

Centralization and collaboration in 5G ultra-dense network architecture

WEI Hong-jing¹, GUO Bao¹, ZHANG Yang²

(1. China Mobile Group Shanxi Co., Ltd., Taiyuan 030032, China;

2. China Mobile Group Co., Ltd., Beijing 100033, China)

Abstract: An ultra-dense network scenario is a scene where a large number of people assemble in a limited area to generate centralized broadband data traffic requirements. Because ultra-dense networks generate enormous traffic pressure, traditional network capabilities are not enough to accommodate the user's needs. Based on the description of ultra-dense network architecture, we analyze millimeter wave radio spectrum, high gain beam forming, physical layer frame structure, resource concentration and edge computing technology. In addition, the cooperative technology required by overlay and interference symbiosis in the dense network architecture as well as the access control technology of centralized access is analyzed and discussed comprehensively.

Key words: ultra-dense network architecture; millimeter wave; edge computing; multi-point collaboration

CLD number: TN915.02

doi: 10.3969/j.issn.1674-8042.2020.01.009

0 Introduction

5G can be considered that people-centric communication extends to both people-centered and object-oriented communication. Its system requirements are very broad. Now the 5G systems provide communication services mainly including extreme mobile broad band (eMBB), massive machine type communication (mMTC) and ultra-reliable Machine Type Communication (uMTC). The eMBB provides extremely high data rate, low latency communication and extremely high coverage. The mMTC provides wireless connection to tens of billions of network devices. Compared with data rate the mMTC is more concerned with the scalability of connection, efficient delivery of small data, broad area coverage and deep coverage; The uMTC offers ultra-reliable high-availability and low-latency communication network services. To ensure these services, 5G has designed four key technologies: dynamic radio access network (DyRAN), minimal system control plane, content localization and spectrum toolbox. Among them, DyRAN mainly provides wireless access networks that suit user needs and 5G service combinations. The main purpose of the system is to provide services to ultra-dense networks (UDNs), nomadic and relay nodes, antenna beams, terminal equipment as temporary

access nodes for access and return communication^[1].

We focus on the research of ultra-dense networks. The main challenge of the ultra-dense network scenario is to provide each user with data transmission requirements that meet a certain bandwidth in a relatively limited space, while supporting high connection density and high flow requirements, and maintaining service connections to keep the user in good condition experience^[2]. Based on the description of UDN architecture, we analyze millimeter wave (MMW) radio spectrum, high-gain beam forming, physical layer frame structure, resource concentration and edge computing technology^[3].

1 UDN architecture

Increased network density can directly increase network capacity. While network density can be enhanced by deploying dense microcells in design of 5G network system. The goal rate of UDN designed is 10 Gbit/s above. In order to achieve this design goal, we must consider three key technologies: MMW spectrum, high gain beam forming and physical layer subframe structure of UDN^[4].

1.1 Millimeter waves

The propagation of MMW reduces the coverage to

Received date: 2019-11-27

Corresponding author: WEI Hong-jing (85358838@qq.com)

a smaller area, therefore NDN network is the inevitable result of high-band spectrum selection, which results in a huge increase in service capacity in the coverage area. The increase of spectrum

efficiency is mainly caused by sharp reduction in interference signals as well as to the high-gain beam forming^[5]. The centimeter wave (CMW) MMW and 5G spectrum range are shown in Fig. 1.

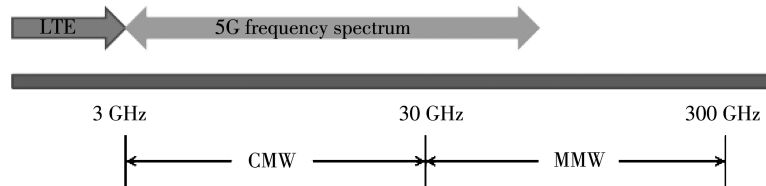


Fig. 1 Centimeter waves, millimeter waves and 5G spectrum

The MMW band presents a high challenge for wireless communications. The large-scale loss of visible-distance path generally follows free-space loss and is proportional to the radiation attenuation in all directions and the square of the increase in operating frequency^[6]. At different frequencies, the constant coupling loss can remain frequency-independent if the aperture size of the transmitting or receiving antenna remains constant. The attenuation of frequency radiation in all directions, that is, the free space loss, is usually more than the compensation used in high-gain antenna for transmitters and receivers. In the MMW band, any moving wireless system will need to adapt the antenna array or beam-forming of warn sector^[7].

