Mobile Edge Computing and Field Trial Results for 5G Low Latency Scenario

Jianmin Zhang, Weiliang Xie, Fengyi Yang, Qi Bi
China Telecom Co., Ltd.Technology Innovation Center, China Telecom Beijing Information Sci.&Tech. Innovation Park, Future Sci.&Tech. City, Beiqijia Town, Changping District, Beijing, 102209, China

Abstract: Through enabling the IT and cloud computation capacities at Radio Access Network (RAN), Mobile Edge Computing (MEC) makes it possible to deploy and provide services locally. Therefore, MEC becomes the potential technology to satisfy the requirements of 5G network to a certain extent, due to its functions of services localization, local breakout, caching, computation offloading, network context information exposure, etc. Especially, MEC can decrease the end-to-end latency dramatically through service localization and caching, which is key requirement of 5G low latency scenario. However, the performance of MEC still needs to be evaluated and verified for future deployment. Thus, the concept of MEC is introduced into 5G architecture and analyzed for different 5G scenarios in this paper. Secondly, the evaluation of MEC performance is conducted and analyzed in detail, especially for network end-to-end latency. In addition, some challenges of the MEC are also discussed for future deployment.

Key words: mobile edge computing (MEC); 5G; network architecture; low latency

I. INTRODUCTION

With the advances in mobile Internet and the continuous emergence of various services and applications, the wireless communication has experienced explosive growth over the past decade. Meantime, this rapid growth has triggered the investigation of 5G for future mobile network in both the academic and industrial communities, which mainly includes METIS (Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society), 5GNOW (5th Generation Non-Orthogonal Waveforms for Asynchronous Signaling), 5GPPP (5G Public-Private Partnership), 5G Research Center, 5G Forum, IMT-2020 and 3GPP, etc[1-5].

On the basis of the worldwide research, a broad consensus on the usage scenarios and requirements of the future 5G network was reached, in which the key usage scenarios defined by ITU-R include enhanced mobile broadband, massive Machine Type Communications (MTC) and ultra-reliable and low latency communications[6]. Particularly, the requirements mainly include higher traffic volume, lower latency and huge number of connected devices, etc [7]. For example, it is predicted that the continuous growth of mobile Internet applications and services will trigger 1000-fold mobile traffic demand over the next decade [8]. Meanwhile, there is another prediction of a total of 50 billion connected devices by 2020, these huge number of MTC will replace the communication scenario of human-centric [9]. Besides, the most challenging latency requirement comes from the tactile action of human limbs with visual or audio feedback, which is on the order of ms [10].

In order to fulfill the diverse requirements mentioned above, both radio technologies and network architecture should be taken into account at the same time. For the 5G mobile network architecture, our previous work [11-12] pointed out that service localization, local breakout and caching can decrease the requirement of backhaul bandwidth, thereby supporting the high-capacity hot-zone scenario. Meanwhile, by migrating computational intensive task to network, computation offloading can prolong the battery life of resource-constrained (computation and storage resource) MTC devices in massive MTC scenario. Besides, service localization and caching can also reduce the end-to-end latency by managing some specific user requests directly at the network edge, instead of forwarding all traffic to remote Internet service. In addition, the network context information, especially for RAN context information, should be exposed to third-party application or content providers for user Quality of Experience (QoE) improvement and service innovation.

It can be concluded that future 5G RAN network should support the IT and cloud computation capacities and provide the functions of service localization, local breakout, caching, computation offloading, network context information exposure. Therefore, the concept of mobile edge computing has attracted more attention, which is proposed by ETSI (European Telecommunication Standard Institute) [13-15]. According to the definition, MEC can provide the IT and cloud computation capacities for RAN, through deploying general IT servers at the edge of RAN. However, there are few work related to MEC and its performance still needs to be evaluated and verified. Thus, how to exploit the MEC for future 5G network and evaluate its performance becomes the main focus of our work.

Motivated by the above discussion, this paper firstly introduced the concept of MEC into 5G network architecture and analyzed according to three main usage scenarios, i.e. enhanced mobile broadband scenario, ultra-reliable and low latency communications scenario, and massive MTC scenario. Considering the air-technologies of 5G are still in developing state, the function and performance of MEC are evaluated in the real LTE network and analyzed in detail, especially for network end-to-end latency. Finally, some challenges of the MEC are also discussed for future deployment, including bypassing function, local content charging, security, etc.

