Current and Future R&D at NTT DOCOMO for the 5G Era

Current State and Progress in Each Area

Core Network for Social Infrastructure in 5G Era

Core Network Development Department Hiroyuki Oto
Research Laboratories Wataru Takita

In the 5G era, services will continue to advance and diversify and new value will be created through collaboration among various industries and businesses, so increasing demand for networking as social infrastructure is expected. This article describes the principal technologies required in the core network to realize this, based on 5G use cases, and also introduces initiatives for more advanced operations, required for efficient operation of this increasingly complex network.

1. Introduction

5th Generation mobile communications systems (5G) will meet various needs beyond those addressed with 4G, which focused on mobile phones and smartphones. It will evolve in various ways, not limited to increased speed and capacity. For example, requirements also include low latency, increased reliability, and the ability to connect to large numbers of devices.

To realize a network as social infrastructure that satisfies these requirements, advances in the core network[*1], in addition to advances in wireless communications, are needed, and various technologies have been specified for 5G. This article describes network development as social infrastructure for the 5G era, together with related principal technologies.

*1 Core network: The part of the network comprising equipment for switching, for subscriber information management and other functions. Mobile terminals communicate with the core network via a radio access network.

2. Network Deployment for the 5G Era

In the 5G era, all kinds of objects, such as automobiles, homes, and wearable devices, will connect to wireless networks, and the Internet of Things (IoT), which automatically and intelligently gathers information and manages and controls these "things," is expected to spread. On the other hand, as IoT develops, various services composed of large numbers of IoT devices, such as smart meters*2 and environmental sensors, will emerge. To support them, communications requirements such as high speed and capacity, low cost, and low latency will also be needed. As such, NTT DOCOMO's vision of the 5G world is to provide an advanced, flexible network to which all of these things can connect, and that can be used by services confidently and without stress. We are cooperating with various industries to develop such a social infrastructure network. Below, we introduce some concrete 5G use cases and describe technologies that will be needed to support them.

3. 5G Use Cases

The main usage scenarios for 5G discussed at the 3GPP are 1) high speed and capacity (enhanced Mobile Broadband (eMBB)), 2) high reliability and low latency (Ultra-Reliable and Low Latency Communications (URLLC)), and 3) Machine-type communications*3 with many simultaneous connections (massive Machine Type Communications (mMTC)). These are summarized in the specifications.

1) High Speed and Capacity: eMBB Use Case

Video On Demand (VOD) viewing on smartphones and tablets was a major use case for 4G, and accounts for a large share of the traffic.

As the amount of high-resolution video content, such as 4K/8K broadcasts and other rich content increases in the future, demand is also expected to increase for high-capacity communications on mobile networks.

2) High Reliability, Low Latency: URLLC Use Case

Advanced driver support and self-driving vehicles are expected to become more common in the future and they are basically expected to operate autonomously, using the vehicle's own sensors and processing system, and not utilizing external networks. However, there are use cases for using mobile networks, such as to get information beyond the scope of onboard sensors (nearby blind spots, information beyond sensor range, wide-area information), or for inter-vehicle communication, and these need to occur reliably and with low latency. Reliable and low latency communication is also needed when the mobile network is used in smart factories, such as for remote control of industrial robots, to ensure control and safety.

3) Many Simultaneous Connections: mMTC Use Case

Agricultural sensors, smart meters and plant equipment only produce small amounts of data, but use cases with many sensor devices connected simultaneously are envisioned. In such cases, a mechanism to manage the large numbers of terminals, and a specialized architecture to transmit small amounts of data efficiently (such as sending user data over control signals) could be needed. For large numbers of devices, reduced cost and power consumption for each device also becomes a requirement.

---

*2 Smart meter: A device that enables real-time measurement and visualization of electricity usage.

*3 Machine-type communication: A collective term for 3GPP machine communication with no intervening communication operations performed by humans.

4. Technologies for Creating New Value in the 5G Era

4.1 Initiatives to Advance Current Networks for the 5G Era

1) Network Function Virtualization (NFV)*4

As smartphones have spread rapidly in recent years, data traffic has increased, but even under these conditions, a network that users can connect to at any time must be built at reasonable cost. NTT DOCOMO has implemented NFV by applying virtualization technology, which is used widely in IT, to our communication carrier network, and we are expanding the scope of equipment being virtualized.