Radio frequency (RF) module performance decreases with frequency increasing. In general, every 10 times the frequency increases, the power amplifier capability for a given integrated circuit technology is roughly 15 dB down, and higher operating frequencies require small geometries, which inevitably also leads to lower operating power.

1.2 High-gain beam forming

The increase in spectral efficiency is mainly due to sharply decrease of interference relative to useful signal, which results from high-gain beam forming. Microcell deployments enable short and multi-view wireless links, lower output power and new spectrum access. High-gain beam forming with a large number of antenna elements provides additional energy efficiency, compensates for high path loss at higher frequencies and reduces interference from the same physical resources^[8-9].

The multiple input multiple output (MIMO) network utilizes the degree of freedom (DoF) provided by antennas to transmit information to multiple users on the same time-frequency resource,

to focus the radiated signals on the intended users and to minimize intercell and intercell interference. By transmitting the same signal from multiple antenna points, applying different phase shifts to each antenna (different phase shifts may be applied to different parts of the system bandwidth) makes the signals coherently overlap at the intended target location.

Due to the size of antenna array, large-scale MIMO technology used in the existing LTE networks is difficult in practice, however, the MMW technology greatly reduces the size of antenna elements and makes it possible to implement large-scale MIMO in UDNs^[10].

There are n_t transmitting antennas and n_r receiving antennas, when the channel is narrowband, the MIMO channel capacity is defined as

$$C = \sum_{i=1}^{n_{\min}} \log \left(1 + \frac{P}{n_t N_0} \lambda_i^2 \right), \quad (1)$$

where $n_{\min} = \min(n_t, n_r)$, λ_i is the singular value of the channel matrix, P is the effective power of signal, and the signal-to-noise ratio is $SNR = \frac{P}{N_0}$.

When the number of antennas tends to infinity, $n \rightarrow \infty$, the capacity of $n \times n$ point-to-point MIMO links can be approximated as

$$\lim_{n \rightarrow \infty} \frac{C_m(SNR)}{n} = C(SNR) \rightarrow C_m(SNR) \approx nC(SNR). \quad (2)$$

For a large number of antennas and users, scaling of the capacity can be achieved by increasing the number of transmitting and receiving antennas, but large-scale MIMO also faces various challenges that must be overcome. For example, for pilot pollution, pilot overhead should be reasonable for large-scale MIMO with many channel components that need to be estimated. In the subsequent optimization work,

it is necessary to consider pilot power control based on pilot power coded pilots using sparse channel attributes or random use of pilot sequences in a cell to overcome pilot pollution.

1.3 UDN physical layer subframe structure

UDN requires continuing to reduce time division duplex (TDD) handoff time and enhance digital signal processing performance on the basis of 4G standard. It is possible to quickly and fully flexibly switch between sending and receiving using a shorter frame length and a physical frame structure for the microcell design^[11-12].

The bi-directional control portion is embedded in each subframe, as shown in Fig. 2, which allows a device in the network to receive and send control signals in each subframe, such as scheduling requests and scheduling grants. In addition to scheduling-related control information, the control section may further include reference signals and synchronization signals.