II. MOBILE EDGE COMPUTING (MEC) IN 5G NETWORK ARCHITECTURE

2.1 Concept of MEC

As the definition described, MEC can provide the IT and cloud computation capacities for RAN, by deploying general IT servers at the edge of RAN. Compared with the traditional network, whose services are all hosted in remote datacenter, MEC makes it possible to deploy and provide services at the edge of RAN. That is to say MEC has the following advantages:

(1) RAN can offer a service environment with ultra-low latency and high-bandwidth, alleviating the challenges of bandwidth and latency for 5G network.

(2) Service localization, local breakout and caching can efficiently decrease the network load and requirement of backhaul bandwidth, thereby reducing the network operation and maintenance cost.

(3) Service localization makes the services and applications are more closer to RAN and users, then RAN context information (cell ID, location, network load, wireless resource efficiency, etc) can be exposed and exploited

by service and content providers to improve the QoE of users.

(4) Through migrating computational intensive task to MEC server, computation offloading can prolong the battery life of resource-constrained MTC devices.

(5) RAN exposure to third-party partners (content or service provider) can accelerate the applications and services innovation towards mobile subscribers, enterprises and other vertical segments, which can be translated into value.

Due to the advantages mentioned above, MEC can be applied to the various scenarios of mobile Internet and Internet of Things (IoT), which have diverse requirements, e.g. low latency, high bandwidth, location awareness, RAN context information awareness, etc. Therefore, MEC can be referred to as a potential technology for future 5G network architecture and will be analyzed in the following section.

2.2 MEC in 5G network architecture

Based on our previous work related to 5G network architecture [11-12], which consists of three clouds, we further introduce the concept of MEC into the future 5G mobile network architecture and analyze the MEC potential in different scenarios, as illustrated in Fig.1. It should be noticed that only the information related to MEC is presented in Fig.1, due to the limited space and main focus of this paper. More information of our proposed three-cloud 5G mobile network architecture can refer to the previous work.

2.2.1 MEC potential in 5G usage scenarios

1) Enhanced Mobile Broadband Scenario

In order to satisfy the challenges of 1000 times higher traffic volume and 100 times higher user data rate, both the evolution of existing technologies and new radio concepts should be taken into account at the same time, e.g. massive MIMO, millimeter Wave (mmWave), Ultra Dense Network (UDN)[20], etc. The main objective of these technologies is to improve the RAN capacity. However, the explosive growth of traffic volume in RAN also bring the enormous pressure to backhaul of traditional LTE Core Network (CN), which centralizes the data-plane function at the boundary with the Internet and forces all traffic to flow through the P-GW.

Therefore, through enabling the IT and cloud computation capacities at RAN, i.e. MEC, RAN can obtain the capacities of local breakout, services localization, caching,