The virtualization technology used for NFV removes restrictions due to a particular physical configuration by logically integrating and partitioning hardware resources (CPU, memory, HDD) and treating them as resource pools*5, so various communication software can coexist and operate on general purpose hardware.

This technology reduces costs by enabling use of general purpose hardware rather than dedicated hardware, and reduces the amount of equipment by utilizing hardware resources efficiently. When the network is stressed by sudden increases in traffic, such as during a disaster, the capacity of network equipment can be scaled*6 automatically to maintain connectivity, and if new services need to be implemented quickly, it can be done through instantiation*7 of communications software on general purpose hardware.

2) DEdicated CORe Network (DECOR)*8

Recently, IoT devices have become inexpensive and are spreading rapidly and the number and types of devices connecting to LTE are increasing steadily. If separate networks could be created to accommodate the resulting traffic having differing characteristics and priorities, the networks could be controlled more flexibly and appropriately than previously, with more efficient localization and accommodation design. As such, NTT DOCOMO is currently introducing technology called DECOR to partition the core network and accommodate terminal traffic separately according to its characteristics.

At the 3GPP, there have been earlier schemes to partition the network itself, but due to implementation of terminal functionality, it could not be applied to widespread existing terminals. With DECOR, the network partitions terminals based on a terminal identifier*9, so the networks they connect to can be separated without having an impact on the terminals themselves. In this way, networks can be designed according to the characteristics and priorities of the traffic from the terminals they accommodate. This should help ensure reliability and reduce costs.

4.2 5G Network Technical Elements

This section describes 5G core network technologies for implementing the use cases mentioned above (eMBB, URLLC, and mMTC).

1) High Speed and Capacity: Technologies for Implementing eMBB

(a) Initial method for providing 5G

Methods to increase speed and capacity in the core network include extending the existing Evolved Packet Core (EPC)*10, and introducing a newly specified 5G architecture. NTT DOCOMO will be providing the

*4 NFV: A technology that uses virtualization technologies to implement processing for communications functionality in software running on general-purpose hardware.

*5 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.

*6 Scaling: The optimization of processing power by increasing or decreasing Virtual Machines (VMs) that configure communications software whenever processing power is insufficient or excessive according to hardware and VM load conditions.

former for its initial 5G deployment (**Figure 1**). This is because by using EPC, high speed and capacity can be implemented while maintaining stable quality in areas where LTE/LTE-Advanced has already been deployed. Thus, 5G can be introduced earlier and with more stability than if a new 5G network was introduced.

(b) Content delivery technology for full operation of 5G

For full operation of 5G, supporting distribution of richer content will require increased capacity in the wireless component, but it will also require technology that can distribute the rich content efficiently in the core network. To realize this use case, we will apply a Content Delivery Network (CDN)*11 on a Mobile Edge Computing (MEC) server, which places a gateway*12 relatively close to the wireless access point and performs service processing at a location close to the terminal. This can increase practical communication speed by reducing both traffic and latency on the wired segments.

For 4G and earlier, no standard method had been established to perform MEC and Internet access at the same time according to traffic characteristics, while also maintaining terminal mobility in this way. However, new methods have been added to allow MEC and mobility at the same time with 5G, such as UpLink CLassifiers (UL CL), which identify specific traffic and offload it near the terminal (Local Breakout*13), and Session and Service Continuity (SSC), which is described below.

UL CL is a technology that identifies packets sent by a terminal to a specific IP

