Received November 3, 2019, accepted November 9, 2019, date of publication November 12, 2019,
date of current version November 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953172

Device-Enhanced MEC: Multi-Access Edge Computing (MEC) Aided by End Device Computation and Caching: A Survey

MAHSHID MEHRABI[1], DONGHO YOU[1], VINCENT LATZKO[1], HANI SALAH[1], MARTIN REISSLEIN[2,3], (Fellow, IEEE), AND FRANK H. P. FITZEK[1,3]

[1]Deutsche Telekom Chair, Technische Universität Dresden, 01062 Dresden, Germany
[2]School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287-5706, USA
[3]Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, 01062 Dresden, Germany

Corresponding author: Martin Reisslein (reisslein@asu.edu)

This work was supported in part by the European Union H2020 Research and Innovation Program under Grant H2020-MCSA-ITN-2016-SECRET 722424 [1] and in part by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) as a part of the Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden.

ABSTRACT Multi-access edge computing (MEC) has recently been proposed to aid mobile end devices in providing compute- and data-intensive services with low latency. Growing service demands by the end devices may overwhelm MEC installations, while cost constraints limit the increases of the installed MEC computing and data storage capacities. At the same time, the ever increasing computation capabilities and storage capacities of mobile end devices are valuable resources that can be utilized to enhance the MEC. This article comprehensively surveys the topic area of device-enhanced MEC, i.e., mechanisms that jointly utilize the resources of the community of end devices and the installed MEC to provide services to end devices. We classify the device-enhanced MEC mechanisms into mechanisms for computation offloading and mechanisms for caching. We further subclassify the offloading and caching mechanisms according to the targeted performance goals, which include throughput maximization, latency minimization, energy conservation, utility maximization, and enhanced security. We identify the main limitations of the existing device-enhanced MEC mechanisms and outline future research directions.

INDEX TERMS Caching, computation offloading, device-to-device (D2D) communication, mobile edge computing (MEC).

I. INTRODUCTION

A. MOTIVATION

The Multi-access Edge Computing (MEC) paradigm, which is also known as Mobile Edge Computing, has been introduced to bring computing and storage resources in close physical proximity of the wireless end devices [2], [3]. For instance, MEC resources can be co-located with the base stations (BSs) or backhaul entities of cellular wireless communications [4], as illustrated in the left half of Figure 1. The MEC thus helps to provide low-latency services requiring intensive computations or large data volumes to mobile wireless end devices [5]–[7]. The number of wireless end devices, such as user equipment (UE) nodes in cellular wireless networks, is expected to further grow and substantially contribute to the overall Internet traffic growth [8]. Also, the computing and data demands of the wireless end devices are projected to grow substantially over the coming years. This growth is in part due to newly emerging service paradigms, such as the Tactile Internet [9] requiring millisecond latency responsiveness, e.g., for robotic control applications, the Internet of Things (IoT) [10] connecting enormous numbers of devices, Machine-Type-Communications (MTC) [11], online gaming, as well as virtual or augmented reality. The increasing computing and data demands due to these emerging service paradigms which will likely be utilized by large numbers of wireless end devices may overwhelm the installed MEC computing and storage infrastructure. Moreover, cost pressures in the telecommunication industry may limit the installation of higher and higher MEC compute and storage capacities.

The associate editor coordinating the review of this manuscript and approving it for publication was Huan Zhou.

