A Survey on Mobility Management for MEC-enabled Systems

Mahshid Mehrabi*, Hani Salah*, Frank H. P. Fitzek*†

*Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden, 01062 Dresden, Germany
Email: first.last@tu-dresden.de
† Centre for Tactile Internet with Human-in-the-loop (CeTi), Technische Universität Dresden, 01062 Dresden, Germany

Abstract—**Mobile Edge Computing (MEC) is a new network architecture concept. Aiming to mitigate the root cause of latency as well as long distance between devices and central cloud, MEC brings the computation and storage resources as close as possible to end devices. However, mobility of end devices brings challenging issues for service continuity and Quality of Service (QoS) of these systems. In this paper, the state-of-the-art research efforts towards handling user mobility in a MEC environment for different kinds of services are surveyed. Also, open research problems are discussed.**

Index Terms—**latency, cloud computing, Mobile Edge Computing, mobility**

I. INTRODUCTION

Recent advances in mobile devices in terms of portability and computation capabilities have made them an inseparable part of our life. This has also led to the emergence of new applications such as augmented reality/virtual reality (AR/VR), online gaming and on-device image processing. However, the limited battery life time of mobile devices along with low latency requirements of these applications has raised the need for a new networking paradigm [1], [2].

Mobile Edge Computing (MEC), firstly introduced by the European Telecommunications Standard Institute (ETSI), is a new platform that "provides IT and Cloud computing capabilities within the Radio Access Network (RAN) in close proximity to mobile subscribers" [3]. MEC brings the computation and storage resources from the core network towards the network edge. Therefore it can minimize the latency of communication between mobile devices and cloud servers.

Mobility is an essential part of most MEC applications. Some applications are independent of the state of the users and their previous interactions are not useful for them, while others are particularly related to users' movements. The MEC system needs to support the service continuity and the mobility of application as the consequences of user mobility [4]. Figure 1 shows user mobility within MEC network and the movement of a user within communication range of one small cell base station towards the second one. This movement not only brings issues for the connection between user and edge server, but also for the Device to Device (D2D) links between the adjacent devices, and thus for the services that the user receives from edge server or other devices.

We survey the most notable current research efforts towards mobility management in different aspects of MEC systems and finally conclusion and open issues are stated.

II. MOBILITY MANAGEMENT FOR EDGE OFFLOADING AND CONTENT CACHING SCENARIOS

Mobility is one key challenging topic in MEC systems, which have effects on decisions in several domains such as caching, computation offloading and connected vehicles. The parameters that measure the efficiency of mobility management algorithms are time delay including both computation and communication, the energy consumption of devices, throughput and communication reliability. These are the parameters that have been analyzed in most research papers in the next section.

A. Caching

The upcoming increase in demand for mobile data traffic, driven by video streaming, will saturate the backhaul capacity of networks. This issue can be tackled by caching or storing the data at the edge. In the following, we summarize some recent works towards handling mobility for caching in mobile edge cloud environment.

In [5] the first mobility-aware coded caching scheme for video delivery in MEC-enabled Small Cell Networks is proposed to optimize the throughput. This method is based on a trade-off among the user mobility, content diversity, and channel selection. User mobility is modeled by discrete random jumps [6]. Heuristic algorithms are deployed to numerically optimize the solutions due to the complexity of throughput expression. To further consider user's mobility and social behaviors for MEC-enabled small cell networks, Liu et al. [7] proposed a context-aware caching method. The aim is maximizing the offloaded backhaul data, which integrates contextual information into a new model that characterizes the influence of social networks on users' behaviors based on a multi-armed bandit algorithm. This information include user mobility patterns, preference towards demanded files and the consumed traffic amount per time.

