VoLTE | IMS Platform | ALL IP

Technology Reports

VoLTE for Enhancing Voice Services

NTT DOCOMO has ranked VoLTE as a necessary function from the viewpoint of enhancing voice services for users and reducing facility costs, and in anticipation of a migration from the conventional circuit switching method to VoLTE, it has come to convert the voice network to an IP-based network and introduce an IMS platform. Meanwhile, on the global stage, it was agreed at GSMA in February 2010 that diverse methods would be unified through VoLTE using IMS as a method for providing voice services and SMS over LTE. This article presents background to VoLTE development at NTT DOCOMO, provides an overview of VoLTE architecture using IMS, and describes its functional features and basic control methods.

Core Network Development Department — *Kazuhito Tokunaga*, *Shinichi Minamimoto*, *Mana Kaneko*, *Yoichi Kamiya*†
Radio Access Network Development Department — *Rika Kito*
Communication Device Development Department — *Hideyuki Sakuramoto*
Core Network Division, DOCOMO Technology, Inc. — *Issei Imamura*

1. Introduction

Voice over LTE (VoLTE) is a technology for providing voice services over LTE. The functions needed to provide such services have been prescribed at the GSM Association (GSMA)*1, an industry association [1]. The provision of VoLTE services is already underway by some carriers in Korea and elsewhere (as of March 2014) and it appears that 200-plus carriers around the world that have adopted LTE are about to introduce VoLTE.

Conventional LTE-capable terminals have been connecting to the 3G network to provide voice services, and as a result, the quality of voice services has been held to the 3G level. This scheme has also meant that the time needed for originating and terminating calls by switching radio systems is relatively long and that packet communications during voice services are limited to 3G transmission speeds. At NTT DOCOMO, we decided to introduce VoLTE as a means of resolving these issues and commenced the provision of VoLTE services at the end of June 2014. Since a terminal can become camped within an LTE area even while a call is in progress, the provision of VoLTE means that the user can enjoy quick call originating and call terminating at that time while also being able to use high-speed multi-access and receive Area Mail*2 while a call is in progress. Furthermore, by providing new high-quality voice and video calls in combination with VoLTE, users can expect an enhanced sense of usability in a variety of ways. At the same time, by improving spectrum efficiency*3 for the voice-service portion, an efficient frequency spectrum can be diverted to data traffic to provide users with smooth pack-

† Currently Core Network Division, DOCOMO Technology, Inc.

*1 **GSMA:** An association that supports and manages activities of the mobile industry such as formulating roaming rules. World's largest industry association in the mobile communications domain. About 800 mobile operators in 219 regions and more than 200 IPX operators, terminal, equipment and software vendors have joined the Association.

*2 **Area Mail:** A service for instantly broadcasting alerts such as earthquake early warnings from the JMA.

et communications. NTT DOCOMO's VoLTE implements a series of functions prescribed by GSMA and 3GPP such as Quality of Service (QoS)*4 control on the IP network and a handover*5 function for call connection from LTE to 3G.

In this article, we present the background leading to VoLTE development at NTT DOCOMO, provide an overview of VoLTE architecture, and describe the functional features and basic control methods of VoLTE.

2. Background to VoLTE Development

NTT DOCOMO has come to convert the core network*6 to an IP-based network in a stepwise manner to provide a network having even higher transmission speeds and levels of quality, to deal with ever-changing volumes of traffic in voice and data communications, and to reduce the cost of facility investment (**Figure 1**).

In voice communications, the approach taken to convert voice services executed in the Circuit Switched (CS) domain*7 to IP (hereinafter referred to as "CS-IP*8 network") was to use the IP Multimedia Subsystem (IMS)*9 [2]. A key feature of IMS is its ability to provide common voice services (such as voice mail and automatic voice translation) in any type of access network, that is, 3G, LTE, wireless LAN, etc. NTT DOCOMO proceeded to introduce IMS even in networks such as LTE not having a CS domain with an eye to providing voice communications.

In data communications, high-speed

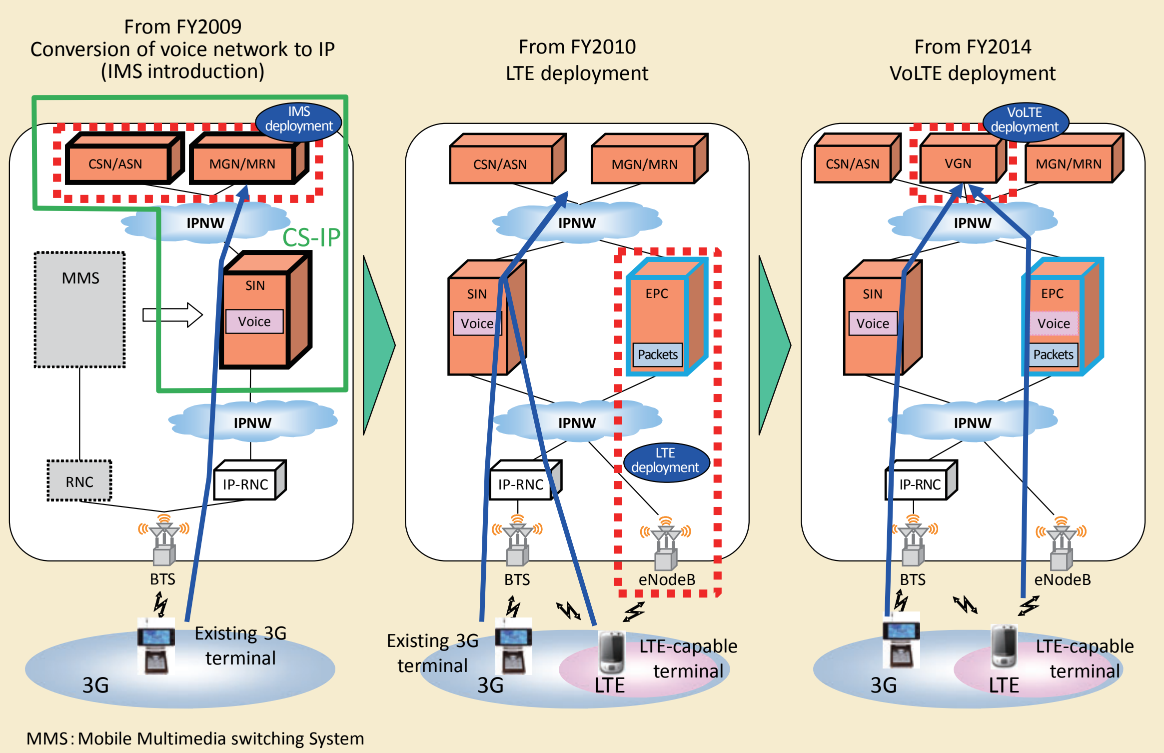

Figure 1 Transition of core network (blue arrow indicates flow of voice communications)

*3 **Spectrum efficiency:** The number of data bits that can be transmitted per unit time over a particular frequency spectrum.

*4 **QoS:** A level of quality on the network that can be set for each service. The amount of delay or packet loss is controlled by controlling the bandwidth that the service can use.

*5 **Handover:** A technology for switching base stations without interrupting a call in progress when a terminal straddles two base stations while moving.

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

*7 **CS domain:** A functional block of a network offering circuit switched services.

*8 **CS-IP:** An IP-based core network for controlling and transmitting voice traffic using the IMS standardized by 3GPP.

mobile broadband and multimedia services have come to be provided on Evolved Packet Core (EPC)*10 that accommodates LTE.

On the other hand, a variety of methods have been proposed elsewhere in the world to provide LTE voice services and SMS such as Circuit Switched FallBack (CSFB)*11 [3], IMS, and Voice over LTE via Generic Access (VoLGA)*12 [4]. However, considering that different methods should be unified for the sake of interconnectivity and international roaming, an agreement was reached at GSMA in February 2010 that VoLTE using IMS would be adopted [5]. Following this, a document called IR.92 [1] was standardized at GSMA as specifications for a set of minimum required functions to achieve an IMS-based VoLTE, and using this document as a basis to work from, worldwide discussions were held on the details of a method for achieving VoLTE.

In this way, the migration envisioned by NTT DOCOMO came to match world trends, and for this reason and the fact that method unification under VoLTE would mean no issues in interconnectivity, we began VoLTE development in earnest.

3. Overview of NTT DOCOMO VoLTE Architecture

The network architecture of VoLTE developed by NTT DOCOMO and that of standard VoLTE are compared in **Figure 2**. In either case, VoLTE architecture can be broadly divided into an LTE network configured by eNodeB and EPC and the core network configured by IMS.

As for the core network, NTT DOCOMO had already introduced IMS architecture when introducing the CS-IP network. One advantage of using IMS is that the same voice services can be provided even for different IP Connectivity Access Network (IP-CAN)*13 types by connecting to the same IMS. For this reason, an architecture that could make use of NTT DOCOMO's existing IMS network has been adopted to provide VoLTE services. This means that, other than introducing a VoLTE Gateway Node (VGN) as new equipment, existing equipment could be used. The VGN constitutes gateway*14 equipment for connecting a VoLTE terminal to the IMS network—it corresponds to the standard Proxy Call Session Control Function (P-CSCF)*15 and IMS-Access GateWay (IMS-AGW)*16. In response to Session Initiation Protocol (SIP)*17 Control Plane (C-Plane)*18 signals received from a terminal, this equipment performs processing to absorb any differences between the terminal and the NTT DOCOMO network and functions as security equipment to prevent any abnormal signals from being drawn into IMS equipment. It also operates as User Plane (U-Plane)*19 control equipment and it has a mechanism for controlling U-plane signals when making a transition from LTE to 3G during a call as described later. Other equipment in the core network has the same functions as those in the existing NTT DOCOMO network. Specifically, the Call Session control Node (CSN) performs session*20 control, the Application Serving Node (ASN) controls voice services, the Media Gateway Node (MGN) controls connections with other networks, and the Media Resource Node (MRN) controls the delivery of announcements. With regard to the MGN, while differences exist between call-connection signal processing in existing 3G voice services and call-connection signal processing in VoLTE, there is no change in connection control to other networks, so this equipment also has the role of absorbing these differences.

