5GC via LTE. In December 2018, the specifications for a terminal that simultaneously communicates with multiple RAN nodes (LTE-NR Dual Connectivity (DC)*8 and NR-NR DC) with 5GC were released.

2.2 Bearer Types in Non-standalone Operation

The bearer type in non-standalone operation is based on what was specified for DC in Release 12 LTE and was extended to facilitate bearer control with greater flexibility. **Figure 2** shows the bearer types prescribed for non-standalone operation.

In LTE-DC, a split bearer is defined as one that simultaneously uses the radio resources*9 of multiple base stations (Master Nodes (MN)*10 and Secondary Nodes (SN)*11). However, in this case, the node terminating the bearer on the network side is limited to MN [2]. In non-standalone operation,

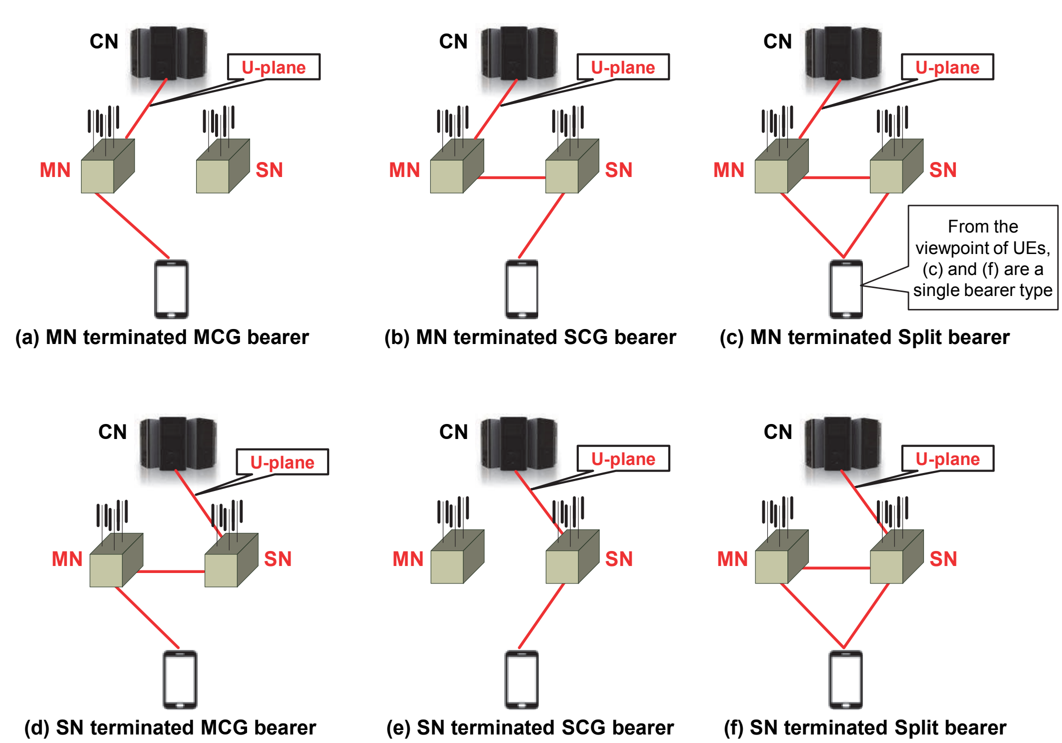


Figure 2 Bearer types in non-standalone operation

*6 **Higher layer:** All layers positioned above the physical layer, namely, layers such as MAC (see *44), PDCP (see *22), RLC (see *20), S1AP, and X2AP.

*7 **Network interface:** An interface used in RAN, between RAN-CN, and in CN (X2, S1, Xn, NG, F1, E1, etc.).

*8 **DC:** A technology that achieves wider bandwidths by connecting two base stations in a master/secondary relationship and performing transmission and reception using multiple component carriers supported by those base stations.

*9 **Radio resource:** General term for resources needed to allocate radio channels (frequencies). This can include radio transmission power, TRX resources, BB channels, and RLC (see *20) resources.

*10 **MN:** In DC, the base station that establishes an RRC connection with the UE. In LTE-NR DC, this could be the LTE base station (eNB) or the NR base station (gNB).

eNB is used as the MN to ensure connectivity between the terminal and the network via LTE wireless communication. However, if the same definition as the split bearer in LTE-DC is used for non-standalone operation, all the U-plane*12 data transmitted and received in LTE and NR will have to be processed and routed by eNB. In the standardization discussions, it was pointed out that this approach is likely to limit the wireless transmission rate due to a bottleneck in the processing performance of eNB equipment. To mitigate this issue, a new split bearer that terminates the bearer in SN (the Secondary Cell Group (SCG) split bearer in Fig. 1 of Reference [2]) was specified. In order to introduce this new bearer type, further discussions were held regarding the unification of bearer types in order to reduce the variation of bearer types from the viewpoint of the User Equipment (UE) and further expansion to facilitate more efficient operation.

1) Bearer Type Unification

With the introduction of a new bearer type, the specifications now define four bearer types from the viewpoint of the UE. When different operators can choose different bearer types and multiple bearer types are implemented to meet the needs of these multiple operators, the cost of the UE increases. Therefore, in order to minimize the variation of bearer types, an attempt was made to unify the types. Specifically, the U-plane protocol stacks*13 used in split bearers terminated by MN (Fig. 2 (c)) and SN (Fig. 2 (f)) are almost identical in terms of the functions and operations provided at Layer 2. In this sense, from the viewpoint of UE, there is no need to distinguish between them and they can be regarded as the same bearer type. By specifying these as a single split bearer type in the standard specification, it becomes possible to reduce the variation of bearer types as seen from the viewpoint of UE.

2) Configuration of Bearers in the Network for Efficient Operation

In a split bearer terminated with SNs (Fig. 2 (f)), EN-DCs*14 are set up and released when the user enters or leaves an SN area, whereupon the bearer termination nodes on the network side are switched between MN and SN. Every time this happens, control signaling for the path switch*15 is generated between the base station and CN, which is one of the issues discussed in the standardization process. In particular, during the early stage of the introduction of NR, since it is assumed that NR would be operated at hotspots*16 using high-frequency millimeter wave*17 signals, this sort of switching is likely to occur frequently. In an attempt to avoid these path switches, a plan was considered whereby even if a user leaves an SN area, the base station does not release the EN-DC configuration from the UE and keeps the bearer termination in the SN on the network side (with the U-plane data being transmitted and received on the LTE side). However, in this case, since the UE continues to search for an NR cell*18 outside the NR area, its battery power will be wasted.

Therefore, an inter-network procedure*19 was specified such that bearer termination nodes can be freely set on the network side, regardless of whether an EN-DC is configured for the UE. For example, when using a split bearer (Fig. 2 (f)) terminated by SN in EN-DC, the radio bearer is ter-

---

*11 SN: A base station that provides a UE in DC with radio resources in addition to those provided by the MN. In LTE-NR DC, the SN is an NR base station (gNB) if the MN is an LTE base station (eNB) and an LTE base station (eNB) if the MN is an NR base station (gNB).

*12 U-plane: The protocol for transmitting user data.

*13 Protocol stack: Protocol hierarchy.

*14 EN-DC: An architecture for NR non-standalone operation. Performs RRC connection with LTE wireless, and also uses NR as an additional wireless resource.

*15 Path switch: The process of changing paths taken by routes used to transmit/receive data between CNs and base stations.

*16 Hotspot: A place where traffic is generated in concentrated form, such as the plaza or square in front of a train station.

*17 Millimeter waves: Radio frequency band with wavelengths in the range of 1 to 10 mm.

minated at the SN. When the UE leaves the SN area, the base station deletes the radio resource setting (up to the Radio Link Control (RLC)*20 layer) on the SN side of the UE, whereby the UE can only use LTE for data transmission and reception. However, from the viewpoint of the network, since the SN continues to be used as a bearer termination node (as shown in Fig. 2 (b)), it does not generate control signals for switching paths to the CN.

In this way, since the flow of U-plane data can adopt a flexible configuration in non-standalone operation from the viewpoint of the network, nomenclature rules have been prepared to deal with this situation. Specifically, the U-plane bearer type is represented by combining two elements as follows.

- Distinguishing between MN-terminated and SN-terminated bearers according to the type of bearer termination node (MN or SN).
- Distinguishing between master cell group (MCG) bearers, SCG bearers, and split bearers according to which node's radio signals are used to transmit and receive data (MN radio only, SN radio only, or both).

As examples, Fig. 2 (c) shows an MN terminated split bearer, and Fig. 2 (d) shows an SN terminated MCG bearer.

3. Main Features of the Higher Layer in NR

This section describes the NR C-plane*21 and U-plane protocol stacks specified in 3GPP standards, and the functions of these protocol stacks.

3.1 C-plane

Figure 3 shows the NR C-plane protocol stack. The C-plane protocol stack is similar to that of LTE, and the protocols below the Packet Data Convergence Protocol (PDCP)*22 use the same protocols as the U-plane. The NR Radio Resource Control (RRC)*23 protocol introduces a number of improvements based on the LTE RRC protocol. **Table 1**

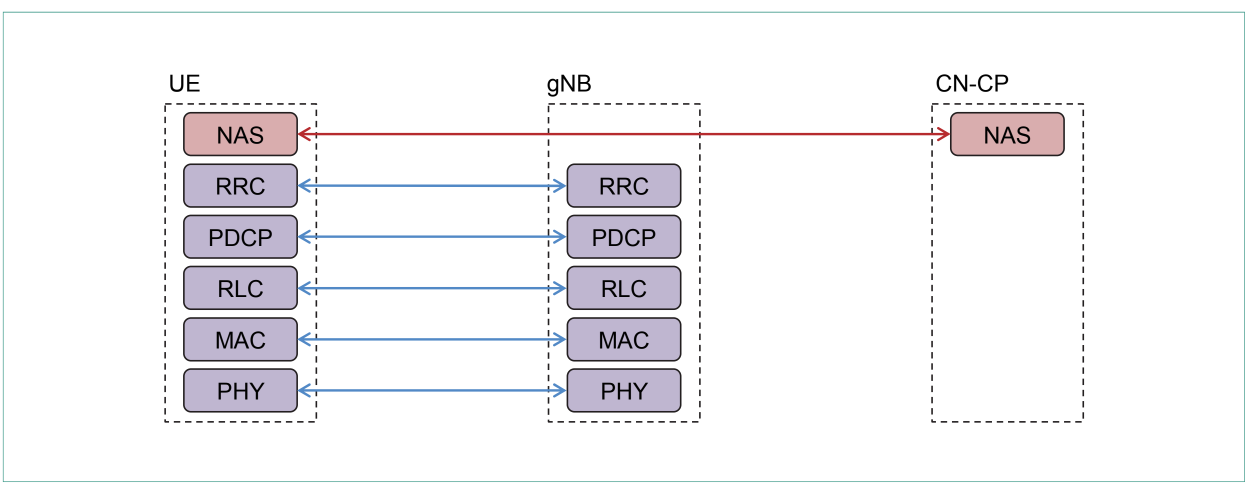


Figure 3　NR C-plane protocol stack

*18 **Cell:** The smallest area unit for sending and receiving radio signals between mobile communication network and mobile terminals.

*19 **Procedure:** A signal processing procedure implemented between base stations, between a base station and a CN, or between a base station and a terminal.

*20 **RLC:** A sublayer of Layer 2 (see *43). A protocol that performs services such as retransmission control.

*21 **C-plane:** A Protocol controls radio resource in RAN.

*22 **PDCP:** A sublayer of Layer 2 (see *43). A protocol that performs ciphering, integrity check, reordering and header compression, etc.

*23 **RRC:** A protocol that controls radio resources in a radio network.

Table 1 List of functions in NR RRC protocol

Function	Sub function	Non-standalone	Standalone	Difference from LTE RRC
System information	Broadcast of minimum system information	✓	✓	—
	Broadcast of other information		✓	Introduction of on-demand/area provision
Connection control	Bearer/cell settings	✓	✓	Introduction of split SRB/direct SRB
	Connection establishment with CN		✓	Introduction of RRC_INACTIVE state
	Paging		✓	Introduction of RAN paging
	Access control		✓	Introduction of unified access control
Mobility	Handover		✓	—
	Cell selection/reselection		✓	—
Measurement	DL quality measurement/reporting	✓	✓	Introduction of beam measurements
	Cell identifier measurement/reporting	✓	✓	—

shows the functional classification in NR RRC, the availability of each function for non-standalone/standalone operation, and the differences from the LTE RRC functions. In the following, the NR RRC functions will be described in terms of how they differ from the LTE RRC functions. For the details of split SRB*24/direct SRB*25 and RRC_INACTIVE*26 state, see the 2017 special article [2].

1) On-demand/area Provision of System Information*27

The RAN broadcasts system information that is needed when the UE is camped on a serving cell (e.g., frequency information of the serving cell and access control information) and other common control information (e.g., cell reselection information, frequency information of the neighboring cell, and public warning information). The UE acquires the system information from the RAN and stores this information during a certain time period.

In LTE, the eNB broadcasts all the information at regular intervals, but the UE may not necessarily acquire and store all of this information. For example, a stationary UE does not need information about neighboring cells, and a UE that does not support inter-Radio Access Technology (RAT) cell reselection*28 does not need the corresponding frequency information. Based on this background, the gNB needs to broadcast only the Minimum System Information (MSI) and is available for on-demand provision of Other System Information (OSI). Specifically, the gNB can use MSI to indicate whether each OSI is currently being broadcast or providing in an on-demand way. Thus, a UE that needs OSI provided on-demand transmits a system information

*24 Split SRB: A bearer for duplicating RRC messages generated by MN for terminals in DC, and transmitting them via SN.

*25 Direct SRB: A bearer whereby the SN can send RRC messages directly to terminals in DC.

*26 RRC_INACTIVE: A terminal state in RRC where the terminal does not have cell level identification within the base station, and where the context of the terminal is held in the base station and the core network.

*27 System information: Information that includes the location code, which is required to decide whether location registration is needed for a mobile terminal, serving and neighboring cell information, access control information, and so on. Base stations broadcast this in each serving cell.

*28 Inter-RAT cell reselection: Reselection process of cells belonging to different RATs.

request to the gNB and can then obtain the corresponding OSI from the gNB by dedicated signaling*29. Since broadcasting continuously consumes radio resources, the use of on-demand to provide the OSI with low frequency can reduce the usage of radio resources for system information.

Furthermore, although system information in LTE is broadcast on a cell-by-cell basis and has to be acquired by UE every time the serving cell changes, NR RRC allows different cells to share the same OSI. This means the gNB can use MSI to indicate an area using the same OSI. Thus, when the UE moves the serving cell within the area, it is possible to skip the system information acquisition process when the UE has already stored valid OSI acquired from another cell. This is expected to reduce the power consumption of UE.

2) RAN Paging*30

In LTE, the CN records the UE position information over an area consisting of one or more cells, called a Tracking Area (TA)*31. When downlink data occurs for a specific UE in the RRC_IDLE*32 state, the CN transmits paging messages simultaneously via all the eNBs in the TA (**Figure 4** (a)).

In NR, it is assumed that more UEs can be accommodated than in LTE and that the newly defined RRC_INACTIVE state can be applied. When the same method as LTE is applied to UEs in the RRC_INACTIVE state, there is likely to be an increase in the signaling overhead between the CN and RAN and in the consumption of radio resources for sending multiple paging messages. To resolve this issue, it is specified that for UEs in the RRC_INACTIVE state, the RAN should record the UE

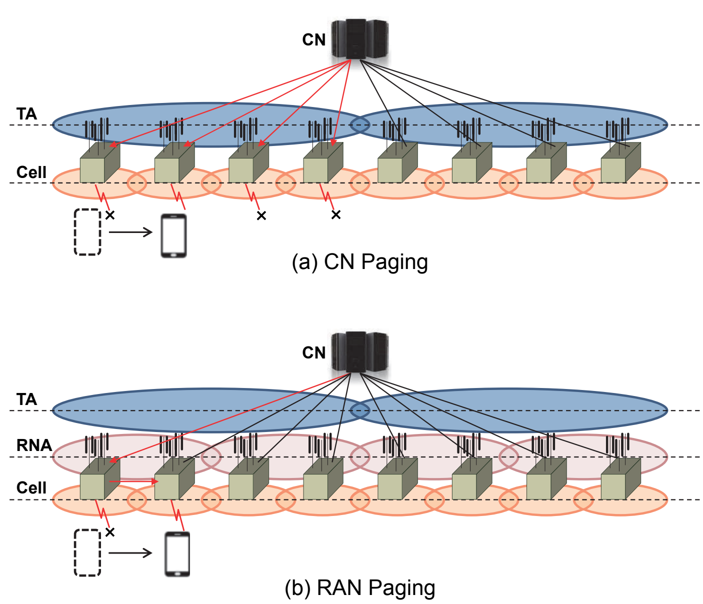


Figure 4 CN paging and RAN paging

*29 Signaling: Control signals used for communication between terminals and base stations.

*30 Paging: A procedure and signal for calling a UE while camped in a cell in standby mode at the time of an incoming call.

*31 TA: A cell unit expressing the position of a mobile terminal managed on a network, and composed of one or more cells.

*32 RRC_IDLE: A UE RRC state in which the UE has no cell-level identity within the base station and the base station stores no UE context. The core network stores UE context.

position information instead of the CN. Specifically, NR RRC defines a new area that is a subdivision of a TA, called a RAN Notification Area (RNA). A gNB that has received downlink data for a UE in the RRC_INACTIVE state from the CN transmits paging messages simultaneously via all the gNBs in the RNA, which is expected to reduce the signaling overhead between the CN and RAN and the usage of radio resources (Fig. 4 (b)).

3) Unified Access Control

In LTE, the eNB performs access control for restricting connection request signals from UEs and ensuring the connectivity of emergency calls and other important calls while maintaining network equipment by burst traffic. We expect to adopt a similar method in NR. However, since the access control in LTE has been extended along with the functional expansion of the standard specifications, it now includes a mixture of access control specifications, including Access Class Barring (ACB)*33, Service Specific Access Control (SSAC)*34, Extended Access Barring (EAB)*35, and Access Control for general Data Connectivity (ACDC)*36 [5]. This has made the specification much more complicated.

To address this issue, NR defines a Unified Access Control (UAC) specification that unifies the access control specification of LTE. In UAC, each access attempt in UE is mapped to a single access category and one or more access identities, and the RAN is able to control each combination separately.

Table 2 shows the access categories and access identities specified by NR.

- An access category is a service type identifier. Different numbers are assigned to each type of access attempt in the UE, and a region

Table 2 Access categories and identities

Access Category	Type of access attempt
0	Mobile originating signaling resulting from paging
1	Delay tolerant access
2	Emergency call
3	Mobile originating signaling resulting from other than paging
4	Voice call
5	Video call
6	SMS
7	Mobile originating data that do not belong to any other access category
8 ～ 31	Undefined
32 ～ 63	Operator specific definitions

Access Identity	Call type
0	Other than the following
1	MPS call
2	MCS call
3 ～ 10	Undefined
11	AC11
12	AC12
13	AC13
14	AC14
15	AC15

*33 ACB: A method for restricting burst-like connection-request signals during natural disasters or major events (e.g., New Year celebrations, fireworks festivals). The terminal itself evaluates whether or not restriction is required using network-notified access control parameters for each Access Class (AC) that individual terminals belong to, and refrains from sending connection-request signals if its access is indeed being restricted.

*34 SSAC: A method for restricting connection request signals for non-emergency voice calls and video calls.

*35 EAB: A method for restricting connection request signals in a Machine Type Communication terminal.

*36 ACDC: A method for restricting connection request signals from individual applications in a terminal.

that can be independently defined by the operator is also specified.

- An access identity is a call type identifier. Different numbers are assigned to Multimedia Priority Service (MPS) calls*37, Mission Critical Service (MCS) calls*38, and high priority calls (access classes 11–15). Other types of call are uniformly assigned the number zero.

For more details of UAC, the RAN can use MSI to notify the access control parameter, including the combination of an access category and one or more access identities to be restricted. The UE then judges whether or not restriction is required on the basis of the access category and access identity every time an access attempt is requested. If restriction is required, the UE refrains from sending the corresponding message during a specified time period calculated by the access control parameter in MSI. In LTE, UEs in the RRC_IDLE state are mainly subject to access control, but in NR, UEs in all RRC states can be restricted to enable more finely tuned access control.

4) Beam Measurement

The RAN instructs a UE to measure the downlink quality of the cell where the UE is located and of its neighboring cells. On the basis of the quality information reported by the UE, the RAN can perform handover*39 to another cell with better quality or adjust the cell settings to improve the quality and/or throughput.

In LTE, downlink quality measurement and reporting is performed in cell units. In NR, we assume a beamforming*40 environment with massive Multiple Input Multiple Output (MIMO)*41 for high speed and large capacity [6]. In NR, we can also instruct and report in beam units as well as in cell units. This makes it possible to implement handover to appropriate cells and/or beams even in beamforming environments.

