



下一代新通话网络媒体面演进及关键技术研究

翟振辉¹, 郝倩², 刘蕾¹, 吴倩¹, 刘晨³

1. 中国移动通信集团设计院有限公司, 北京 100080;
2. 北京京东方显示技术有限公司, 北京 100176;
3. 中国移动通信集团有限公司, 北京 100032)

摘要: 基于5G和IP多媒体子系统(IP multimedia subsystem, IMS)网络架构, 通过引入数据通道(data channel, DC)和人工智能(artificial intelligence, AI), 提出了下一代新通话的网络架构。随着网络架构的日益复杂, 媒体节点类型多、路由迂回、业务时延高等问题愈发突出。为简化网络、提升效率, 需要明确面向未来的网络演进思路。首先提出了新通话媒体面未来的演进目标和方向, 然后针对网络演进过程中的关键技术进行了深入的研究, 为新通话媒体面向目标架构演进提供了技术支持。

关键词: 新通话; 网络演进; 数据通道; 人工智能

中图分类号: TN915

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025037

Research on the evolution and key technology of the next generation new calling network media node

ZHAI Zhenhui¹, HAO Qian², LIU Lei¹, WU Qian¹, LIU Chen³

1. China Mobile Group Design Institute Co., Ltd., Beijing 100080, China
2. Beijing BOE Display Technology Co., Ltd., Beijing 100176, China
3. China Mobile Communications Group Co., Ltd., Beijing 100032, China

Abstract: Based on 5G and IP multimedia subsystem (IMS) network architecture, the network architecture of the next generation new calling was put forward by introducing data channel (DC) and artificial intelligence (AI). As the network architecture became more complex, the problems such as many media nodes, circuitous routing, and service delay were particularly prominent. In order to simplify the network and improve efficiency, it is necessary to define the network evolution ideas. Firstly, the future evolution goals of the new call media node was outlined, and then the key technologies in the process of network evolution was conducted in-depth analysis. It provides a reference to the evolution of the new calling media node.

Key words: new calling, network evolution, data channel, AI

收稿日期: 2024-12-07; 修回日期: 2025-02-13

通信作者: 翟振辉, zhaizhenhui@cmdi.chinamobile.com



0 引言

回顾移动通信的发展历程,从最初的模拟语音到数字语音,从音频通话到高清视频通话^[1],语音一直是人们最基础的通信业务需求,同时也是运营商最重要、最核心的基础业务,具有不可估量的商业价值和社会价值。

随着移动数据业务的快速发展,各种互联网应用也得到了快速发展。其中,以微信为代表的语音类应用对运营商的传统语音业务造成了极大冲击。相比之下,运营商现有的语音业务缺乏创新性、趣味性和交互性,难以满足年轻用户群体的个性化需求。因此,运营商迫切需要一个全新的、强大的、面向未来的新通话网络,以扭转在语音领域的颓势。

1 下一代新通话技术的发展趋势

随着5G技术的发展和成熟,作为5G标准语音解决方案的新空口承载语音(voice over new radio, VoNR),借助5G空口大带宽和低时延的网络优势,为用户带来了优质的语音和视频体验。然而,VoNR依然仅停留在可听和可视的业务体验上,与4G时代的长期演进语音承载(voice over long-term evolution, VoLTE)语音解决方案相比,并未发生实质性的变化。

为满足用户更丰富的多元化通话需求,新通话技术成为业界公认的下一代通话技术。相比于现有通话技术所具备的基础音视频能力,新通话技术具有更好的互动性和智能化特点,其突出的特性是引入数据通道^[2](data channel, DC)能力和人工智能(artificial intelligence, AI)能力。

(1) 引入DC能力。在现有音视频通道基础上增加一条数据通道,用来传递交互消息等内容,从而实现了用户在通话过程中进行点击、共享、标记等一系列基础交互功能。DC能力的应用需要网络和终端芯片同时支持,才能使用户

在可听、可视的基础上增加交互性,极大地丰富了业务场景,给用户带来了颠覆性的通话新体验。

(2) 引入AI能力。AI能力可以在通话过程中对语音和画面进行智能识别和处理等,如对通话背景进行渲染、对人物动作或者语音进行识别和响应。基于AI能力,可以衍生出丰富的通话业务形式^[3]。