We have so far described NTT DOCOMO's IMS network, but with regard to the IP-CAN LTE network, it can be introduced without having to add new equipment since NTT DOCOMO has already begun the provision of LTE services. However, in contrast to data services, there is a need here in the LTE network to provide a voice bearer*21 for voice control in LTE packet communications and to perform bandwidth control to ensure voice quality.

---

*9 **IMS:** A communication system standardized by 3GPP for implementing multimedia services. IMS used IP and the SIP protocol used for Internet telephony to integrate the communication services of the fixed telephone and mobile communication networks.

*10 **EPC:** A core network consisting of MME, S-GW, P-GW, and PCRF and providing functions such as authentication, mobility control, bearer management, and QoS control.

*11 **CSFB:** A procedure for switching to a radio access system having a CS domain, when a terminal sends/receives a circuit switched communication such as voice while camped on an LTE network.

*12 **VoLGA:** A technology to offer virtual circuit switched voice services by accommodating the LTE radio in circuit switched networks.

*13 **IP-CAN:** A network that provides an IMS terminal with a means of accessing an IMS network operator (home network or destination network while roaming).

*14 **Gateway:** A node having functions such as protocol conversion and data relaying.

*15 **P-CSCF:** A function deployed at the connection point with EPC and at the connection point with S-CSCF and I-CSCF. It has the roles of linking with EPC to initiate QoS control and of relaying SIP signals between the mobile terminal and S-CSCF and I-CSCF.

4. VoLTE Functional Features

4.1 Ensuring Voice Quality in VoLTE

1) Overview

The introduction of VoLTE means the introduction of voice packets into the LTE network. Here, if voice packets were to be given the same priority as data communication packets in sp-mode, they could also be discarded at times of congestion*22 in the LTE network thereby degrading voice quality. There is therefore a need in VoLTE for bit rate control (QoS control) to guarantee a minimum bit rate. There is also a need within a VoLTE service to change priority control in the system, the level of quality in the IP network, etc. for each type of signal, each type of call, and each type of service as in C-Plane and U-Plane signals, voice and video calls, and general and emergency calls. For these reasons, IMS that controls VoLTE services will coordinate with EPC that provides a bearer (VoLTE signal transmission path), and the information so exchanged will be reflected in each system's control operations. At the same time, bearer control will be performed whenever an appropriate QoS setting is made in the IP network.

2) Bearer Control

Guidelines for QoS control for each service type are stipulated in 3GPP stand-

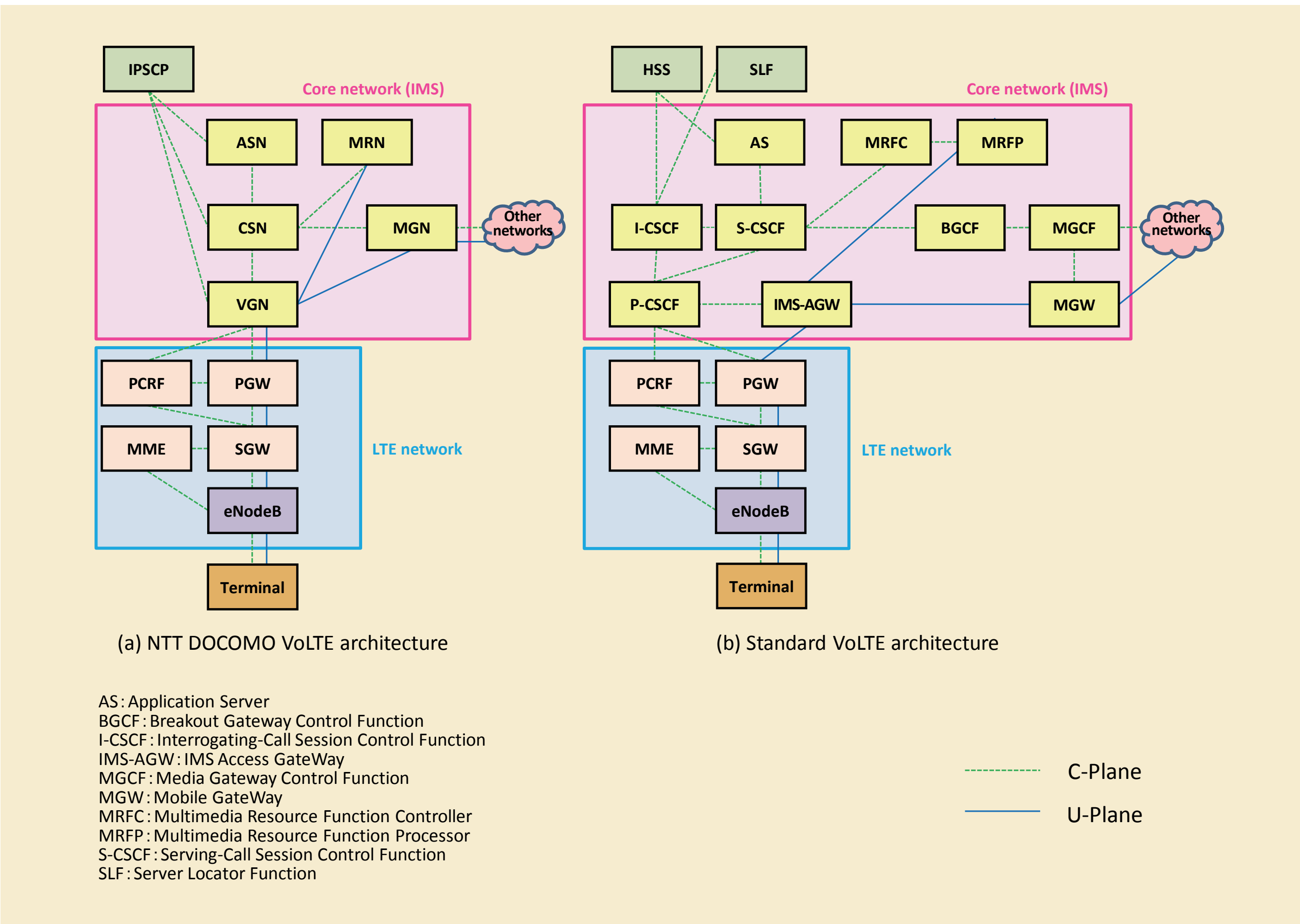

Figure 2 NTT DOCOMO VoLTE architecture and standard VoLTE architecture

---

*16 **IMS-AGW:** A gateway deployed at the same position as P-CSCF with the role of relaying voice signals in contrast to P-CSCF's role of relaying SIP signals. It links with P-CSCF to control voice signals and provide security functions.

*17 **SIP:** A call control protocol defined by the Internet Engineering Task Force (IETF) and used for IP telephony with VoIP, etc.

*18 **C-Plane:** Transmission path for control signals such as establishing and disconnecting communications.

*19 **U-Plane:** A transmission path for user data in contrast to the C-Plane, which is a transmission path for control signals. From the viewpoint of IMS, SIP signals are carried on the C-Plane and RTP/RTCP on the U-Plane, but from the viewpoint of EPC, both are carried on the U-Plane.

*20 **Session:** An interactive exchange of information between a server and a client or between two servers. Here, "session" is used as a series of exchanges in a call control sequence.

*21 **Bearer:** A logical user-data packet transmission path established along P-GW, S-GW, eNodeB, and UE.

*22 **Congestion:** Impediments to communications services due to communications requests being concentrated in a short period of time and exceeding the processing capabilities of the service control server.

ard specifications [6]. For each QoS Class Identifier (QCI)*23, these guidelines specify its bit rate control status (target or non-target of guaranteed bit rate), priority, allowed delay and packet loss rate (**Table 1**). In conventional bearer control prior to the provision of VoLTE, a best-effort type of bearer was provided as a Default Bearer*24 and no 2nd or later bearer (Dedicated Bearer*25) was operated on the same Access Point Name (APN)*26. However, to provide VoLTE, a maximum of three bearers need to be constructed on an IMS specific APN: a bearer for C-Plane use based on SIP protocol (Default Bearer), a U-Plane bearer for voice use (Dedicated Bearer) and a U-Plane bearer for video use (Dedicated Bearer) based on Real-time Transport Protocol (RTP)*27 and RTP Control Protocol (RTCP)*28. Additionally, the bearer for C-Plane use based on SIP protocol has a Non-Guaranteed Bit Rate (Non-GBR) while the bearer for U-Plane use based on RTP/RTCP has a GBR.

4.2 Video Calls

1) Overview

NTT DOCOMO provides video calls conforming to IR.94 [7] based on IR.92. In the 3G videophone service, video calling is achieved through a mechanism that transmits both voice and video streams within a bit rate limited to 64 kbps, but in video calls on VoLTE, picture quality has been improved by transmitting voice and video separately, the latter with a minimum bit rate of 384 kbps. Moreover, in addition to using the series of functions specified in IR.94, NTT DOCOMO also uses a variable rate method (rate adaptation) as described below to further improve picture quality.