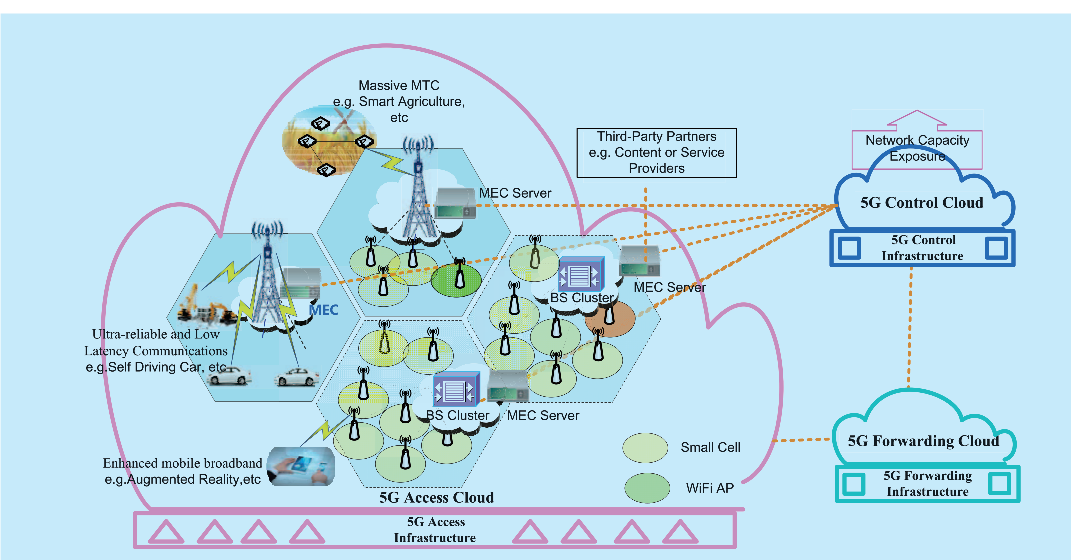


Fig.1 *MEC in 5G network architecture*

computation offloading, network context information exposure, etc. In which, local breakout, service localization and caching can dramatically decrease the requirement of backhaul bandwidth, thereby alleviating the pressure of CN and decreasing the further investment. In other words, local breakout and service localization can be seen as the efficient ways to realize the distributed data plane, where the most functions of control plane is still centralized in the control cloud.

For example, in the enterprise/campus high capacity scenario, through service localization and local breakout, MEC can provide local service and connectivity for enterprise/campus users. Thus, RAN can support virtual Local Area Network (LAN) with high transmission bandwidth and low latency, realizing the motto of MEC, which is *local problems should be solved locally*[15].

2) Ultra-reliable and Low Latency Communications Scenario

Ultra-reliable and low latency communications scenario refers to the scenario, which is extremely sensitive to latency and strict with reliability. e.g. healthcare, security, vehicle-to-vehicle and mission-critical control. In which, the most challenging demand of latency is on the order of 1ms [16]. Regarding to the requirement of network latency, both the radio technologies (e.g. Generalized Frequency-Division Multiplexing, GFDM) and network technologies (e.g. service localization, caching) must be considered simultaneously [17].

As mentioned above, based on the IT and cloud computation capacities deployed at the RAN, some traditional service or content hosted by Internet services or CDNs can be shifted to the network edge. Thus, the local service or extremely popular content can be deployed or cached in close proximity to where they are usually requested. Then the network end-to-end latency can be decreased efficiently.

Therefore, service localization and caching provided by MEC can dramatically decrease the network end-to-end latency, and satisfy the requirement of 5G network latency to a certain extent, which will be evaluated and verified in the following content.

3) Massive Machine Type Communications Scenario

As predicted above, there will be a total of 50 billion connected devices by 2020, these huge number of MTC will replace the communication scenario of human-centric [9]. Besides the challenge of the huge number of on-line connections should be support by 5G network, another challenge is the battery life of these low-cost and resource-constrained (computation and storage resource) MTC devices.

Considering the low processor and storage capacity embedded in many MTC devices, computation offloading can save the process power and energy of MTC devices through migrating computational intensive task to remote datacenter in traditional way, thereby prolonging the battery life. However, offloading computation to the remote public cloud may involve the long latency for data exchange between the public cloud and MTC devices.

Therefore, offloading computation to the edge of RAN, i.e. MEC server, can overcome the above latency problem. In other words, computation offloading supported by MEC has the potential to prolong the battery life of these low-cost and resource-constrained MTC devices. Meanwhile, MEC server can also act as local gateway for MTC devices, in which the original data received from MTC devices can be stored locally and implemented immediately, thereby decreasing the requirement of MTC storage capacity and network response time (e.g. monitoring service, etc). In addition, local storage provided by MEC can also decrease the traffic volume transmitted to remote datacenter and the corresponding network load.

4) Other Potential of MEC

Clearly, service localization make the applications and services are more closer to the radio network and user, then real-time network context information (cell ID, location, network load, wireless resource efficiency, etc) can be exploited by application and services to offer context-aware services, which can differentiate the mobile broadband experience and improve the QoE.

In addition, network operator can partly expose the RAN capacities and allowed the radio network edge to be handled by third-party partners, this will allow to rapidly deploy new applications and services to the mobile user and enterprises. Therefore, based on the network context information exposure, MEC has the potential to improve the QoE of users and accelerate the service innovation.

According to the above discussion, it can be seen that the service localization is most important function of MEC, which results in RAN with high transmission bandwidth and low latency. Local breakout as the precondition of service localization should first be realized. Therefore, before the advantage of MEC in terms of network latency, the local breakout function of MEC is firstly evaluated and verified in following content.

2.2.2 Deployment scenarios

As illustrated in Fig.1, through deploying MEC server or MEC function in RAN, RAN is able to support service localization, local breakout, caching, computation offloading, etc. However, the deployment of MEC function or MEC server depends on the usage scenarios, mainly including outdoor macro base station and indoor small cell scenarios.