新通话技术已成为业界公认的下一代新通话技术^[4]。无论是相关标准的建立,还是终端和芯片厂商对DC的支持,新通话技术均获得了前所未有的快速发展^[5]。

2 新通话网络标准架构

3GPP Rel-18标准中定义了基于IP多媒体子系统(IP multimedia subsystem, IMS)标准架构的新通话网络架构,主要引入了两个新的功能实体:数据通道信令功能(data channel signalling function, DCSF)和数据通道媒体功能(data channel media function, DCMF)^[6]。3GPP Rel-18定义的增强IMS数据通道架构如图1所示。

DCSF主要负责DC控制逻辑、控制上传和下载DC应用、订阅和处理媒体事件通知等。DCMF主要负责管理DC的媒体资源,建立和终止DCMF与终端之间的DC等。DCSF和DCMF共同完成语音网络对新通话DC能力的升级。在实际网络应用中,DCMF功能实体又称为新通话媒体面。

此外,新通话网络架构还支持向第三方DC应用服务器开放,可以基于网络的音频、视频、DC等原子能力开放,从而实现更多业务的灵活部署。

3 新通话网络演进研究

3.1 基本业务流程

新通话基本呼叫业务流程如图2所示,该流

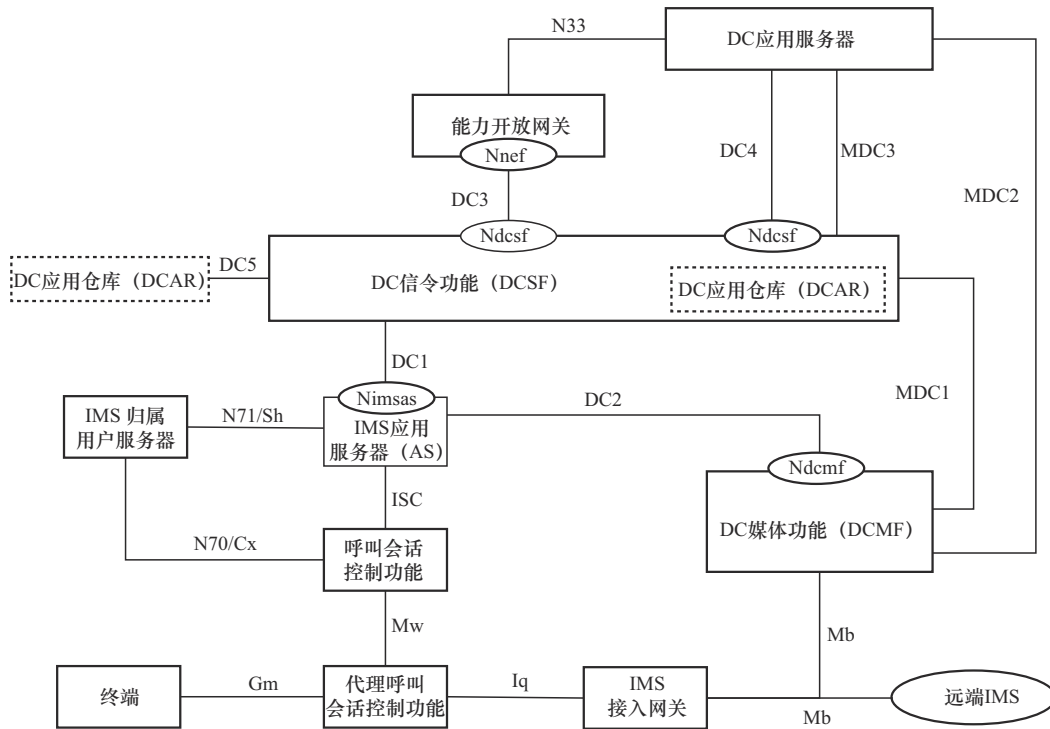


图1 3GPP Rel-18定义的增强IMS数据通道架构

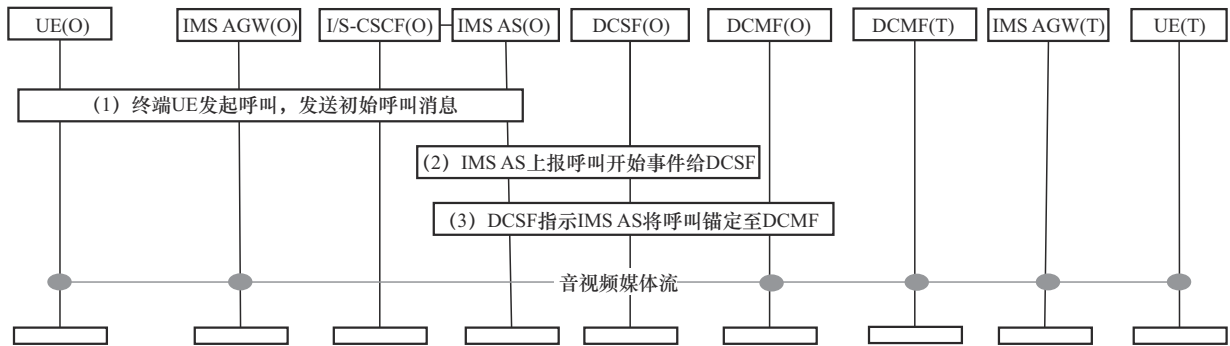


图2 新通话基本呼叫业务流程

程重点描述了DCSF和DCMF在呼叫建立过程中的作用。当签约新通话业务的用户发起新通话呼叫时，IMS AS上报呼叫开始事件给DCSF，DCSF根据呼叫是否为视频通话或者终端是否支持DC等条件进行判断，如果需要触发，则指示IMS AS将该呼叫的媒体流由IMS接入网关（access gateway，AGW）锚定至DCMF，新通话的媒体处理将由IMS AGW和DCMF共同完成。新通话的业务AS和DCSF可以直接对新通话媒体面进行媒体资源控制，完成各种针对媒体资源的渲