2) Rate Adaptation

To provide video calls with stable image quality even in environments with poor radio quality, NTT DOCOMO has adopted a variable rate method (rate adaptation) instead of a fixed rate method. As the name implies, a fixed rate method performs communications at the same rate, but under difficult radio quality conditions, communications at that rate can result in picture distortion owing to packet loss. In rate adaptation, however, communications are performed while changing the bit rate in accordance with current conditions, which enables the continuous provision of a relatively stable picture in visual terms. There are two methods for implementing

Table 1 QCI and bearer type

QCI	Bit Rate Guarantee Type	Priority	Allowed Delay	Packet Loss Rate	Commercial Services
1	GBR	2	100ms	10^{-2}	Call (voice)
2		4	150ms	10^{-3}	Call (video) [Live Streaming]
3		3	50ms	10^{-3}	Real-time games
4		5	300ms	10^{-6}	Non-calling (video) [Buffered Streaming]
5	Non-GBR	1	100ms	10^{-6}	IMS Signalling
6		6	300ms	10^{-6}	Video communications [Buffered Streaming] TCP-based packets (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video)
7		7	100ms	10^{-3}	Voice communications Video communications [Live Streaming] Interactive games
8		8	300ms	10^{-6}	Video communications [Buffered Streaming] TCP-based packets (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video)
9		9			

*23 **QCI:** QoS classes specified by the 3GPP for bearers in the LTE/EPC. There are values of 1 to 9, and the smaller the value is the more the bandwidth is guaranteed and the smaller the transmission delay.

*24 **Default Bearer:** The 1st bearer established under an APN; used in IMS-APN to send and receive SIP.

*25 **Dedicated Bearer:** The 2nd or later bearer established under an APN, used in IMS-APN to send and receive RTP/RTCP.

*26 **APN:** The name of a connection point; the name of a network connection point prepared by a corporate user as a connection destination.

*27 **RTP:** A transmission protocol for streaming video and voice media. As a UDP type of protocol, no countermeasures to packet loss are performed. Generally used in combination with a communications status report provided by RTCP.

*28 **RTCP:** A protocol for controlling RTP data streams used in combination with RTP. Exchanging information such as bandwidth and delay time over RTCP enables quality management to be performed.

rate adaptation: Audio Video Profile (AVP)*29 [8] and Audio-Visual Profile with Feedback (AVPF) [9]. NTT DOCOMO has adopted AVPF because of its higher accuracy and its recommendation in standard specifications. The following describes the simple mechanism behind rate adaptation (**Figure 3**).

To begin with, the originating side calculates the maximum bit rate at which it should be able to receive data via RTP on the downlink based on the radio quality of the interval between its terminal and the network and transmits that information to the terminating-side terminal by RTCP (Fig. 3 (1)). Similarly, the terminating side calculates the maximum bit rate at which it should be able to transmit data on the uplink based on the radio quality of the interval between its terminal and the network. It then compares that value with the value received from the originating terminal and sets the smaller of the two as the bit rate to be used in the direction from the terminating terminal to the originating terminal (Fig. 3 (2)). Next, the terminating terminal calculates the maximum bit rate at which it should be able to receive data on the downlink and transmits that information to the originating terminal (Fig. 3 (3)). The originating terminal then calculates the maximum bit rate at which it should be able to transmit data on the uplink and sets the smaller of the two values to the bit rate to be used in the direction from the originating terminal to the terminating terminal (Fig. 3 (4)). This exchange of information can be repeated periodically to provide a method of communication in which the bit rate is constantly being updated depending on current radio conditions.

4.3 Improvement of Spectrum Efficiency on the Radio Link

1) Overview

In LTE, data transmission/reception between User Equipment (UE) and eNodeB is achieved through the use of a Physical Downlink Control CHannel (PDCCH)*30 and Physical Downlink Shared CHannel (PDSCH)*31/Physical Uplink Shared CHannel (PUSCH)*32. The frequency resources and transmission format used on PDSCH/PUSCH are specified on PDCCH (**Figure 4**).

VoLTE uses the Adaptive MultiRate (AMR)*33 codec in which voice data of approximately 300 bits is generated every 20 ms as a RTP packet (**Figure 5**). In other words, the transmitting and re-

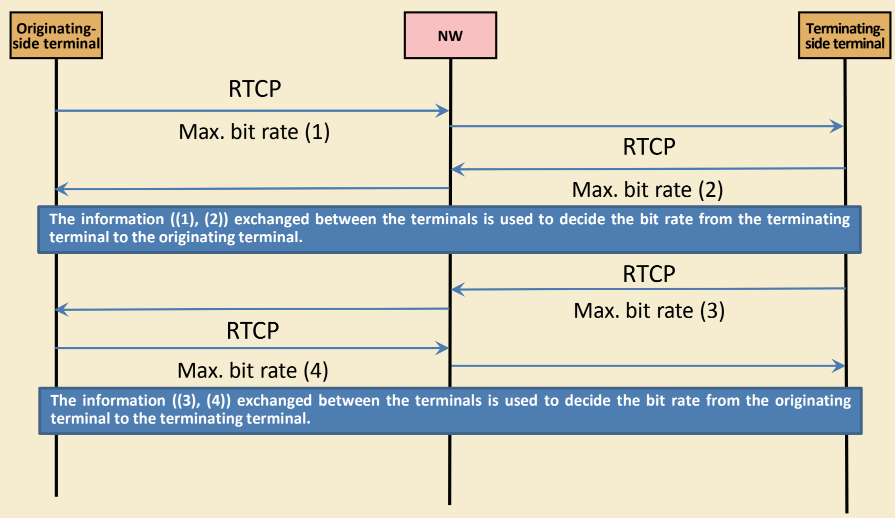

Figure 3 Overview of rate adaptation (AVPF)

*29 **AVP:** A variable communication rate method used in RTP and the most basic of those methods.

*30 **PDCCH:** A channel used to indicate data transmission format and timing in LTE.

*31 **PDSCH:** A shared channel used in downlink data transmission in LTE.

*32 **PUSCH:** A shared channel used in uplink data transmission in LTE.

*33 **AMR:** A speech coding method prescribed as essential in 3GPP standard specifications.

ceiving of voice data is carried out every 20 ms.

Since PDCCH is essential for specifying frequency resources and transmission format, PDCCH usage can be excessive for data generated in short periods as in the case of voice data. This can limit the number of voice users that can be accommodated by the system. Two techniques for improving PDCCH usage efficiency here are delay packing and Transmission Time Interval (TTI) bundling.

Furthermore, with respect to the amount of voice data transmitted, VoLTE deals with the large protocol headers in RTP, User Datagram Protocol (UDP)*34, and IP by using the RObust Header Compression (ROHC)*35 technique [10].

2) PDCCH Usage Efficiency Improvement and Header Compression

(a) Delay packing

Instead of transmitting voice data every time it's generated, the delay packing technique buffers a voice packet and combines it with the subsequent voice packet when the latter is generated. It then generates a Media Access Control Protocol Data Unit (MAC PDU)*36, a unit of data transmission on the MAC layer, to transmit these voice packets together (**Figure 6**). This scheme decreases the number of times that voice data needs to be transmitted/received and reduces PDCCH resource usage.

However, since the amount of data that can be transmitted or received per unit time depends on the radio environment, the number of packets to be transmitted in batch must be adaptively controlled according to

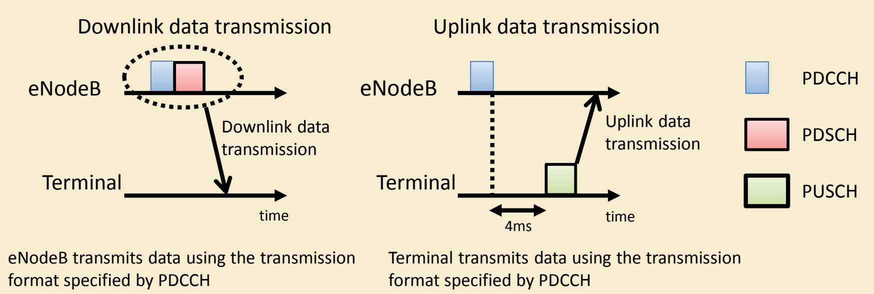

Figure 4 Downlink/uplink data transmission in LTE

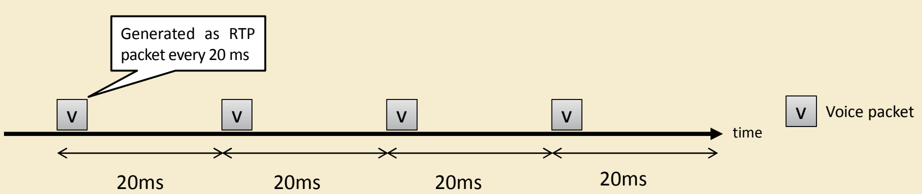

Figure 5 Voice-packet spacing in AMR codec

*34 **UDP:** A protocol on the transport layer used for real-time communications since it features light processing without delivery confirmation, congestion control, etc. and few issues even if some data should go missing.

*35 **ROHC:** A technique for compressing the RTP/IP/UDP header.

*36 **MAC PDU:** A protocol data unit on the MAC layer. PDU expresses protocol data including header and payload.

the radio environment of each mobile terminal. Specifically, delay packing will be applied to UE with a good radio environment and will not be applied to UE with a poor radio environment as may be found at a cell edge.

(b) TTI bundling

One method that could be considered for securing a fixed level of throughput in a poor radio environment such as at a cell edge would be to widen the frequency bandwidth of the transmission signal. However, there is an upper limit to UE transmission power, so widening the frequency bandwidth on the uplink would have the effect of reducing transmission power per unit bandwidth thereby preventing required receive quality from being satisfied.