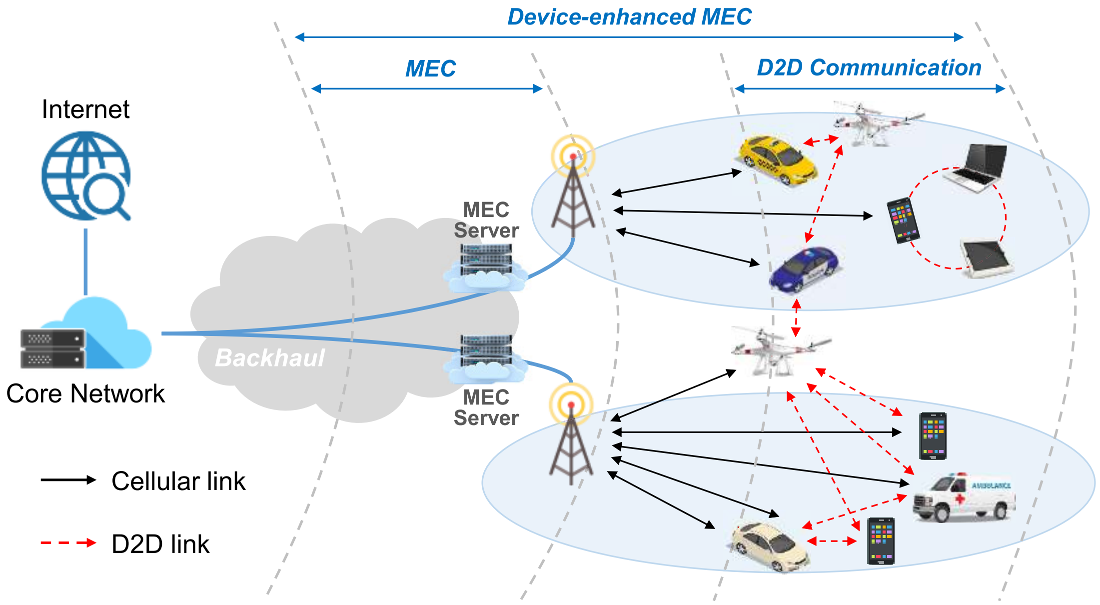

FIGURE 1. Illustration of device-enhanced MEC: The conventional MEC infrastructure extends to the base stations (BSs, cell towers), e.g., through compute and storage capacities installed at the BSs and along the network path from the BSs via the backhaul network towards the core network and Internet. This survey covers device-enhanced MEC where the mobile end devices with their computing and storage capacities collaborate with the conventional MEC infrastructure to provide services to other mobile end devices. End devices reach the MEC resources through cellular communication links and the other nearby end devices through D2D communication links.

A possible solution to this dilemma is to utilize the increasingly powerful processing units, e.g., central processing units (CPUs) and special-purpose processing units, and increasing storage capacities of modern wireless end devices for providing services. That is, the community of wireless end devices, which is also referred to as mobile device cloud, can contribute its aggregate computing and storage resources to provide services to individual end devices *jointly* with the MEC. Effectively, the end devices share their resources and collaborate with the MEC to quickly provide compute- and data-intensive services to their fellow end devices. This sharing among end devices is facilitated by recent advances in Device-to-Device (D2D) communication [12]–[15]. D2D communication enables an end device to exploit the resources of the end devices in its proximity via direct D2D connections, as illustrated by the red links in the right half of Figure 1; thus, potentially reducing the traffic load on the cellular network and MEC infrastructure.

The collaboration of (*i*) the MEC, which has installed resources up to the BSs, with (*ii*) the sharing of resources among end devices, which is enabled through D2D communication, gives rise to the paradigm of *device-enhanced MEC*. As illustrated in Figure 1, device-enhanced MEC encompasses conventional MEC and D2D communication enabled end device resource sharing and thus extends across the entire scope of Figure 1.

B. RELATED SURVEYS ON MEC AND D2D COMMUNICATION

This section gives an overview of the existing surveys on the related topics of MEC and D2D communication. To the best of our knowledge, the present survey is the first to comprehensively cover the topic area of device-enhanced MEC, which builds on and combines the MEC and D2D communication concepts.

MEC surveys have covered mechanisms for offloading compute- and data-intensive service provisioning from the end devices to the installed MEC server infrastructure; the offloading to other end devices has not been considered. In particular, the existing surveys have approached the MEC topic area from a variety of perspectives, including applications and use cases [16]–[20], opportunities and challenges [21], [22], security threats and mechanisms [23]–[25], computation offloading [26]–[30], caching [31], [32], communication perspective [33], [34], service migration [35], architecture and orchestration [36]–[40], technological developments [41], and edge computing for IoT [42]–[45]. Closely related to the MEC surveys are surveys on fog computing. Fog computing generally considers a slightly wider set of devices than MEC for providing computing and storage resources, i.e., fog computing typically encompasses switches, routers, access points (APs), BSs, as well as dedicated compute and storage nodes [46]–[48]; however, the sharing of end device resources is generally not considered in fog computing. Architectural and algorithmic perspectives of fog computing have been surveyed in [49], [50], network applications and the design of fog computing have been covered in [51], and access control for UEs in fog computing focusing on security aspects has been reviewed in [52].