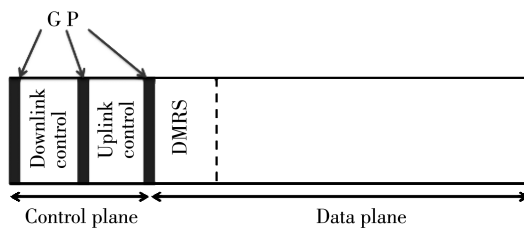


Fig. 2 Physical layer subframe structure of UDN system

The data portion in one subframe contains data symbols for transmission or reception. The demodulation reference signal (DMRS) for channel and interference coordination is located in the first orthogonal frequency division multiplexing (OFDM) symbol in the dynamic data portion and can be used for precoding for the same vector/matrix.

From the point of protection and control overhead, the short subframe length can achieve 0.25 ms at CMW frequency assuming that subcarrier spacing is 60 kHz, which follows the harmonized OFDM principle. When the frequency is used in MMW, the frame number parameter is further scaled, resulting in shorter frame lengths, such as about 50 μ s.

2 Resource concentration and content sinking

2.1 Resource concentration

Interference is a major issue in UDNs. To do this, more coordination and cooperation are needed

between base stations. In addition, the increase of network density has put forward higher requirements on mobility management. Mobile users want to continue seamless switching, but from a network perspective, they want to minimize management overhead caused by frequent handoffs in ultra-dense networks. How to design a more efficient switching mechanism and mobility management is an urgent need for a future wireless access network^[13].

Centralized radio access network (C-RAN) is based on distributed and remote base station. It uses remote radio unit (RRU) and base band unit (BBU) architecture networking to concentrate all or part of the baseband processing resources to form a baseband resource pool, manage and dynamically allocate them. While improving resource utilization and reducing energy consumption, network performance is enhanced through effective support for collaborative technologies^[14]. In the 5G network, C-RAN addresses high-frequency, high-bandwidth, multi-antenna, mass connection and low-latency requirements by introducing functional reconfiguration of the centralized unit/ distributed unit (CU/DU) and forwarding architecture of the next-generation front-haul interface (NGFI).

The BBU function of 5G is reconstructed into two functional entities, CU and DU. The function segmentation of CU and DU is distinguished based on timeliness of processing content. The CU device mainly includes a non-real-time wireless high-layer protocol stack function, and also supports the deployment of some core network functions and edge application services. The DU device mainly processes physical layer functions and Layer 2 functions of real-time requirements. To save the transmission resources between the RRU and the DU, some physical layer functions can also be moved up to the RRU.

In terms of specific implementation solution, the CU device adopts common platform, which not only can support wireless network function but also has the ability to support core network function and edge application. The DU device can be implemented by using a dedicated device platform or a universal + dedicated hybrid platform, and supports high-density mathematical computing capabilities. After the introduction of network functions virtualization (NFV) framework, unified management and orchestration (MANO), combined with software defined networking (SDN) controller and traditional

operating and maintenance center (OMC) functional components, can achieve end-to-end flexible resource scheduling and configuration capabilities including CU/DU to meet operators' rapid on-demand business deployment needs^[15].

In order to solve the transmission problem between CU/DU/RRU, we can introduce NGFI framework. The CU is connected to the remote DU through a switching network. The technical feature of this architecture is that the functional units can be deployed flexibly according to scenario needs. When transport network resources are sufficient, we can centralize the deployment of DU functional units to achieve physical layer collaboration, but when the transport network resources are insufficient, we can still distribute the deployment of DU processing unit.

The CU functions are to centrally concentrate part of the functions of the original BBU. It is fully compatible with both centralized deployment and distributed DU deployment. This can maximize the ability to ensure collaboration, at the same time, compatible with different transport network capabilities.

C-RAN centralization in 4G networks is that a certain number of BBUs are centrally placed in a large central equipment room. With the introduction of CU/DU and NGFI, C-RAN in 5G networks gradually evolves into a concept of logical two-level concentration. The first level concentrates on the concept of BBU placement to realize the concentration of physical layer processing. This has the obvious advantages of reducing the difficulty of site selection, reducing the number of equipment rooms, and sharing auxiliary equipment (such as air conditioners). We can select the appropriate application scenarios, selectively small-scale concentration (such as the level of hundreds of carriers). The second level of concentration is the concentration of wireless high-level protocol stack functions after the introduction of CU/DU. The existing eNodeB functions are segmented, and some wireless high-level protocol stack functions are deployed centrally.