This dilemma could be solved by segmenting the voice packet over several subframes to reduce the number of transmission bits per subframe, but the number of PDCCH transmissions here would be equal to the number of segments (**Figure 7**). In addition, each segment of data would require a MAC/Radio Link Control (RLC)*37/ Packet Data Convergence Protocol (PDCP)*38 header thereby degrading the radio-channel usage efficiency. Furthermore, in Semi-Persistent Scheduling (SPS)*39, only one resource can be specified per period, so segmentation cannot be used.

TTI bundling is prescribed in the LTE standard as a technology for resolving these issues [11]. This technique enables one voice packet to be transmitted over four consecutive subframes. At this time, only one PDCCH will be transmitted to instruct the UE on this data transmission thereby reducing PDCCH resource usage (**Figure 8**).

(c) ROHC

An IP packet of voice data in VoLTE has the structure shown in **Figure 9**. As shown, the RTP/UDP/IP header takes up approximately 65% of the packet. Transmitting packets in this form invites a drop in radio resource*40 usage efficiency. To deal with this issue, VoLTE applies header compression (ROHC) control. ROHC control, while not being specifically prescribed in a 3GPP standard [12], is described in a RFC document [10],

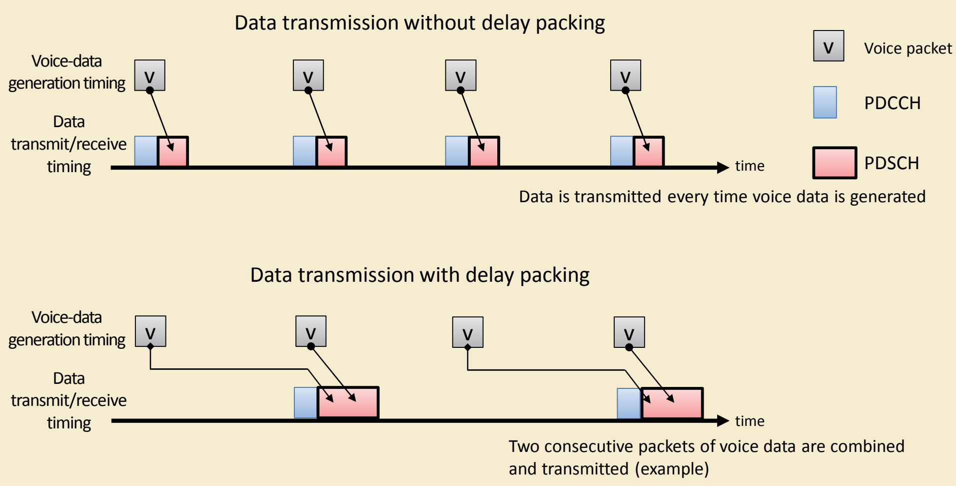

Figure 6 Comparison of data transmission with and without delay packing

*37 **RLC:** One of the sublayers in Layer 2 of the radio interface in LTE that provides protocols for retransmission control, duplicate detection, reordering and the like.

*38 **PDCP:** One of the sublayers in Layer 2 of the radio interface in LTE that provides protocols for ciphering, integrity protection, header compression and the like.

*39 **SPS:** A scheduling technique that performs semi-persistent resource allocation in LTE.

*40 **Radio resource:** Unit of time or frequency range allocated to each user for communication purposes.

which is referenced in the former. It performs compression/decompression using variation patterns in individual information elements making up the RTP/UDP/IP headers. For example, given an information element like an IP address whose value normally goes unchanged while communications are in progress, it can be omitted once the decompression side has received it successfully thereby reducing the amount of header information. Moreover, for information elements like RTP sequence numbers and RTP timestamps whose values change packet by packet, the fact that such values change according to a fixed rule can be exploited to send only a minimum amount of data and reduce the total amount of header information. Decompression uses variation patterns in each header information element in the same way to restore compressed headers. ROHC control can reduce RTP/UDP/IP header size from approximately 60 bytes to a minimum of 3 bytes.

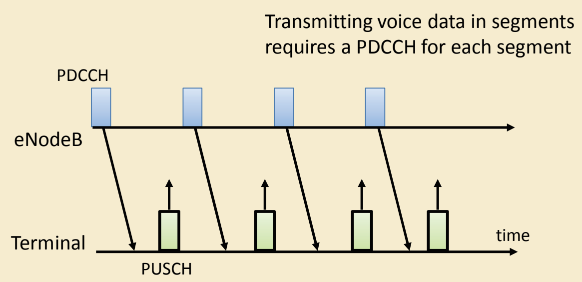

Figure 7 Segmentation

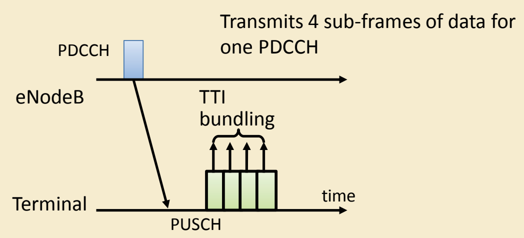

Figure 8 TTI bundling

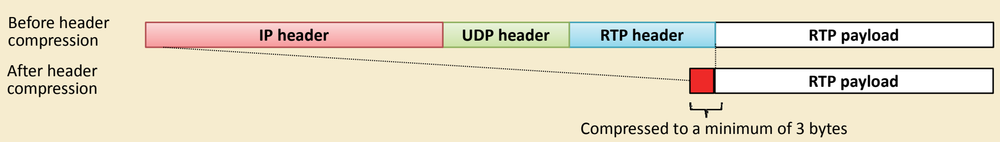

Figure 9 Header compression by ROHC

5. VoLTE Basic Control Methods

In this section, we describe the VoLTE location registration control process and the basic call originating/terminating control process while referring to the network architecture shown in **Figure 10**.

5.1 Location Registration Control

An overview of location registration processing by a VoLTE-capable terminal is shown in **Figure 11**. Location registration of a VoLTE-capable terminal consists of Attach*41 processing for the LTE layer and Registration*42 processing for the IMS layer.

1) Attach Processing

For a terminal equipped with the W-CDMA function, turning its power ON initiates Attach processing for the 3G and LTE network (Fig. 11 (1)).

Since the VoLTE-capable terminal does not specify the connection APN at this time, the Mobility Management Entity (MME)*43 connects to the default APN (IMS) received from the Home Subscriber Service (HSS)*44 (Fig. 11 (2) (3)). Next, the EPC Serving and PDN GateWay (ESPGW)*45 issues a connection request to the Policy and Charging Rule control Function (PCRF)*46, which initiates P-CSCF discovery processing to select the VGN address and pass it to the terminal by Protocol Configuration Options (PCO)*47 (Fig. 11 (4) (5)). We point out here that P-CSCF discovery processing is executed by ESPGW in standard specifications, but in the NTT DOCOMO network, this process is executed by PCRF. This is because P-CSCF discovery processing by PCRF is more flexible to equipment failures and congestion in the network. Furthermore, to attach to the 3G network, the originating terminal sends a location registration request signal to the IP Service Control Point (IPSCP)*48

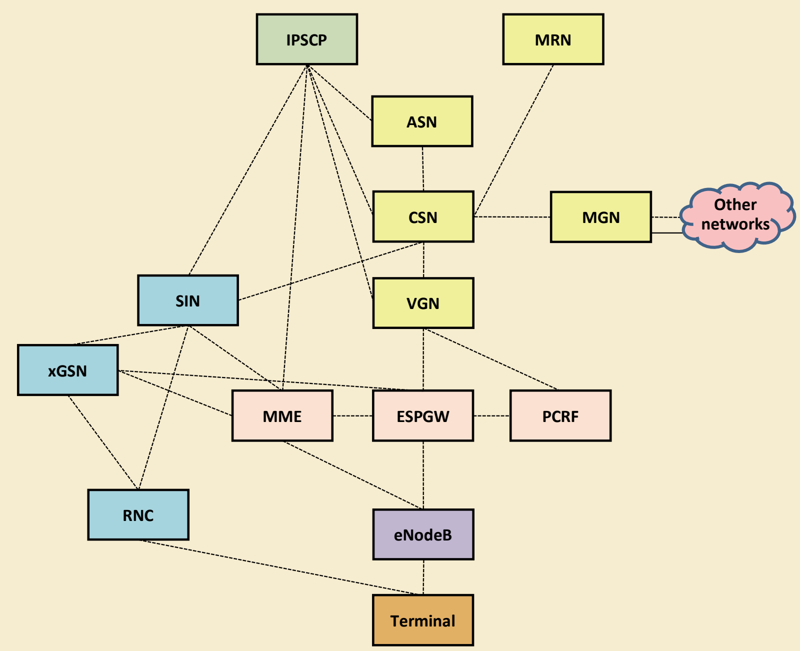

Figure 10 VoLTE network architecture at time of location registration control and basic originating/terminating control

*41 **Attach:** A procedure to register a terminal on the network when, for example, its power is switched on.

*42 **Registration:** In IMS, mobile terminals register current location data in HSS with SIP.

*43 **MME:** A logical node accommodating a base station (eNodeB) and providing mobility management and other functions.

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

*45 **ESPGW:** Equipment having the capabilities of S-GW and P-GW.

*46 **PCRF:** A logical node for controlling user data QoS and charging.

*47 **PCO:** Transmits the various protocol options in the bearer-established signal.

*48 **IPSCP:** A node with functions to manage subscriber service information (contract and configuration information) and for service control.

via SIN/CSN/ASN (Fig. 11 (6)).

Now that the IMS Default Bearer has been established, the MME sends a response signal indicating Attach completion to the terminal (Fig. 11 (7)).