The authors in [8] introduced a novel model to optimize content migration, which are located at the edge of cloudified mobile network. This end-to-end architecture fully exploits Information-Centric Networking (ICN) and the Mobile Follow-Me Cloud approaches. There are several criteria to
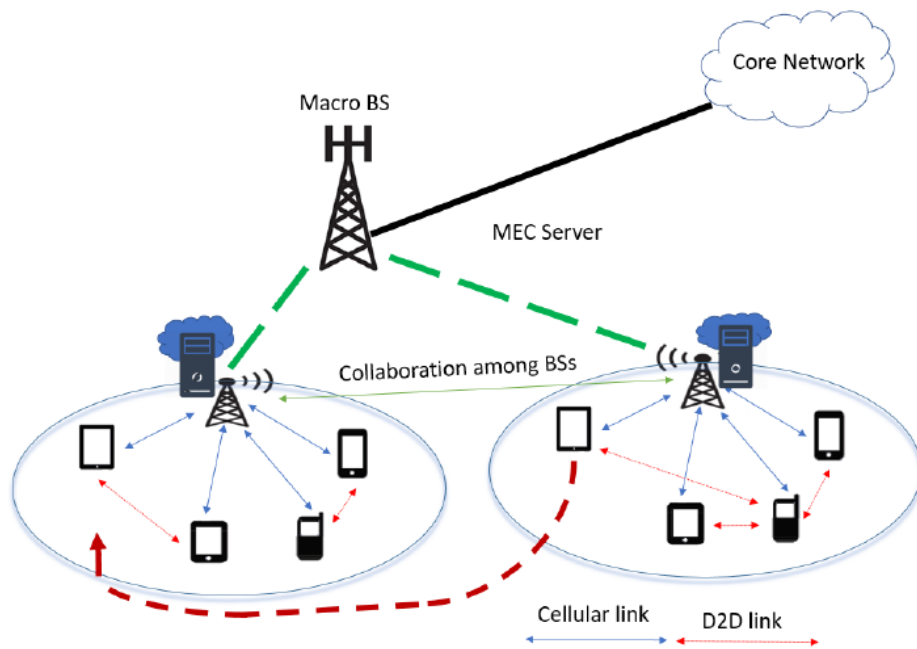
Figure 1. User mobility.

optimize the problem of content migration such as content popularity, content size and the caching capacity of different routers. Multiple Attribute Decision Making (MADM) algorithms are deployed to tackle this issue. Since the implementation of actual user traces is not available, they used a data set for mobility prediction, collected by Nokia for academic research during the Nokia Mobile Data Challenge (NMDC) [9].

As the authors in [10] discussed, in order to relieve the burden on backhaul links and enhance spectrum efficiency when traffic grows, the data storage that exists at users' side can be used to help the MEC systems. A computation as well as traffic offloading schemes in cache-aided device-to-device multicast networks are proposed. They firstly propose a cluster head selection scheme, which jointly considers the social behaviors, existing energy, and transfer rate of D2D users in order to provide stable links and increase computing resources. Then, a novel multicast-aware coded and cooperative caching scheme is proposed to enhance the efficiency of content distribution and the energy consumption of content delivery. Finally, a model optimized for computation offloading is formulated based on D2D users association, power allocation for D2D cluster head transmission using uplink full duplex and MEC computation resource scheduling.

Previous research on caching strategies have not considered mobility and industrial features together. In [11], a three-layer cache architecture based on edge computing and traditional networks is presented to tackle this issue. Despite traditional networks with random movements of devices, a fixed path

must be followed to complete a specific task. This architecture is based on clustered networks due to their higher efficiency for real-time data delivery in industrial domains. Since edge cloud servers cannot provide continuous wireless link service, to store fragmented parts of big data, sensor nodes are employed; therefore, to guarantee real time, moving nodes can fetch data from these sensors. This proactive caching strategy is proposed by leveraging the sojourn time, capacity of edge cloud servers and neighboring nodes. In addition, a distributed Hungarian algorithm is employed to solve data fetching problem.

To address the mobility issue of both the content caching nodes and end devices, the authors in [12] designed a new cooperative edge caching architecture for 5G networks, where content tasks are shared with base stations by means of smart vehicles as collaborative caching agents. To increase the storage resource utilization of the edge nodes, contents are compressed in MEC servers. The caching and computing resources of base stations and smart vehicles are jointly optimized by this method.

B. Computation Offloading

Computation offloading is one key use case for MEC systems. In order to fulfil the low latency requirements of applications and to save the battery life time of the devices, computation-intensive tasks can be offloaded to the more powerful edge servers. However, the mobility of users with frequent handovers between base stations can negate the benefit of offloading; therefore this should be considered in offloading schemes to satisfy the low latency demands.

Traditional approaches for computational offloading lead to high latency since the the data task is only fetched after completion of the handover process. In order to solve this issue, the authors in [13] suggested an online prefetching method based on task-level computation and prediction of users' trajectory, referred to as live prefetching. This means that the data transport time is minimized by prediction and prefetching data that is part of a possible future task, while the current task is being processed. To minimize mobile energy consumption, an optimal prefetching policy has been designed using a stochastic optimization method, which can select possible tasks for prefetching and control their data sizes.

The authors in [14] proposed a lightweight algorithm that jointly considers task information, small base station and the user mobility information to reduce execution delay. To make offloading decisions more accurate, a delay estimation method is employed. The problem is then formulated as a constraint satisfaction problem and is solved by utilizing a lightweight heuristic algorithm. Minimizing the energy consumption together with satisfying the delay requirements, a novel dynamic mobility-aware algorithm for partial offloading is proposed in [15] by exploiting short-term mobility prediction. Tthe time to next handover and users' movement are predicte. Then, a data size is assigned to each time slot to fulfill the energy consumption and delay reduction goals.