Corresponding to the concept of two-level concentration, the first level focuses on the small-scale physical layer concentration, and physical layer technologies such as coordinated multi-point (CoMP) and distributed MIMO can be introduced to achieve jointly transmitting and jointly receiving between

multi-cell / multi-data transmission points, and to improve cell edge spectral efficiency and average throughput. The second level of concentration is the concentration of large-scale wireless high-level protocol stacks, which can serve as anchor points of the control plane and user plane in wireless services. With the introduction of 5G air interfaces in the future, it enables collaborative capabilities such as multi-connectivity, seamless mobility management and efficient coordination of spectrum resources.

2.2 Content sinking

The existing applications and services in 4G networks also place higher demands on low latency. Simply increasing the speed is not sufficient to meet low latency requirements in various scenarios. Therefore, content resources need to be sunk in the edge network. The technical characteristics of mobile edge computing (MEC) are low latency of communication, platformization of application and openness of interaction layer, which can realize the transformation of traditional wireless communication network architecture as well as discovery and release of the potential of wireless network services. The traditional wireless communication network from the terminal to the service ends include user end, eNodeB, transport bearer network, evolved packet core (EPC) and business platform. In the logical position order, the traditional wireless communication network clearly distinguishes radio bearer network from service network (or service platform). After the service platform is in the core network, the data that the user accesses from the service platform need to be transmitted through radio access network, transport network and evolved packet core.

The MEC technology has brought forward the demand for the change of traditional network architecture, and introduced the MEC equipment platform between the traditional wireless base station and EPC. MEC equipment platform hardware consists of servers, switches and memory, as shown in Fig. 3. MEC software architecture includes control unit, calculation unit and exchange unit in three parts. Based on the hardware and software architecture of MEC platform, operators can support various business applications by embedding in house virtual machines and set up a service platform behind the MEC. Through local data forwarding and API interface opening, the service data are forwarded to

the corresponding the third parties business platform to support rich and varied business types. At the same time, because the business source sinks to the wireless access network through the MEC platform,

the delay of the end-to-end service is shortened, the resource consumption of core network by large connection and massive service is reduced, and service fee cost is also reduced.