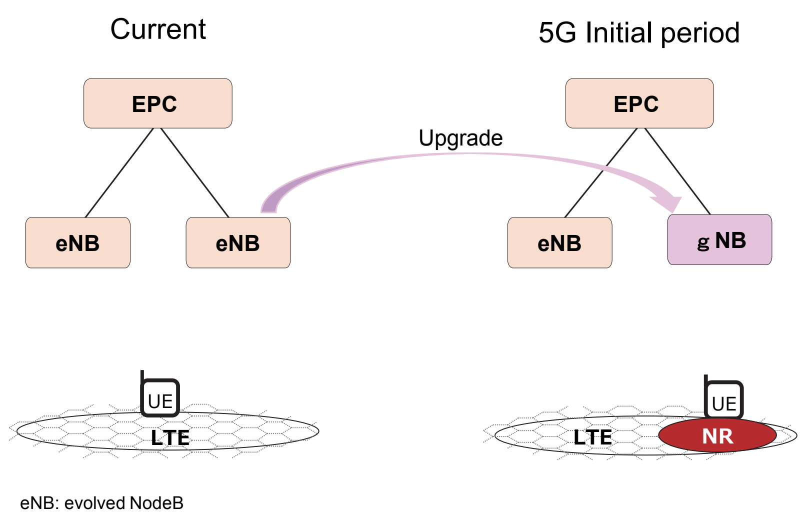

Figure 1 Accommodating 5G access in the initial period

---

*7 **Instantiation:** A procedure for launching a communications application in a cloud environment.

*8 **DECOR:** A dedicated core network that is partitioned according to terminal-type identifiers that express the type of each terminal, and gathers terminals with the same terminal identifier together and accommodates them.

*9 **Terminal identifier:** An identifier in the subscriber information that indicates the type of terminal or device and its use.

*10 **EPC:** A core network accommodating a radio access network such as LTE. It consists of MME, S-GW, P-GW, and PCRF and provides functions such as authentication, mobility control, bearer management, and QoS control.

address, and routes them differently than other packets. It is generally used to connect to an MEC server as shown in **Figure 2**.

2) High Reliability, Low Latency: Technologies for Implementing URLLC

Besides MEC, which reduces delay that is dependent on the distance between the terminal and the server, we are making changes for implementing services that require high reliability and low latency, such as improving the functionality that prevents interruption during hand over*14, which is another cause of delay, and adding extensions to the routing control functionality to enable communication over the shortest route between terminals.

These functions are called SSC. SSC is a new technology that, as with MEC connections, the network and server to which a terminal connects could change as the terminal moves from one area to another. SSC has the following three modes (**Figure 3** (a)-(c)).

(a) SSC mode 1 is similar to 4G and earlier, connecting to a single place in the network regardless of area. It is used for ordinary Internet connections.

(b) With SSC mode 2, if a terminal is connected to an MEC server associated with a given area, A, and then moves to a new area, B, it then changes its connection to the MEC server associated with area B.

(c) With SSC mode 3, if a terminal is connected to an MEC server associated with a given area, A, and then moves to a new area, B, it connects to the MEC server for area B while maintaining the connection with the former server from area A. With SSC mode 3, a new "Make Before Break" option has been added, to avoid interruptions when reconnections are made during hand over. The connection to the old area A network is kept while connecting to the new area B network and communication with the area A network is only terminated after the new

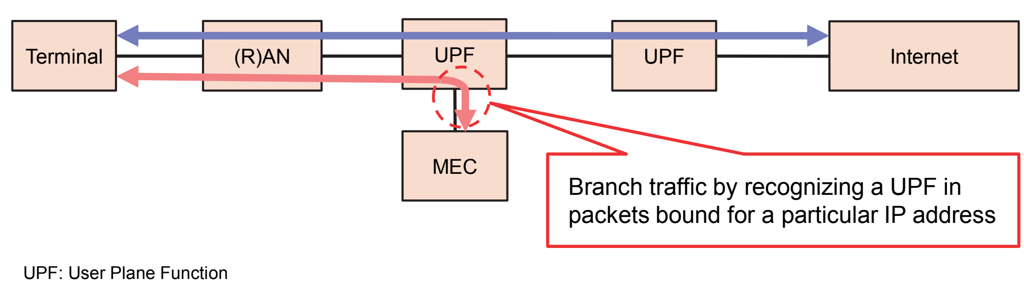

Figure 2 UL CL

*11 CDN: A network solution optimized for fast and stable distribution of large files such as images and video.

*12 Gateway: An intermediate device that has functions such as protocol conversion and data transfer to allow communication between devices.

*13 Local Breakout: A method that allows specific traffic to flow from each base directly to the Internet, in order to avoid stress on lines connecting the central base with other bases.

*14 Hand over: The technique of switching from one base station to another without interrupting communication when a terminal moves between base stations.