An online user-centric mobility management scheme is proposed in [16] to maximize the edge computation performance while keeping the energy consumption of user's communication low by using Lyapunov optimization and multi-armed bandits theories. The authors considered letting users make mobility decisions instead of using traditional centralistic approaches, like using base stations or Evolved Packet Core (EPC) as the decision maker. The method depends on the BS-side state information for users and whether there is full state information in each period. It is developed in a way that future system state information are not required. This approach is extended in [17] to minimize the total delay of computation migration and communication while satisfying the long-term energy consumption constraint of the user. However, the authors only considered a single-user service placement case and neglected the more real multi-user scenario.

Beside energy consumption and delay, the MEC system performance can be further improved by utilizing the recent developments of social networks and energy harvesting methods. In [18], the authors proposed a dynamic offloading scheme leveraging a game-theoretic approach by considering these two factors aimed to minimize the social group execution cost. To model delay and energy cost, three different queue models are applied at mobile devices, fog node and central cloud. Figure 2 shows the system model for utilizing social relationships among devices into the design of computational offloading. In this kind of systems, the architecture is divided into two layers, namely social and physical layer, and there is one corresponding end device per user in the physical layer. Users' behaviors in social networks such as Facebook, Instagram and Twitter affects the real social connections among

them. To achieve a better decision for computation offloading or content caching, both social ties and the physical locations of users should be exploited [19].
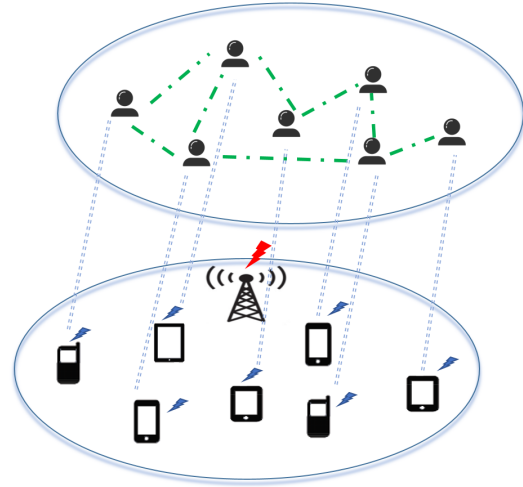

Figure 2. Utilizing social relationships among devices.

To consider the mobility of users in fog computing networks with limited cellular service coverage of fog computing nodes, a generic three-layer architecture is designed in [20]. This architecture jointly optimizes the offloading decisions and computation resource allocation. The problem is divided into two parts, and then, a Gini coefficient-based fog computing nodes selection algorithm (GCFSA) and a distributed resource optimization algorithm based on genetic algorithms (ROAGA) are utilized to solve the offloading decision and resource allocation issues respectively. The mobility of users is represented by the sojourn time, which follows the exponential distribution and is handled by reduction of the migration probability.

III. CONNECTED VEHICLES

Nowadays, vehicles have become a considerable part of people's daily life and connected devices to the network. Due to the powerful communication and computation units, vehicles can be important agents to control traffic or other interactive services [21]. In such a network, edge servers are connected to roadside units (RSUs), which act as gateways to provide wireless access for vehicles from infrastructure and are installed next to the road [22], as shown in Figure 3. Considering the higher speeds of vehicles, mobility management becomes a crucial part to fulfill low latency requirements of applications.

A novel predictive offloading scheme for MEC-based vehicular networks is proposed in [23] based on the mobility of the vehicles and computation task execution time, aiming to improve the transmission efficiency while satisfying the delay constraint of the computation tasks. This method chooses the proper way of offloading the task to the edge server, which means deciding whether the tasks should upload directly or through relay transmissions. It considers various types of
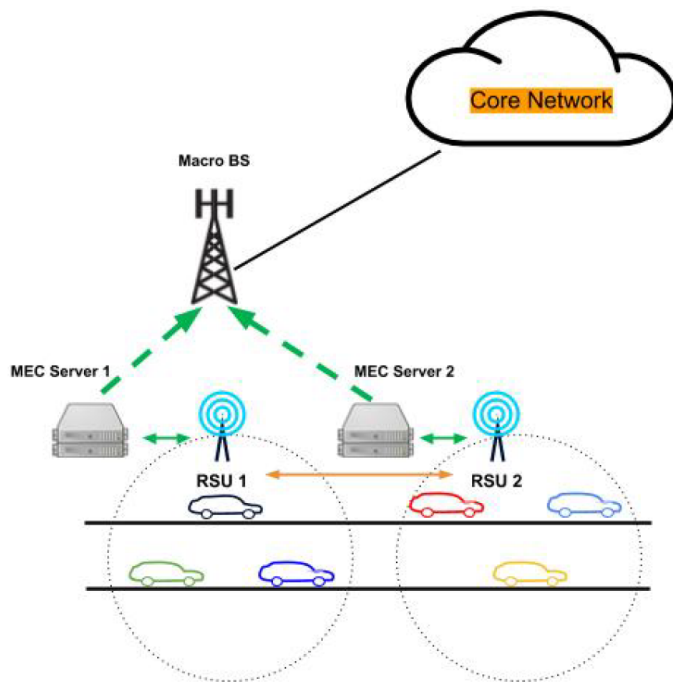
Figure 3. MEC-enabled connected vehicles.