1) Outdoor macro base station

Considering the capacities of macro base station in terms of computation and storage, the above MEC function can be (partly) embedded into the macro base station. This deployment method is more beneficial to decrease the network latency, improve the utilization of network infrastructure, obtain real-time network context information and support disruptive vertical services (e.g. vehicle-to-vehicle scenario with low latency requirement).

2) Indoor pico base station

Due to the limited coverage and number of serving users of pico base station in indoor scenario, MEC server can take the form of local gateway, where dedicated intelligence serves local purpose. Thus, multiple services can be deployed in this local gateway, e.g. aggregation gateway for IoT scenario, local gateway or application server for enterprise/campus scenario, and big data analytics.

2.2.3 MEC server platform

It can be seen that the key element of MEC is the MEC server, which should base on the general IT server and provide computation resources, storage capacity, connectivity, access to user traffic and network context information. Thus, the function of MEC can be seen as individual software module operating in isolated virtual machine (VM), which can be installed or uninstalled according to the usage scenario of MEC.

Therefore, we propose the detailed MEC server platform based on the high-level overview of MEC server platform in [13], as illustrated in Fig.2. Clearly, MEC server platform mainly consists of three layers, MEC platform infrastructure layer, MEC application platform layer and MEC application layer.

1) MEC Platform Infrastructure Layer:

Providing the physical resource for MEC application platform, e.g. computation, storage and connectivity, etc.

2) MEC Application Platform Layer:

Consisting of the MEC virtualization management of application and application platform services. Where the MEC virtualization management can support a flexible and efficient, multi-tenancy, run-time and hosting

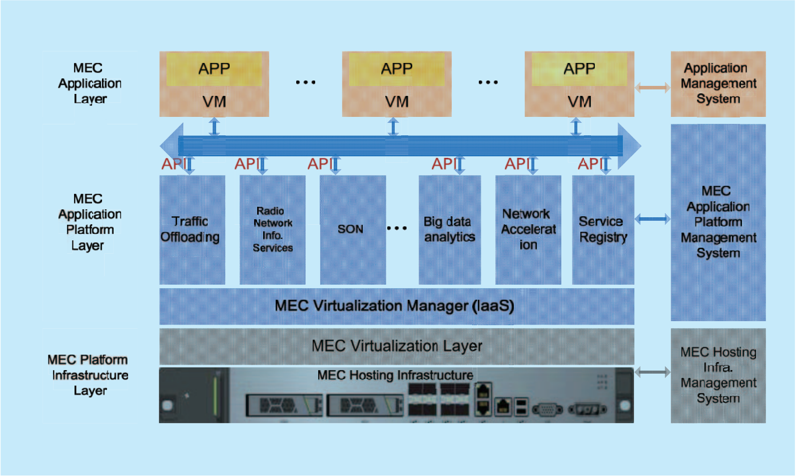


Fig.2 *MEC Server Platform*

environment for application layer by providing Infrastructure as a Service (IaaS). MEC application platform services mainly includes traffic offloading, Radio Network Information Services, Self-Organizing Network (SON), big data analytics, network acceleration, and service registry, etc. These middleware application services and infrastructure services can expose to MEC application layer through standard API interface.

3) MEC Application Layer:

Through combing different MEC middleware application services, MEC application (service localization, local breakout, caching, location, etc) can be formed and exposed to the third-party partners using standard API interface. These MEC applications should be deployed and executed within isolated VMs.

Besides, MEC hosting Infrastructure management system, MEC application platform management system, and application management system can manage the IT physical resource, MEC application platform services and MEC applications, respectively.

Based on the above discussion, MEC has the potential to satisfy the requirements of 5G network to a certain extent, due to its functions of services localization, local breakout, caching, computation offloading, network context information exposure, etc. Especially, MEC can decrease the end-to-end latency dramatically through service localization, which is key requirement of 5G low latency scenario. As described above, we will evaluate and verify the basic function (local breakout) of MEC and potential gain in low latency usage scenario in the next section.