3.2 U-plane

The U-plane protocol stack of NR is shown in **Figure 5**. It is based on LTE, and a Service Data Adaptation Protocol (SDAP)*42 layer is provided

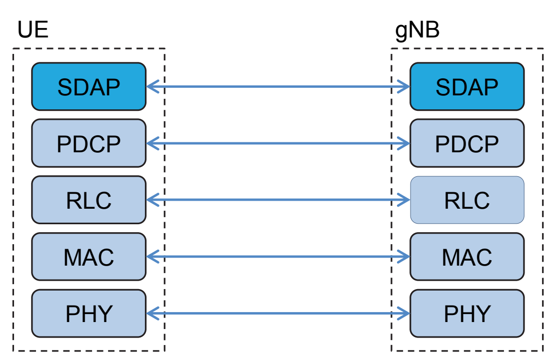


Figure 5　NR U-plane protocol stack

*37　**MPS:** A service that allows specific traffic to be provided to terminals even when the network is congested.

*38　**MCS:** Critical services whose failure or interruption could have serious adverse effects on life and society.

*39　**Handover:** The technique of switching from one cell to another without interrupting communication when a terminal moves between base stations.

*40　**Beamforming:** A technique for increasing or decreasing the gain of antennas in a specific direction by controlling the amplitude and phase of multiple antennas to form a directional pattern with the antennas.

as a new layer 2 protocol on the PDCP layer for 5GC. **Table 3** shows a functional classification of the functions in the other U-plane layer 2*43 protocol, detailing whether or not each function is compatible with standalone or non-standalone operation, and the differences between these functions and the functions in LTE.

This section describes how the SDAP layer is controlled, which is one way in which the U-plane functions differ from LTE. Details of beam control in the Medium Access Control (MAC)*44 layer can be found in a separate item in this special article [6], and details of RLC layer not supporting packet reordering function and PDCP layer packet duplicated transmission control for highly reliable communication such as Ultra-Reliable and Low Latency Communication (URLLC)*45 can be found in an item published in the 2017 special article [2].

Table 3　NR U-plane Layer 2 protocol function list

Layer	Function	Non-standalone	Standalone	Difference from LTE
MAC	Mapping between logical channels and transport channels	✓	✓	—
	Multiplexing/de-multiplexing data of same and different logical channels	✓	✓	New support for data multiplexing/demultiplexing in the same logical channel
	Scheduling	✓	✓	—
	Error correction by HARQ	✓	✓	—
	Priority control between logical channels	✓	✓	—
	Beam management control	✓	✓	New support
RLC	Packet segmentation and reassembly	✓	✓	Packet reordering is not supported
	Duplicate detection	✓	✓	—
	Protocol error detection	✓	✓	—
	Lossless transmission by ARQ	✓	✓	—
PDCP	Header compression	✓	✓	—
	Packet reordering and duplicate detection	✓	✓	—
	Security (ciphering and integrity protection)	✓	(✓)	Integrity protection for U-plane is supported in SA
	Timer-based packet discard	✓	✓	—
	Duplicate packet transmission	✓	✓	New support
SDAP	Mapping between QoS flow and radio bearer		✓	New support

*41　Massive MIMO: MIMO systems that transmit radio signals overlapping in space by using multiple antenna elements for transmission and reception. Massive MIMO systems aim to achieve high-speed data communications with greater numbers of simultaneous streaming transmissions while securing service areas. They achieve that aim by using antenna elements consisting of super multi-element arrays to create sharply formed radio beams to compensate for the radio propagation losses that accompany high-frequency band usage.

*42　SDAP: A sublayer of Layer 2 (see *43). A protocol that performs mapping between QoS flows and radio bearers.

*43　Layer 2: The second layer (data link layer) in the Open Systems Interconnect reference model.

*44　MAC: A sublayer of Layer 2 (see *43). A protocol that performs functions such as radio resource allocation, mapping data to TBs, and HARQ retransmission control.

During the initial investigation of 5GC, particularly due to requests from European operators, there were discussions of the concept of "access agnostic" specifications with inclusive support for not only Next Generation (NG)-RAN*46 but also non-3GPP access schemes such as Wireless Local Area Network (WLAN) access and fixed access.

In the Quality of Service (QoS) control*47 of a conventional EPC, per-call QoS-related Tunneling End-point IDentifiers (TEIDs)*48 are dispensed with, and QoS control is performed based on TEID (**Figure 6 (a)**). However, considering the affinity with non-3GPP

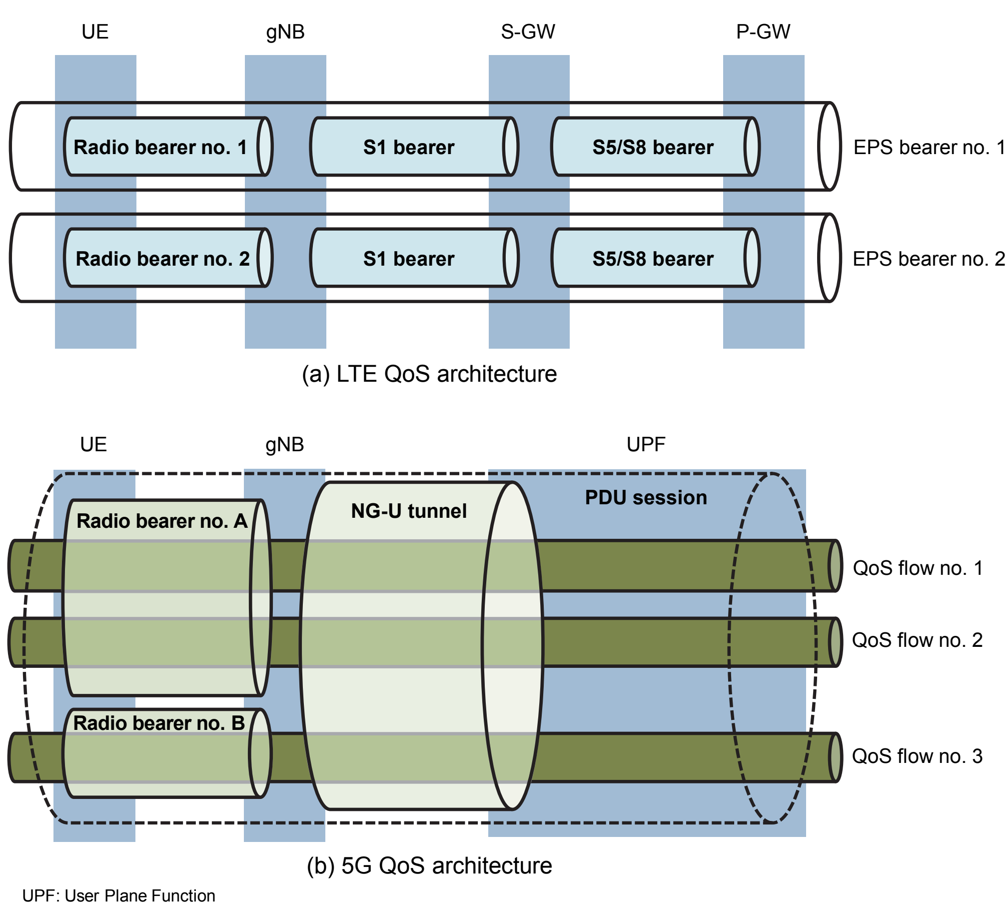


Figure 6 LTE QoS architecture (top) and 5G QoS architecture (bottom)

*45 URLLC: Generic terminology for communication requiring low delay and high reliability.

*46 NG-RAN: A RAN connecting to the 5G core network using NR or Evolved Universal Terrestrial Radio Access as radio access technology.

*47 QoS control: Technology to control communication quality such as priority packet transfer.

*48 TEID: A connection path identifier used in GRPS Tunneling Protocol.

access networks that do not perform QoS control based on the per-call TEID, it became necessary to introduce a new QoS framework for 5GC. Specifically, a QoS flow*49 method was adopted whereby a dedicated pointer to multiple QoS characteristics is included in each IP packet so that packets with multiple QoS characteristics can be handled without using TEID on individual calls. In the QoS flow method, multiple QoS flows passing through a Protocol Data Unit (PDU)*50 session tunnel set up between the CN and base station are mapped to individual radio bearers*51 established between the base station and the terminal equipment (Fig. 6 (b)).

This mapping between QoS flows and radio bearers is carried out at the SDAP layer newly introduced in NR. Although this mapping can also be set up in the C-plane, the setting and modification of mappings in the C-plane requires the use of RRC signals, resulting in setting delays and overheads*52.

Therefore, reflective QoS control is introduced to implement the setting and modification of mapping in a more dynamic fashion. Specifically, when a change is needed in the mapping between a QoS flow and a radio bearer, the CN assigns a new QoS flow identifier to each packet and sends it to the base station. When the base station detects a new QoS flow identifier, it includes its information in an SDAP header corresponding to a packet and transmits to the terminal. When the terminal receives this packet, it detects the identifier assigned to the corresponding SDAP, updates the mapping information managed by this terminal, and performs transmission of the uplink packet with the corresponding radio bearer. In this way, the base station makes settings and modifications to the mapping of transmitted packets and reports this information to the terminal, thus enabling the implementation of QoS modifications suited to the IP flow more dynamically than when relying on control signals.

4. Network Interface

In LTE networks, X2 and S1 are defined as interfaces between RAN nodes and between RAN and CN. In 5G, X2 and S1 extensions for non-standalone operation and new interface for standalone operation have been discussed. Fig. 1 shows the interface used between nodes.

1) Interface between RAN Nodes (X2/Xn)

Different interfaces are used between RAN nodes for non-standalone and standalone operation. The X2 interface used between eNBs in LTE is reused for interfaces between RAN nodes in non-standalone operation (between eNB and en-gNB in Fig. 1 (a)), and the Xn interface is newly specified between RAN nodes in standalone operation (between ng-eNB and ng-eNB/gNB and gNB/ng-eNB and gNB, as shown in Fig. 1 (b)). The X2 extension functions and the Xn functions are shown in red in **Table 4**.

The extensions of X2 include functions adopting EN-DC and flow control for split bearers for non-standalone operation. The flow control function, which was defined for LTE-DC split bearers in Release 12, is used for appropriately split downlink data when using the radio resources of multiple RAN nodes. Although functions and interfaces just for basic flow control were specified for LTE-DC, the information exchanged between RAN nodes

*49 **QoS flow:** A unit of flow used in QoS control.
*50 **PDU:** A unit of data processed by a protocol layer/sublayer.
*51 **Bearer:** The path to transfer user data packets.
*52 **Overhead:** Control information needed for transmitting/receiving user data, plus radio resources used for other than transmitting user data such as reference signals for measuring received quality.

Table 4 Main functions of X2/Xn interface

	Function	X2	Xn	Overview
C-plane functions	Interface management	✓	✓	Interface setup, reset, configuration update, error indication
	UE context management	✓	✓	Function for UE context management
	Mobility management	✓	✓	Management for e.g. UE inter-node mobility (handover)
	DC within same RAT	✓	*1	Function to operate DC in same RAT
	Load management	✓*2		Function to indicate resource load status to other nodes
	Energy saving	✓	✓	Function for energy saving by turning cells on/off
	Message transfer	✓		Function for forwarding messages to other eNBs
	UE context retrieval	✓		Function for retrieving UE context from other nodes
	EN-DC	✓	*1	Function for operating DC between LTE and NR
	Secondary RAT data usage report	✓		Function for reporting data volume of a RAT when using several RATs (such as EN-DC)
	Inactive management		✓	Function for managing inactive UEs
U-plane functions	User data transfer	✓	*1	Transferring user data during DC
	Flow control	✓	*1	Additional functions: polling, discarding duplicate data, status indication for retransmission data, etc.

*1 To be introduced in December 2018

*2 Notifications in non-standalone operation are under discussion

is further enhanced to optimize the flow control for non-standalone operation. Although Xn is based on the X2 function, the UE context management function is chiefly enhanced for adopting the above-mentioned new QoS flow framework and network slices*53 [3].

2) Interface between RAN Node and CN (S1/NG)

Like the interfaces between the RAN nodes, the interfaces between RAN nodes and CNs also differ between non-standalone and standalone operation. In non-standalone operation, the S1 interface used between eNBs and EPC is reused for interfaces between RAN node (eNB/en-gNB in Fig. 1 (a)) and EPC. On the other hand, the NG interface is newly specified between RAN nodes (ng–eNB/gNB in Fig. 1 (b)) and 5GCs in standalone operation. The S1 extension functions and the NG functions are shown in red in **Table 5**.

The extensions of S1 include a function that reports data volume for a specific RAT in non-standalone operation. In the standardization discussions, there were demands from operators for charging based on the data volume of each RAT (i.e., LTE and NR) in non-standalone operation. Thus, this function was introduced for calculating the amount of data volume via NR. In non-standalone

*53 Network slice: One format for achieving next-generation networks in the 5G era. Architecture that optimally divides the core network in units of services corresponding to use cases, business models, etc.

Table 5 Main functions of S1/NG interface

	Function	S1	NG	Overview
C-plane functions	Interface management	✓	✓	Interface setup, reset/configuration update, error indication
	Bearer/session management	✓	✓	Function for managing UE bearers and sessions
	UE context management	✓	✓	Function for UE context management
	Mobility management	✓	✓	Management for e.g. UE inter-node mobility (handover)
	Paging	✓	✓	Function for paging UEs
	Roaming/access restrictions	✓		Function for notifying RAN node roaming and access restrictions
	NAS signalling transport	✓	✓	Function for exchanging NAS signals between core network and UEs
	Radio quality notification	✓		Function for notifying core node radio quality measured by UEs
	Location reporting	✓	✓	Function for notifying core node current location of UEs
	Warning message information transfer	✓	✓	Function for transferring UE warning message (earthquake, tsunami, etc.)
	Function for secondary RAT data usage reporting	✓	*1	Function for reporting data volume for a RAT when using different RATs simultaneously (such as EN-DC)
	Function for core node selection		✓	Function for selecting core network node according to load status
U-plane functions	Function for user data transfer	✓	✓	Transfer of user data between core network and UEs

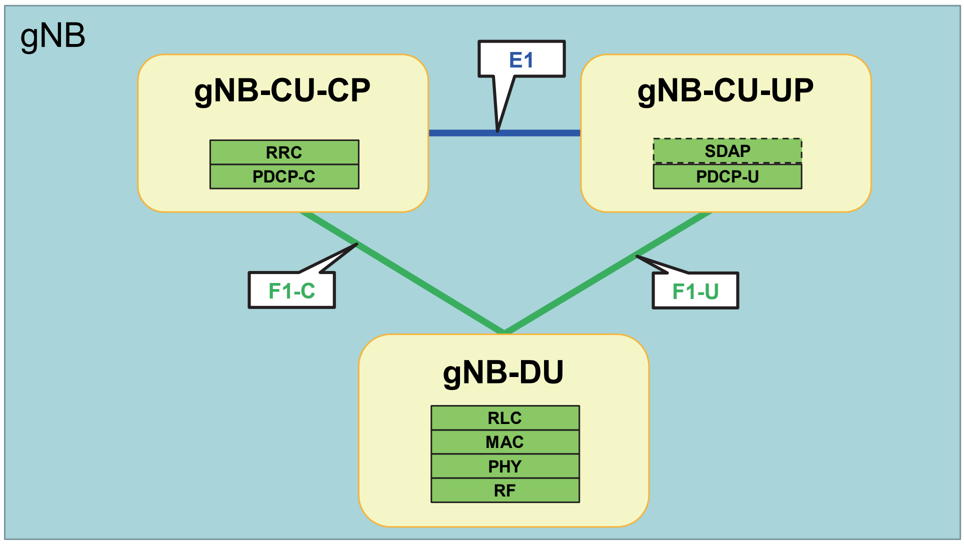
*1 To be specified in December 2018

operation, since the S1-C interface is only established between the MN and CN, the data volume through MN terminated bearers is counted by the MN itself and reported directly to the CN via S1, while the data volume from SN terminated bearers is counted by SN and reported to MN via X2, and then reported by MN to CN via S1. Similar with Xn for X2, although NG is based on the S1 function, the bearer/session management functions and UE context management functions have been chiefly enhanced for adopting a new QoS framework and network slices.

3) Functional Split and Open Interfaces (F1, E1) within RAN Nodes

In 3GPP standardization, the functional split within a RAN node, which places part of the functions in separate logical nodes, and an open interface between logical nodes has been discussed. **Figure 7** shows the functional split of gNB and the interfaces between the logical nodes.

In 3GPP, to address the issue of explosive increases of the bandwidth required for the transport



*SDAP layer is used only for standalone operation.

Figure 7 Functional split of gNB and interfaces that are used

between the Central Unit (CU) and Distributed Unit (DU) by the introduction of massive MIMO and extending the frequency bandwidth using Centralized RAN (C-RAN)*54 deployment, the new functional split between CU (gNB-CU) and DU (gNB-DU) within gNB and the corresponding open interface between these nodes were discussed [2]. Specifically, a functional split was adopted where the PDCP layer and above are located in the gNB-CU, and the RLC layer and below are located in the gNB-DU. The standard interface between them is specified as F1. The functions of the F1 interface are shown in **Table 6**.

In addition to the functional split between gNB-CU and gNB-DU, the functional split of C-plane termination and U-plane termination in gNB-CU was discussed [4]. For example, when the C-plane terminating parts are placed near gNB-DU and the U-plane terminating parts are placed near CN, the RRC signal used by the C-plane can be controlled without long delays and the U-plane functions can be placed in the cloud. Conversely, when the C-plane termination is placed near CN and the U-plane termination is placed near gNB-DU, it is possible to reduce the delays of U-plane signals for edge computing applications, and the C-plane functions can be migrated to the cloud. In the past, this sort of functional split has only been achieved within the same

*54 C-RAN: A radio access network having a configuration that consolidates the baseband processing sections of base station equipment and controls the radio sections of that equipment through optical fiber connections vendor. However, with 3GPP standardization, an open interface between the C-plane termination parts and U-plane termination parts of gNB-CU has been specified so that this sort of functional separation can be achieved even between different vendors. A node that terminates the C-plane of gNB-CU is called gNB-CU-CP, and a node that terminates the U-plane of the gNB-CU is called gNB-CU-UP. The standard interface between these nodes is specified as E1. The functions of E1 are shown in **Table 7**.

5. Conclusion

This article has described the main specifications of the 5G higher layer specified by 3GPP. The 3GPP is continuing its efforts with the aim, by December 2018, of finalizing the standard specifications for a terminal that simultaneously communicates with multiple RAN nodes with 5GC. For Release 16, further enhancements for functional expansion and the study/specification for new functions are planned. Specifically, 3GPP will discuss further enhancement of the higher layers for URLLC, as well as a function for minimizing interruption time during handover and a function for collecting data from gNB for the self-optimization of the network. NTT DOCOMO will continue to contribute to standardization actively by submitting technical proposals in each 3GPP RAN working group (WG).

Table 6 Main functions of F1 interface

	Function	Overview
C-plane functions	Interface management	Interface setup, reset/configuration update, error indication
	System information management	Function for broadcasting system information
	UE context management	Function for UE context management
	RRC message transfer	Function for exchanging RRC signaling between gNB-CU and UE
	Paging	Function for operating paging for UEs
	Warning message information transfer	Function for transferring UE warning message (earthquake, tsunami, etc.)
U-plane functions	User data transfer	Transfer of user data between gNB-CU and gNB-DU
	Flow control	Flow control for gNB-DU

Table 7 Main functions of E1 interface

Function	Overview
Interface management	Interface setup, reset/configuration update, error indication
Bearer management	Function for managing bearers for UEs

REFERENCES

[1] A. Minokuchi, et al.: "5G Standardization Trends at 3GPP," NTT DOCOMO Technical Journal, Vol.19, No.3, pp.5–12, Jan. 2018.

[2] A. Umesh, et al.: "5G Radio Access Network Standardization Trends," NTT DOCOMO Technical Journal, Vol.19, No.3, pp.36–47, Jan. 2018.

[3] 3GPP TR38.806 V15.0.0: "Study of separation of NR Control Plane (CP) and User Plane (UP) for split option 2," 2018.

[4] A. Minokuchi, et al.: "5G Core Network Standardization Trends," NTT DOCOMO Technical Journal, Vol.19, No.3, pp.48–54, Jan. 2018.

[5] K. Aoyagi, et al.: "Access Class Control Technology in LTE/LTE-Advanced Systems," NTT DOCOMO Technical Journal, Vol.17, No.2, pp.65–76, Oct. 2015.

[6] K. Takeda, et al.: "NR Physical Layer Specifications in 5G," NTT DOCOMO Technical Journal, Vol.20, No.3, pp.49–61, Jan. 2019.

Technology Reports | 5G | NR | Radio Performance/Management Specifications

Special Articles on Release 15 Standardization
—Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

5G Radio Performance and Radio Resource Management Specifications

Radio Access Network Development Department Yosuke Sano Suguru Okuyama
Naoto Iizasa Takuma Takada
Communication Device Development Department Kei Ando† Naoki Fujimura

3GPP has made specifications for a new radio technology called NR, which satisfies the requirements for a 5G RAN. In addition to the frequency bands below 6 GHz already used by LTE/LTE-Advanced, NR is able to use sub-millimeter and millimeter wave frequencies, realizing high throughput with wideband communication. This article discusses frequency trends for 5G and describes specifications for base station and terminal RF performance and radio resource management.