computation tasks and their different requirements and Urban Mobility type has been chosen to simulate the road traffic. However, to simplify the offloading process, the wireless interference and transmission capacity limitation have not been considered.

The authors in [24] proposed MEC-assisted network slicing, which provides vehicle-related services on demand, isolation and a traffic scheduling policy to reduce the latency and enhance service delivery flexibility. To satisfy the high mobility and customization requirements, a smooth inter-cell handover mechanism for vehicles is designed based on the available road information and the enhanced capacity of the edge server.

Considering a practical use case scenario, an software defined networking (SDN) based architecture is presented in [25] to reduce the latency and enhance the communication reliability. MEC technology assists this architecture to offload the traffic load from the backbone network. To obtain reliable Vehicle to Everything (V2X) communication, different types of access technologies have been used. An SDN architecture is proposed in [26] to dynamically manage groups of neighboring vehicles in a 5G environment which separates the entire network into three planes, namely social, data and control plane. To enable vehicle neighbor group networking, universal plug and play (UPnP) standard is exploited. As it is shown in these works, the combination of MEC and SDN provides flexibility for network architecture.

IV. SATELLITE COMMUNICATIONS

Beside using MEC in conventional cellular networks, the mentioned techniques can be also leveraged in high speed satellite communications as a new concept called satellite

mobile edge computing (SMEC) as proposed in [28]. The authors proposed to implement MEC techniques in high-speed satellite-terrestrial networks to improve the quality of service (QoS) of mobile users. To integrate the network resources, a dynamic Network Function Virtualization (NFV) technique is used, where a dynamic resource monitor is added in the orchestrator layer to handle the mobility of Low Earth Orbit (LEO) satellites. Then, a cooperative scheme for computation offloading is proposed to complete the User Equipment task on the SMEC servers. However, this work is more focused on the performance of task scheduling models and more research needs to be done to address the architecture design issues in dynamic scenarios.

V. FINAL COMMENTS AND FUTURE WORKS

In this paper we surveyed the most notable research efforts towards considering user mobility in Mobile Edge Cloud environment as a crucial factor in emerging applications. There are still various open issues and challenges regarding mobility management such as incentives for content caching, group-based mobility [27], and security and privacy issues. Furthermore, most works have considered the theoretical aspects and neglected the practical deployments. Due to the dynamic mobility pattern of end users, one research challenge is designing adaptive methods for caching and offloading. Employing machine learning algorithms seems to be a necessary element to enable usage in real scenarios. In addition to user mobility, the edge servers can be also mobile, therefore the mobility consequences of servers should be also considered [29].

ACKNOWLEDGMENT

REFERENCES

[1] Othman, Mazliza, Sajjad Ahmad Madani, and Samee Ullah Khan. "A survey of mobile cloud computing application models." IEEE communications surveys & tutorials 16.1 (2013): 393-413.

[2] Ahmed, Arif & Ahmed, Ejaz. (2016). A Survey on Mobile Edge Computing. 10.1109/ISCO.2016.7727082.

[3] Patel, Milan, et al. "Mobile-edge computing introductory technical white paper." White Paper, Mobile-edge Computing (MEC) industry initiative (2014): 1089-7801.

[4] ETSI, MECISG. "Mobile edge computing (MEC); Technical Requirements GS MEC 002 V1. 1.1." (2016).

[5] Liu, Xinwei, et al. "Mobility-aware coded probabilistic caching scheme for MEC-enabled small cell networks." IEEE Access 5 (2017): 17824-17833.

[6] Shin, Seungjae, et al. "Analysis of cell sojourn time in heterogeneous networks with small cells." IEEE Communications Letters 20.4 (2016): 788-791.