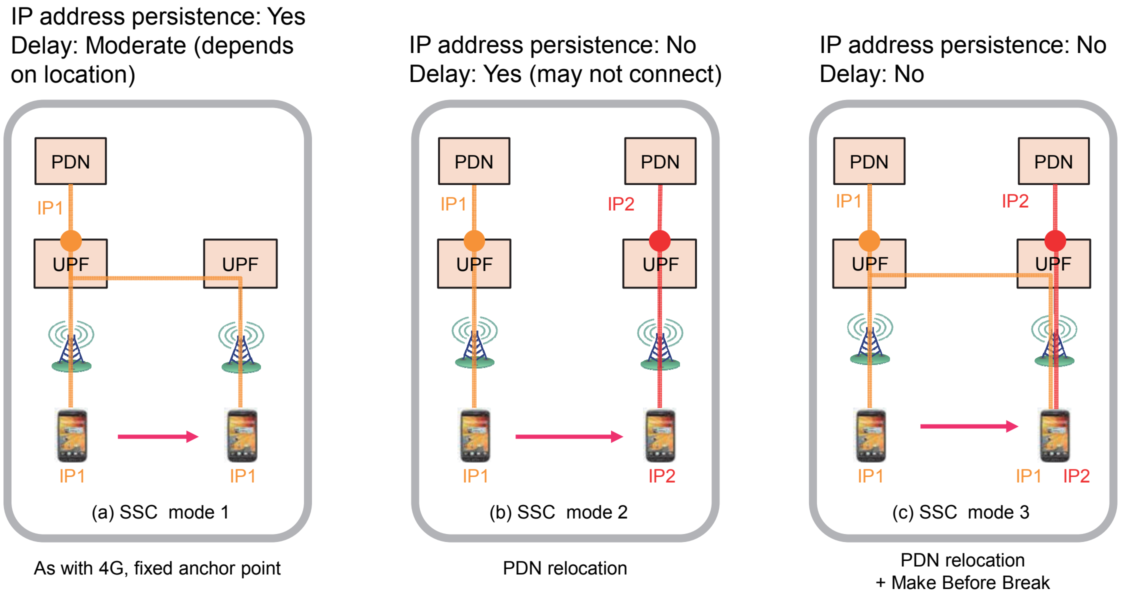

Figure 3 SSC

connection is completed.

3) Many Simultaneous Connections: Technologies for Implementing mMTC

One way to manage many devices efficiently is to not perform registration of location for devices that are installed at fixed locations, such as sensors. Also, to reduce the cost of sensor devices, security and other functionality can be offloaded to the core network, and to avoid concentration of very large numbers of sessions*15 in one place, the data collected from sensors can be processed with MEC.

Another technology being studied to improve utilization of the network is to layer data onto control signals when the amount of data is small.

4.3 Accommodating Diverse Networks Efficiently

Performance requirements are more strict than ever before for the diverse use cases of 5G and aspects such as latency, bandwidth, and number of terminals, but each of these use cases can be realized, even if all of these capabilities cannot be maximized at the same time.

It is also possible to use different Radio Access Networks (RAN)*16 depending on the use case, and it is expected that future networks in the 5G era will accommodate multiple radio technologies including 4G wireless access, 5G wireless access and Wi-Fi®*17, according to the characteristics of the service. By also using virtualization technology, logical networks specialized to the requirements

*15 Session: A series of communications exchanged between a client and a server or between servers.

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

*17 Wi-Fi®: A registered trade mark of the Wi-Fi Alliance.

of each service, whether high efficiency or low latency, can be built as network slices*18, while sharing physical network devices such as servers and transport devices (**Figure 4**). A technology called Service Based Architecture (SBA) is being introduced to deploy network functions on network slices, adding flexibility for building networks and reducing the time required to provide services.

4.4 SBA

In the 5G era, a platform able to provide various types of services quickly is needed. To realize various requirements quickly, the core network must do more than simply transfer data, adding functionality suited to the characteristics of the application. It must also be possible to add such functionality quickly, to meet the development cycles of terminal and server applications. However, current core networks have a monolithic architecture*19, and due to the tight coupling between devices a long development cycle is needed to ensure overall consistency in the system, even when adding small-scale functionality to the core network.

In contrast, for terminal applications themselves, and the server applications linked to them, as with smartphone applications, a wide range of services are being developed, rapidly adding functionality to both of them. SBA has been standardized so that functionality can also be added rapidly to the core network in this way.