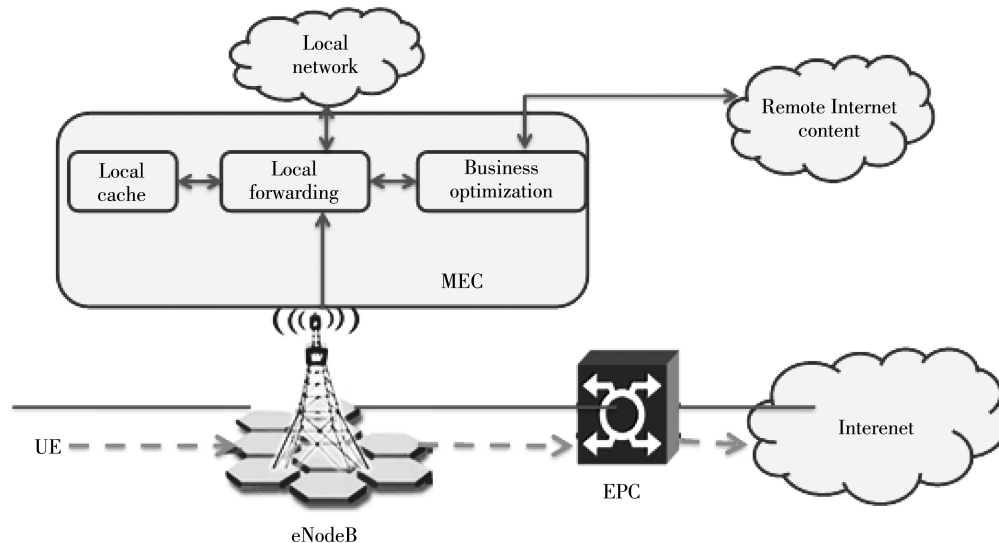


Fig. 3 MEC wireless communication network architecture

MEC technology also has the technical characteristics of open level of interaction, and mining and release of wireless networks on business support potential. At the mobile network edge, through the opening of MEC device API interface, operators provide application developers and content providers with cloud computing capabilities, IT service environments and real-time access to wireless network information. Thus the wireless access network from a purely data pipeline changes to have the value-added service capabilities of wireless resource network. Through this wireless resource network, combining functions such as positioning, service diversion and localization, and marketing strategies, operators can diversify their value-added services and enrich the content and applications of digital services.

When implementing a local cache, the MEC needs to have a cache scheduling function and a content storage function. When a user initiates a service request, the MEC senses the content of the user service request through packet parsing and determines whether the content is cached. If it is already cached, it is routed to local storage server. If there is no cache, the packet is transparently forwarded to core network. The local forwarding needs MEC with route resolution and address translation capability to implement local forwarding. Unlike local caches, local forwarding requires that

the terminal's IP address should be translated to gain access to the local network.

3 Access control and interference coordination

3.1 Access control

In large-scale sports events, concerts, universities and other UDN scenarios, a large number of users in a short time to start the business, serious cases will lead to a sharp decline in wireless connectivity and user experience significantly lowers. 5G network designs a large-capacity call attempt per second (CAPS) to be responsible for the sudden impact of large-capacity business. CAPS capability measures the base station's ability to handle concurrent signaling. In general, terminal access to the network consists of two processes: first access to the network and then allocation of resources. The former is a random access process, which is the control channel and traffic channel allocation process. Random access is a necessary process for establishing a radio link between a terminal and a network. Only when the random access process is completed, the base station and the terminal can perform regular data transmission and reception. Through the random access procedure, the UE can implement uplink synchronization with the evolved node B (eNB) and

apply for uplink resources.

Taking into account the wireless resource of the air interface of base station and limited hardware resources of the device, the base station can only support a certain number of connected users, cell activation users and bearers. Therefore, when the terminal randomly accesses the network, the base station needs to access the access request control. In the case of ensuring the number of users and quality of service (QoS), as many users as possible are allowed to access the network. To ensure the new access bearer QoS, we need to improve system capacity and resource utilization.

After the terminal accesses the network randomly, the base station allocates the uplink and downlink resources according to data transmission request. The data transmission in the LTE system includes two processes of uplink scheduling and downlink scheduling. During uplink scheduling, the UE sends a scheduling request through an uplink control channel, and requests an uplink resource from the eNB. The eNB notifies the UE of the resource allocation result through a downlink control channel. The UE can know at which time which carrier transmits the uplink data, and uses the modulation and coding scheme. The eNB dynamically allocates resources to a UE at each transmission time interval (TTI) and transmits a corresponding cell-radio network temporary identifier (C-RNTI) on a downlink control channel. During downlink scheduling, the eNB allocates downlink resources for the UE according to the downlink channel quality reported by the UE, and populates the data on the PDSCH according to the resource allocation result.

CAPS refers to the ability of the device to handle concurrent signaling. The processing capabilities of the base station control board and baseband board are limited. Each signaling message consumes the CPU resources of the control board or baseband board. The CPU load reflects the use of CPU resources. When the terminal accesses the TD-LTE cell uniformly and the signaling inbound and outbound traffic per unit time is lower than the CAPS capability of the base station, the signaling traffic speed is smooth and the user can access the cell quickly. When a large number of terminals in the cell frequently initiate a service and trigger a random access, the concurrent signaling traffic leads to signaling congestion, thus the CPU load increases and the signaling processing capability decreases.