2) Registration Processing

Once Attach processing has been completed, signal transmission from the terminal to the IMS layer becomes possible and Registration processing by SIP protocol between the terminal and IMS can be performed. Here, the terminal sends a SIP_Registration request via VGN to CSN (Fig. 11 (8)), which provides an Authentication and Key Agreement (AKA)*49 authentication function (TS24.229, TS33.203) in coordination with IPSCP and sends an authentication request to the terminal (Fig. 11 (9) (10)). The terminal, in turn, performs a computation based on the vector set in the authentication request signal and sends the result of that computation back to CSN (Fig. 11 (11) (12)). The CSN now compares that result with its own computation result (Fig. 11 (13)), and if they match, considers that authentication has been completed and continues with location registration processing by sending the SIP_Registration request to ASN. On receiving this SIP_Registration request, the ASN issues a location registration request to IPSCP. Each of the above nodes stores a SIP_Registration response together with profile*50 information (Fig. 11 (14)).

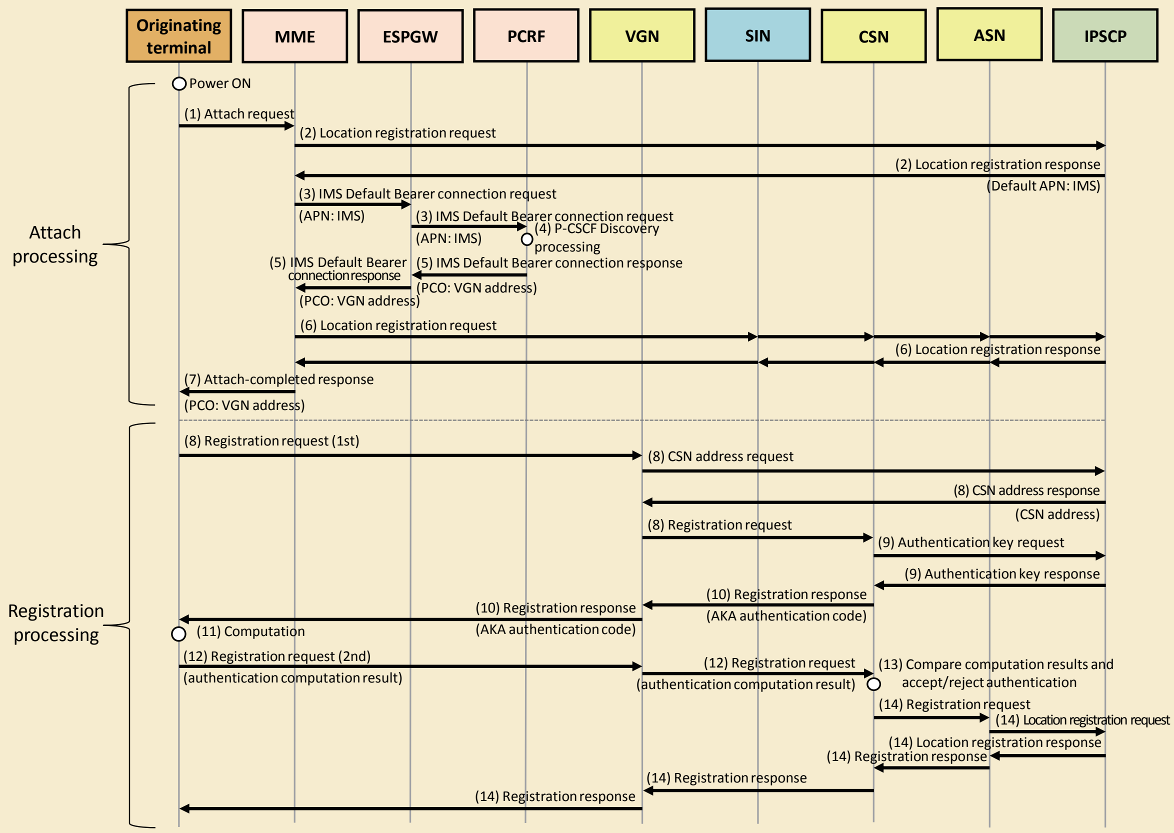

Figure 11 Overview of location registration processing

*49 **AKA:** An authentication process combined with key agreement. The USIM computes a cipher key and integrity key based on parameters supplied by the network and checks the validity of those parameters.

*50 **Profile:** Basic information covering contract, user setting, camp-on, etc.

5.2 Basic Communication Control

1) Overview of Originating and Terminating Call Control

Basic originating/terminating call processing following IMS Registration is shown in **Figures 12 – 14**. This processing is divided into (a) originating-side control part, (b) terminating-side control part and (c) processing from session establishment to call establishment.

(a) Originating-side control part

Key features of the originating-side control part are obtaining location information from the LTE radio access network via EPC and interworking*51 between the terminal and IMS at VGN.

On receiving a call-initiation request from the originating terminal (Fig. 12 (1)), VGN performs a location-information acquisition process (Fig. 12 (2)). This process makes use of Network Provided Location Information for IMS (NetLoc)*52 as specified in the 3GPP standard [13]. In VoLTE, there is no direct connection between the radio access network and IMS so that VGN cannot obtain location information in the same way as CS-IP. Instead, it obtains location information from the cells of the LTE radio access network via EPC (Fig. 12 (2)-1, (2)-2). The originating VGN then sends a call-initiation request to the originating CSN. At this time, the originating VGN performs signal interworking the same as the call-initiation request generated by the originating Signaling Interworking Node (SIN)*53 in 3G voice-call initiation. This makes it possible to apply the same 3G processing performed by the equipment positioned after the originating VGN (Fig. 12 (3)).

Specifically, the call-initiation request processing functions of CSN and ASN and the terminating-user location query processing functions of IPSCP are used both by the VoLTE network and 3G network (Fig. 12 (4) – 12 (8)).

(b) Terminating-side control part

Key features of the terminating-side control part are initiating terminating-side location identification at the mutually terminating ASN and coordinating with IPSCP, xGSN and MME, and interworking between the terminal and IMS at VGN.

The terminating ASN that receives an incoming-call request from the terminating CSN initiates terminating-side location verification (Fig. 13 (1)). This process makes use of Terminating Access Domain Selection (T-ADS)*54 as specified in the 3GPP standard [14]. In this way, the terminating ASN determines whether the last network in which the terminating user was located in was LTE or 3G (Fig. 13 (1)) and sends a call-initiation request to that network (Fig. 13 (2) – (6)). The specific method for identifying the terminal's last location is as follows. The IPSCP sends a query to MME (LTE network) and the Serving General packet radio service Support Node (SGSN)*55 (3G network) asking for the time of the terminating terminal's last access to that equipment. The most recent location comparing the obtained data from two different nodes is taken to be the terminal's last location (Fig. 13 (1)-1, (1)-2). However, if the terminal is located in an area in which the radio base station (eNodeB) and EPC do not support VoLTE even though that area might be in the LTE domain, the incoming call will be handled not by VoLTE but by CSFB.

After identifying the location of the terminal for terminating the call, CSN and ASN apply the same processing as that performed at the time of a 3G voice-call initiation (Fig. 13 (2) – (5)). Finally, the terminating VGN deletes information used only for control purposes within the IMS network from signals directed to the terminating terminal and issues a call-initiation request to the terminal (Fig. 13 (6)).

(c) Processing from session establishment to call establishment

A key feature of session establishment is the use of a Precondition

---

*51 **Interworking:** Interaction between communications systems.

*52 **NetLoc:** A procedure used by IMS equipment for obtaining a terminal's location information (the cell in which the terminal is currently camped) from the network. Location information is obtained from the network since that passed from the terminal could have been tampered with.

*53 **SIN:** Equipment equipped with a function for accommodating a 3G radio access network and a function for converting between CC and SIP to enable 3G circuit switching to connect to IMS.

*54 **T-ADS:** A function that specifies the access network in which the terminal is currently camped.

*55 **SGSN:** A logical node in 3GPP standard specifications providing functions such as mobility management for a mobile terminal performing packet switching and packet communications.

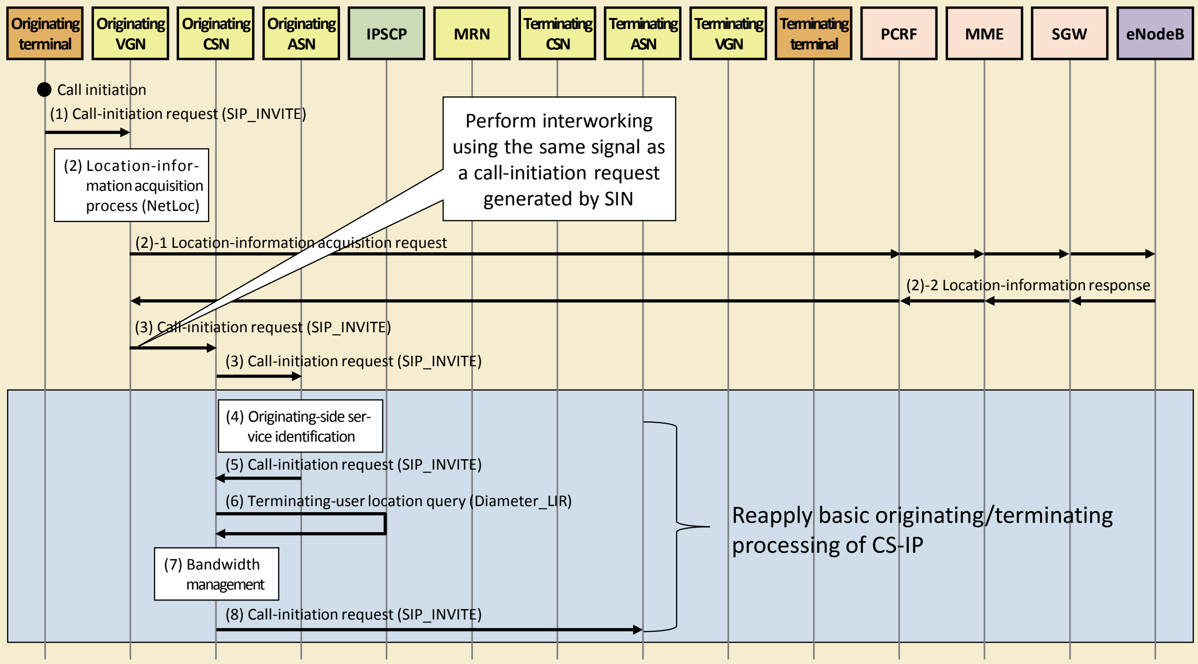

Figure 12　Overview of basic originating/terminating processing in VoLTE (originating-side control part)