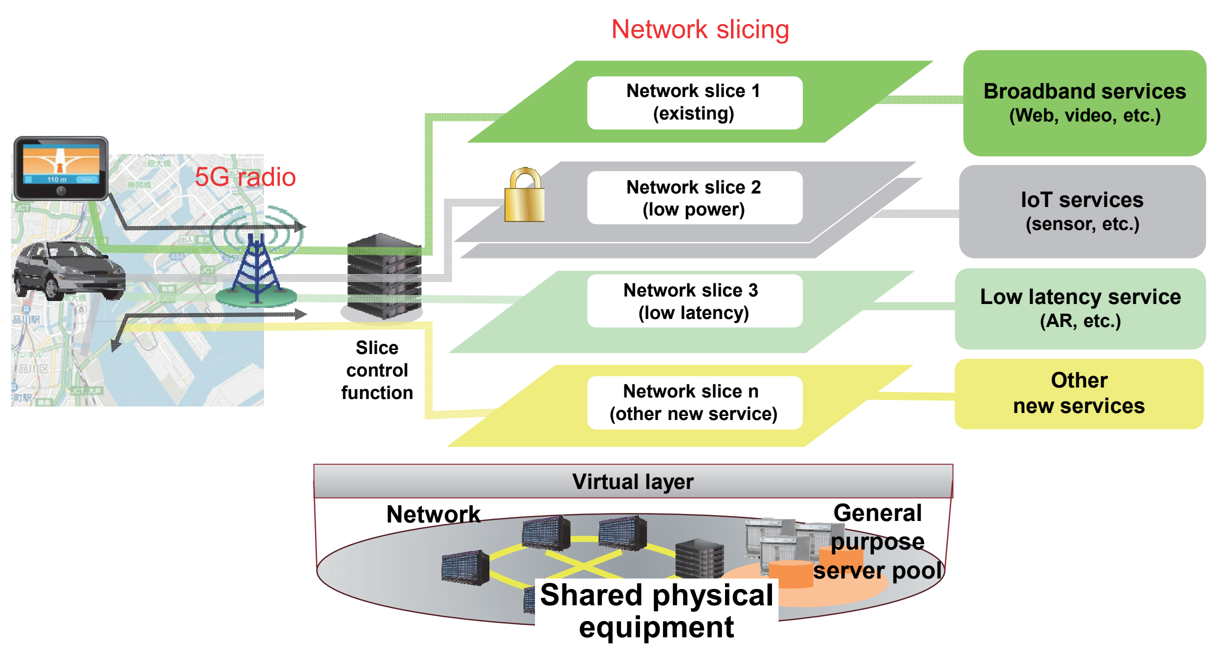

Figure 4　Network slicing

*18 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.

*19 Monolithic architecture: An architecture in which the functional components are tightly coupled and the overall system operates as a single module.

SBA is an application of micro-service architecture in the 3GPP standards. A micro-service architecture separates the application realized by a service into functions called micro-services. By loosely coupling the interfaces of each of the micro-services, the effects of adding or changing functionality are localized, reducing the time required to implement and test services and accelerating the process (**Figure 5**).

SBA is a principal technology for realizing a cloud-native[*20] core network, and the goal of a cloud-native core network combined with network slicing is to reduce operating costs and fulfill the various requirements at the same time.

4.5 Network Operation Expansion

With the implementation of 5G in the future, mobile networks are expected to offer more diverse services, which will lead to more complexity in networks than ever before. Virtualization technology will allow for hardware configurations similar to or even simpler than before, but logical configurations providing services, such as network slices, will increase in complexity, using multiple virtualized logical resources. We can expect an increasing number of issues with manual maintenance and operation of these configurations. As such, NTT DOCOMO is working to make conventional operations systems more economical and efficient using automation and Artificial Intelligence