Table I *Network Parameters of Field Trial*

Parameter	Value
System Bandwidth	15MHz
PRB Number	75
Maximum Tx. Power of UE	23dBm
eNB Tx. Power (indoor)	15dBm
Uplink central frequency	1927.5MHz
Downlink central frequency	2117.5MHz
Frequency reuse factor	1
Test Services	FTP, web browsing, ping, etc

III. FIELD TRIAL FOR MEC

In this section, we will present the field trial for MEC from three aspects:

- The influence of MEC on traditional network service performance e.g. downlink/uplink throughput, network end-to-end latency.
- The performance of service localization in terms of downlink/uplink throughput, network end-to-end latency.
- The performance of MEC in terms of network end-to-end latency compared with traditional network.

Considering the 5G system is still in developing state, we conduct and MEC field trial in real LTE network, as illustrated in Fig.3.

3.1 Field trial setup

Fig.3 presents the network topology of MEC field trial in real indoor LTE network. In which, MEC server has the capacity of local breakout according to the destination IP address, and simple local services (e.g. FTP and web servers) are deployed in local network. Meanwhile, based on the rule of local breakout (destination IP address), different services will have different path:

- Local service: UE can access the local network directly, the local traffic is not necessary to flow through the CN.
- Public service: The public service traffic will be flow through the MEC server transparently without any processing.

Notably, although the local and public data-plane service traffic flows (S1-U) have different paths, their control plane traffic (S1-C) is still forced to flow CN without any processing in MEC server. No processing of control plane traffic guarantees the MEC is transparent to user and CN in this environment of field trial.

Besides, the network parameters of Field Trial are indicated in Table I.

3.2 Field trial results

As mentioned above, the MEC server in this field trial environment only offload the data

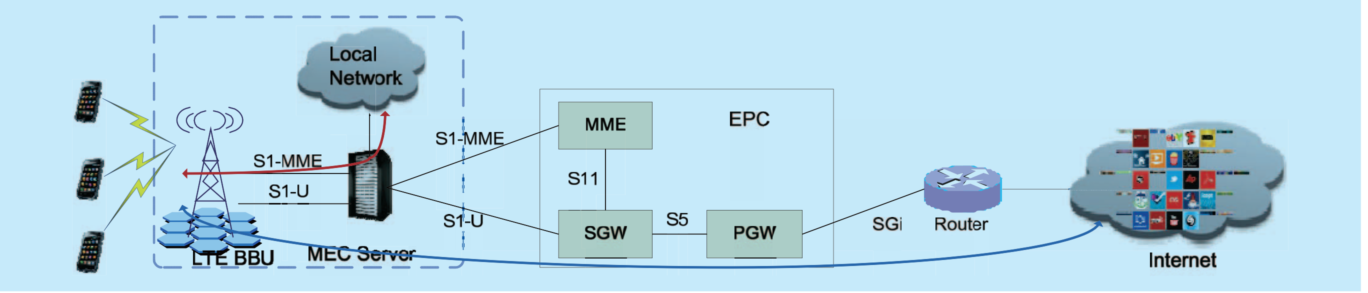


Fig.3 *Network Topology of MEC Field Trial*

plane traffic flow according to the rules of local breakout, the control plane traffic all flow through the MEC server transparently, which guarantees there is no influence of MEC on user registration, de-attach, handover, paging, service request, etc. Before the throughput and latency test, these functions have already been evaluated and verified. Due to the limited space, these simple function test results will not be presented in this paper. Only the throughput and latency test results will be presented and analysis in detail in the following subsection.

3.2.1 Network MAC throughput

In order to evaluate the influence of MEC on network throughput, we conduct the test of downlink and uplink FTP service in three scenarios: public service with FTP server deployed in Internet without MEC server, public service with FTP server deployed in Internet with MEC server, local service with FTP server deployed in local network. During the test, the UE position is fixed to minimize the influence of radio condition. Fig.4 illustrates the field trial results of network MAC throughput, and following conclusion can be obtained:

● Public Service: there is nearly no difference between the scenario with and without MEC server in terms of downlink and uplink MAC throughput. That is to say, the MEC introduced has no influence on network throughput.

● Local Service: Compared with the public service, the local service has no difference in terms of downlink and uplink network throughput.