3.2 Interference coordination

The frequency reuse factors of 4G and 5G networks are usually 1 or very close to 1. In this case, the system is mainly interference-limited, and performance cannot be improved by simply increasing the transmission power. Especially in UDN, dense overlapping of signals and interference is obviously sub-optimal if the interference is simply treated as white noise, neglecting the characteristic that the interference signal can be used to improve the quality of the received signal. The effects of interference can be mitigated by various receivers, such as interference rejection combining (IRC), where multiple receiving antennas and subsequent receiving filters attenuate interference to some extent. Interference can be decoded and canceled, such as network assistant interference canceling (NAIC). On the transmitter side, interference can be partially avoided by performing interference aware precoding.

In summary, interference coordination is the joint transmission of signals by various nodes in the downlink. The signal is coherently overlapped at the intended receiver and destructively overlapped at the interfered reception, while signals from other cells can be regarded as useful signal energy instead of interference. In the uplink, multiple nodes may jointly receive and decode signals from multiple UEs. This kind of ending is collectively referred to as CoMP, meaning that multiple nodes in the network coordinate or cooperate to mitigate the impact of interference.

Uplink multi-point cooperation (UL CoMP) function selects eligible UEs for joint reception of multi-cell antennas, so for UL CoMP-enabled UEs, it is similar to increasing the number of receive antennas and gaining the gain from increasing the number of antennas. There are two sources of gain for UL CoMP when the UE is located in different geographical locations; signal combining gain (UE is at cell edge) and interference suppression gain (UE is at any cell location).

Signal merger gain diagram is shown in Fig. 4(a). The UE is located at the edge of a cell or at the overlap of two cells. The transmitted signal can be simultaneously received by the antennas of different cells. The joint reception by the UL CoMP improves the received signal quality, and obtains obvious signal combination gain.

Interference suppression gain diagram is shown in

Fig. 4(b). The eNodeB chooses to perform UL CoMP due to interference from cell-edge users. When co-reception of UL CoMP is performed, co-interference suppression algorithm is used to suppress co-channel interference so as to obtain interference suppression gain for cell edge user (UE1). The magnitude of the interference suppression gain obtained by UE0 is only related to the location of UE1. The greater the interference from UE1 to UE0, the more obvious the interference suppression effect.

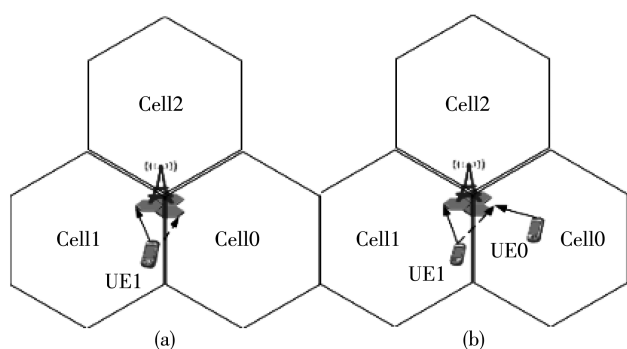


Fig. 4 UL CoMP gain diagram

4 Conclusion

UDN scenarios usually involve a large number of people gathering in limited areas, and visitors want to share high-definition pictures and videos with family and friends in real time, which creates huge traffic pressure. The traditional network capacity is far from enough to meet the needs of users. The main challenge of the UDN scenario is to provide each user with data transmission that not only meets certain bandwidth requirements in a relatively limited space, but also supports high connection density and high traffic requirements. There is also a need to safeguard service connections so that users maintain a good experience.

UDNs need to dramatically increase network density and decrease the number of active users in a single station, which requires the use of a new collaborative-based air interface for flexible spectrum utilization. On the basis of using millimeter waves to realize the centralized use of large-scale MIMO antennas. For large wireless resources, we can use resource-intensive C-RAN technology. For services concentrated in the demand of ultra-dense scenes, we can use MEC technology to sink content resources to the wireless device to shorten service response delay. Due to the signal and interference generated by the

dense antenna, access control CAPS and interference coordination CoMP need to be considered. In addition, we need to plan reasonable mobility management, prediction technology and switching optimization.