染、合成、插播等操作。

对于VoLTE/VoNR呼叫，与现有呼叫流程一致^[7]，即媒体流仅由IMS AGW处理，不涉及DCMF^[8]。因此，VoLTE/VoNR呼叫和新通话呼叫的基本媒体流程不同，对媒体节点的选择也不同。

3.2 媒体面演进方向

目前，运营商现网均已部署了大量承载VoLTE/VoNR业务的IMS AGW^[9]。IMS AGW主要实现的功能包括VoLTE/VoNR业务媒体资源管



理、服务质量 (quality of service, QoS)、音视频媒体编解码和媒体转发等。当用户发起 VoLTE/VoNR 呼叫时, 由 IMS 信令面选择用户接入地的 IMS AGW, 实现呼叫媒体流的建立。

IMS AGW 在现网实际部署中主要存在以下两种模式: 模式一是信令和媒体合设方式, 即媒体设备 IMS AGW 和信令设备 P-CSCF 作为融合网元部署; 模式二是信令和媒体分设方式, 即媒体设备 IMS AGW 和信令设备 P-CSCF 作为两个独立网元分别部署。对于模式一, 由于信令和媒体已采用合设方式, 因此 IMS AGW 不再作为独立的媒体设备进行演进。本文重点基于模式二, 即 IMS AGW 作为独立的功能实体, 研究媒体面未来的演进方向。

新通话网络引入了新的媒体面 DCMF, 主要实现的功能包括新通话业务的媒体处理、DC 资源管理以及背景替换、视频合成等 AI 功能。在引入 DCMF 功能实体后, 新通话媒体流首先由主叫侧 IMS AGW 锚定至主叫侧 DCMF, 再至被叫侧的 IMS AGW 和 DCMF 共 4 个媒体节点。不同节点分别对媒体进行转发、编解码、媒体 AI 渲染和合成等操作, 大大提高了媒体处理时延, 导致业务体验变差等问题出现^[10], 因此, 媒体面成为网络演进的研究重点。

从网络架构远期演进分析, 存在融合演进和独立演进两种思路。融合演进思路主要指将同样作为媒体处理的 IMS AGW 和 DCMF 向融合的网络架构进行演进, 不同业务的媒体均由统一的融