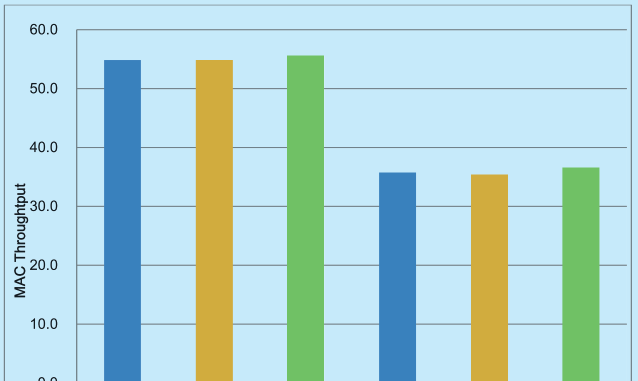
Therefore, it can been say that MEC can provide the local service, connectivity, etc, without influence on any public service in terms of network throughput.

3.2.2 Network end-to-end latency

Regarding to the network end-to-end latency, we adopt the ping test with different package sizes (32byte and 1500byte) in different scenarios with/without MEC server. Besides, three different public services IP address (www.189.cn, www.sina.com.cn, yjt.189.cn) are also pinged for evaluating the delay brought by MEC server and comparing with the network delay of local service. In addition, we allocate the local and public IP address for local server to compare the difference between local access with MEC and tradition access through CN. All the above results are presented in Table II.

Obviously, we can obtain the following conclusion:

● For the public service, the additional delay



	Downlink Public Service (with MEC)	Downlink Public Service (w/o MEC)	Downlink Local Service	Uplink Public Service (with MEC)	Uplink Public Service (w/o MEC)	Uplink Local Service
Mbit/s	54.8	54.8	55.6	35.8	35.3	36.6

Fig.4 *Network MAC Throughput in different scenarios*

Table II *Network End-to-End Latency Field Trial Results in different scenarios*

Package Size	Scenario	Local Server (Local IP)	Local Server (Public IP)	www.189.cn	www.sina.cn	yjt.189.cn	AverageDelay(ms)
32byte	MEC	15	17	39	35	86	-0.5
	w/o MEC	not accessable	15	38	39	87	
	Delay		2	1	-4	-1	
1500byte	MEC	17	18	43	36	91	1.0
	w/o MEC	not accessable	18	40	38	88	
	Delay		0	3	-2	3	
Average Delay brought by MEC Server							0.25

brought by MEC server is about 0.25ms.

● Compared with the delay of public service, the delay of local service can save about 60%~91%, which mainly depends on the distance between application server and local application server.

● For the local server, the delay of local access using local IP address can be saved about 1.5ms compared with traditional access through CN using public IP address, which is the processing delay of P-GW.

Therefore, it can be say that MEC can provide local access with the network delay of less than 17ms, without obvious influence on network delay of public service. In other words, MEC can support the low-latency services, whose requirement is not smaller than 17ms. More stricter requirement of latency should be solved from the new radio technologies or D2D communication, which will be further investigated in our following work.

IV. CHALLENGES

As the potential technology, MEC can satisfy the requirements of 5G network to a certain extent, due to its functions of services localization, local breakout, caching, computation offloading, network context information exposure, etc. Especially, MEC can decrease the end-to-end latency dramatically through service localization. However, there are still many challenges that need to be further investigated, which can be summarized as following:

1) Bypassing function of MEC server

As illustrated in Fig.1 and Fig.3, the MEC server is deployed between the eNB and CN, then the MEC server should support the bypassing function. That is to say, when MEC breakdown occurs (e.g. power error, hardware error, software error, etc), MEC server should switch to the backup connection automatically. In other words, bypassing function aims to avoid the influence of MEC breakdown on network operating. Meanwhile, it can also decrease the complexity of operation and maintenance during MEC updating, debugging, etc.

2) Charging policy of local service

MEC brings the virtual LAN for enterprise/campus users without forcing local traffic flow through the CN, yields the P-GW has no information of local service and cannot provide billing to charging gateway. Therefore, the charging policy of local service should be considered for future deployment.

3) Security problem

As described above, the network context information, especially for RAN context information should be exposed to third-party application or content provider for user Quality of Experience (QoE) improvement and service innovation. However, the network capacity exposure yields the problem of wireless network security and information security, which is an important aspect in our future work.

V. CONCLUSIONS

In this paper, we firstly introduce the concept of MEC into 5G architecture and analyzed for different 5G scenarios. Secondly, the evaluation of MEC performance is conducted and analyzed in detail, especially for local breakout and network end-to-end latency. It can concluded that MEC can support the low-latency services, whose requirement is not smaller than 17ms.

More stricter requirement of latency should be solved from the new radio technologies or D2D communication. In addition, some challenges of the MEC are also discussed for future deployment.