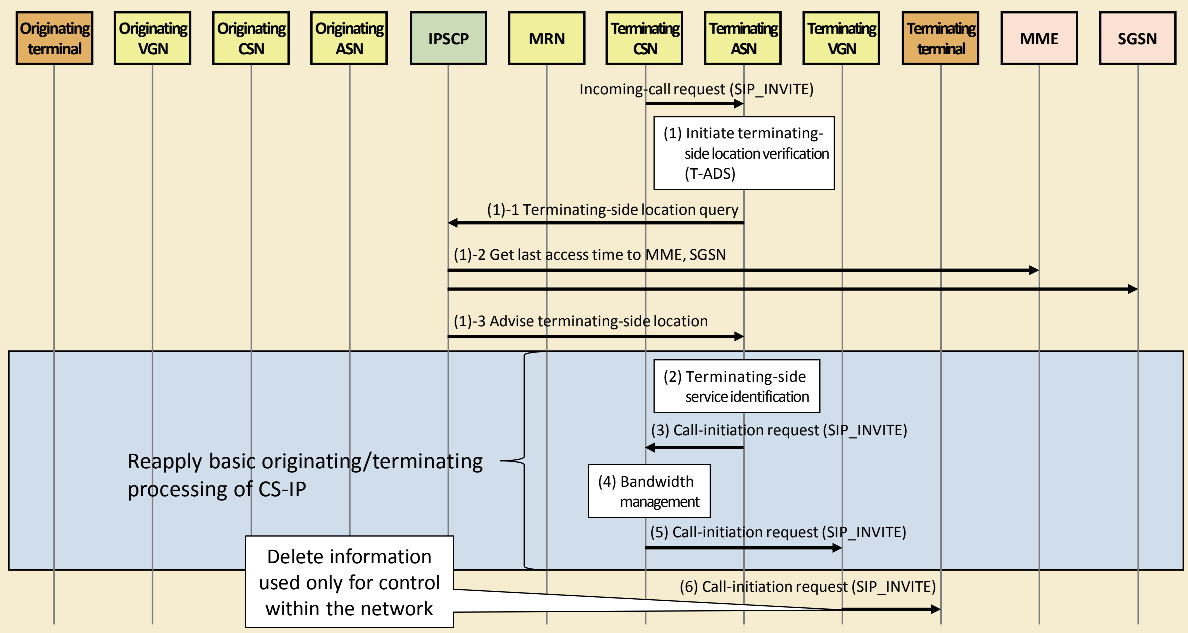

Figure 13　Overview of basic originating/terminating processing in VoLTE (terminating-side control part)

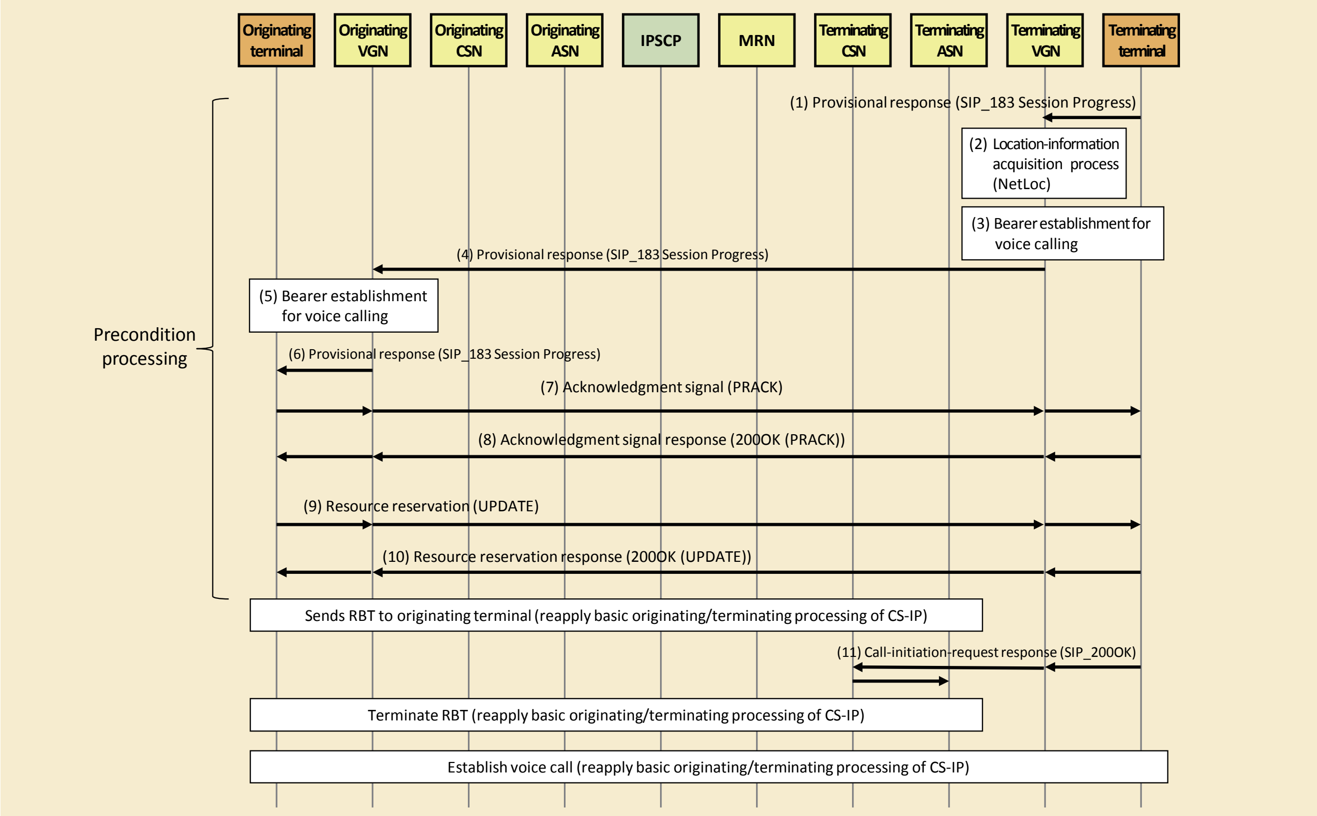

Figure 14 Overview of basic originating/terminating processing in VoLTE (session establishment and U-PLANE establishment)

[15] function in which a terminal must fulfill certain resource conditions (such as conditions associated with U-Plane establishment) before call establishment can proceed. The advantage provided by this Precondition function is that unnecessary call-control processing is eliminated since the incoming call is received only after resource acquisition has been completed.

On receiving a call-terminating request, the terminating terminal sends to the terminating VGN a provisional response that sets resource conditions in Session Description Protocol (SDP) [16], which contains session-related performance information such as the codec*56 to be used (Fig. 14 (1)). At this time, the terminating terminal simultaneously performs processing in accordance with the resource conditions included in the provisional response. Now, on receiving the provisional response, the terminating VGN gets location information of the terminating subscriber (using NetLoc) and establishes a bearer for voice calling between the terminating terminal and network (Fig. 14 (2) (3)). Following this, the terminating VGN sends the provisional response back to the originating VGN, which establishes a bearer for voice calling between the originating terminal and network (Fig. 14 (4) (5)). Next, after receiving the provisional response from the originating VGN, the originating terminal performs control processing in accordance with the specified resource conditions.

*56 **Codec:** Technology for encoding/decoding data as in speech processing.

It then sends an acknowledgment signal (PRACK) to the terminating terminal in response to the received provisional response, and the terminating terminal sends a response to that PRACK signal (Fig. 14 (6) – (8)). Resources can now be reserved among the network, originating terminal, and terminating terminal (Fig. 14 (9) (10)).

Following the above Precondition processing, CS-IP call processing is reapplied to control the sending of a Ringing Back Tone (RBT) to the user, terminate the RBT when the terminating subscriber performs an off-hook operation (Fig. 14 (11)), and establish a call between the originating and terminating terminals [2].