合媒体面完成处理, 融合媒体面兼具 IMS AGW 和 DCMF 两种功能实体能力。独立演进思路主要指按照现有标准的方式将 IMS AGW 和 DCMF 作为两个独立功能实体分别进行演进。

下面对两种演进思路进行对比分析。网络架构演进思路对比分析见表 1。

经过测试得出, 每减少一次媒体节点的转发可节省时延约 150 ms, 端到端最多可节省约 300 ms。传输资源节省情况主要取决于新通话用户的规模, 用户规模越大, 节省的媒体传输资源越多。基于上述业务时延、传输资源、组网复杂度和故障定位等维度的分析, 融合演进思路优于独立演进思路, 能够较好地解决时延、传输等问题, 并简化了网络架构。

虽然新通话在信令面同样引入了新的功能实体 DCSF, 但并未对业务体验、网络资源等方面造成大的影响。同时, 考虑信令流程交互相对复杂、影响范围大, 因此, 信令面暂不需要对网络架构进行调整。

4 关键技术研究

4.1 AI 能力测算方法

相较于传统音视频业务, 新通话业务首次在网络中引入媒体 AI 技术, 由 DCMF 配置图形处理单元 (graphics processing unit, GPU) 卡来提供相应的 AI 能力。GPU 的 AI 能力作为 DCMF 媒体硬件最关键、最重要的能力之一, 将直接影响设备性能和网络容量。因此, 新通话业务对 AI 能

表 1 网络架构演进思路对比分析

比较维度	融合演进思路	独立演进思路
业务时延	低: 媒体流在节点内部处理, 不需要经过外部传输, 时延较低	高: 媒体流要经过不同节点分别进行处理, 时延较高
传输资源	少: 媒体在节点内完成处理, 不需要占用额外的传输资源	多: 需要在 IMS AGW 和 DCMF 间占用传输资源
组网复杂度	低: 媒体节点类型和数量少	高: 媒体节点类型和数量多
故障定位	易: 媒体仅在融合节点处理, 媒体故障定位简单	难: 涉及不同的媒体处理节点, 媒体故障定位难度相对较大

力需求的测算方法,是需要研究和解决的关键技术之一。

目前,新通话的各类业务主要应用AI的推理能力。例如,在视频通话中进行背景替换时,需要DCMF内置的AI能力对画面中的人像轮廓进行识别和标记,然后基于轮廓匹配替换背景素材,从而完成对视频背景的替换^[11]。因此,背景替换对AI的能力需求主要为整个通话过程中的背景图像推理能力。这一过程通过深度神经网络技术对视频流进行处理,大致分为如下4个步骤。

步骤1: 使用基础网络提取特征,获得特征向量。

步骤2: 使用空洞卷积组合增加感受野。空洞卷积是一种卷积操作,其可以通过在卷积核中插入空洞(即稀疏化卷积核)来增加感受野,从而提升网络的性能及用户体验。感受野是指在当前尺寸的特征图上,一个像素点能够映射回原图像的区域大小,即该像素点在原图像的覆盖范围。

步骤3: 优化人像轮廓边缘。

步骤4: 图片替换及视频编码^[12]。

此外,其他新通话业务同样会应用到AI技术。例如,手势动效借助AI技术对画面中的特定手势进行分析和识别,并在视频画面上叠加对应的手势动效。因此,手势动效对AI能力需求主要是动效的叠加显示过程中的AI图像推理能力。再如,翻译字幕业务通过AI技术对语音进行识别并翻译显示为字幕。因此,翻译字幕对AI的能力需求主要是在通话视频中叠加显示字幕的AI图像推理能力^[13]。不同业务对AI能力的需求不尽相同,因此,需要针对不同业务分别测算所需的AI资源。

GPU数量具体测算方法如下所示。

$$N = \left(\sum_{i=0}^n s_i \times c_i \times t_i \times a_i \right) \div G \quad (1)$$