1. Introduction

The 3rd Generation Partnership Project (3GPP) has studied a New Radio (NR) communication technology satisfying the requirements for a 5G Radio Access Network (RAN)*1 [1], and has made specifications for it as the Release 15 specifications. The RAN Working Group 4 (RAN4), which has responsibility for making specifications for base station and terminal Radio Frequency (RF)*2 performance and radio resource management, established the new specifications not only for the frequency bands below 6 GHz already used by LTE/LTE-Advanced but also for sub-millimeter*3 and millimeter*4 wave frequency bands, which have not been used for commercial cellular systems previously.

Specifically, RAN4 defined new frequency bands with consideration for the frequency allocation plans

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently Network Department

*1 RAN: The network consisting of radio base stations and radio-circuit control equipment situated between the core network and mobile terminals.

*2 RF: Referring to the radio frequency analog circuit.

*3 Sub-millimeter wave: Radio signals of frequencies in the millimeter wave range, from approximately 10 to 30 GHz.

*4 Millimeter wave: Radio signals of frequencies in the range from 30 GHz to 300 GHz.

for 5G in various countries, and specified base station and terminal RF performance specifications for each frequency band. Radio Resource Management specifications, such as the specifications for terminals to measure the cell quality in the current cell and neighboring cells, were also specified mainly to ensure mobility performance of terminals.

This article describes these new frequency bands and specifications for RF performance and radio resource management.

2. New Frequency Bands for NR

In Japan, the new frequencies being considered for initial 5G deployment are the 3.7 GHz band (3.6-4.2 GHz), the 4.5 GHz band (4.4-4.9 GHz), and the 28 GHz band (27.0-29.5 GHz) [2]. As shown in **Figure 1**, parts of these ranges overlap with 5G frequency bands being considered in other countries. RF modules are manufactured and incorporated in terminals by frequency band (or groups of adjacent frequency bands), so terminal manufacturing costs can be reduced by harmonizing frequencies—defining bands that include the bands several countries—so that terminals for different countries can share RF modules. However, as the bandwidth of an RF module is increased, the fractional bandwidth (bandwidth divided by center frequency[*5]) increases and it becomes necessary to adjust power (impedance matching) over a wide frequency range, generally making it more difficult to design. As such, bands must be defined considering both their effect in harmonizing frequencies and feasibility of RF devices. The frequency bands defined and specified in Release 15 are described below.

Note that in the NR frequency band specifications, band names are prefixed with an "n" to distinguish them from existing LTE bands. The

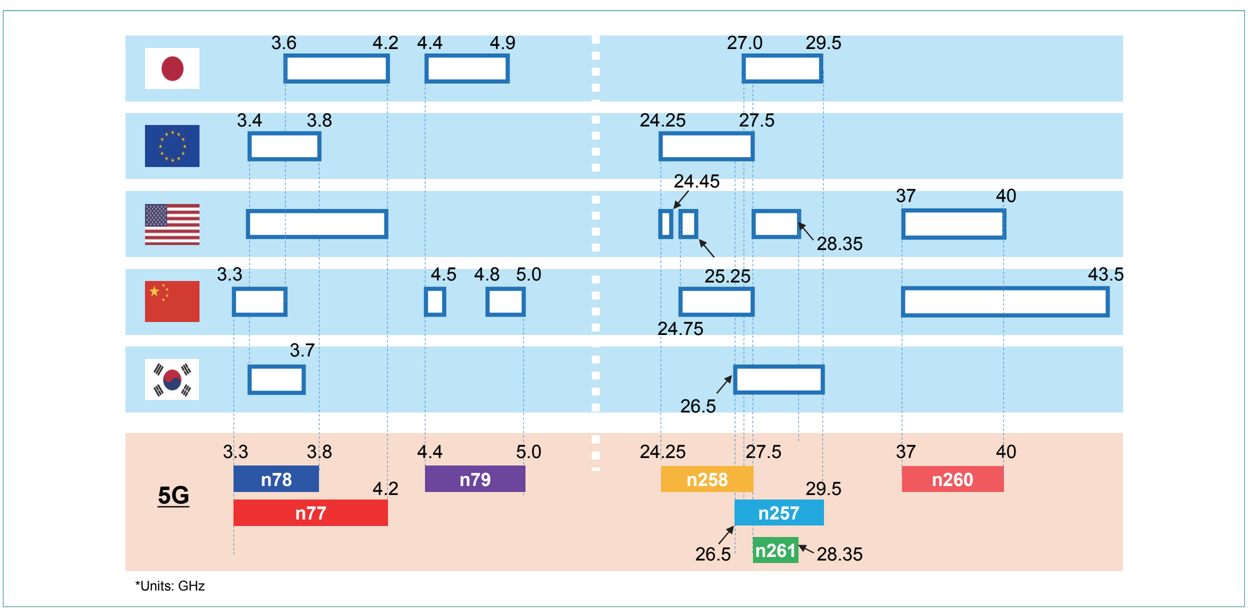


Figure 1 5G candidate frequencies in Japan and other countries

*5 Center frequency: The frequency within a frequency band at the center of the range used for communication.

frequency bands are also grouped into two ranges in Release 15, as follows.

- Frequency Range 1 (FR1): 450 to 6,000 MHz
- Frequency Range 2 (FR2): 24,250 to 52,600 MHz

2.1 3.7 GHz Bands (Bands n77 and n78)

It would be desirable to define a range from 3.3 to 4.2 GHz, to harmonize the bands used in Europe, the USA, China, and South Korea. However, this range has a fractional bandwidth 4.2 times that of Band 42 (3.4 to3.6 GHz) in the 3.5 GHz band, which was introduced with LTE/LTE-Advanced. This raises issues, particularly regarding power amplifier efficiency. As a result, two bands were defined: n77 (3.3 to 4.2 GHz), which will harmonize frequencies including Japan, and n78 (3.3 to 3.8 GHz), which is a narrower band encompassing frequencies used in Europe, the USA, China, and South Korea.

2.2 4.5 GHz Band (Band n79)

In November 2017, China indicated plans to use the 4.8 to 5.0 GHz band for 5G and that they are studying use of the 4.4 to 4.5 GHz band. The 4.4 to 4.9 GHz band is also a candidate in Japan, so Band n79 (4.4 to 5.0 GHz) was defined to include these ranges and harmonize with China.

2.3 28 GHz Band (Bands n257, n258, and n261)

The 27.0 to 29.5 GHz band is a candidate in Japan, so it is desirable to harmonize this band with the USA and South Korea. RF devices supporting a wide bandwidth of 5.25 GHz would be required to include this range of 5G candidate frequencies from all regions (24.25 to 29.5 GHz), hence for practical considerations, two bands were defined: band n257 (26.5 to 29.5 GHz) for Japan, North America (USA, Canada, etc.), and South Korea, and n258 (24.25 to 27.5 GHz) for Europe and China. The narrower n261 band (27.5 to 28.35 GHz) was also defined for operators in North America because the RF modules needed to support both the 28 GHz (n257) and 39 GHz (n260) bands at the same time, while maintaining RF performance, could be larger than desired.

2.4 Other Frequency Bands

The USA and China are also considering the 39 GHz band (n260: 37 to 40 GHz), as shown in Fig. 1. Note that besides the above, frequency bands used by existing LTE systems can also be used with NR. These are called LTE-refarming bands and for Japan, the 700 MHz (n28), 800 MHz (n5), 900 MHz (n8), 1.5 GHz (n74), 1.7 GHz (n3), 2 GHz (n1), and 2.5 GHz (n41) bands have been specified. See references [3] and [4] regarding other bands regulated in Release 15.

3. NR Radio Parameters

As mentioned above, NR is capable of using a wide range of frequencies from existing LTE bands to sub-millimeter and millimeter-wave frequencies. However, radio characteristics significantly vary depending on frequency range and band, so it is desirable to be able to configure radio parameters, such as subcarrier*6 spacing, channel bandwidth, and effective communication bandwidth in Orthogonal Frequency Division Multiplexing (OFDM)*7, according to the configured frequency band.

*6 Sub-carrier: Individual carrier for transmitting signals with multi-carrier transmission such as OFDM.

*7 OFDM: A digital modulation system developed to improve resistance to multi-path interference. It converts a signal with a high data rate to multiple low-speed narrow-band signals and transmits those signals in parallel along the frequency axis. OFDM enables signal transmission with high spectral efficiency.

3.1 Subcarrier Spacing

In FR1, the subcarrier spacings that can be used for data signals are 15, 30, and 60 kHz. In FR2, they are 60 and 120 kHz. All terminals must support all of these except 60 kHz in FR1. In NR, the Cyclic Prefix (CP)*8 insertion ratio (the CP length relative to the OFDM symbol*9 length) is the same for all available subcarrier spacings. As such, as the subcarrier spacing increases, the absolute CP length decreases and resistance to multi-path signal delay decreases. This can result in reduced coverage for each base station. On the other hand, use of wider subcarriers can increase resistance to frequency shift, e.g., due to Doppler shift*10. FR2 is also quite susceptible to phase noise*11 due to oscillator error*12 in terminals and base stations, but there are components that can generally be mitigated by setting a wider subcarrier spacing.

Subcarrier spacing is configured by the network using higher layer*13 signals. The optimal subcarrier spacing can be applied according to the deployment scenario and considering the trade-offs described above.

3.2 Channel Bandwidth

1) Channel Bandwidth in LTE

In LTE (Release 8), channel bandwidth can be set in the range of 1.4 to 20 MHz. In LTE-Advanced (Release 10), Carrier Aggregation (CA)*14 technology was introduced, realizing wide bandwidths up to 100 MHz by using up to five LTE carriers, called Component Carriers (CC)*15 at the same time [5]. In Release 13, the number of CCs that can be used with CA was expanded to a maximum of 32, for a maximum bandwidth of 640 MHz [6].

2) Channel Bandwidth in NR

CA is also supported in NR, and Release 15 supports intra-band contiguous and non-contiguous CA, inter-band CA, and combinations of these (**Figure 2**). It also supports Dual Connectivity (DC)*16, communicating on both NR carriers (including CA) and LTE carriers simultaneously.

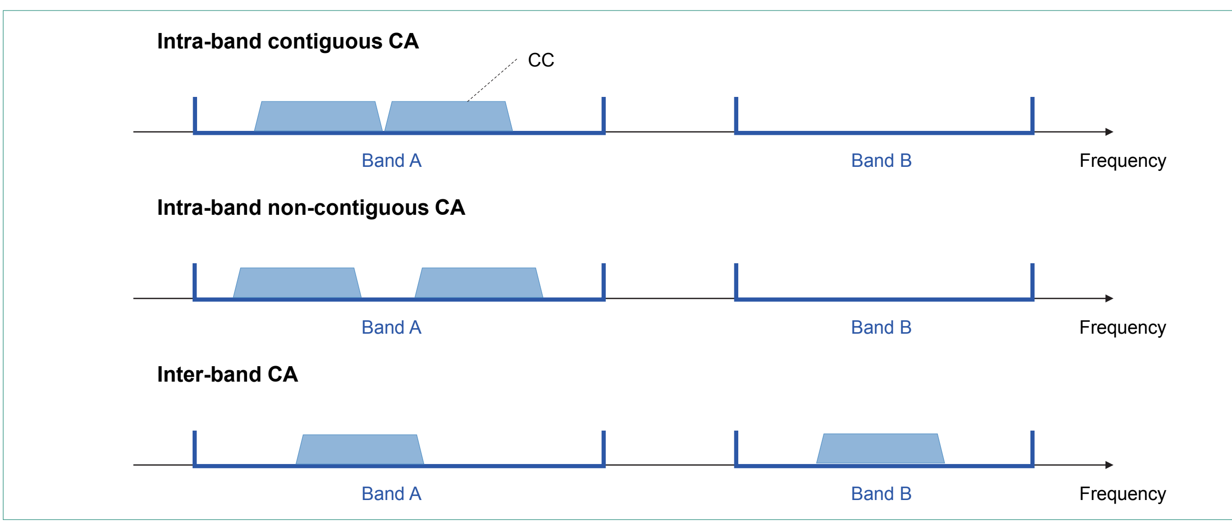


Figure 2 CA supported in Release 15

---

*8 CP: A guard time (also called "guard interval") inserted between symbols in OFDM signals to minimize interference between prior and subsequent symbols due to multipath effects. Usually, this part of the signal is copied from the latter-part of the symbol.

*9 OFDM symbol: The unit data of transmission, consisting of multiple subcarriers for OFDM. A CP is inserted at the front of each symbol.

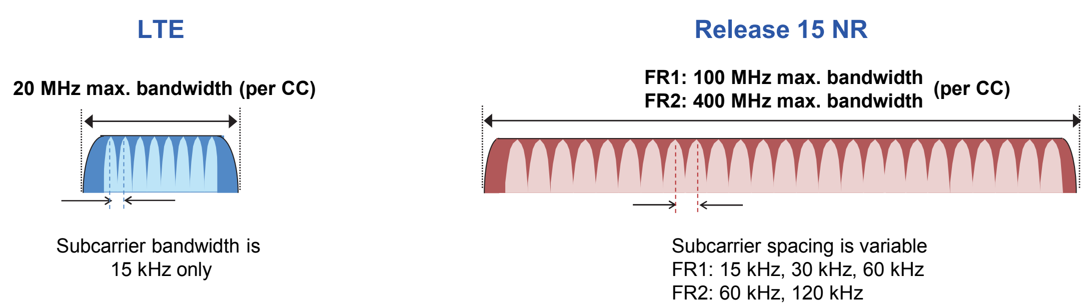
*10 Doppler shift: Shift in carrier-wave frequency due to the Doppler effect.

*11 Phase noise: Any fluctuation in phase that is not needed for communication, originating in the RF devices due to oscillator error and other causes. Can result in interference between subcarriers or common phase error, degrading the quality of communication.

CA is an important technology for realizing wideband communication, but as mentioned above, NR defines frequency bands that are much wider than LTE. With an LTE-equivalent channel bandwidth of 20 MHz (per CC), a huge number of CCs would be needed for CA. This has led to concern for the complexity of controlling the CCs, and resulting increased terminal manufacturing costs, and increasing control-signal overhead. For NR, the channel bandwidth per CC is increased to a maximum of 100 MHz in FR1 (when a subcarrier spacing of 30 or 60 kHz is used), and 400 MHz in FR2 (when a subcarrier spacing of 120 kHz is used). Note that supporting 400 MHz in FR2 is an optional feature, and terminals are only required to support up to 200 MHz. Specific channel bandwidths that can be configured in each frequency band are shown in **Figure 3**.

3.3 Effective Bandwidth During Communication

With LTE, only the central 90% of the channel bandwidth is used for data communication, to protect adjoining frequencies. For example, a channel



Frequency range	Frequency band		Subcarrier spacing (kHz)	CC width (MHz)*1														
	Band no.	Frequency band		5	10	15	20	25	30	40	50	60	70	80	90*2	100	200	400*2
FR1	n77/n78	3.7 GHz band	15		○	○	○			○	○						–	–
			30		○	○	○			○	○	○		○	○	○	–	–
			60		○	○	○			○	○	○		○	○	○	–	–
	n79	4.5 GHz band	15							○	○						–	–
			30							○	○	○		○		○	–	–
			60							○	○	○		○		○	–	–
FR2	n257	28 GHz band	60	–	–	–	–	–	–	–	○	–	–	–	–	○	○	
			120	–	–	–	–	–	–	–	○	–	–	–	–	○	○	○

*1 "○" indicates configurable.
*2 Optional function

Figure 3 NR channel bandwidths

*12 Oscillator error: The difference in frequencies produced by oscillators on the base station and the terminal. An oscillator is an RF device that produces a particular frequency and is used to convert the baseband signal to be transmitted to a signal in the carrier frequency band, or vice versa.

*13 Higher layer: Refers to any layer positioned above the physical layer, namely, layers such as Medium Access Layer (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), and Application Protocols (e.g.: S1AP, X2AP, etc.).

*14 CA: A technology for increasing bandwidth and data rate by simultaneously transmitting and receiving signals for one user using multiple carriers.

*15 CC: Term denoting each of the carriers used in CA.

*16 DC: A technology that achieves wider bandwidths by connecting two base stations in a master/slave relationship and performing transmission and reception using multiple CCs supported by those base stations.

with 20 MHz bandwidth would have an effective bandwidth of 18 MHz, excluding the 1 MHz guard bands*17 at each end. The guard bands at each end are inserted to reduce interference on other frequencies, but NR also applies preprocessing to normal OFDM signals, as shown in **Figure 4**, to reduce interference on other frequencies (such as in [7]). Accordingly, ways to increase the effective bandwidth have also been studied. As a result, except for some of the narrower bandwidths (e.g.: 5, 10 MHz), greater than 90% of the bandwidth can be used for communication (up to 98%, depending on the channel bandwidth), further increasing spectrum efficiency*18. Note that the width of the guard bands has been reduced, but this assumes the preprocessing to reduce interference as mentioned above, and values specified regarding interference on neighboring frequencies (leakage power, etc.) remain the same as for LTE.

4. Base Station and Terminal RF Performance Specifications

RF configurations and RF performance specifications for base stations and terminals specified for NR by frequency band are described below.

4.1 Base Stations

1) Base Station RF Configurations

The configurations specified for NR base station equipment are shown in **Table 1**. These include configurations specified earlier for LTE, which are BS type 1-C, having the base station and antennas connected by coaxial cable*19, and BS type 1-H, having an integrated Active Antenna System (AAS)*20 with radio transceivers connected to antennas using a Transceiver Array Boundary (TAB) connector*21. In addition to these, a new configuration called BS type 1-O/2-O is defined with a connector-less AAS

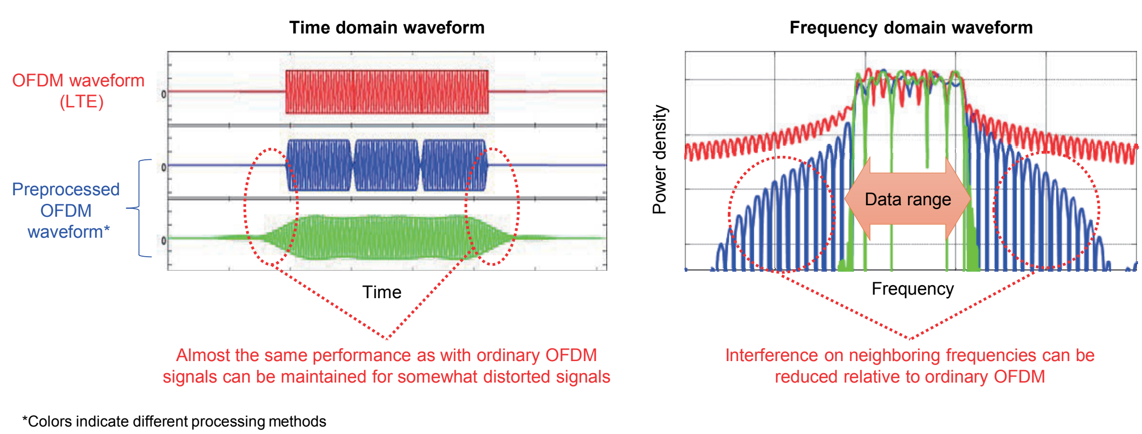


Figure 4 OFDM signal preprocessing

*17 **Guard band:** A frequency band set between the bands allocated to different wireless systems to prevent interference between the RF signals of those systems.

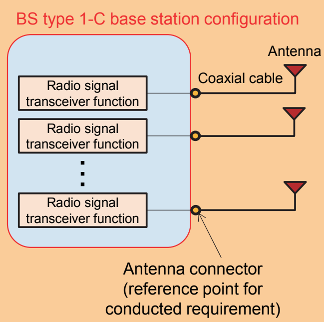
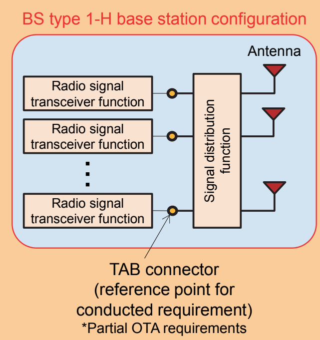
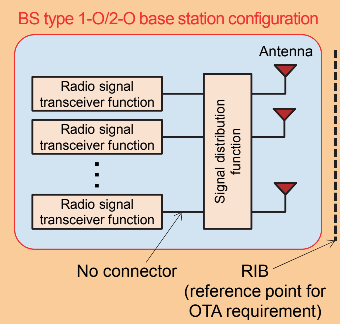
*18 **Spectrum efficiency:** The number of data bits that can be transmitted per unit time per unit frequency.

*19 **Coaxial cable:** A type of cable with an outer shielding layer, used mainly for transmitting high frequency signals. Used to transmit signals between antennas and the base station.

*20 **AAS:** A system that integrates antenna elements and RF circuits that have traditionally been separated thereby providing a more efficient system.