Several surveys have covered the general principles and mechanisms of D2D communications [53]–[59]. Moreover, specific aspects of D2D communication have been surveyed, namely D2D communications architectures [60], device discovery for D2D communications [61], relay assisted D2D [62], and mobility [63]. The interference management for D2D links has been surveyed in [64], [65], while D2D channel models were covered in [66]. The offloading of cellular network traffic to D2D links has been surveyed in [67]. D2D communication for 5G wireless networks has been the focus of [14], while security aspects were covered in [68], [69]. Recently, the relationships between social networks and D2D communications have been surveyed in [70], [71], while smart city aspects have been covered in [72]. Directly building on the D2D communication links, the so-called mobile ad-hoc cloud paradigm [73]–[76] supports the service computing and caching for a given end device through the neighboring end devices (without the involvement of an installed MEC and possibly also without the involvement of BSs or APs). The scope of this survey article is the area of device-enhanced MEC, i.e., we do not survey mobile ad hoc cloud studies without the involvement of MEC servers. Instead, we comprehensively survey device-enhanced MEC studies that collaboratively involve MEC servers and resources at other end devices (i.e., in a sense the local D2D communication ad hoc cloud) for providing services to a given end device.

To the best of our knowledge, only the conference paper [77] and the computing oriented survey article [78] have provided an overview of device-enhanced MEC, which lies at the intersection of MEC and D2D communications. Specifically, the conference paper [77] gives a general overview of the concept of device-enhanced MEC, covering application scenarios, benefits of proximity, user incentives, as well as the challenges of D2D communications and Quality of Service (QoS). Specific mechanisms and individual studies on device-enhanced MEC have not been discussed in [77]. In contrast to [77], we provide a comprehensive up-to-date survey of the existing mechanisms and studies on device-enhanced MEC. The survey article [78] describes in detail the various forms of cloud computing, including forms where resources are shared among end devices. However, the survey only reviews mechanisms where either MEC resources or resources from other mobile devices are utilized. In contrast, we focus on mechanisms that jointly utilize MEC and shared end device resources.

C. CONTRIBUTION AND ORGANIZATION OF THIS SURVEY ON DEVICE-ENHANCED MEC

This article provides a comprehensive survey of device-enhanced MEC, i.e., the enhancement of MEC services for a particular end device (or set of end devices) through the resources of other end devices. The resources of the other end devices are reached through direct D2D communication. We provide background on the enabling technologies for device-enhanced MEC, namely conventional MEC and D2D communications in Section II. Our literature search of the device-enhanced MEC area indicated that the existing studies on device-enhanced MEC have focused on enhancing two main MEC services, namely MEC computation offloading and MEC caching.

MEC computation offloading [79] transfers tasks that require high computational and storage resources, such as image or video processing and interactive gaming, from an end device to an MEC server. Computation offloading speeds up the computation process while extending the end device battery life time [80]. The existing mechanisms for device-enhanced MEC computation offloading are summarized in Table 1 and comprehensively surveyed in Section III. The device-enhanced MEC computation offloading studies have exploited the collaboration of MEC computation resources and the computation resources at nearby end devices (reached via D2D communication) to achieve three main objectives: Minimization of the service latency for applications, minimization of the end device energy consumed for computing application requests, and enhancement of the security. A few existing studies have jointly considered the minimization of latency and consumed energy.

MEC caching supports low-latency data-intensive services, such audio and video streaming, to the end devices. MEC caching stores popular content items in caches that are located close to the end devices, e.g., at BSs [81]. Cached content items can be delivered to the end devices without involving distant origin servers, thus reducing the service latency and the traffic load on the network path to the origin servers. The existing mechanisms for device-enhanced MEC caching are summarized in Table 2 and comprehensively surveyed in Section IV. The existing studies have developed and evaluated methods for placing content items at caches at MEC servers and end devices; these MEC and end device caches collaborate in the overall device-enhanced MEC caching systems. The existing studies have also examined the collaborative delivery of the content items from the MEC and end device caches to the requesting end device.