2) SMS Control

A VoLTE terminal located in an LTE area and successfully registered with IMS performs SMS sending and receiving by the SMS over IP [17] method using SIP protocol on the Default Bearer. One function in SMS over IP specified in standards documents is IP-Short-Message-Gateway (IP-SM-GW)*57. In the NTT DOCOMO network, this function is achieved in the ASN, which performs a similar function in 3G. The 3G network uses SMS protocol between the terminal and SIN and Mobile Application Part (MAP) protocol between SIN and ASN, but VoLTE uses SIP protocol between the terminal and ASN. VoLTE adopts a mechanism whereby differences between 3G and VoLTE are absorbed by ASN with the result that there are no differences in 3G and VoLTE processing in nodes positioned above ASN (**Figure 15** (1) – (3)). Furthermore, at the time of an SMS transfer, ASN assesses the area in which the terminating user is located

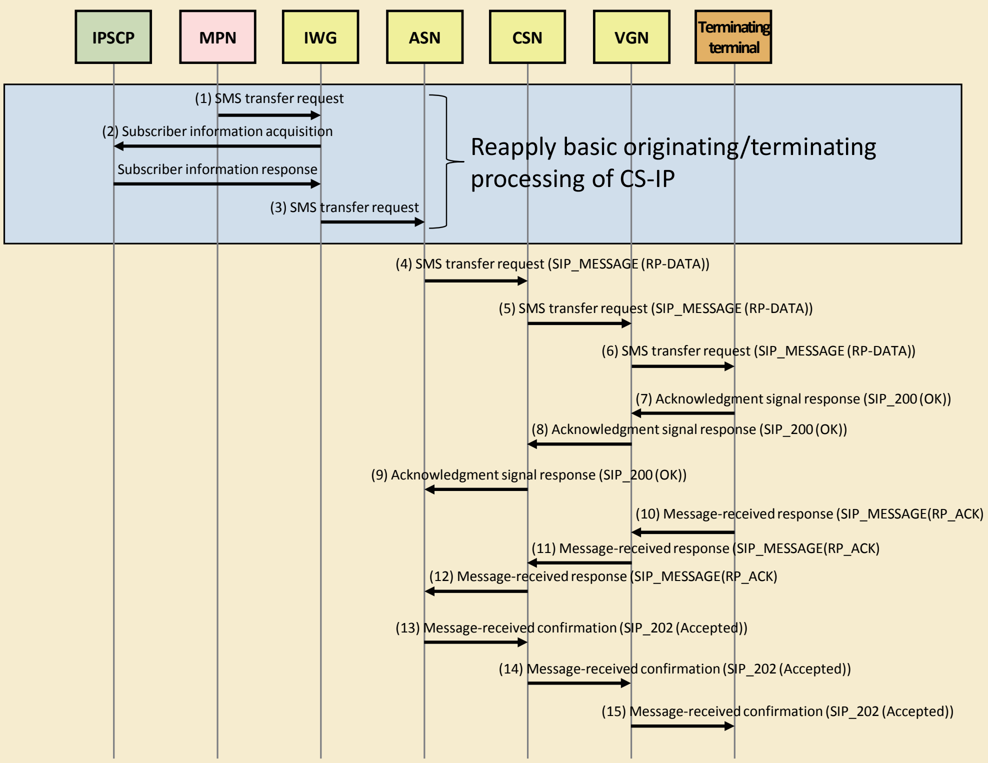

Figure 15 Overview of SMS terminating processing

*57 **IP-SM-GW:** A logical node that manages SMS transmitting/receiving in conventional circuit switching and IP.

and decides on the protocol and bit rate to use. For example, ASN would select SMS over IP for the case of a terminating user located in LTE and registered with IMS, SMS over SGs*58 for the case of a terminating user located in LTE but not yet registered with IMS, and a 3G route for the case of a terminating user located in 3G. If an SMS over IP route is selected, ASN sends CSN an SMS transfer request (SIP_MESSAGE (RP-DATA)), which is then forwarded to VGN and the terminating terminal (Fig. 15 (4) – (6)). On receiving the SMS transfer request, the terminating terminal sends an acknowledgment signal (SIP_200 (OK)) back toward ASN to indicate that the SMS-transfer-request signal itself has been received (Fig. 15 (7) – (9)). The terminal then determines the content of the SMS and sends a response (SIP_MESSAGE (RP-ACK)) back toward ASN to indicate that a valid message has been received (Fig. 15 (10) – (12)). Finally, on receiving the response (SIP_MESSAGE (RP-ACK)), ASN sends a confirmation signal (SIP_202 (Accepted)) back toward the terminal (Fig. 15 (13) – (15)).

3) SRVCC

Single Radio Voice Call Continuity (SRVCC) [18] refers to a function for handing over a VoLTE voice call to 3G. The processing for this handover is shown in **Figure 16**.

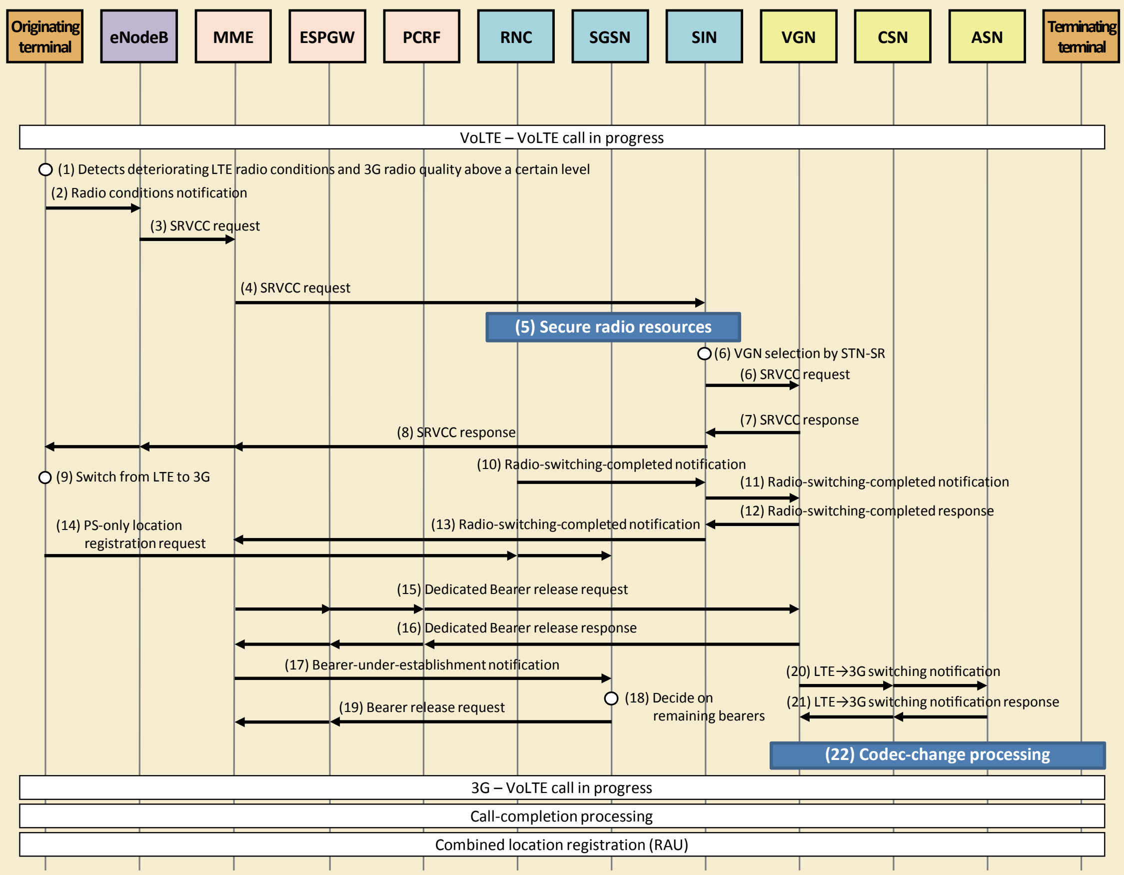

Figure 16 Overview of SRVCC processing

*58 **SMS over SGs:** SMS performed over interface SGs connecting MSC and MME. A non-VoLTE-capable LTE terminal performs SMS transmitting/receiving by SMS over SGs.

To begin with, the originating terminal notifies eNodeB that LTE radio conditions are deteriorating and 3G radio quality is above a certain level. The eNodeB then issues a SRVCC request to MME (Fig. 16 (1) – (3)). The MME, in turn, issues a SRVCC request to SIN (Fig. 16 (4)). Here, MME may select any SIN for the following reason. At the time of Attach/Tracking Area Update (TAU)*59, MME notifies the SIN of the Session Transfer Number for SRVCC (STN-SR)*60 (VGN address) received from IPSCP, so there is no need to select the same SIN selected at the time of combined location registration*61/TAU. Now, the SIN that has received the SRVCC request from MME secures 3G radio resources with the Radio Network Controller (RNC)*62 (Fig. 16 (5)) and sends a SRVCC request indicating the handover-call identifier to the VGN derived from the STN-SR (Fig. 16 (6)) in order to establish an IMS session (Fig. 16 (6)). Then, once 3G radio resources have been secured from the RNC side, a radio-switching request to 3G (SRVCC response) is sent to the terminal (Fig. 16 (7) (8)). The terminal now switches the network from LTE to 3G (Fig. 16 (9)) and executes a Packet-Switching (PS)-only location registration (Routing Area Update (RAU))*63 (Fig. 16 (14)). A PS-only location registration is performed at this time since the call is in progress, but a combined location registration will be performed after the call completes. Now, on receiving a notification from RNC that radio switching to 3G has been completed (Fig. 16 (10)), the SIN first issues a radio-switching-completed notification to VGN and then issues the same to MME (Fig. 16 (11) – (13)). At this time, MME releases the Dedicated Bearer used to transmit and receive voice signals during the VoLTE call (Fig. 16 (15) (16)) and sends SGSN a notification containing information on the bearer under establishment (Fig. 16 (17)). The SGSN now decides which bearers are to remain (Fig. 16 (18)) and releases all other bearers (Fig. 16 (19)). For example, when executing SRVCC in a state in which the terminal is connected to IMS APN and sp mode APN in LTE, the bearer for voice in SGSN is deemed unnecessary and the IMS APN is cut off since connection to the CS domain is continuous in 3G.