*21 **TAB connector:** Used in BS type 1-H NR base stations and a reference point used for radio characteristics.

Table 1 Base station configurations specified for NR

BS type	BS type 1-C (equivalent to earlier base stations)	BS type 1-H (equivalent to AAS in LTE-Advanced Release 13)	BS type 1-O/2-O (newly introduced AAS from NR)
Frequency range	FR1	FR1	FR1/FR2
Base station configuration			

for each frequency range [8].

(a) FR1 RF configuration

One of the RF configurations for FR1 does not require connectors between radio signal transceivers and antennas according to BS type 1-O specifications, so smaller equipment size and improved power efficiency can be expected, compared with AAS in LTE Advanced Release 13 [9].

(b) FR2 RF configuration

FR2 has the advantages of wideband communication in high-frequency bands, but in terms of RF configuration, the shorter wavelengths result in higher power losses in connectors and cables; and higher radio propagation losses and reduced coverage due to lower power density over the wider frequency band. High antenna gain*22 is necessary to maintain coverage, but it will be difficult to implement the radio signal transceivers and antennas with high density in a limited space for FR2 if conventional RF configurations with connectors are used. For this reason, only the BS type 2-O RF configuration with no connectors is defined. With the specification for BS type 2-O, it should be possible to implement beam forming*23 over a wide band and maintain coverage while achieving high-speed communication.

2) Base Station RF Performance Requirements

The RF performance specifications for BS type 1-C/1-H are based on the LTE-Advanced specifications, applied to the NR radio parameters described above. However, BS type 1-O/2-O have integrated radio signal transceiver and antennas and there are no connectors to take measurements, so Over-The-Air (OTA)*24 specifications have been extended so that in the overall RF performance specifications, a reference position in the radiated space (the Radiated Interface Boundary (RIB)*25) can be defined.

---

*22 Antenna gain: Radiated power in the direction of maximum radiation usually expressed as the ratio of radiated power to that of an isotropic antenna.

*23 Beam forming: A technology that gives a directionality to a transmitted signal, increasing or decreasing the signal power in a particular direction. Analog beam forming works by controlling the phase in multiple antenna elements (RF devices) to create directionality, while digital beam forming controls phase in the baseband module.

*24 OTA: A method for measuring radio characteristics transmitted or received from a base station or terminal, by positioning it opposite to a measurement antenna. Configurations have been defined for NR base stations and terminals that have no antenna connectors, so this method has been established for regulating such devices.

(a) New TRP specifications in OTA requirements

In addition to the Equivalent Isotropic Radiated Power (EIRP)*26 and Equivalent Isotropic Sensitivity (EIS)*27 [8] including the antenna characteristics in the beam direction, which are defined in LTE-Advanced Release 13, the Total Radiated Power (TRP) is introduced as a new metric definition. Defining TRP makes it possible to specify OTA requirements for power related requirements, including base station output power, spurious emissions*28 and so on. The definitions for EIRP, EIS and TRP used in the OTA requirements are illustrated in **Figure 5**.

(b) Features of FR2 RF performance specifications

A comparison of the main base station RF performance specifications is given in **Table 2**. The RF performance specifications for FR1 are based on the LTE Advanced specifications, with the maximum channel bandwidth expanded to 100 MHz. For FR2, the specifications reflect wider bands, lower latency and fast response, with maximum channel bandwidth expanded to 400 MHz and transmitter transient period of 3 μs, which is the ON/OFF switching time for Time Domain Duplex (TDD)*29. The specifications also consider deterioration of RF device characteristics in the millimeter-wave bands, compensating for it using beam forming with high-gain antennas, and relax requirements for radio characteristics such as the Adjacent Channel Leakage Ratio (ACLR)*30 in the transmission characteristics and Noise Figure (NF)*31 in the reception circuits. This helps ensure feasibility in manufacturing. With OTA requirements, transmit power deviation*32 is regulated as TRP accuracy for

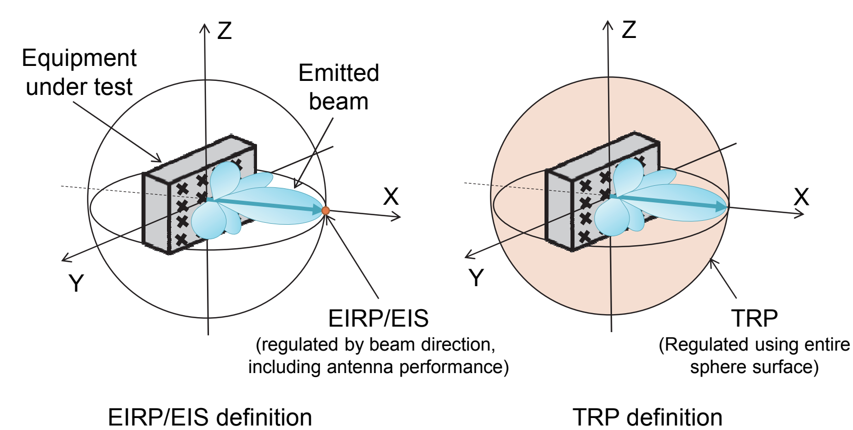


Figure 5　Definitions used for NR base station and mobile terminal RF requirements

*25 **RIB**: A standard point used for regulating OTA in base station RF performance specifications. This is the point where characteristics of radio emissions and reception space are measured.

*26 **EIRP**: The transmission power at the reference point in radio radiation space.

*27 **EIS**: The received power at the radiated requirement reference point in radio reception space.

*28 **Spurious emissions**: Unneeded radio emissions outside the desired band. Causing interference with neighboring frequencies.

*29 **TDD**: A bidirectional transmit/receive system. It achieves bidirectional communication by allocating different time slots to uplink and downlink transmissions that use the same frequency band.

*30 **ACLR**: When transmitting a modulated signal, the ratio between the transmitted signal band power and undesired power generated in the adjacent channels.

Table 2　Comparison of main base station RF performance specifications

	LTE	NR FR1 BS type 1-O	NR FR2 BS type 2-O
Max. channel bandwidth	20 MHz	100 MHz	400 MHz
Transmitter Transient period	< 17 μs	< 10 μs	< 3 μs
ACLR	45 dB	45 dB	28 dB*2
NF	5 dB*1	5 dB*1	10 dB
Transmit power deviation	+/- 2.0 dB	+/- 2.2 dB (EIRP accuracy)	+/- 3.4 dB (EIRP accuracy)
		+/- 2.0 dB (TRP accuracy)	+/- 3.0 dB (TRP accuracy)

*1　Specified values for wide-area base stations
*2　Specified values for 24.25 to 33.4 GHz

per-carrier total power deviation and EIRP accuracy, accounting for antenna performance.

4.2 Terminals

1) Terminal RF Configuration

In contrast with the NR base stations described above, terminal implementations only apply beam forming for FR2, and the RF front end configuration for FR1 NR terminals is not significantly different from earlier LTE, except for the new 3.7 and 4.5 GHz band implementations.

When NR is first introduced, non-standalone operation linked with LTE is assumed, so terminals supporting it will need to have conventional LTE radio equipment as well as the NR radio equipment.

2) Terminal RF Performance Requirements

For FR1, RF performance is specified for new radio parameters such as maximum transmit power and receiver sensitivity, to support NR as described above. As with conventional LTE-Advanced, Conducted requirements*33 for antenna connectors are applied.

On the other hand, for FR2, transceivers and antennas are integrated and measurements cannot be made at connectors, as with the base stations, so OTA requirements have been introduced. Requirements have been adopted for EIRP maximum transmit power in FR2 using cumulative distributions*34 of each EIRP value obtained when performing beam forming in a full sphere with the terminal at the center (**Figure 6**). The purpose of introducing these requirements was to statistically guarantee that the beam could be aimed correctly in the intended direction (toward the communicating base station) and with the necessary range. The Min peak value is defined as the value that at least one of the measured EIRP values must exceed, and spherical coverage is defined as the value on the cumulative distribution at X%; in other words, the value that must be maintained on (100-X)% of the area of the surface of the sphere. The Max value is defined as the value that the maximum

*31　NF: The level of noise power generated inside a piece of equipment. It is defined as the ratio between the Signal-to-Noise (SN) ratio of input signals and the SN ratio of output signals.

*32　Deviation: Dispersion or fluctuation from standard values.

*33　Conducted requirements: A type of test requirement for base stations and terminals, or a method for conducting tests with wired connections to the terminal.

*34　Cumulative distribution: The probability that the property being evaluated will be at or below a particular value.

measured EIRP value must not exceed, and was decided with consideration for requirements in various countries and regions, by organizations like the Federal Communications Commission (FCC)*35. Besides smartphones and other mobile terminals, which are the main consideration, Release 15 also considers ways to increase communication distances and data rates for other types of terminals such as fixed wireless terminals*36. These types of terminals are categorized according to Power class in the standard specifications, and the four Power classes are specified below, distinguished by required transceiver power and differences in spherical coverage (**Table 3**).

- Power class 1

  This class conforms with transportable stations as defined by FCC 16 to 89. In contrast with Power class 3 described below for

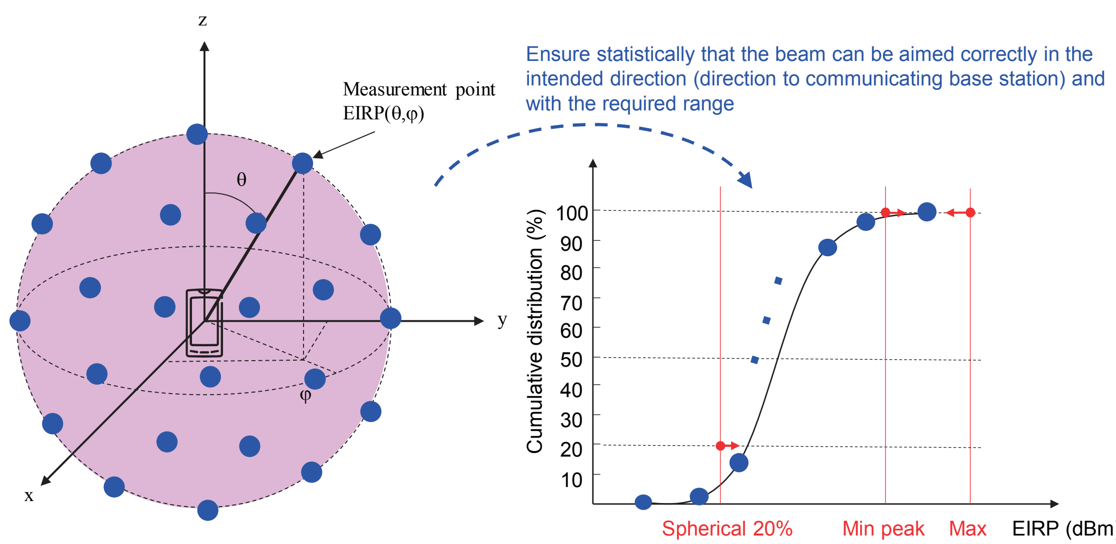


Figure 6　OTA EIRP evaluation using a cumulative distribution

Table 3　FR2 power class

FR2 power class	UE type example	EIRP (dBm)			TRP (dBm)
		Max	Min peak	Spherical	Max
1	Fixed wireless terminal	55	40	32 (85%)	35
2	Vehicle-mounted terminal	43	29	18 (60%)	23
3	Mobile terminal (Smartphone, etc.)	43	22.5	11.5 (50%)	23
4	Fixed wireless terminal	43	34	25 (20%)	23

*35　FCC: The Federal Communications Commission of the USA. Has the authority to approve and license industries including television, radio, telegraph and telephone.

*36　Fixed wireless terminal: A communication method in which both base station and terminal have a fixed location. Applications include use for backhaul communication between base stations.

handheld terminals, this Power class permits a maximum TRP of 35 dBm with maximum permitted TRP of 23 dBm*37. It assumes use cases with fixed equipment able to emit a strong signal with a narrow beam in a particular direction.

- Power class 2

  This class of terminals is intended mainly for use in vehicles. Its target for spherical coverage is 60% and maximum TRP is the same as Power class 3 for mobile terminals, but requires a higher EIRP characteristic.

- Power class 3

  This class is intended for smartphones and other mobile terminals. The orientation of the terminal with respect to the base station is random, so the target for spherical coverage is 50%.

- Power class 4

  This class requires wider spherical coverage (20%) and higher EIRP than Power class 2 described above. Since it can be installed without knowing the location of the NR base station, in contrast the narrow-beam Power class 1, it can be used more flexibly in use cases such as in vehicles or trains, and not only in fixed installations.

5. Radio Resource Management Specifications

Specifications regarding Radio Resource Management (RRM)*38, to ensure terminal mobility performance, are described below.

5.1 Measurements for Serving Cell and Neighbor Cell

For operations such as handover to a neighbor cell, or adding a new CC in the case of CA, it is desirable to measure cell quality, such as the Reference Signal Received Power (RSRP)*39 or Reference Signal Received Quality (RSRQ)*40, for neighbor cells. This enables these processes to be performed appropriately, maintaining radio link quality. With LTE, all base stations continually transmit a Cell-specific Reference Signal (CRS)*41, so it is easy for terminals to measure the cell quality from neighbor cells. However, NR has no reference signal like the CRS, to reduce resource overhead and interference to other cells in the case that traffic levels are low.

With NR, the cell quality is measured by using SS/PBCH Blocks (SSB)*42. These are composed of a Synchronization Signal (SS)*43 and the Physical Broadcast CHannel (PBCH)*44 which have a longer transmission periodicity than CRS (**Figure 7** (a)).

The SSB periodicity can be configured for each cell, in the range of 5, 10, 20, 40, 80, or 160 ms. However, terminals do not need to measure cell quality with the same periodicity as the SSB and the appropriate periodicity can be configured according to the channel condition. This is desirable and can help avoid unnecessary measurements and reduce power consumption on terminals.

As such, the new SSB-based RRM Measurement Timing Configuration window (SMTC window)*45 has been introduced to notify terminals of the periodicity and timing of SSBs that the terminals can use for measurements. As shown in Fig. 7, the SMTC window periodicity can be set in the

*37 dBm: Power value [mW] expressed as 10log (P). The value relative to a 1 mW standard (1 mW = 0 dBm).

*38 RRM: A generic term for resource management including mobility operations such as handover or measurement of cell quality via reference signals in order to manage limited radio resources or connect between terminals and base stations smoothly, etc.

*39 RSRP: Received power of a signal measured at a receiver. RSRP is used as an indicator of receiver sensitivity in a terminal.

*40 RSRQ: The ratio of the power of the reference signal to total power within the receive bandwidth.

*41 CRS: Reference signal specific to each cell for measuring cell quality.

*42 SSB: Component including Synchronization signal and Physical Broadcast Channel, which is transmitted periodically. Terminals receive it for not only detecting cell ID and reception timing but also performing measurement the cell quality.

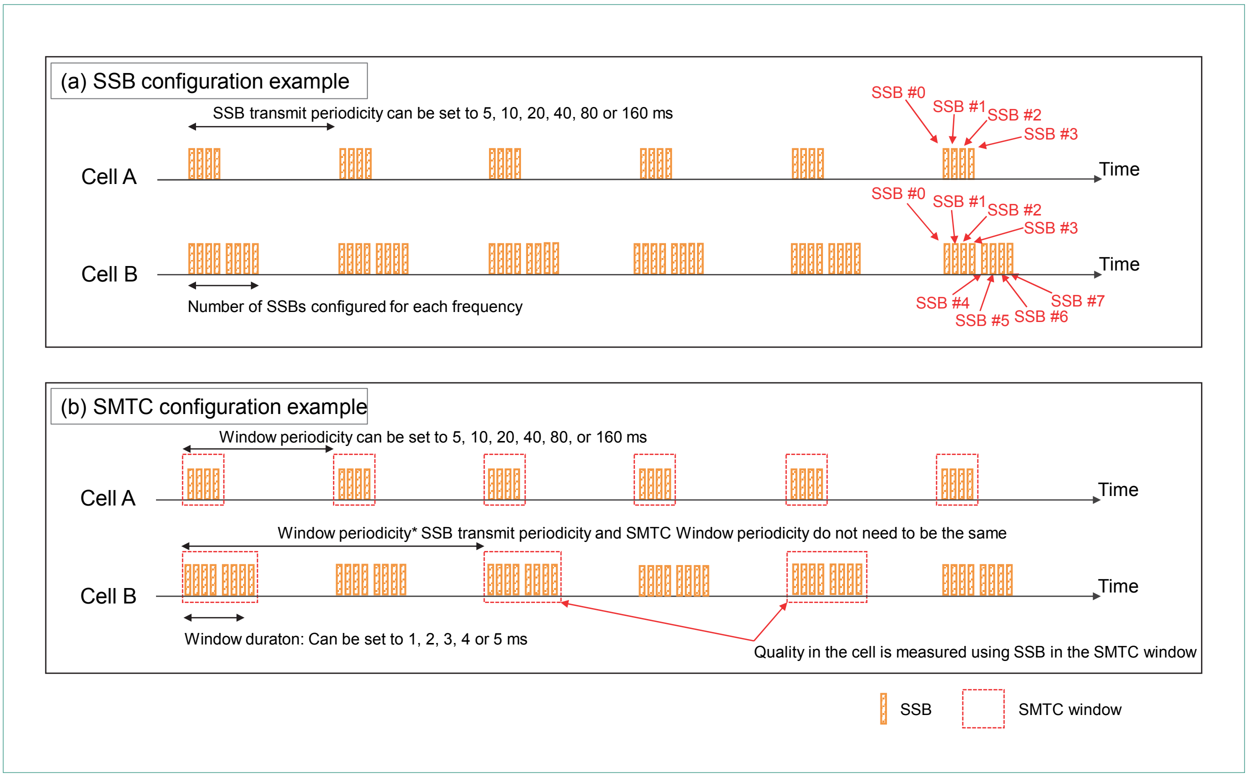


Figure 7 SSB/SMTC relationship in NR

same range as the SSB, of 5, 10, 20, 40, 80, or 160 ms, and the duration of the window can be set to 1, 2, 3, 4, or 5 ms, according to the number of SSBs transmitted on the cell being measured. When a terminal has been notified of the SMTC window by the base station, it detects and measures the SSBs within that window and reports the measurement results back to the base station.

5.2 Measurement Gap

Using the same RF module for measuring cell quality of neighbor cells or other CCs, and also for transmitting and receiving data in the serving cell can make it possible to reduce terminal manufacturing costs. However, this will mean that data cannot be transmitted or received in the serving cell during measurements of cells or CCs of different frequencies. With LTE, data transmission in the serving cell is suspended during a Measurement gap on the terminal, giving UE an opportunity to measure cells and CCs with a different frequency. A Measurement gap procedure similar to that in LTE is also introduced in NR. However, with NR, measurements are done using the SSB as described earlier, and the method for configuring the Measurement gap has been optimized. An overview of this method is given below.

1) Measurement Gap Configuration

(a) LTE

With LTE, the Measurement Gap Length

---

*43 SS: Synchronization signal for detecting carrier frequency of the cell, reception timing, and cell ID in order to connect the cell.

*44 PBCH: Common channel for reporting system information. Terminals receive this channel to get information such as operator code, common channel configuration, neighbor cell information, etc.

*45 SMTC window: Measurement window configured on the terminal for informing the measurement timing, duration and periodicity for each cell being measured when the terminal performs cell quality measurements using the SSB.

(MGL) is fixed, such that at least one SS is included within one gap. LTE SSs are transmitted at 5 ms periodicity, so the MGL of LTE is 6 ms, allowing 0.5 ms for RF retuning at the beginning and end of the MGL. Terminals detect the SS within the MGL, identify the cell ID*46 and reception timing, and thereafter terminals perform measurements with CRS.

(b) MGL extensions on NR

With NR, the SMTC window and window duration can be set to match SSB transmissions as described earlier. However, having a fixed MGL could cause unnecessarily degradation of throughput in the serving cell. For example, if the SMTC window duration is 2 ms and the Measurement gap has MGL of 6 ms, a 4 ms segment would not be available for transmission and reception of data in the serving cell.