其中, N 表示各类新通话业务对GPU卡的数量,

n 表示有 n 种不同业务需要应用AI资源, s_i 表示第 i 种业务的总用户数, c_i 表示第 i 种业务单位时间内平均发生的呼叫次数, t_i 表示第 i 种业务每次呼叫的平均占用时长, a_i 表示AI处理时长占总通话时长的平均比例, G 表示单张GPU卡的AI推理并发通道数能力。

实际应用过程中,首先需要针对新通话各类业务制定相应的业务模型,再按业务逐一进行资源需求测算^[14]。下面仅以手势动效业务为例进行说明。

设定签约手势动效业务的新通话用户为1万户,每位用户忙时平均呼叫次数为1.5 BHCA (busy hour call attempt,忙时试呼),平均通话时长为90 s,即0.025 h;AI处理手势动效的时长占总通话时长的平均比例为10%,则根据AI能力测算方法可以测算手势动效业务对AI图像推理能力需求为37.5路通道数。例如,某GPU卡的AI能力为20路并发通道数/卡,则所需该类GPU卡的数量为1.875块。其他业务计算过程相似,本文不再一一赘述。

4.2 融合组网关键技术

在向全融合部署的目标演进过程中,会存在DCMF、IMS AGW和融合媒体面3种形态共存的过渡阶段。在过渡阶段,由于媒体面网元类型多,因此媒体面混合组网技术复杂^[15],本文研究并提出了以下两种组网方式。

方式一 按照用户业务属性进行独立组网,即按照基础VoLTE/VoNR业务和新通话业务分别组网。IMS AGW单独组网用于承载传统VoLTE/VoNR用户的业务;DCMF和融合媒体面单独组网,用于承载新通话用户的业务。独立组网示意图如图3所示。

方式二 不同类型的媒体面混合组网,即IMS AGW、DCMF和融合媒体面混合组网,同时负责新通话和VoLTE/VoNR两类业务。混合组网示意图如图4所示。

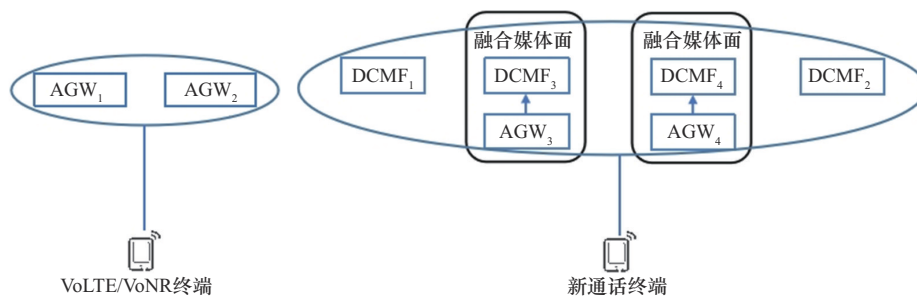


图3 独立组网示意图

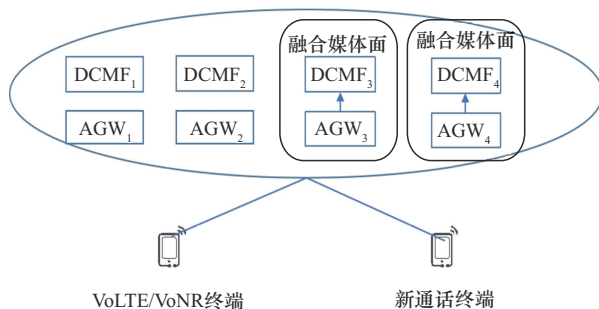


图4 混合组网示意图

对于方式一，在用户注册阶段，系统需要根据用户的签约业务将用户锚定至不同pool内进行注册并提供后续语音业务。对于仅签约VoLTE/VoNR业务的用户，将在IMS AGW pool中完成注册并提供服务；对于签约了新通话业务的用户，则在新通话媒体面pool中完成注册并提供新通话业务。目前，现有技术仅能根据用户所在位置锚定至对应区域的媒体面pool。因此，该方式需要对网络进行功能升级，以支持按用户签约业务类型选择不同pool进行注册等相关流程。同时，随着新通话业务的发展，用户规模越来越大，独立组网需要投资新建融合媒体面以承载新增的新通话用户，然而，该方式无法利用现网已有的IMS AGW资源。