[7] Liu, Xinwei, Chuanhao Sun, and Xing Zhang. "Context-Aware Caching with Social Behavior in MEC-Enabled Wireless Cellular Networks." 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2019.

[8] Gomes, Andre S., et al. "Edge caching with mobility prediction in virtualized LTE mobile networks." Future Generation Computer Systems 70 (2017): 148-162.

[9] Laurila, Juha K., et al. The mobile data challenge: Big data for mobile computing research. No. CONF. 2012.

[10] Wang, Dongyu, et al. "On the Design of Computation Offloading in Cache-Aided D2D Multicast Networks." IEEE Access 6 (2018): 63426-63441.

[11] Li, Xiaomin, and Jiafu Wan. "Proactive caching for edge computing-enabled industrial mobile wireless networks." Future Generation Computer Systems 89 (2018): 89-97.

[12] Zhang, Ke, et al. "Cooperative content caching in 5G networks with mobile edge computing." IEEE Wireless Communications 25.3 (2018): 80-87.

[13] Ko, Seung-Woo, et al. "Live prefetching for mobile computation offloading." IEEE Transactions on Wireless Communications 16.5 (2017): 3057-3071.

[14] Wang, Zi, et al. "User mobility aware task assignment for mobile edge computing." Future Generation Computer Systems 85 (2018): 1-8.

[15] Yu, Fangxiaoqi, Haopeng Chen, and Jinqing Xu. "DMPO: Dynamic mobility-aware partial offloading in mobile edge computing." Future Generation Computer Systems 89 (2018): 722-735.

[16] Xu, Jie, et al. "E 2 M 2: Energy efficient mobility management in dense small cells with mobile edge computing." 2017 IEEE International Conference on Communications (ICC). IEEE, 2017.

[17] Sun, Yuxuan, Sheng Zhou, and Jie Xu. "EMM: Energy-aware mobility management for mobile edge computing in ultra dense networks." IEEE Journal on Selected Areas in Communications 35.11 (2017): 2637-2646.

[18] Liu, Liqing, Zheng Chang, and Xijuan Guo. "Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices." IEEE Internet of Things Journal 5.3 (2018): 1869-1879.

[19] Xu, Chen, et al. "Social network-based content delivery in device-to-device underlay cellular networks using matching theory." IEEE Access 5 (2016): 924-937.

[20] Wang, Dongyu, et al. "Mobility-Aware Task Offloading and Migration Schemes in Fog Computing Networks." IEEE Access 7 (2019): 43356-43368.

[21] Y. He, M. Chen, B. Ge, and M. Guizani, "On WiFi offloading in heterogeneous networks: Various incentives and trade-off strategies," Commun. Surveys Tuts., vol. 18, no. 4, pp. 2345–2385, Apr. 2016.

[22] Salahuddin, Mohammad Ali, Ala Al-Fuqaha, and Mohsen Guizani. "Software-defined networking for rsu clouds in support of the internet of vehicles." IEEE Internet of Things journal 2.2 (2014): 133-144.

[23] Zhang, Ke, et al. "Mobile-edge computing for vehicular networks: A promising network paradigm with predictive off-loading." IEEE Vehicular Technology Magazine 12.2 (2017): 36-44.

[24] Li, Liang, Yunzhou Li, and Ronghui Hou. "A novel mobile edge computing-based architecture for future cellular vehicular networks." 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017.

[25] Liu, Jianqi, et al. "A scalable and quick-response software defined vehicular network assisted by mobile edge computing." IEEE Communications Magazine 55.7 (2017): 94-100.

[26] Huang, Xumin, et al. "Exploring mobile edge computing for 5G-enabled software defined vehicular networks." IEEE Wireless Communications 24.6 (2017): 55-63.

[27] Wang, Chenmeng, et al. "Integration of networking, caching, and computing in wireless systems: A survey, some research issues, and challenges." IEEE Communications Surveys & Tutorials 20.1 (2017): 7-38.

[28] Zhang, Zhenjiang, Wenyu Zhang, and Fan-Hsun Tseng. "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques." IEEE Network 33.1 (2019): 70-76.

[29] Grewe, Dennis, et al. "Information-centric mobile edge computing for connected vehicle environments: Challenges and research directions." Proceedings of the Workshop on Mobile Edge Communications. ACM, 2017.

[30] Rodriguez, J., Radwan, A., Barbosa, C., Fitzek, F. H., Abd-Alhameed, R. A., Noras, J. M., ... & Rayit, A. (2017, September). SECRET—Secure network coding for reduced energy next generation mobile small cells: A european training network in wireless communications and networking for 5G. In 2017 Internet Technologies and Applications (ITA) (pp. 329-333). IEEE.