To reduce such unnecessary degradation of throughput, the MGL for NR was extended to be configurable to 5.5 ms, 4 ms, 3.5 ms, 3 ms, or 1.5 ms in addition to the original 6 ms. This is shown concretely in **Figure 8** (1) and (2). The case in Fig. 8 (1) uses an SMTC

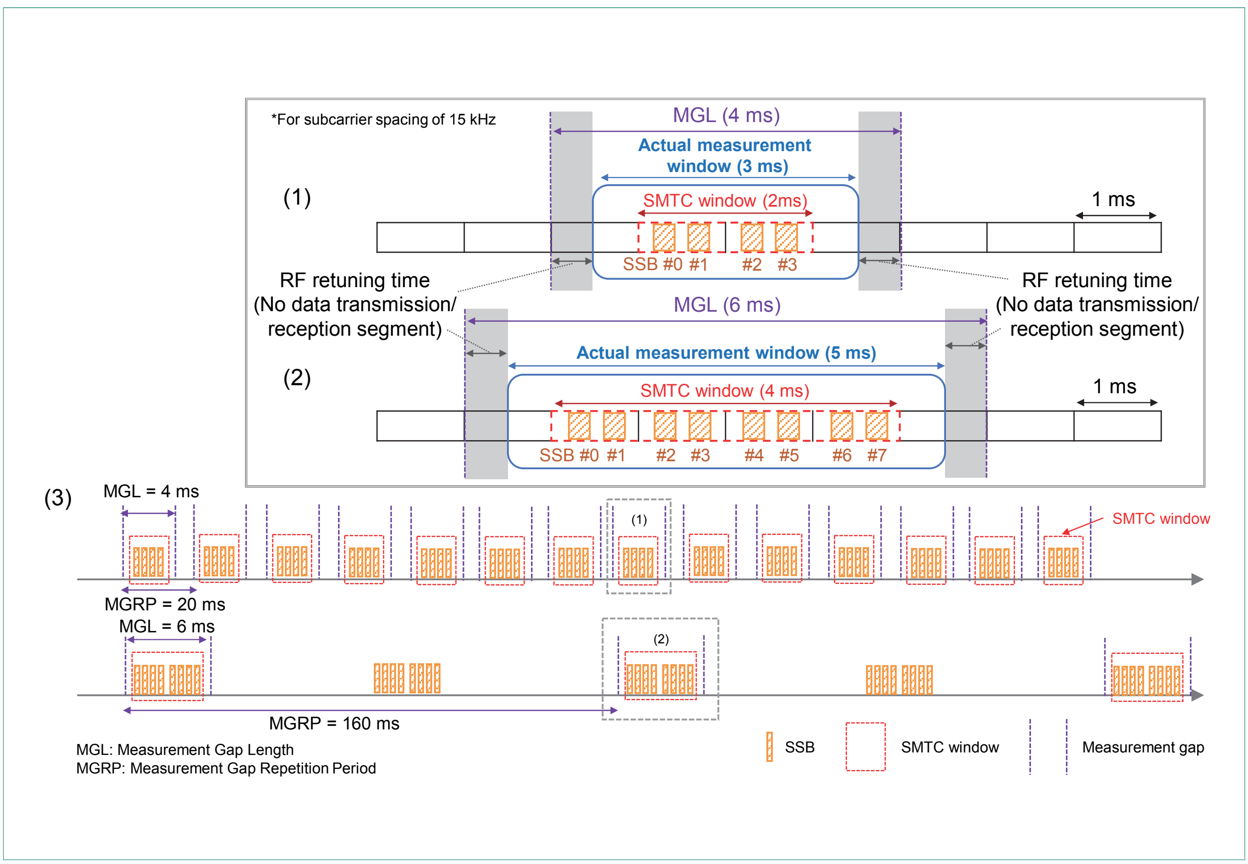


Figure 8 Measurement gap configuration in NR

*46 Cell ID: Identifying information assigned to each cell.

window of 2 ms and gap with MGL of 4 ms, while Fig. 8 (2) uses a 4 ms SMTC window and longer 6 ms MGL. The Measurement Gap Repetition Period (MGRP) can also be set more flexibly than LTE, as shown in Fig. 8 (3) (LTE values: 40, 80 ms; NR values: 20, 40, 80, 160 ms).

2) Measurement Gap Timing Advance

As mentioned earlier, RF retuning time is anticipated at the beginning and end of the configured Measurement gap, and during this time the terminal can neither perform measurements nor transmit or receive data. If, as shown in **Figure 9** (1), the SMTC window and Measurement gap start at the same time, the start of the SMTC window overlaps with RF retuning time and the UE cannot perform measurements in such duration. As such, the new Measurement gap timing advance function was introduced, which enables all of the measurement signal in the SMTC window to be used by advancing the start of the gap by the amount of the RF retuning time, as shown in Fig. 9 (2).

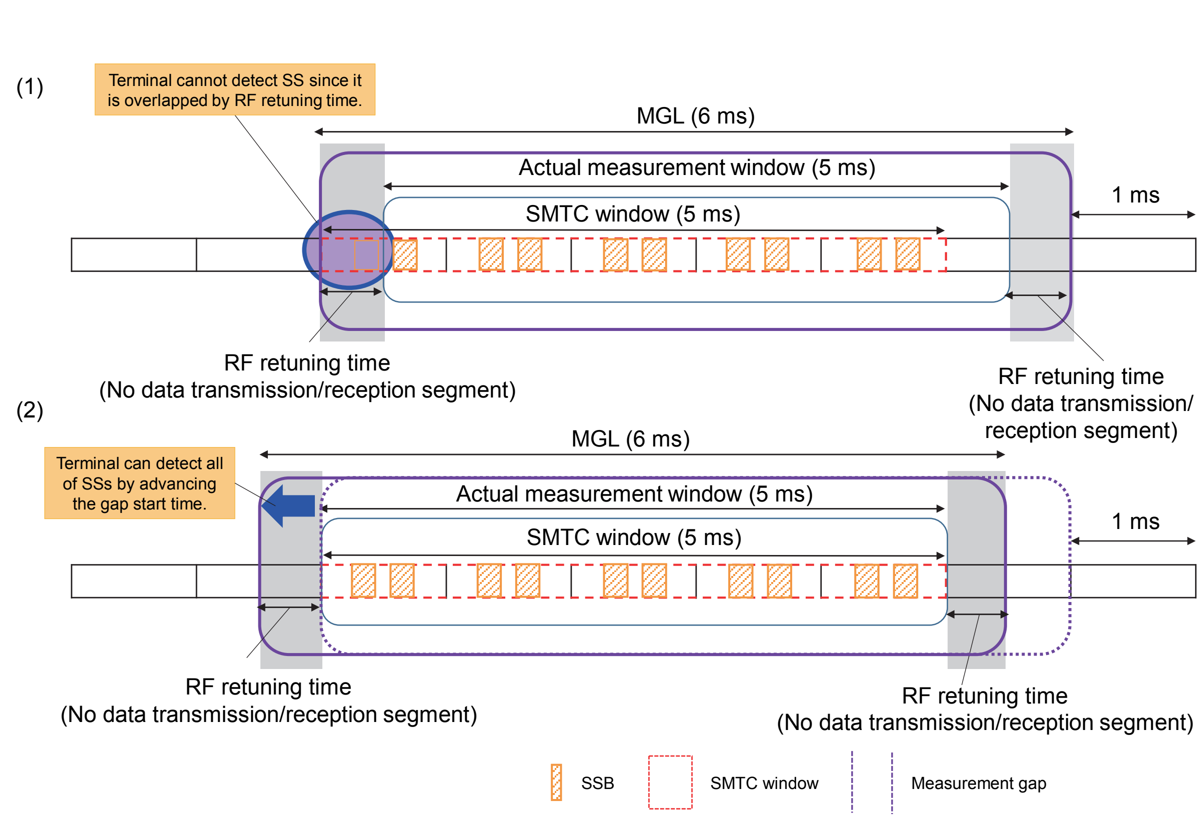


Figure 9 Measurement gap timing advance

5.3 Asynchronous Network Reception Quality Measurement

1) SFN[*47] and Frame boundary Timing Difference (SFTD) Measurement

The SMTC window and Measurement gap are set based on the timing of SSB transmission for the cell being measured. However, when measuring other cells and some other cases, the base station in the serving cell may not have such information. In such case, it is not possible to set SMTC window and Measurement gap appropriately for the terminal. For this reason, a new SFTD measurement function was introduced, in which the terminal measures the timing difference of SFN and frame boundary between the serving cell and NR neighbor cells being measured and reports it to the base station.

2) SFTD Configuration

The base station configures SFTD measurements for a terminal's NR neighbor cells. For a terminal to perform SFTD measurements for NR neighbor cells to which it is not connected, basically the Measurement gap is needed. However if the base station does not know the SSB transmission timing as described above, it cannot be configured appropriately (**Figure 10**). In such a case, the base station would need to configure the Measurement gap repeatedly changing the Measurement gap settings until the terminal detects the SSB of the NR neighbor cell. This would increase the signaling[*48] and delay until the terminal is able to connect to the NR cell. Accordingly, a new procedure for performing SFTD measurement of NR neighbor cells, without

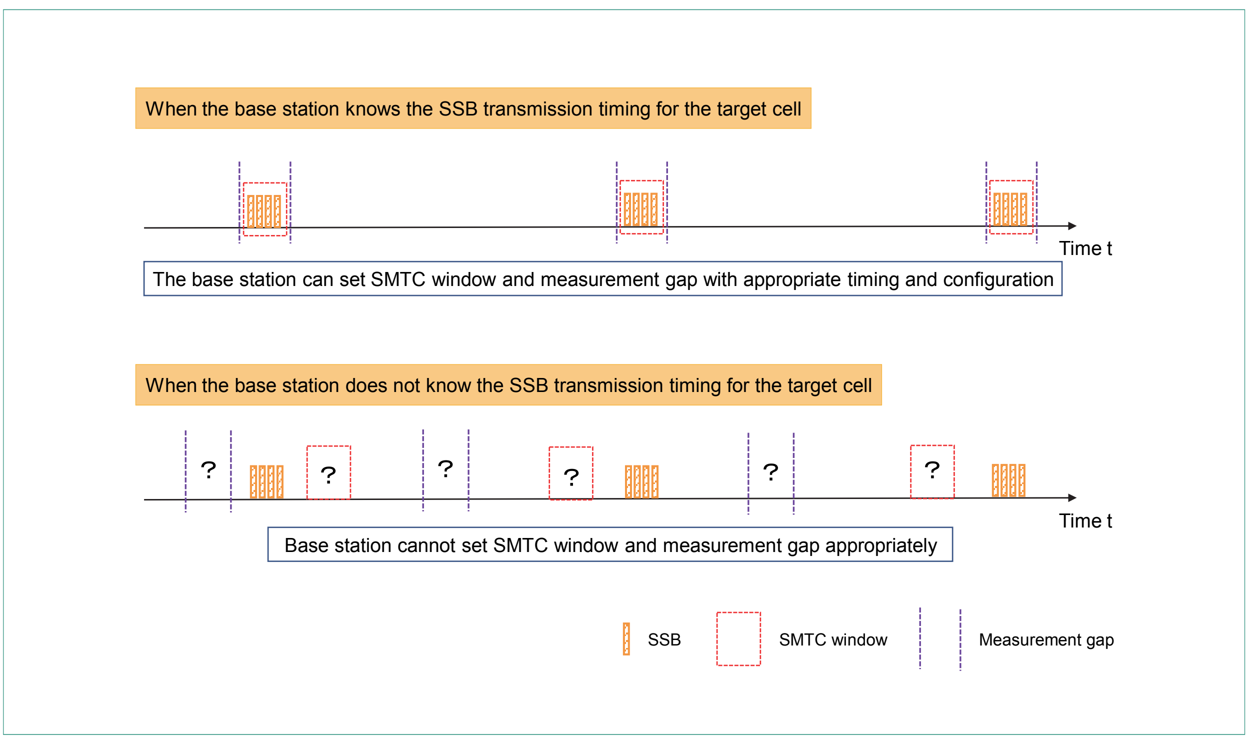


Figure 10 When SSB transmission timing is unknown for the cell being measured

*47 SFN: A number assigned to each radio frame. The value ranges from 0 to 1,023.

*48 Signaling: The sharing of information necessary for connection between base station and terminals (e.g. frequency band, coding and modulation formats, etc.).

using the Measurement gap and while maintaining transmission and reception data in the serving cell was introduced.

A concrete example of the SFTD measurement method is shown in **Figure 11**. Rather than stopping transmission and reception data by the RF device to perform measurements, this method uses a separate RF device on the terminal, which is used in case of CA. When the RF device used for measurement is turned ON or OFF, the RF device being used for transmission and reception data in the serving cell is affected, so the terminal cannot transmit and receive data in one sub-frame before and after the measurement window, but in the rest of the duration, the terminal can transmit and receive data in the serving cell while SFTD measurements for the NR neighbor cells are being performed.

6. Conclusion

This article has described the new radio format (NR) specified in Release 15, mainly regarding RAN 4 specifications for frequency bands, RF performance for base stations and terminals, and specifications related to radio resource management. In the future, baseband*49 performance and other specifications for base stations and terminals, regulating reception throughput, will be decided (planned completion by December, 2019). At the same time, study of Release 16 and later standard specifications will continue, to provide service areas that have even better quality and more advanced functionality.

REFERENCES

[1] 3GPP TS38.913 V14.3.0: "Study on scenarios and requirements for next generation access technologies," Aug. 2017.

[2] Information and Communications Council, Information and Communication Technology Subcommittee, New-generation Mobile Communications System committee (9th meeting): "Committee Report (Draft)," Jun. 2018.

[3] 3GPP TS38.101-1 V15.2.0: "NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone," Jun. 2018.

[4] 3GPP TS38.101-2 V15.2.0: "NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone," Jun. 2018.

[5] N. Miki et al.: "CA for Bandwidth Extension in LTE-Advanced," NTT DOCOMO Technical Journal, Vol.12,

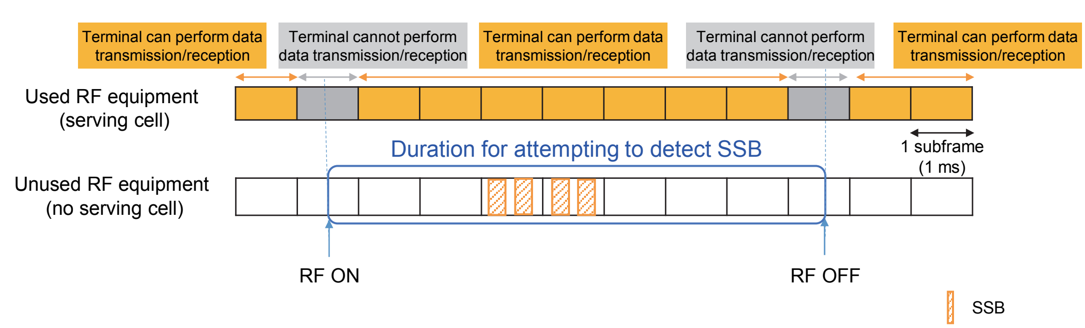


Figure 11 Implementing SFTD measurement without using measurement gap

*49 **Baseband:** The circuits or functional blocks that perform digital signal processing.

No.2, pp.10–19, Sep. 2010.

[6] H. Harada et al.: "Broadband Frequency Technologies in LTE-Advanced Release 13," NTT DOCOMO Technical Journal, Vol.18, No.2, pp.52–61, Oct. 2016.

[7] Qualcomm Technologies, Inc.: "5G Waveform & Multiple Access Techniques," Nov. 2015. https://www.qualcomm.com/media/documents/files/5g-research-on-waveform-and-multiple-access-techniques.pdf

[8] 3GPP TS38.104 V15.2.0: "NR; Base Station (BS) radio transmission and reception," Jun. 2018.

[9] Y. Sano et al.: "LTE-Advanced Release 13 Multiple Antenna Technologies and Improved Reception Technologies," NTT DOCOMO Technical Journal, Vol.18, No.2, pp.62–71, Oct. 2016.

Technology Reports

5G System | Network Slicing | Network Management

Special Articles on Release 15 Standardization
—Advancements in the Completed Initial 5G and LTE/LTE-Advanced Specifications—

Network Management Specifications for 5G Era

Research Laboratories Koji Tsubouchi Shigeru Iwashina

With the introduction of 5G systems that provide for network slicing as a mandatory feature, the existing network management architecture has been revised accordingly. Consequently, 3GPP SA5 WG has produced network-slicing related specifications for management architecture, management model, provisioning, performance management, fault management, etc. in Release 15.

1. Introduction

The 3rd Generation Partnership Project (3GPP) has documented the specifications for network slicing*1 functionality and its management services in Release 15.

Network slicing makes it possible to accommodate communication services having different requirements (Internet of Things (IoT), mobile broadband, etc.) into logically separated networks (i.e., network slices) on a single physical network. This, in turn, enables network operators to offer a variety of such services without mutual interference.

In this article, we describe network management specifications for managing network slices each consisting of Radio Access Network (RAN)*2 functions providing New Radio (NR) and Core Network (CN)*3 functions as overviewed in the leading article on Release 15 standardization in this issue [1].

2. Formulation of Network Management Specifications in 3GPP SA5 WG

The 3GPP Service and System Aspects (SA) 5*4

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 Network slicing: One form for achieving next-generation networks in the 5G era by logically dividing the network into units of services corresponding to use cases, business models, etc.

*2 RAN: The network consisting of radio base stations and radio-circuit control equipment situated between the core network and mobile terminals.

Working Group (WG) has been specifying requirements, architecture, and solutions for managing the 3GPP network (including RAN, CN, IP Multimedia Subsystem (IMS)*5) and services provided by that network.

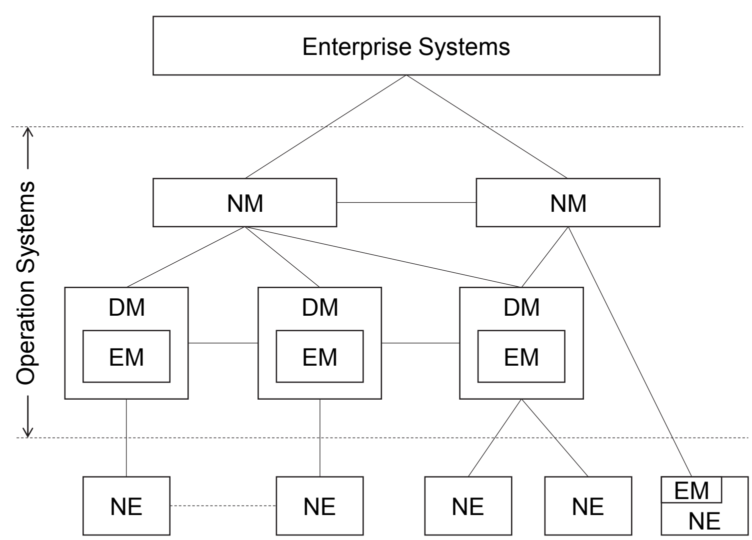
Prior to Release 15, the management architecture consisted of a hierarchical arrangement of functional blocks as shown in **Figure 1** (extracted from [2] Figure 1) and summarized below.

- Network Management (NM)*6 manages instances of Domain Management (DM)*7 that manages a 3GPP domain (RAN, CN)
- Each DM manages instances of Network Element (NE)*8 in the domain targeted for management via Element Management (EM)*9
- EM manages corresponding NE

With the introduction of 5G systems that provide for network slicing as a mandatory feature, 3GPP SA5 WG specified the revised management architecture and produced network-slicing related specifications for management architecture, management model, provisioning*10, performance management, fault management, etc. as network management specifications in Release 15 for the 5G era [3] [4].

3. Network-slice Configuration Specified by Release 15

In the new specifications in Release 15, a Network Slice Instance (NSI)*11 is defined to include all Network Functions (NFs)*12 necessary to provide a certain set of communication services.



Source: 3GPP TS32.101 V14.0.0 Figure 1

* Authors' note: Interface between DM/EM-NE is out of scope of standardization. Interface with another organization is omitted here.

Figure 1 Management architecture prior to Release 15

*3 CN: A network consisting of switches and subscriber information management equipment. Mobile terminals communicate with the core network via the radio access network.

*4 3GPP SA5: 3GPP is a standardization organization concerned with mobile communications systems. SA5 is a working group discussing OSS/BSS relevant topics in 3GPP.

*5 IMS: A subsystem that provides IP multimedia services (e.g., VoIP, messaging, presence) on a 3GPP mobile communications network. Session Initiation Protocol (SIP) is used for the calling control protocol.

*6 NM: Provides functions for managing a network consisting of NEs.

*7 DM: Provides functions for managing a portion of the network consisting of NEs.

*8 NE: Generic term for a base station, switching station, or relay device that make up the system.

Also defined is a particular portion of a slice (such as RAN, CN, etc.) called a Network Slice Subnet Instance (NSSI), which is a concept unique to 3GPP SA5 WG not defined in any other WG.

As shown in **Figure 2**, NFs can be allocated exclusively by a single NSI/NSSI or shared by multiple NSIs/NSSIs. Furthermore, while an NSI/NSSI includes a Transport Network (TN)*13, 3GPP SA5 WG specifies management only for the 3GPP-defined NFs in RAN and CN in Release 15 as has been the case so far. It only specifies input/instructions (configuration information, etc.) as required for the TN and relies on applicable Standards Developing Organizations (SDOs) to stipulate actual management.

4. Provision Formats for Network Slices

3GPP SA5 WG specifies Network Slice as a Service (NSaaS)*14 and Network Slices as Network OPerator (NOP) internals*15 as the possible models for providing network slices as shown in **Figure 3** (a) and (b), respectively [3]. The model in (a) provides a network slice as a network service to a third party (another operator) together with Application Program Interfaces (APIs)*16 for that operator. The model in (b), meanwhile, provides a network slice only as a communication service in a condition that cannot be managed by a third party (another operator), which means that that party is actually unaware of the slice's existence.

5. Management Architecture

5.1 Service Based Architecture

From Release 15 on, the 5G management system will adopt a service-oriented approach instead of conventional point-to-point connections between management functions.