对于方式二，不同类型的媒体面采用混合组网，VoLTE/VoNR用户和新通话用户均由同一融合媒体面pool完成注册和语音业务。对于VoLTE/VoNR用户的媒体处理，由pool内独立的IMS AGW和融合媒体面的AGW模块共同完成处理；对于新通话用户的媒体处理，由pool内独

立的DCMF和融合媒体面的DCMF模块共同完成处理。同时，即使当VoLTE/VoNR用户签约成为新通话用户后，依然可以利用pool内的IMS AGW容量，最大限度地实现资源共享。然而，该方式pool内同时存在3种形态的媒体设备，媒体处理流程不同，参数设置相对复杂。

在现网实际应用过程中，考虑未来新通话业务的发展情况，最大限度利用现网IMS AGW容量，避免投资浪费，建议优先采用方式二进行组网。

4.3 参数设置方法

融合媒体面的媒体处理会优先选择内部的媒体能力进行处理。由于混合组网下不同类型媒体面的媒体处理方式不同，融合组网的技术十分复杂。为避免出现负荷过载引发设备故障，导致网络事故的发生，需要研究混合组网方式下不同媒体面之间的负荷参数设置问题。以图4中的IMS AGW、DCMF和融合媒体面混合组网为例，说明负荷参数设置方法。

(1) IMS AGW的负荷参数设置方法：

$$AGW_1:AGW_2:AGW_3:AGW_4=X_1:X_2:X_3:X_4 \quad (2)$$

其中， X_n 为IMS AGW n 的容量，即按照容量进行等比例设置。

(2) DCMF的负荷参数设置方法：对于IMS AGW $_3$ 和IMS AGW $_4$ ，优选内部的DCMF，即分别选择DCMF $_3$ 和DCMF $_4$ ，不需要设置负荷参数；对于IMS AGW $_1$ 和IMS AGW $_2$ ，可以负荷分担地选择DCMF $_1\sim$ DCMF $_4$ ，不同DCMF应按照如下参数进行设置：

$$\text{DCMF}_1:\text{DCMF}_2:\text{DCMF}_3:\text{DCMF}_4= \\ Y_1:Y_2:(Y_3-X_3 \times R \times Q):(Y_4-X_4 \times R \times Q) \quad (3)$$

其中, Y_n 表示媒体面 n 的容量, R 表示渗透率, 即新通话用户占VoLTE/VoNR用户的比例, Q 表示主叫呼叫比例与非DC终端被呼叫比例之和。

根据终端是否支持DC能力, 可将其分为非DC终端和DC终端。非DC终端作为被叫时, 先选择IMS AGW, 后选择DCMF; DC终端作为被叫时, 会先选择DCMF, 再选择IMS AGW。

媒体面IMS AGW和DCMF的容量, 通常会将硬件能力按照业务模型折算成并发在线用户数。其中, 硬件能力包括中央处理器(central processing unit, CPU)处理会话的能力、GPU的AI图像处理能力、媒体转发带宽等。目前, DCMF的容量主要受限于GPU的AI能力。

5 结束语

新通话是语音领域的一次重要变革, 通过引入DC能力和AI能力, 通话中的业务体验变得更加丰富, 在网络能力增强的同时, 网络架构也变得更加复杂。未来, 随着智能体和人工智能生成内容(artificial intelligence generated content, AIGC)等新技术的应用, 下一代网络应向更简化和更智能的方向演进。本文研究并提出了网络架构演进的思路和关键技术, 对未来网络演进和新技术发展应用有着重要的参考意义。

参考文献:

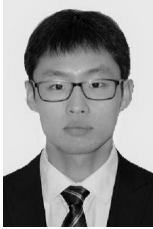
- [1] 边燕南. 构建下一代新通话网络 开启实时通信新时代[J]. 通信世界, 2023(19): 20-21.
BIAN Y N. Constructing the next generation of new telephone network and opening a new era of real-time communication[J]. Communications World, 2023(19): 20-21.
- [2] 陆振锋, 龚敏敏. 5G新通话研究与应用[J]. 江西通信科技, 2024(3): 17-20, 32.
LU Z F, GONG M M. Research and application of 5G new call[J]. Jiangxi Communication Science & Technology, 2024(3): 17-20, 32.
- [3] 褚连杰, 蒋无瑕, 王红线. 基于DC通道和AI技术的5G新通话在适老助残领域的应用[J]. 江苏通信, 2024, 40(1): 6-9.
CHU L J, JIANG W X, WANG H X. Application of 5G new call based on DC channel and AI technology in the field of the elderly and the disabled[J]. Jiangsu Communication, 2024, 40(1): 6-9.
- [4] 刘鸿. 5G新通话将开创个人通信的新时代[J]. 通信世界, 2023(19): 1.
LIU H. New 5G calls will usher in a new era of personal communication[J]. Communications World, 2023(19): 1.
- [5] 王志勤. 产业协同 推动新通话繁荣发展[J]. 通信世界, 2023(19): 12-13.
WANG Z Q. Industrial synergy promotes the prosperity and development of new calls[J]. Communications World, 2023(19): 12-13.
- [6] 3GPP. Study on system architecture enhancement for next generation real time communication: TR 23.700-87[S]. 2022.
- [7] 3GPP. IP Multimedia Subsystem (IMS): TS 23.228[S]. 2020.
- [8] 3GPP. IP multimedia subsystem (IMS) multimedia telephony media handling and interaction: TS 26.114[S]. 2020.
- [9] 3GPP. IMS Application level gateway (IMS-ALG)-IMS access gateway (IMS-AGW): TS 29.334[S]. 2019.
- [10] 李欣然, 孙伟, 马云飞, 等. 5G新通话感知提升策略研究[J]. 通信世界, 2024(2): 25-26.
LI X R, SUN W, MA Y F, et al. Research on the promotion strategy of 5G new call awareness[J]. Communications World, 2024(2): 25-26.
- [11] 冯小芳. 5G新通话技术及应用研究[J]. 电信快报, 2023(9): 11-15.
FENG X F. Technology and application research on 5G VoNR+[J]. Telecommunications Information, 2023(9): 11-15.
- [12] 刘晨. VoNR+新通话关键技术应用研究[J]. 电信工程技术与标准化, 2023, 36(3): 14-20.
LIU C. The study of key technology application for VoNR+ new calling[J]. Telecom Engineering Technics and Standardization, 2023, 36(3): 14-20.
- [13] 倪占玲, 陈永朋, 刘云英. 5G新通话应用场景分析及解决方案[J]. 数字通信世界, 2024(8): 101-103, 115.
NI Z L, HEN Y P, LIU Y Y. Analysis and solution of 5G new call application scenarios[J]. Digital Communication World, 2024(8): 101-103, 115.
- [14] 王海涛. 5G新通话关键技术与应用研究[J]. 广播电视网络, 2024, 31(5): 21-23.
WANG H T. Research on key technologies and applications of 5G new calls[J]. Radio & Television Network, 2024, 31(5): 21-23.



[15] 刘蕾, 刘晨, 吴倩, 等. 5G新通话DC媒体处理兼容方案[J]. 电信工程技术与标准化, 2024, 37(9): 64-71.

LIU L, LIU C, WU Q, et al. 5G new calling DC media processing compatible solution[J]. Telecom Engineering Technics and Standardization, 2024, 37(9): 64-71.

[作者简介]



翟振辉 (1985-), 男, 中国移动通信集团设计院有限公司高级工程师, 主要研究方向为移动通信核心网、5G、IMS等。



郝倩 (1986-), 女, 北京京东方显示技术有限公司工程师, 主要研究方向为移动终端芯片等。



刘蕾 (1977-), 女, 现就职于中国移动通信集团设计院有限公司, 主要研究方向为移动通信核心网、5G、IMS业务等。



吴倩 (1982-), 女, 中国移动通信集团设计院有限公司高级工程师, 主要研究方向为移动通信核心网、5G、IMS等。



刘晨 (1984-), 男, 中国移动通信集团有限公司高级工程师, 主要研究方向为移动通信核心网、5G、IMS等。