As discussed previously, there are some problems need to be further investigated in our future research work.

ACKNOWLEDGEMENT

This work was supported by the National High Technology Research and Development Program (863) of China (No.2015AA01A701).

References

[1] METIS. Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society. EU 7th Framework Programme project. http://www.metis2020.com

[2] 5GNOW. 5th Generation Non-Orthogonal Waveforms for Asynchronous Signalling. http://www.5gnow.eu

[3] 5G-PPP. The 5G Infrastructure Public Private Partnership. http://5g-ppp.eu/

[4] 5G Forum. http://www.5gforum.org/eng/main/index.php

[5] IMT-2020, 5G, promotion association. http://www.imt-2020.org.cn/zh

[6] FUTURISTIC MOBILE TECHNOLOGIES FORESEE "IMT FOR 2020 AND BEYOND". http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx

[7] ICT-317669 METIS Project, Scenarios, Requirements and KPIs for 5G Mobile and Wireless System, Del.D1.1, 2013,https://www.metis2020.com/documents/deliverables/

[8] Cisco, Global Mobile Data Traffic Forecast Update, 2010-2015 White Paper, Feb. 2011.

[9] Ericsson, More than 50 Billion Connected Devices, White Paper,2011,http://www.ericsson.com/res/docs/whitepapers/wp-50-billions.pdf.

[10] G. P. Fettweis, The Tactile Internet-Applications \& Challenges, accepted for the IEE Vehic.Tech. Mag., to appear 2014.

[11] YANG Fenyi, ZHANG Jianmin, XIE Weiliang, WANG Min, etc. Analysis of 5G Cellular Network Architecture. Telecommunications Science [J]. Vol.31. No.5. May. 2015.

[12] YANG Fengyi, WANG Haining, MEI Chengli, ZHANG Jianmin, etc. A flexible three clouds 5G mobile network architecture based on NFV & SDN[J]. China Communications. Dec. 2015. Vol.12. 121-131.

[13] Patel M, et al. Mobile-Edge Computing Introductory Technical White Paper. White Paper, Mobile-edge Computing (MEC) industry initiative (2014).

[14] Beck M, Werner M, Feld S, and Schimper S. Mobile edge computing: A taxonomy. Proc. of the Sixth International Conference on Advances in Future Internet, 2014.

[15] Nunna S, Kousaridas A, Ibrahim M, Dillinger M, et al. Enabling real-time context-aware collaboration through 5g and mobile edge computing. Information Technology-New Generations (ITNG), 2015 12th International Conference on. April 2015, pp. 601–605.

[16] Fettweis G. P. The Tactile Internet-Applications Challenges. accepted for the IEE Vehic. Tech. Mag., to appear 2014.

[17] Wunder G., Jung P., Kasparick M., and etc. 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications. IEEE Communication Magazine, vol. 52, no. 2, pp. 97-105, Feb. 2014.

[18] Network Functions Virtualisation Introductory White Paper. ETSI NFV ISG. October, 2012.

[19] ONF White Paper, Software-Defined Networking: The New Norm for Networks. ONF. April, 2012.

[20] WU zhouyun, etc. User Association in Heterogeneous Network with Dual Connectivity and Constrained Backhaul [J]. China Communications. Feb. 2016. vol.13. 11-20.

Biographies

Jianmin Zhang, received his B.Sc. and Ph.D degrees from Institute of Information and Communication Engineering, Zhejiang University, Hangzhou, China, in 2006 and 2011, respectively. He is currently working as a senior engineer in China Telecom Technology Innovation Center.

Weiliang Xie, received his B.E. degree and M.E. degree in information science and technology from Nankai university, and the Ph.D. degree in information science and technology from Peking university, China. He is a professorate senior engineer of China Telecom Corporation Limited now.

Fengyi Yang, has been a professoriate senior engineer of China Telecom Corporation Limited. He is the vice president of the China Telecom Technology Innovation Center and the Vice President of the Wireless Communications Technical Commitee of the China Communications Standards Association.

Qi Bi, the president of the Technology Innovation Center and the CTO of Beijing Research Institute of China Telecom. He received his M.S. from Shanghai Jiao Tong University and Ph.D. from Pennsylvania State University. Prior to joining China Telecom in 2010 he worked at Bell Labs for 20 years and became a Bell Labs Fellow in 2002. He has published extensively, and holds 42 U.S. patents and 66 European patents.