On receiving the request to release dedicated bearers, the VGN notifies ASN of LTE-to-3G switching (Fig. 16 (20)) and performs codec-change processing with the terminating terminal (Fig. 16 (22)).

In the case of NTT DOCOMO, this processing is necessary since the codec changes when a transition to 3G occurs. Once this change has been completed, communications between the originating and terminating terminals can begin.

4) Video Calling

(a) Overview of video-call originating and terminating

The basic connection procedure for a video call is the same as that of a voice call (**Figure 17**). The differences are that the originating terminal specifies video in its call-initiation-request signal (Fig. 17 (1)), a video bearer is established simultaneously with the establishment of a voice bearer (Fig. 17 (6) (8)), bandwidth is controlled to accommodate a video call since required bandwidth differs from that of a voice call, and control processing for recognizing a video service is performed (Fig. 17 (4)).

(b) Overview of voice-to-video call switching

Call switching between a voice call and video call in a VoLTE video call function can be performed in the same way as the 3G-videophone function. The 3G videophone is achieved using the 3GPP-standard Call Control (CC)*64-based Service Change and UDI Fallback (SCUDIF)*65 function, but a VoLTE video call is achieved using a SIP session update function. As shown in **Figure 18**, the originating terminal sends a video-call switching request to the originating VGN, which sends the same request to the terminating terminal (Fig. 18 (1)). The terminating terminal, in turn, recognizes the video-call switching

---

*59 **TAU:** A procedure for updating location registration in LTE.

*60 **STN-SR:** An element for selecting a VGN at the time of SRVCC.

*61 **Combined location registration:** Performing location registration for both the circuit switching network and packet switching network. For packet switching network, 3G or LTE is possible.

*62 **RNC:** Equipment in the FOMA network for radio access and mobility control defined in 3GPP specifications.

*63 **Location registration (RAU):** A logical node in the 3GPP standard managing the mobility of mobile terminals that perform packet switching and packet communications.

*64 **CC:** A protocol for controlling originating/terminating calls in circuit switching.

*65 **SCUDIF:** A system defined by 3GPP R5 for changing bearers while a circuit-switched call is in progress.

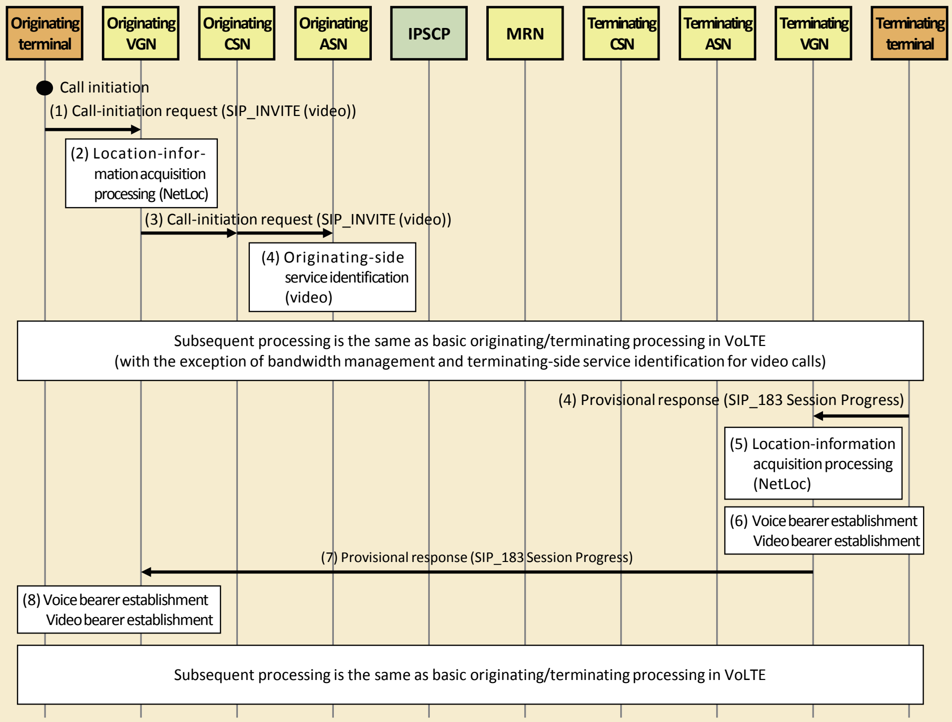

Figure 17 Overview of video-call originating/terminating processing in VoLTE

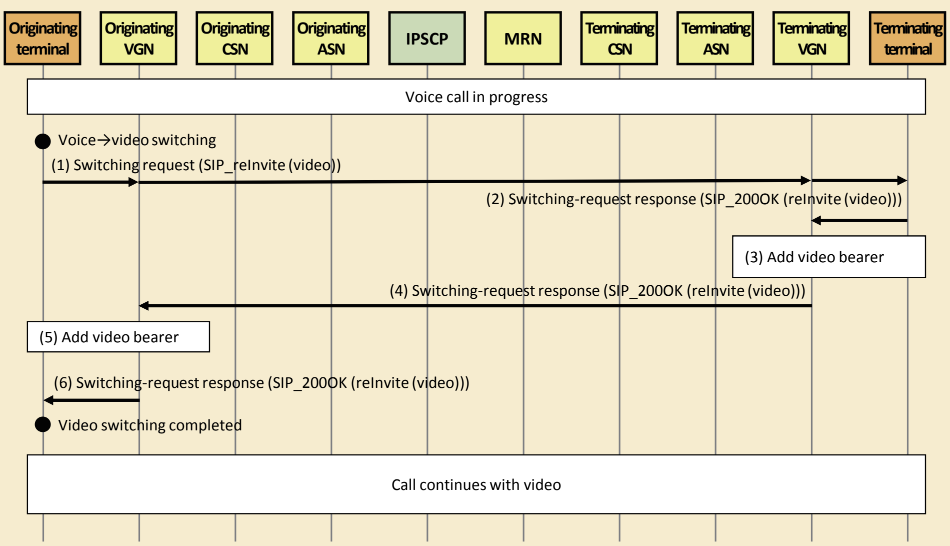

Figure 18 Overview of voice-to-video call switching

request and returns to the terminating VGN a switching-request response that includes information on whether the terminating subscriber is allowed video-call switching (Fig. 18 (2)). If video-call switching is allowed, the terminating VGN adds a video bearer (Fig. 18 (3)). After that, the terminating VGN returns a switching-request response back to the originating VGN, which also adds a video bearer and sends a switching-request response to the originating terminal. This completes voice-to-video call switching (Fig. 18 (4) – (6)). The same procedure is used to switch to a voice call while a video call is in progress.

6. Conclusion

This article described the background leading to VoLTE deployment at NTT DOCOMO, provided an overview of VoLTE architecture and functional features, and described its basic control methods. We expect this development work to enhance the user's sense of usability in a variety of ways, such as high-quality voice calls, quick originating and terminating of calls, use of high-speed multi-access, and use of video calls.

Going forward, we can expect the launch of VoLTE to accelerate the shift to LTE, and with this in mind, we plan to study efficient methods for accommodating and reducing 3G facilities over the medium and long term as 3G traffic decreases. We also plan to study a method for expanding VoLTE services to accommodate VoLTE roaming, interconnectivity, etc.

References

[1] GSMA PRD IR.92: "IMS Profile for Voice and SMS," Mar. 2013.

[2] Y. Shimada et al.: "IP-based FOMA Voice Network toward Enhanced Services and Improved Efficiencies," NTT DOCOMO Technical Journal, Vol. 12, No. 1, pp. 4–14, Jun. 2010.

[3] 3GPP TS23.272 V8.6.0: "Circuit Switched (CS) fallback in Evolved Packet System (EPS); Stage 2," Dec. 2009.

[4] VoLGA Forum website. http://www.volga-forum.com/

[5] I. Tanaka et al.: "Overview of GSMA VoLTE Profile," NTT DOCOMO Technical Journal, Vol. 13, No. 4, pp. 45–51, Mar. 2012.

[6] 3GPP TS23.203 V9.13.0: "Policy and charging control architecture," Sep. 2013.

[7] GSMA PRD IR.94: "IMS Profile for Conversational Video Service," Mar. 2013.

[8] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications," Mar. 2013.

[9] IETF RFC 5104: "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)," Mar. 2013.

[10] IETF RFC 3095: "Robust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," Jul. 2001.

[11] 3GPP TS36.321 V9.6.0: "Evolved Universal Terrestrial Radio Access (EUTRA); Medium Access Control (MAC) protocol specification," Mar. 2012.

[12] 3GPP, TS36.323 V9.0.0: "Evolved Universal Terrestrial Radio Access (EUTRA); Packet Data Convergence Protocol (PDCP) specification," Feb. 2010.

[13] 3GPP TR23.842 V11.0.0: "Study on Network Provided Location Information to the IMS; Stage 2," Dec. 2011.

[14] 3GPP TS23.237 V11.0.0: "IP Multimedia Subsystem (IMS) Service Continuity; Stage 2," Mar. 2011.

[15] IETF RFC 3312: "Integration of Resource Management and Session Initiation Protocol (SIP)," Oct. 2002.

[16] IETF RFC 2327: "SDP: Session Description Protocol," Apr. 1998.

[17] 3GPP TS24.341 V8.5.0: "Support of SMS over IP networks; Stage 3," Dec. 2010.

[18] 3GPP TS23.216 V9.9.0: "Single Radio Voice Call Continuity (SRVCC); Stage 2," Mar. 2012.