This approach is achieved by Service Based Architecture (SBA)*17 in which management service producers provide a variety of management

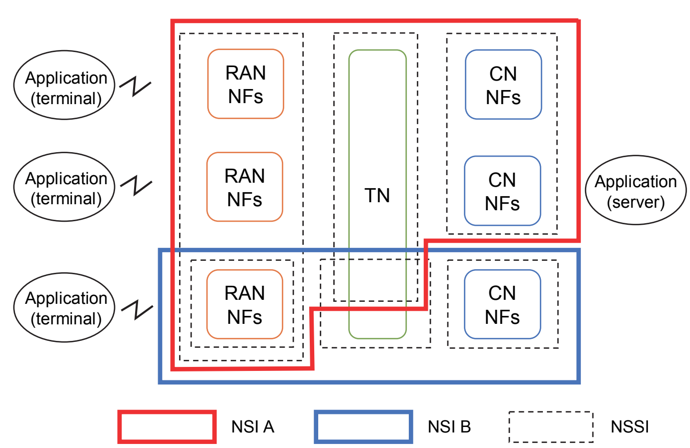


Figure 2 NSI/NSSI examples

*9 EM: Provides functions for managing NEs.

*10 Provisioning: Creation, modification, and deletion of equipment targeted for management such as NFs.

*11 NSI: NFs composing a network slice, and associated resources (e.g., computing resources, storage resources, networking resources).

*12 NFs: Generic name for base stations, switches, and repeaters making up a system. NF is synonymous with NE.

*13 TN: The network connecting the access network with the core network or the network interconnecting equipment within each of those networks.

*14 NSaaS: One business model for providing a network slice. This model provides a network slice as a service to a third party (another operator).

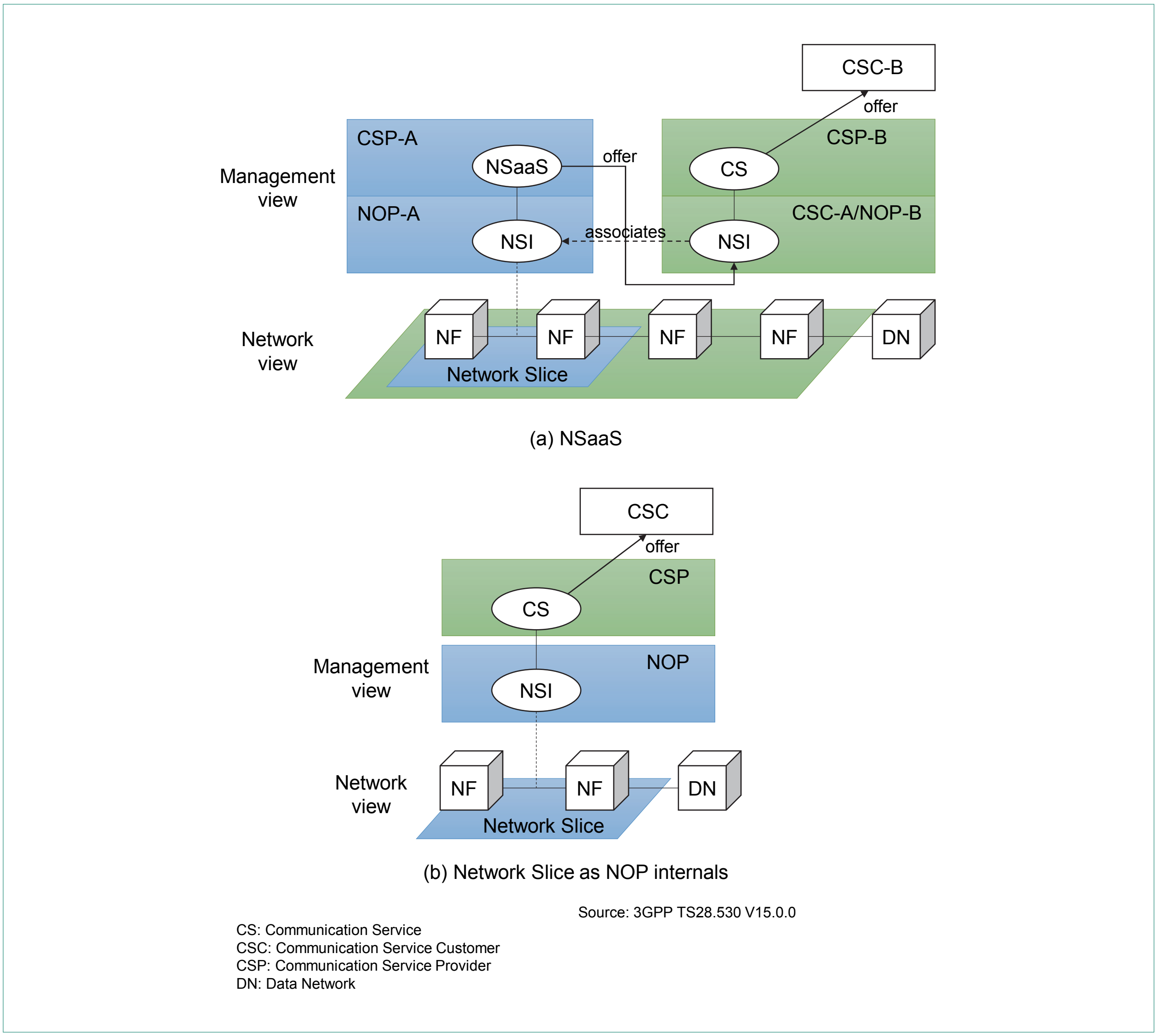


Figure 3 Provision formats for network slices

services to authorized consumers. Here, an implementation of multiple management services for a particular product is defined as a Management Function (MF) as shown in **Figure 4** (extracted from [4] Figure 4.6.1).

In addition, APIs provided by a management service are documented as normative*18 while the MFs themselves, which are implementation-dependent, are documented as informative*19. One advantage of adopting SBA is that each operator can freely select and deploy a combination of products that implement groups of management services that are deemed necessary when making decisions at the time of management-system integration*20.

---

*15 NOP internals: One business model for providing a network slice. In this model, a network operator provides a network slice as a communication service to an end user in a condition that cannot be managed by a third party (another operator).

*16 API: General-purpose interfaces for using functions and data.

*17 SBA: Architecture or framework that prescribes interfaces for using services and provides those interfaces to users.

*18 Normative: Information describing mandatory elements.

*19 Informative: Information describing non-mandatory elements provided mainly to help in understanding normative descriptions.

*20 Integration: The incorporation of equipment or systems in a network operated by an operator.

5.2 Management Services and MF

Management services are broadly specified as follows.

- Provisioning Management Service: Create, modify, get, and delete NSI/NSSI/NF management objects (life cycle management).
- Fault Management Service: Subscribe, notify NSI/NSSI/NF alarm, etc.
- Performance Management Service: Register, list, notify performance measurements, etc.

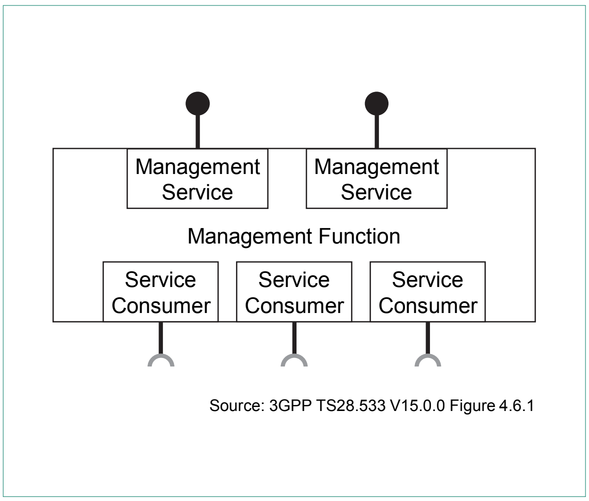


Figure 4 Example of management function and management services

Figure 5 showing an example of implementing MF architecture is documented in an Informative Annex*21 [4]. The MFs included in this figure are explained below.

- NFMF (Network Function MF) provides the management services for one or more NFs and may consume other management services.
- NSSMF (Network Slice Subnet MF) provides the management services for NSSI and may consume other management services.
- NSMF (Network Slice MF) provides the management services for NSI and may consume other management services.
- CSMF (Communication Service MF) consumes the management services provided by other MFs to manage communication services.

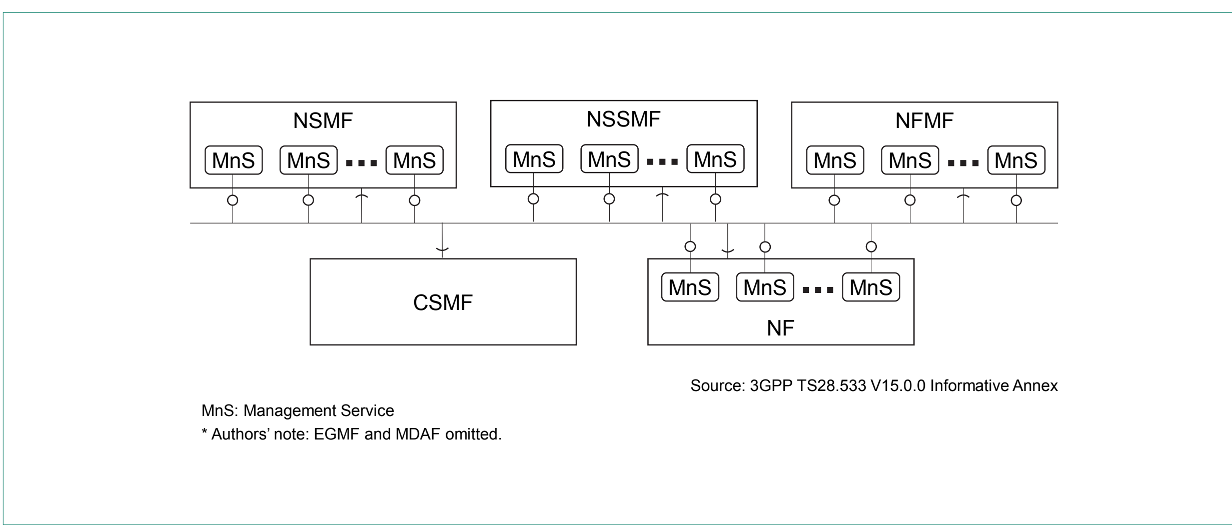


Figure 5 Example of implementing management function architecture

*21 Informative Annex: A supplementary chapter having no mandatory elements that is described mainly to provide information that can aide in understanding normative descriptions.

- NF provides management services such as NF performance management, NF fault supervision, and NF configuration management (does not consume other management services).

An important point reflected by this figure is that, while the interface between EM and NE/NF had not previously been standardized (Fig. 1), the NF interface is disclosed in Release 15 as part of an Informative Annex with expectations of being standardized in the future.

6. Conclusion

We described network management specifications prescribed by 3GPP. Going forward, SA5 WG is expected to specify general-purpose management services specific to no particular NFs in Release 16. Additionally, considering that network slicing is closely related to Network Functions Virtualization (NFV)*22, standardization in this area is likely to proceed in collaboration with European Telecommunications Standards Institute (ETSI)*23 ISG NFV.

REFERENCES

[1] S. Nagata et al.: “Technical Overview of the 3GPP Release 15 Standard,” NTT DOCOMO Technical Journal, Vol.20, No.3, pp.37–48, Jan. 2019.

[2] 3GPP TS32.101 V14.0.0: “Telecommunication management; Principles and high level requirements,” Mar. 2017.

[3] 3GPP TS28.530 V15.0.0: “Management and orchestration; Concepts, use cases and requirements,” Sep. 2018.

[4] 3GPP TS28.533 V15.0.0: “Management and orchestration; Management and orchestration architecture,” Sep. 2018.

*22 NFV: A technology that uses virtualization technologies to implement processing for communications functionality in software running on general-purpose hardware.

*23 ETSI: A standardization organization concerned with telecommunications technology in Europe.

Technology Reports

eSIM | LPA | One Number Service

LPA Application Conforming to GSMA eSIM Specifications

Communication Device Development Department Akio Goda Kumiko Yamaguchi† Koichi Asano

NTT DOCOMO developed and introduced Japan's first eSIM-compatible terminals for consumers in 2017. Going forward, the plan is to expand the lineup of these eSIM-compatible terminals while creating new eSIM-based services and expanding the number of use cases. To this end, it was decided to develop an LPA application for general-purpose operation on Android™*1 smartphones and to incorporate this application in terminals scheduled for sale in 2018 and beyond. This article describes the mechanism of this LPA application developed by NTT DOCOMO.

1. Introduction

Recent years have seen a growing demand for a cellular communications function that can be loaded on various types of Internet of Things (IoT) devices including wearable terminals. To meet this demand, NTT DOCOMO developed and introduced an embedded SIM (eSIM)*2 platform conforming to GSM Association (GSMA)*3 Remote SIM Provisioning (RSP)*4 specifications [1]. This eSIM platform made it easy to activate an NTT DOCOMO circuit on a terminal but the problem remained of having to develop a Local Profile Assistant (LPA)*5 function compatible with the platform for each terminal model.

Given this problem and anticipating an increase in the number of eSIM-equipped terminals and an expansion of services that use eSIM, NTT DOCOMO decided that it would develop an LPA application running on the Android platform and load this

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently Platform Business Department

*1 Android™: A trademark of Google, LLC.
*2 eSIM: Generic name for SIM that can install profiles remotely.

application on terminals. This article describes the mechanism of the developed LPA application.

2. Purpose of LPA Application Development

2.1 Carrier-provided Advantages

In general, an Android terminal manufacturer must develop both hardware and software. The hardware of an Android terminal gives the terminal its special character in terms of specifications (display size, camera performance, etc.), case design, etc. For this reason, hardware is basically different from one terminal to the next. Software, on the other hand, frequently makes up the common portion of terminals. Furthermore, while there is software for terminal functions that directly operates on hardware, such software may be capable of running as a single Android application independent of the terminal in use. These features can be exploited to develop a common Android application equipped with LPA functions for performing eSIM activation (hereinafter referred to as “LPA app”).

It was therefore decided that NTT DOCOMO would develop an LPA app that terminal manufacturers had previously been developing and distribute the LPA app to them (**Figure 1**). This approach reduces the burden of LPA development on terminal manufacturers enabling them to focus on developing hardware and expanding their terminal lineup. At the same time, it enables NTT DOCOMO to make an indirect contribution to the expansion of eSIM-equipped terminals.

The development of the LPA app by NTT DOCOMO also has secondary effects. For example, it can raise efficiency by unifying the development/evaluation of the application with the NTT DOCOMO network and server equipment that it links to, creating a uniform application User Interface (UI) between terminals, and simplifying user guidance.

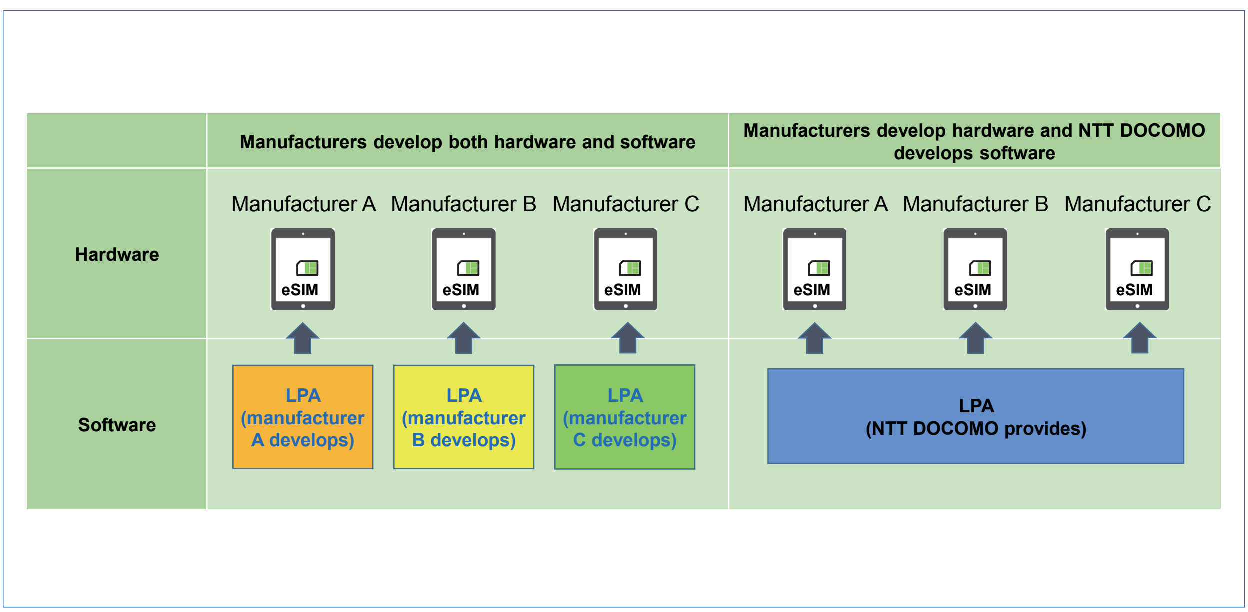


Figure 1 Hardware/software development sharing

*3 GSMA: A global trade body of mobile operators that also includes terminal manufacturers, software companies, and other companies in the mobile industry. In addition to activities such as formulating roaming rules between operators, GSMA leads eSIM-related standardization.

*4 RSP: Generic name for remote profile writing technology for use with eSIM as defined by GSMA.

*5 LPA: Function for linking with SM and downloading profiles to eSIM and profile management functions such as profile switching, deletion, etc.

2.2 Role of LPA App in eSIM Activation

The LPA app has two main functions as shown in **Figure 2** and summarized below.

- Local Profile Download (LPD): Downloads a profile*6 from Subscription Manager (SM)*7 and stores and installs the profile in the terminal's eSIM.
- Local User Interface (LUI): Provides a UI for managing profiles within the eSIM (display, switch, delete, etc.)

When installing LPA functions on a terminal, consideration must be given to the role played by that terminal in eSIM circuit-activation processing and the arrangement of those functions. The following two function-implementation methods are presented in GSMA RSP specifications Version 2 [2] [3] (**Figure 3**).

The first method loads the eSIM and all LPA functions in the terminal targeted for circuit activation so that the terminal itself has eSIM activation ability. This terminal is called a "standalone device" below.

The second method loads the eSIM and a portion of LPA functions in the terminal targeted for circuit activation. This terminal itself does not have eSIM activation ability. In this case, the terminal interfaces with another terminal equipped with the remaining LPA functions to activate the circuit. In the following, the former terminal is called a "companion device" and the latter terminal a "primary device." *1

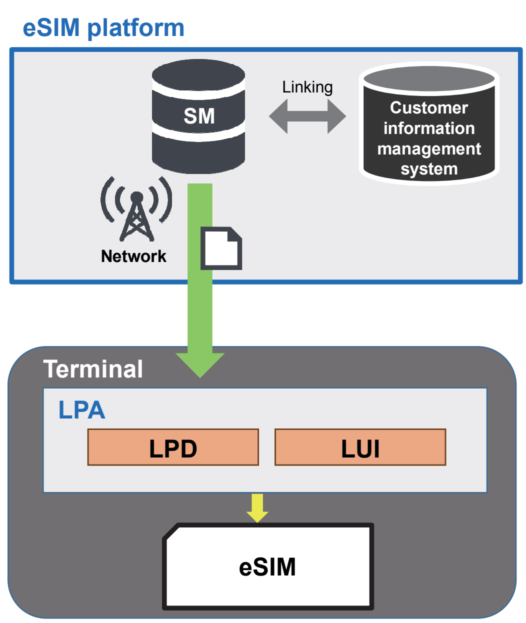


Figure 2　Overview of eSIM platform and LPA app

*6 Profile: UIM software running on eSIM OS consisting of files storing a telephone number and other types of information, an application having a network authentication function, etc.

*7 SM: A server providing a function for generating and saving profiles, a function for downloading and installing profiles in eSIM via LPA, etc.

*1 "Companion Device" and "Primary Device" are defined in GSMA specifications [2] [3].