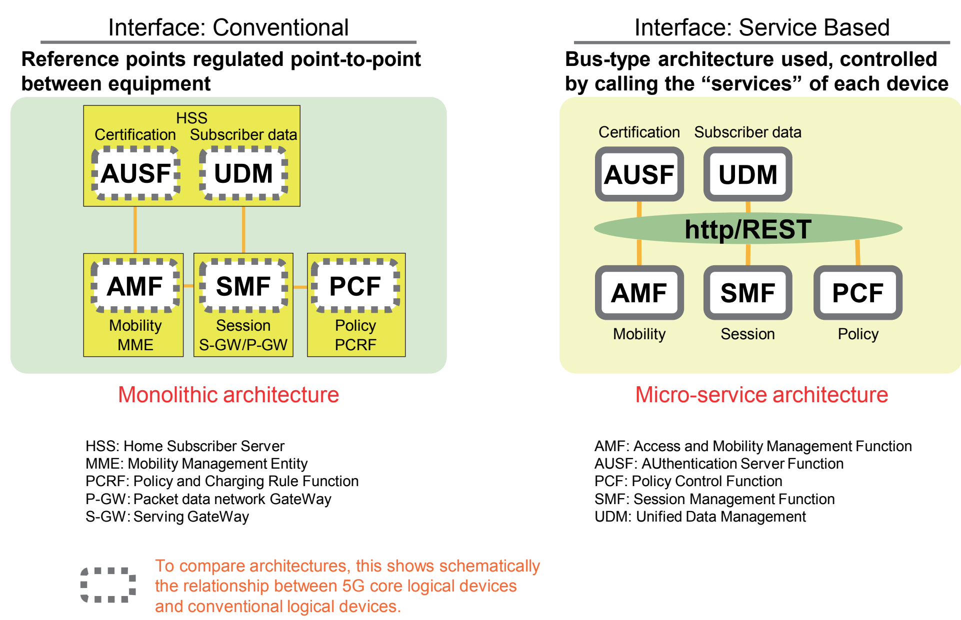

Figure 5 SBA architecture

*20 Cloud native: Refers to systems and services designed for configuration and operation on the cloud, rather than on premises.

(AI), and to generally advance operations work.

In automation efforts, we have introduced a Fulfillment Operation Support System (F-OSS)*21, which automates the increasingly complex and intricate manual work being done by people. Network virtualization has enabled structures that separate hardware and software, but by automating this with an OSS, OPerating EXpenses (OPEX)*22 can be reduced and significant improvements to reliability and availability can be made.

Remarkable progress has also been made in AI in recent years, which is expected to improve work efficiencies in various fields. Increasingly, there are also opportunities to improve efficiencies in mobile-network maintenance and operations work (hereinafter referred to as "operations work") by using AI technologies.

Automation is a key aspect of advancements in operations work, including automation that uses systems and tools to replace work conventionally done by people, and also automation that performs work that could not be done previously by people, such as predicting equipment faults or drops in service quality from current conditions, or preventing degradation in service quality due to equipment fault before it happens. There is hope that using AI technology for the latter will improve efficiencies.

Operations work can be broadly categorized into monitoring, analysis, and taking measures. Advances in operations utilizing AI are described below.

(1) Monitoring

AI technologies such as deep learning*23 and other forms of machine learning*24 can be used together with the large amounts of data that can be collected from the network, such as warnings, the state of equipment, and traffic conditions, to automatically detect symptoms of equipment fault or to predict factors that will affect services in ways not previously possible manually. Detected symptoms can then be visualized in terms of the state of quality by service or area using existing systems.

(2) Analysis

When degradation in service quality is predicted, the suspicious part within the logical network configuration, made more complex due to network slicing, is inferred automatically. Systematic isolation of faults into scenarios, which was previously done manually, is done automatically, and measures to be taken are proposed to operators dynamically, such as changing network resources, according to the details of the fault.

(3) Taking measures

Depending on the measures to be taken, automatically determined in the analysis, measures to recover automatically and maintain service continuity are taken at the discretion of the operator for phenomena that can be dealt with remotely. For phenomena that require on-site attention, such as replacing equipment due to a server fault, information such as a description of the required work, the parts, management of the work scheduling, and efficient travel routes are presented to support the on-site worker.

Implementing the above enables a change from performing maintenance conventionally, after the

*21 F-OSS: A system that manages data for building networks centrally and automates design and construction work to make network virtualization work more efficient.

*22 OPEX: Amount of money expended for maintaining and operating facilities.

*23 Deep Learning: Machine learning using a neural network with a many-layer structure.

*24 Machine learning: A framework that enables a computer to learn useful judgment standards through statistical processing from sample data.

equipment has issued warnings, to performing it based on predictions of the future. This promises to improve service continuity, preventing any effects on services to customers before they occur.

5. Conclusion

This article has described principal technologies needed in the core network and developments in network operations based on 5G use cases. NTT DOCOMO is providing connectivity for all kinds of "things" through 5G technology, and we will continue to contribute to deploying our network as social infrastructure.