References

- [1] Zhang J M, Xie W L, Yang F Y. Network architecture and implementation of 5G ultra dense network. *Telecommunications Science*, 2016, 32(6): 36-43.
- [2] Liu Y M, Li X, Ji H. Key technologies of network self-organization for 5G ultra dense scenes. *Telecommunications Science*, 2016, 32(6): 44-51.
- [3] Jiang J S, Jiang L P, Zhu X R. Modeling and analysis of mobile performance in hyper dense clustering networks based on Markov model. *Information and Communication Technology*, 2017, 11(1): 78-84.
- [4] Li Y Z, Jiang T, Cao Y. 5G green, ultra dense wireless heterogeneous networks: concepts, technologies and challenges. *Telecommunications Science*, 2017, 33(6): 34-40.
- [5] Cheng M. New technology of network planning for 5G ultra dense networks. *Mobile Communications*, 2016, 40(17): 28-29.
- [6] Zhou X. Study on broadband spectrum fast sensing in ultra-dense networks. *Mobile Communications*, 2016, 40(8): 52-56.
- [7] Ge L, Wang Z X. Energy optimal high efficiency dense network deployment of. *Telecom Science*, 2017, 33(3): 44-51.
- [8] Zhan J, Yan Q F, Tang X H. 5G oriented cache assisted multi antenna relay strategy. *Telecommunications Science*, 2017, 33(6): 2-10.
- [9] Fang Z, Li Y, Li H T. Study on beam forming of air-space in ultra-dense networks. *Journal of University of Electronic Science and Technology of China*, 2016, 45(2): 185-190.
- [10] Lei Q Y, Zhang Z Z, CHENG F. C-RAN based 5G wireless access network architecture. *Telecommunications Science*, 2015, 31(1): 106-115.
- [11] Wang J X, Tang S Y, Sun C Y. Distribution of network resources in super intensive residential district based on user clustering. *Journal of Xi'an University of Posts and Telecommunications*, 2016, 26(1): 16-20.
- [12] Zhang J M, Xie W L, Yang F Y. Mobile edge computing technology and its local shunting scheme. *Telecommunications Science*, 2016, 32(7): 132-139.
- [13] Bai L, Liu T T, Yang C Y. Interference coordination method and performance analysis in ultra-dense networks. *Signal Processing*, 2015, 31(10): 1263-1271.
- [14] Chen G, Cai F E, Cheng L. Research on interference coordination technology based on service characteristics in ultra-dense networks. *Telecommunication Engineering Technology and Standardization*, 2016, 29(3): 75-78.

- [15] Zhu X R, Zhu W R. Algorithm of cell clustering and power allocation based on interference coordination in ultra-dense small nest networks. Journal of Electronics and Information, 2016, 38(5): 1173-1178.

5G 密集组网架构中的集中与协同

卫鸿婧¹, 郭 宝¹, 张 阳²

(1. 中国移动通信集团山西有限公司, 山西 太原 030032;

2. 中国移动通信集团公司, 北京 100033)

摘 要: 超密集网络场景指大量的人聚集在有限的区域形成集中的宽带数据业务需求的场景。超密集网络会产生巨大的流量压力, 传统的网络能力远远不足以承载用户的需求。本文在阐述超密集网络组网架构的基础上, 对毫米波无线频谱、高增益的波束赋型、物理层帧结构以及资源集中与边缘计算技术进行了分析, 并对在密集组网架构中必然产生的覆盖与干扰共生所要求的协作技术以及集中接入的接入控制技术进行了综合的分析探讨。

关键词: 超密集网络; 毫米波; 边缘计算; 多点协作

引用格式: WEI Hong-jing, GUO Bao, ZHANG Yang. Centralization and collaboration in 5G ultra-dense network architecture. Journal of Measurement Science and Instrumentation, 2020, 11(1): 70-77. [doi: 10.3969/j.issn.1674-8042.2020.01.009]