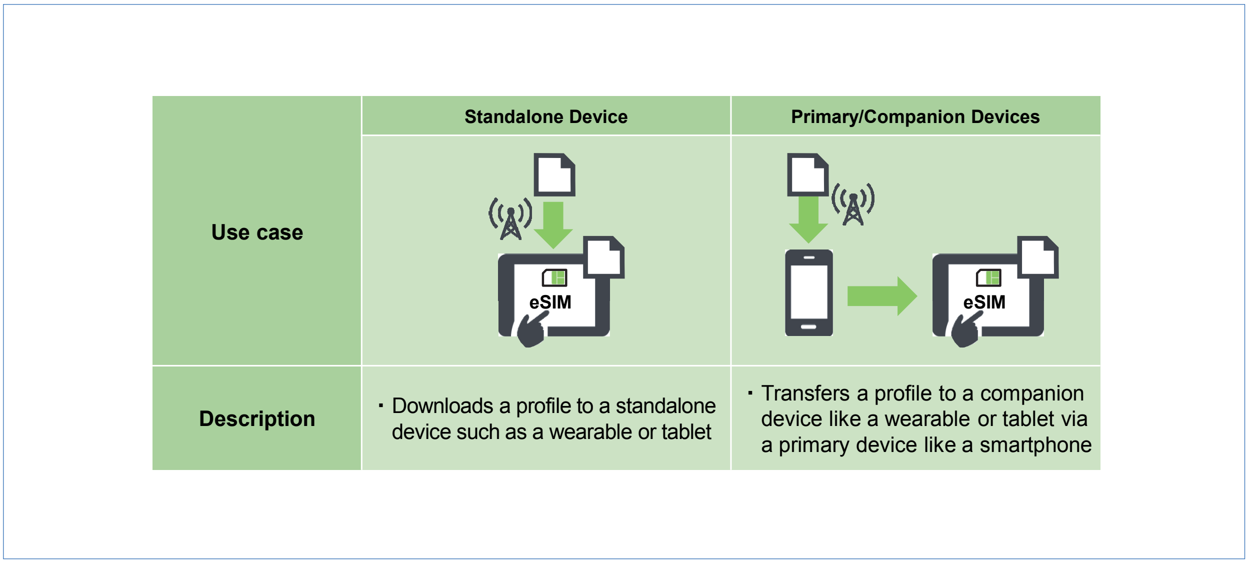


Figure 3　Standalone device and primary/companion devices

The LPA app developed by NTT DOCOMO supports all of the above devices: standalone, companion, and primary.

3. Standalone Device Model

3.1 Functional Configuration of LPA App for Standalone Device

In the standalone device model, only that one terminal is capable of eSIM circuit activation, so the LPA app for a standalone device contains all LPD and LUI functions. The function configuration is the same as that shown in Fig. 2.

This LPA app developed by NTT DOCOMO conforms, of course, to GSMA specifications and was designed to be an application that users could easily use on any Android terminal. Specifically, we improved ease-of-use by making user authentication extremely simple while ensuring security and designing the UI in conformance with Android Material Design*8 and Setup Wizard.

3.2 Mechanism of Circuit Activation in Standalone Device

The main components of the eSIM platform other than the terminal and LPA app are eSIM and SM. The following provides a brief explanation of these two components. Details can be found in a past article [1] in NTT DOCOMO Technical Journal.

- eSIM: The conventional scheme was to distribute SIM cards to end users after writing data to them “beforehand” such as internal files (telephone number, various user identifiers, etc.) and an Operational Profile (OP) consisting of a network authentication function and various applications. In contrast, an eSIM features the capability of adding or modifying a profile “afterward” by remotely and securely downloading an OP from the SM server.
- SM: This component generates and saves profile data for each user. Given a profile

*8 Material Design: Comprehensive design guidelines recommended for Android applications covering graphics, screen configuration, operational feeling, etc.

download request from an LPA, SM prepares the profile for that user and downloads it securely to the target eSIM.

The linking of these two components enables eSIM circuit activation. The sequence of operations performed for downloading an OP to an eSIM is shown in **Figure 4** and summarized below.

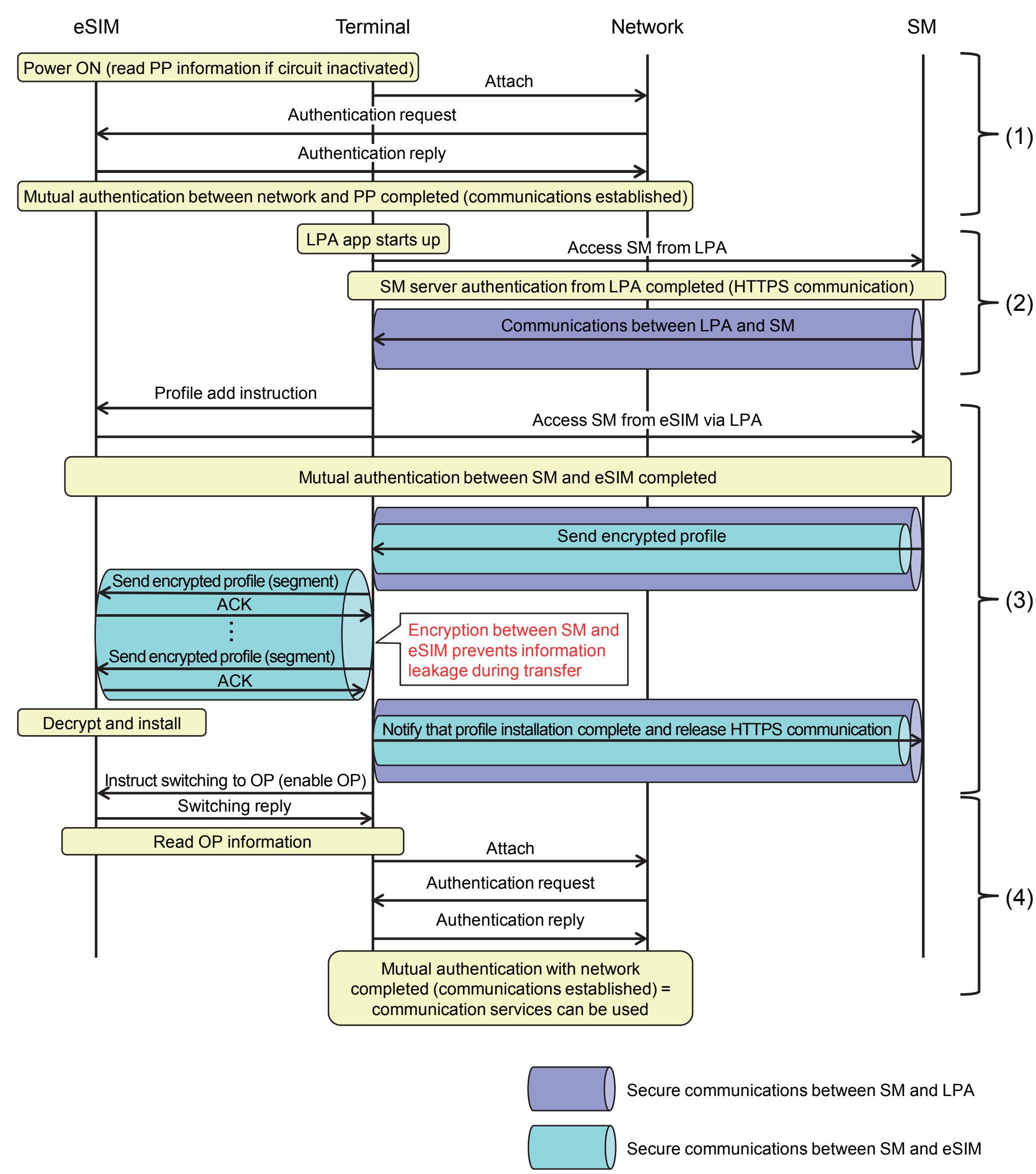


Figure 4 Activation sequence for standalone device

(1) At the time of power ON, the terminal reads profile information from within the eSIM. If there are no OPs at all in the eSIM, that is, if no circuit has been activated yet, the Provisioning Profile (PP)*9 is automatically read. The PP enables a packet communications call to be made to enable the terminal to connect to SM.

(2) The LPA app starts up (in general, the LPA app is automatically called from the Setup Wizard in the terminal). The LPA app connects to SM and establishes an HTTPS communication session*10. At this time, the LPA app performs a server certificate*11 verification using its own certificate to verify the validity of SM. This process prevents accidents such as the uploading of user data from the terminal to an unauthorized SM through impersonation (spoofing) or the downloading of unauthorized profile data likewise from an unauthorized SM.

(3) The eSIM and SM perform mutual authentication via the communication path created by the LPA app in step (2) thereby establishing an end-to-end secure encrypted communication path between eSIM and SM. The eSIM now proceeds to download a profile from SM on this communication path. Then, once the profile has been downloaded to the LPD function on the LPA app, the LPD function transmits the profile in segments to the eSIM. This is due to the limited message size in the communication interface between the terminal and eSIM. These profile segments are combined within the eSIM and the profile is installed after decryption.

(4) Finally, once the installed profile app has been enabled from the LPA, the terminal uses the OP to again attach*12 itself to the network making a variety of services available such as voice calls and packet communications.

4. Primary/Companion Device Model

4.1 Advantages of Equipping Companion Device with eSIM

A typical use case of the primary/companion device model features 1) a terminal (companion device) such as a wearable or on-vehicle device for which eSIM circuit activation processing is difficult on its own and 2) a smartphone terminal (primary device) oriented to UI operations such as input and screen display required for a performing a circuit activation procedure. In this use case, the latter performs eSIM circuit activation on behalf of the former.

Companion devices are often equipped with an embedded SIM fixed within the terminal. In contrast to conventional removable SIMs, the embedded type negates the need for an SIM slot, which means a device with less hardware and reduced volume having an affinity with waterproof/dustproof mechanisms. Furthermore, by having the primary device shoulder the burden of eSIM-circuit-activation operations and procedures, the companion device can omit input devices such as a touch panel or keyboard and reduce display size thereby achieving a hardware/software design specific to the purpose of that device.

Additionally, since circuit activation can be performed online via the primary device, “obtaining a terminal” and “activating a cellular circuit” can be performed separately, which is not a feature found

*9 PP: A dedicated profile used when downloading an OP. It cannot be used in other applications.

*10 Session: A series of communications exchanged between a client and server.

*11 Server certificate: A digital certificate containing server-owner information, cryptographic key, and signature data of certificate issuer.

*12 Attach: The process of registering a mobile terminal to a network when the terminal's power is turned on, etc.

in past terminals. This approach opens up the possibility of distributing companion devices through a variety of sales channels instead of being restricted to carrier shops or volume retailers.

4.2 Companion Device Development at NTT DOCOMO

NTT DOCOMO has developed "One Number*13 Phone*14 ON 01" as a 2018 winter model (**Photo 1**). One Number Phone is a companion device equipped with an eSIM that is capable of downloading an OP to the eSIM via an Android smartphone as a primary device. This scheme makes it possible to give the One Number Phone a compact design specific to the telephone function.

NTT DOCOMO's One Number Phone is the world's first eSIM companion device approved as a GSMA Compliance Process [4]*2.

4.3 Functional Configuration of LPA App for Companion Device

We developed an LPA app for the primary device (hereinafter referred to as "primary app") and an LPA app for the companion device (hereinafter referred to as "companion app") based on the LPA app for a standalone device. Dividing up main LPA functions into a primary app and companion app required a mechanism for real-time linking between these two apps, which we accomplished through Bluetooth®*15 communications. The components of this primary/companion device model are shown in **Figure 5**.

4.4 Activation Mechanism in Companion Device

In addition to the components appearing in standalone-device activation, eSIM activation for the companion device requires linking anew with the customer information management system. This system has functions for managing the One Number Service®*16 contract status of the companion-device circuit user, performing an immediate contract-creation procedure online if no contract exists, etc.

Photo 1 One Number Phone ON 01

*13 **One Number:** A service that enables a single phone number to be shared by a primary device like a smartphone and a companion device like an accessory terminal and that enables either type of terminal to make voice calls or use data communications.

*14 **One Number Phone:** A trademark or registered trademark of NTT DOCOMO, INC.

*2 As of October 23, 2018 based on an NTT DOCOMO survey.

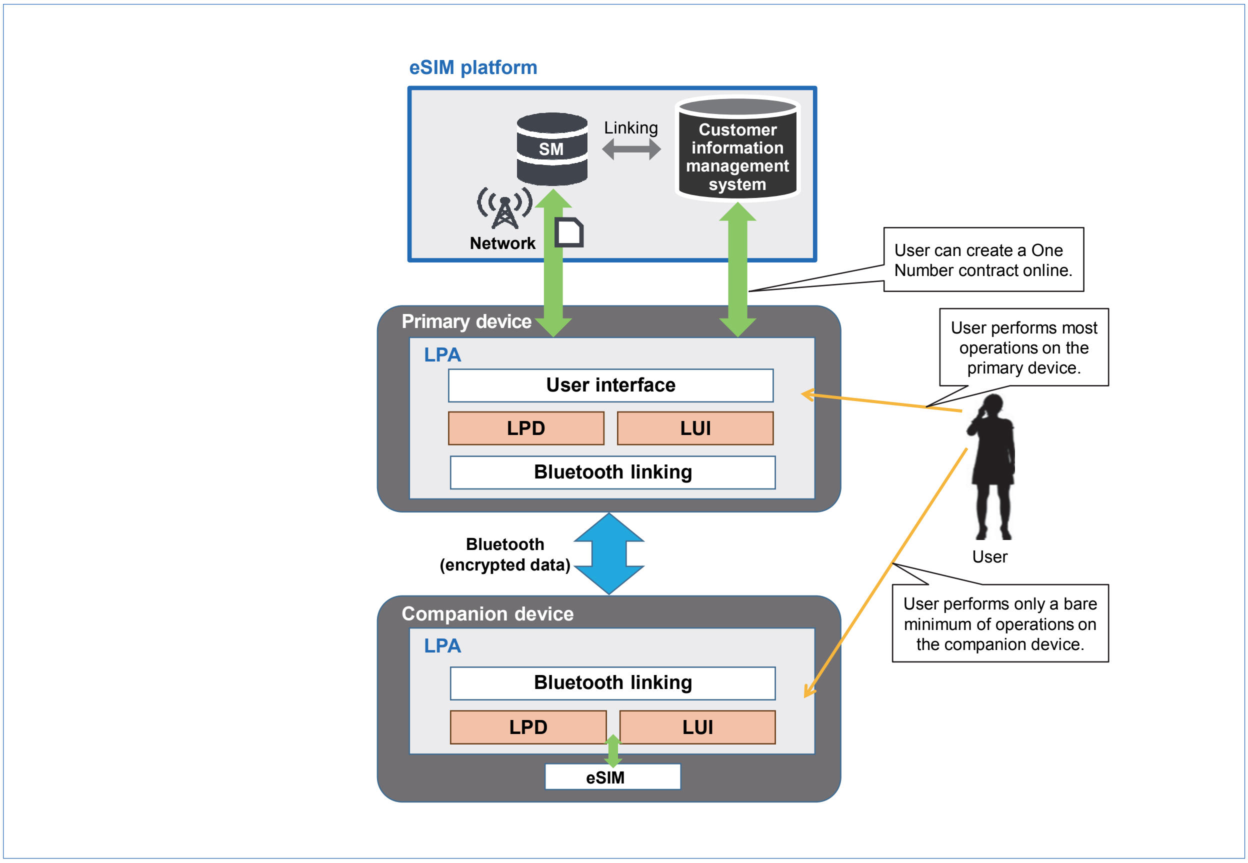


Figure 5　LPA app configuration for primary/companion device model

The activation sequence for the companion device is shown in **Figure 6**. This process assumes that an SIM card capable of cellular communications is already inserted in the primary device and that a packet-communications call has been established.

(1) Both the primary app and companion app start up.

(2) The primary app uses the Android Bluetooth function to perform Bluetooth pairing*17 between the primary device and companion device. This pairing enables the primary app and companion app to directly and securely interact with each other during the Bluetooth session.

(3) The primary app connects to the customer information management system and checks One Number Service contract status. If no contract has been created, a contract-creation procedure begins.

(4) The LPD function of the primary app accesses SM and establishes secure HyperText Transfer Protocol Secure (HTTPS)*18 communications.

(5) Making use of the previously established Bluetooth connection between the primary app and companion app and HTTPS communications between the primary app and

---

*15 Bluetooth®: A short-range wireless communication specification for wireless connection of mobile terminals, notebook computers, PDAs and other portable terminals. Bluetooth is a registered trademark of Bluetooth SIG Inc.

*16 One Number Service®: A registered trademark of NTT DOCOMO.

*17 Pairing: A procedure in Bluetooth for connecting two devices. Once pairing has been successfully performed between two devices, subsequent connections between them can be automatically or semi-automatically established.

*18 HTTPS: A method for performing secure HTTP communications using TLS protocol to prevent spoofing, intermediary attacks, and eavesdropping.

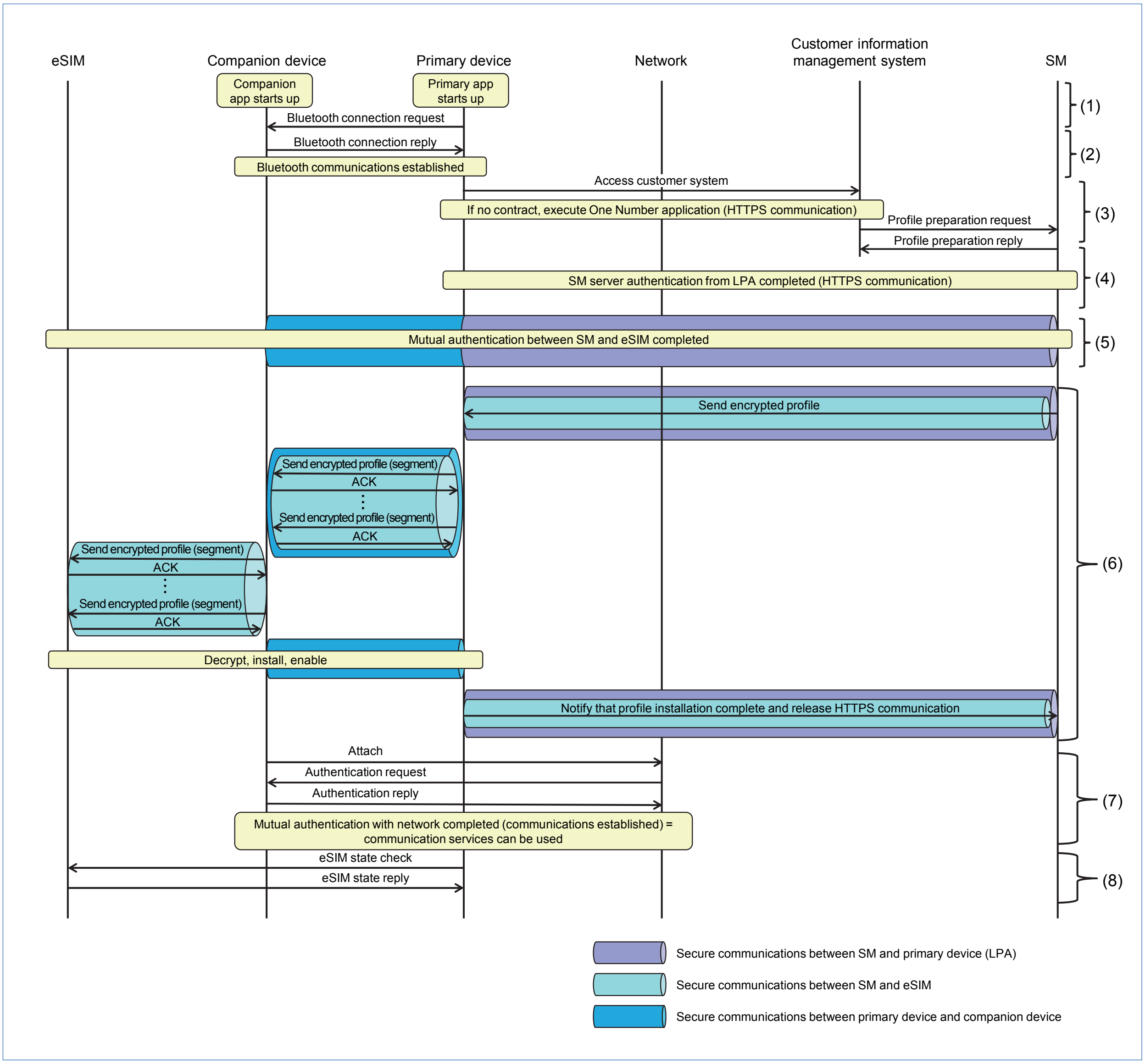


Figure 6 Activation sequence for companion device

SM server, eSIM on the companion device and SM perform mutual authentication and establish a secure encrypted communication path between them. This, of course, prevents the eavesdropping or tampering of profile information on the Internet circuit of that transmission path as well as in the primary-app or companion-app while also guaranteeing the authenticity of the profile to be downloaded to the eSIM.

(6) The LPD function of the primary app downloads the encrypted profile information from SM and transmits the profile in segments to the LPD function of the companion app.

This is due to the limited message size in Bluetooth communications between the primary and companion devices. The LPD function of the companion app now transmits the received profile in segments to eSIM and installs an OP in the eSIM.

(7) If the OP is correctly installed, the companion device attaches itself to the network enabling it to perform cellular communications on its own.

(8) The internal state of the eSIM can be checked from the LUI function of the primary app by accessing the eSIM via the LUI function of the companion app (for example, the user can view installed OP information).

The One Number Phone incorporates a function for copying the phonebook data in the primary device to the companion device after the above procedure has been completed. The One Number Service has been designed to allow for the seamless use of calls and SMS on either the primary-device or companion-device terminal.

5. Conclusion

NTT DOCOMO has developed an LPA application conforming to global GSMA RSP specifications with an eye to expanding eSIM-compatible devices. The LPA app for a standalone device is expected to promote the equipping of conventional Android smartphones and tablets with eSIM. The LPA app for primary/companion devices, meanwhile, should contribute to the bestowing of cellular communication ability to wearables and other types of IoT devices and to the creation of device-unique use cases.

Looking to the future, we plan to continue with LPA application development to conform to next-generation RSP specifications now being drafted at GSMA. We will also undertake the provision of increasingly convenient communication services using eSIM technology for consumer devices including the expansion of eSIM-equipped IoT devices and creation of new use cases.

REFERENCES

[1] T. Sasagawa et al.: "eSIM for Consumer Devices toward Expanded eSIM Usage—Secure Installation Conforming to GSMA—," NTT DOCOMO Technical Journal, Vol.19, No.2, pp.5–13, Oct. 2017.

[2] GSMA SGP.21: "RSP Architecture Version 2.1," Feb. 2017.

[3] GSMA SGP.22: "RSP Technical Specification Version 2.1," Feb. 2017.

[4] GSMA SGP.24: "RSP Compliance Process Version 1.1," May 2017.

Providing VoLTE Interconnectivity

Core Network Development Department Takayuki Sato Tomonori Kagi†1
Aki Ohashi†2 Yuu Momma

High quality voice calling, which to date was only available between terminals belonging to the same Mobile Network Operator (MNO) in Japan, became available between terminals of different MNOs from October 2018* - an improvement in usability. High-quality voice calling refers to voice calling covering a wide bandwidth of audio frequencies, and is enabled by using Adaptive Multi-Rate Wide-Band (AMR-WB)*1 and Enhanced Voice Services (EVS)*2 for the audio codec.

Currently, connections between the MNOs are routed via existing voice communications networks with Synchronous Transfer Mode (STM)*3 lines, but VoIP standards cannot be applied to these existing call processing systems (Common Channel Signaling Systems*4), it has not been possible to achieve high quality voice calling with VoLTE. Thus, to use VoIP between different MNOs, NTT DOCOMO developed and implemented interworking equipment and peripheral equipment to process the relevant signals. In addition, NTT DOCOMO established DOCOMO Technical Requirements for interconnection [1], and newly organized its interface conditions such as various parameters for Session Initiation Protocol (SIP)*5, so that they fall within the scope of compliance with international standards [1].

This article describes functions involved in the operations to open connections to provide VoLTE interconnectivity.

1) Connection Route (Point Of Interface (POI)*6) Determination

Conventionally, interconnection between MNOs for voice calling is routed through a POI with the STM line (STM-POI) on the existing voice communications network, whereas VoLTE interconnection using SIP is routed through a POI on the IP connection (IP-POI) (**Figure 1**).

For this reason, it is necessary to decide which POI to use to connect to another MNO network when making phone calls. This is performed on the originating side by distinguishing whether the

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

†1 Currently, DOCOMO CS, Inc., Solution Integration Department
†2 Currently, R&D Strategy Department

* Initially, interconnectivity was only applied to a portion of originating calls, and target users being expanded sequentially.

*1 AMR-WB: An audio codec standardized by 3GPP with higher quality than AMR-NB, and used for voice telephony services, etc.

*2 EVS: An audio codec used for digitizing voice.

*3 STM: A type of multiplexing system for simultaneous communications that entails splitting on a physical line into several logical lines. A time-division multiplexing system used on networks with fixed communications speeds.

*4 Common Channel Signaling System: A method of sending and receiving control signals on a physically separate transmission route to the calling route. A common example is SS7 (Signaling System No. 7).

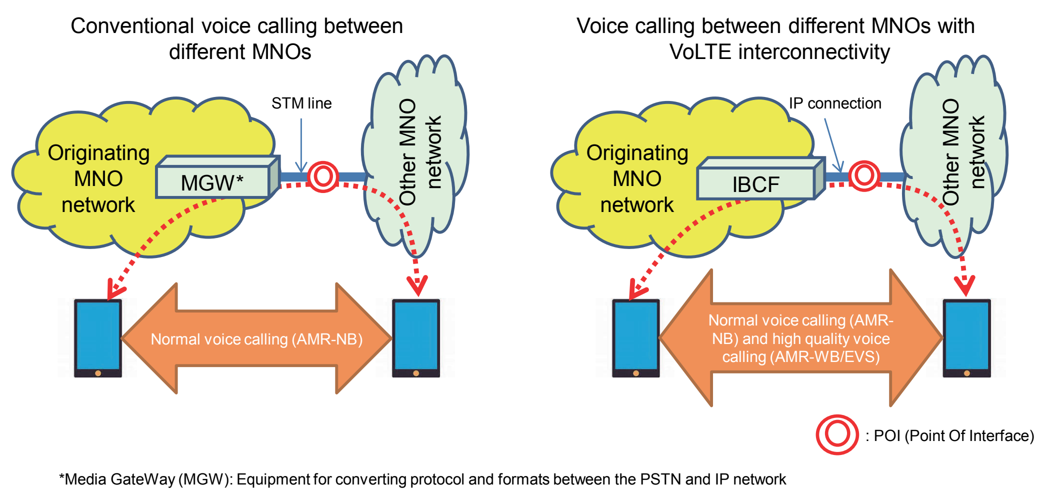


Figure 1　Difference of POIs between conventional connection and VoLTE interconnection

call was a VoLTE call or a 3G call. The Call State Control Function (CSCF)*7 determines which POI to use based on 3G/VoLTE information attached by the Application Server (AS)*8 (**Figure 2**).

2) Mobile Number Portability (MNP)*9 Support

MNP is available between MNOs in Japan, therefore it is difficult to determine the terminating MNO from the phone number alone. In conventional Common Channel Signaling System using STM-POI, the originating side recognizes the Original Number Range Holder (ONRH)*10 from the destination phone number, and sends a connection signal to it. If the destination for the signal is a different operator due to MNP, the originating side receives a disconnect signal with information to redirect to the actual destination MNO to which the number is ported, and then sends a reconnect signal to that MNO (**Figure 3**). In contrast, in interconnection via IP-POI, the number ported operator is determined by a resolution scheme using E.164 NUmber Mapping (ENUM)*11 [2] before sending the SIP request to the destination MNO.

*5 SIP: A call control protocol important in achieving VoIP defined by the Internet Engineering Task Force (IETF) for IP telephony, etc.

*6 POI: Connection points that mutually connect the communication devices of each company, and are also the responsibility boundary points of each company.

*7 CSCF: A central function in an IP multimedia subsystem. Functional equipment that performs call control with mobile terminals and capability exchange, etc.

*8 AS: Network Entity responsible for telephony service control, etc.

*9 MNP: A service which allows users to continue to use the same phone number when they change their contracted MNOs. Although MNOs who can assign mobile phone numbers are determined by the Ministry of Internal Affairs and Communications, this service enables continuation of services with a different MNO to the one that assigned the phone number.

*10 ONRH: A network operator who is assigned telecommunications numbers from the Ministry of Internal Affairs and Communications. Even when a phone number is ported to a different network operator with Number Portability, the ONRH retains the new operator information and notifies the information to the originating operator to achieve an appropriate call connection.

*11 ENUM: A method for converting phone numbers to domain names based on E.164, an ITU-T recommended numbering plan for telephone networks. Destinations parties and supported application information are derived from the URI acquired by DNS functions.

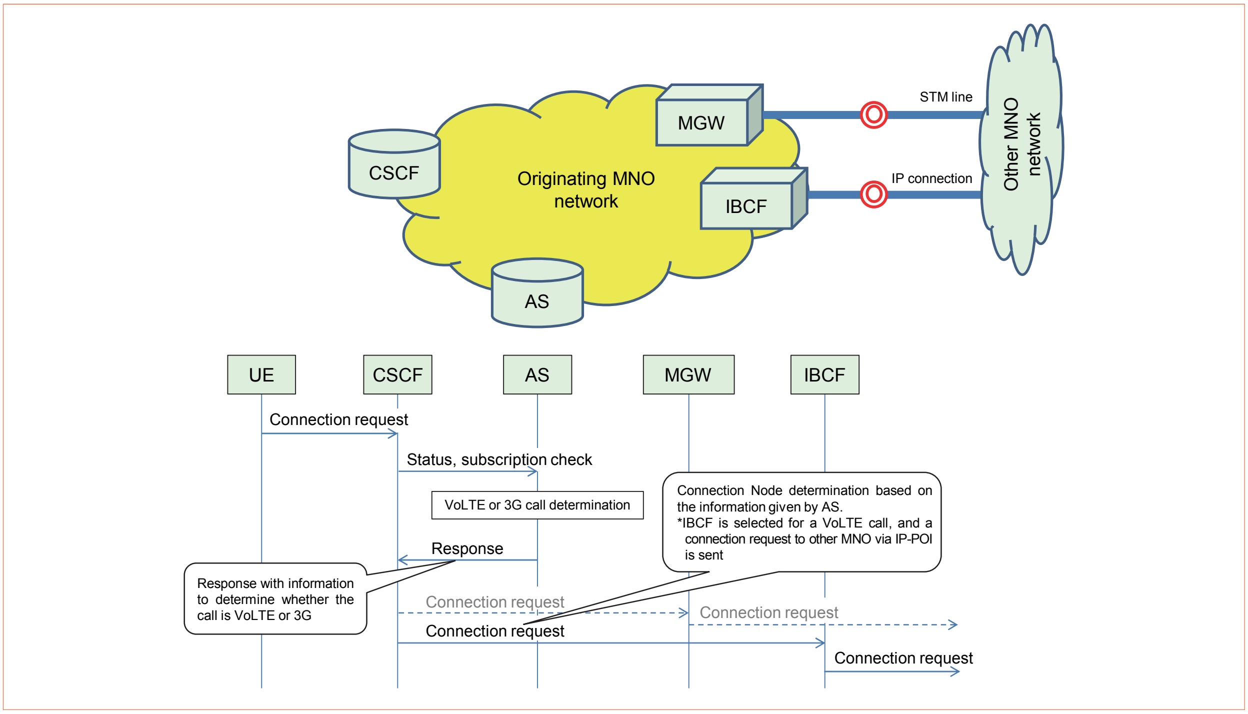


Figure 2 Overall procedure for interconnection with other MNOs, and POI determination

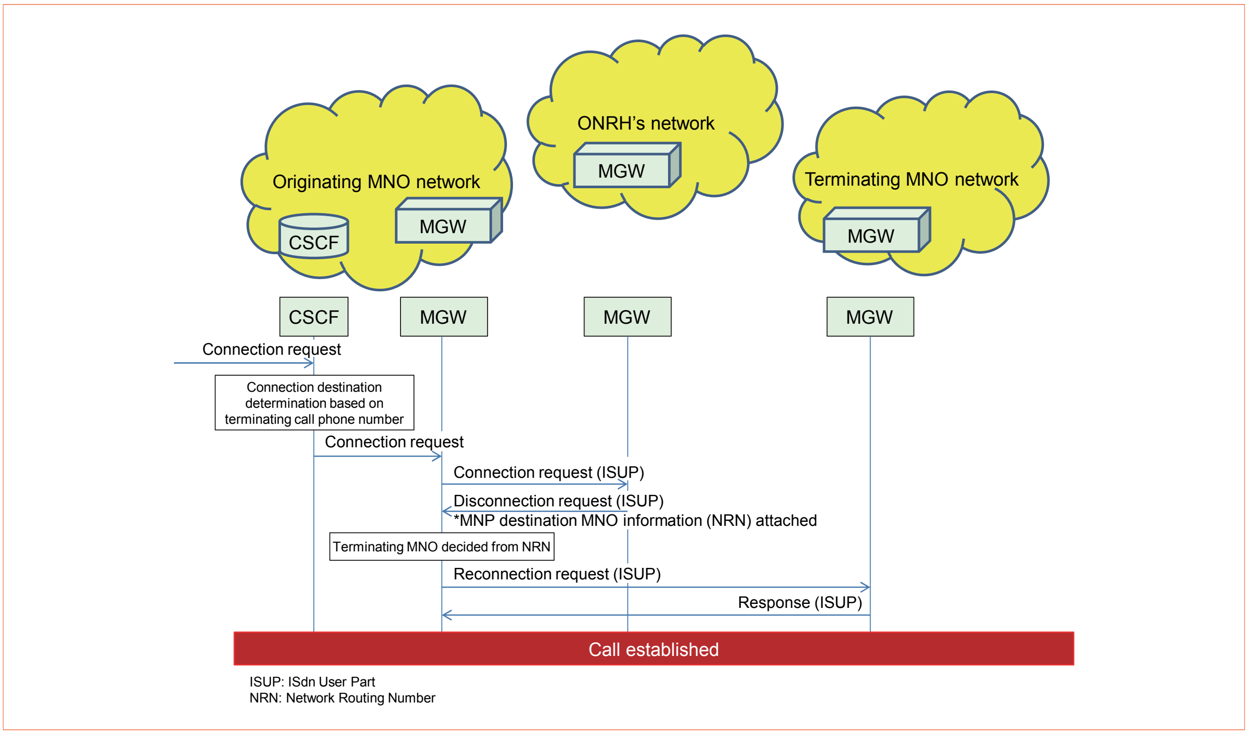


Figure 3 Connection example with MNP for existing STM-POI (Terminating call originating from NTT DOCOMO to an MNP terminal on another MNO network)

(Figure 4).

The ENUM server has information about the phone numbers assigned to the operator and where the numbers are ported. The ENUM client determines which ONRH's ENUM server to query for each number range (Fig. 4 (2)), and performs the

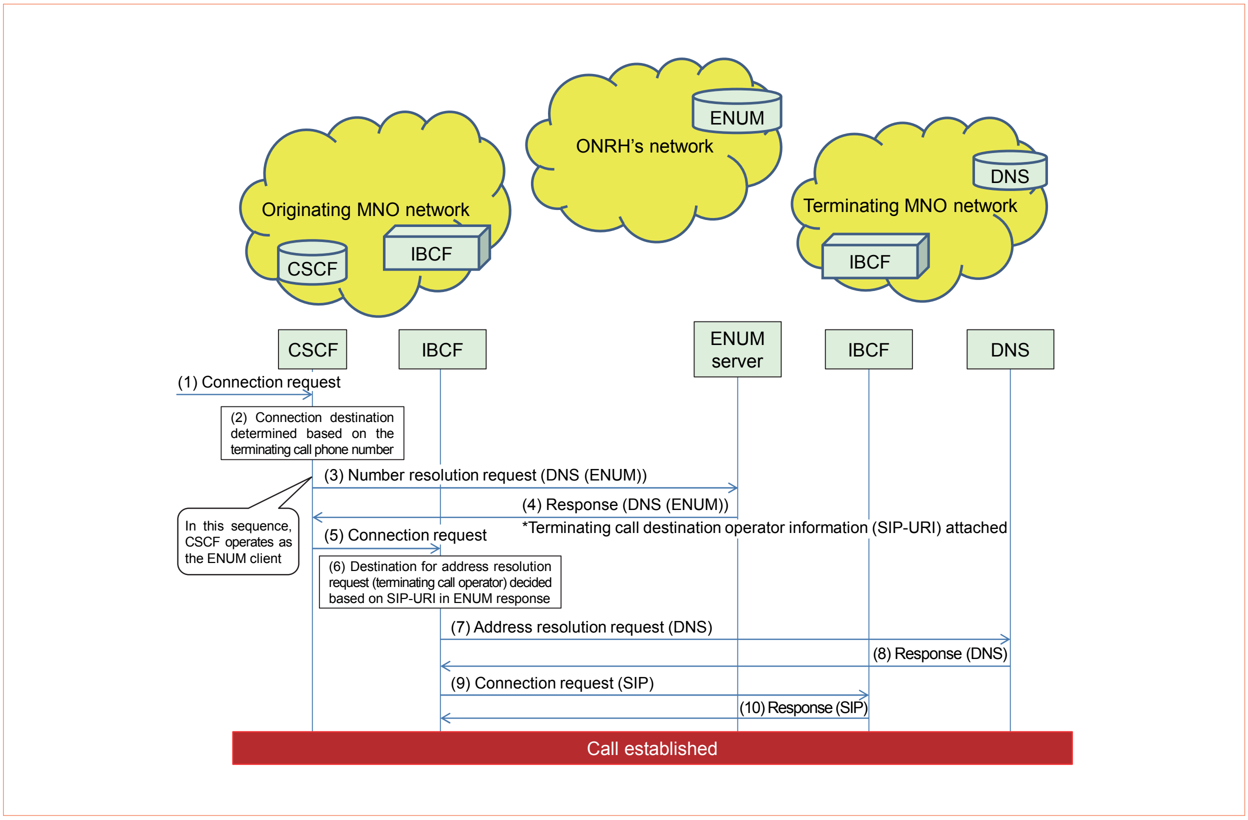


Figure 4 Connection example with MNP for VoLTE interconnection (IP connection) (Terminating call originating from NTT DOCOMO to an MNP terminal on another MNO network)

query (Fig. 4 (3)). The ENUM server responds to the query from the ENUM client with the terminating call destination MNO and SIP signal connection destination, called SIP Uniform Resource Identifier (SIP URI)*12 (Fig. 4 (4)). After that, CSCF selects the appropriate Interconnection Border Control Function (IBCF)*13 from the SIP URI, and makes a connection request (Fig. 4 (5)). The IBCF that receives the connection request determines the connection destination MNO from the SIP URI, and makes a connection request to the IP address for the appropriate destination via that determined in (3) below (Fig. 4 (6) to (9)). Because NTT DOCOMO already has a database containing MNP information, NTT DOCOMO built an ENUM server using that database.

3) Connection Destination Determination

As discussed above, it's possible for the originating MNO to determine the destination MNO from the SIP URI. However, to actually start communications using SIP transactions, the IP address of the IBCF, which is the destination linked with the SIP URI, is required.

There is another method that entails exchange of the terminating IBCF IP address linked to the SIP URI information between MNOs in advance. However, this method requires the originating side to have full control of IBCF selection, even though the terminating MNO may have preferences in traffic control upon network/node problems. Therefore, for VoLTE interconnection, NTT DOCOMO applied a system using the Domain Name System (DNS)*14

---

*12 SIP URI: The SIP addressing scheme used when making a telephone call via SIP protocol.

*13 IBCF: Network entity that acts as a gateway for external network, and performs topology hiding to conceal internal network architecture by screening SIP messages.

*14 DNS: A system that resolves IP addresses by hostname on IP networks.

mechanism in which the terminating side dynamically responds with the IBCF IP address to the originating MNO. This enables the terminating side to control which IBCF is to be selected. This procedure operates in compliance with GSM Association (GSMA)*15 IR.67 [3] (Fig. 4 (7), (8)).

This method enables the terminating side to isolate its own IBCF nodes upon malfunctions without the risk of the originating MNO accessing the failed IBCF.

This article has described typical operations and functions involved in applying VoLTE interconnection.

Nippon Telegraph and Telephone Corporation (NTT) announced "PSTN migration - the general outlook -" [4] in November 2010, then in November 2015, the company announced "The future of "fixed line" telephones" [5]. Not limited to connections between the MNOs introduced in this article, it's assumed that there will also be migration from conventional STM-POI connections to IP-POI connections for interconnection with Fixed Network Operators (FNO). Going forward, we will continue our studies described in this article to enable smooth migration with interconnectivity with the aforementioned FNOs.

REFERENCES

[1] NTT DOCOMO: "Interconnection information, technical requirements collection," (In Japanese).
https://www.nttdocomo.co.jp/corporate/disclosure/interconnection/requirement/

[2] TTC standard JJ-90.31: "Carrier ENUM interconnectivity common interface," Aug. 2015. (In Japanese).
http://www.ttc.or.jp/jp/document_list/pdf/j/STD/JJ-90.31v1.pdf

[3] GSMA IR.67 V8.0: "DNS/ENUM Guidelines for Service Providers & GRX/IPX Providers," Nov. 2012.
https://www.gsma.com/newsroom/wp-content/uploads/2012/11/IR.67-v8.0.pdf

[4] Nippon Telegraph and Telephone East Corporation, Nippon Telegraph and Telephone West Corporation: "PSTN migration - the general outlook -," Nov. 2010 (In Japanese).
https://www.ntteast.co.jp/release/detail/pdf/20101102_01_01.pdf

[5] Nippon Telegraph and Telephone Corporation: "The future of "fixed line" telephones," Nov. 2015 (In Japanese).
http://www.ntt.co.jp/news2015/1511jwbw/xddh151106d_01.html

---

*15 GSMA: An organization active in the development of the entire mobile telecommunications industry, and which consists of MNOs and mobile telecommunications industry-related members from around the world.

NTT DOCOMO
Technical Journal Vol.20 No.3

Editorship and Publication
NTT DOCOMO Technical Journal is a quarterly journal edited by NTT DOCOMO, INC. and published by The Telecommunications Association.

Editorial Correspondence
NTT DOCOMO Technical Journal Editorial Office
R&D Strategy Department
NTT DOCOMO, INC.
Sanno Park Tower
2-11-1, Nagata-cho, Chiyoda-ku, Tokyo 100-6150, Japan
e-mail: dtj@nttdocomo.com

Copyright
© 2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.