Open challenges and limitation of the surveyed research studies and the resulting future research directions in the area of device-enhanced MEC are outlined in Section V. Section VI concludes this survey article.

II. BACKGROUND ON EDGE COMPUTING AND D2D COMMUNICATION

A. EMERGENCE OF MULTI-ACCESS EDGE COMPUTING (MEC)

The demands of popular applications running on mobile end devices have brought several challenges for network operators. The limited battery lifetimes as well as the limited computational and storage resources of mobile end devices have motivated network operators to modify their existing infrastructures. The Mobile Cloud Computing (MCC) paradigm was introduced to extend cloud computing features to mobile end devices with the aim of centralizing the management of the computational and storage resources in the core network [33], [82]–[84]. The MCC benefits mobile end devices by expanding the available computation and storage resources as well as the flexibility to support multiple platforms. However, the MCC fails to fulfill the low-latency requirements of emerging mobile applications due to the long distances to the devices and the backhaul bandwidth limitations [32]. To tackle this problem, computing and storage resources should be placed as close as possible to the mobile end devices, e.g., by deploying cloud servers inside cellular BSs or APs depending on the network architecture. This trend of deploying cloud servers close to the mobile end devices was initially called Mobile Edge Computing (MEC) and standardized by the European Telecommunications Standards Institute (ETSI) Industry Specification Group (ISG). In order to extend the MEC usage to heterogeneous networks technologies, e.g., WiFi and fixed access, ETSI ISG has renamed Mobile Edge Computing to Multi-access Edge Computing in September 2016 [85], [86].

Compared to the centralized MCC, the MEC paradigm with distributed computing and caching resources being placed in close physical proximity to the mobile end devices, e.g., by placing compute and caching servers at BSs, brings several advantages for future low-latency networking, such as the Tactile Internet and IoT applications with millisecond-scale latency requirements. Besides reducing communication delay as the main goal, the MEC paradigm reduces the backhaul data traffic (compared to sending all UE service requests to the core network) [21], [87], extends the UE battery life times by offloading compute intensive tasks to edge servers [88], and provides real-time information of UE locations and behaviors, which are helpful for enabling context-aware services [5], [6]. Also, the MEC can support the wireless power transfer to mobile end devices [89]–[91].

B. KEY TECHNOLOGIES FOR IMPLEMENTATION OF MEC CONCEPT

To implement the MEC paradigm and make it operational, multiple integrative technologies are involved [92], mainly Software Defined Networking (SDN), Network Function Virtualization (NFV), and Information Centric Networking (ICN), as outlined next.

1) SOFTWARE DEFINED NETWORKING (SDN)

The main idea for introducing SDN was to enable the use of commodity and off-the-shelf hardware to create intelligent networks that are programmable and application aware [93], [94]. This is achieved by separating the control plane, which manages the network, from the data plane, which transfers actual data streams. Key to assuring interoperability between various equipment manufacturers and vendors is a well-defined open interface between the two planes. Logically centralized SDN controllers help to solve classical networking problems, such as routing, tunneling, and IP address translation, as well as new challenges in future 5G applications, such as UE mobility, adaptation to service degradation, as well as security and integrated protection for IoT systems [44], [95]. Through SDN, network traffic flows can be flexibly steered to and from the MEC [96], [97] so as to seamlessly integrate MEC computations and caching into the provisioning of network services to mobile applications.

2) NETWORK FUNCTION VIRTUALIZATION (NFV)

NFV leverages virtualization techniques to enable the flexible design, deployment, and management of network functions, independent of the underlying physical network equipment [98]–[100]. These network functions may include classical functions, such as firewalls, deep packet inspection, the elements of the Evolved Packet Core (EPC), which is a framework to provide converged voice and data on LTE networks, but also innovative functions, including network coding, data aggregation, or computation as a service. An intuitive extension of the NFV concept combines single virtual network functions in a sequence to modularize complex functionalities in so-called Service Function Chains (SFCs) [101]–[104].

3) INFORMATION CENTRIC NETWORKING (ICN)

The Internet, which was originally designed for host-to-host communications, is mainly used today for content distribution. The Information Centric Networking (ICN) paradigm aims to narrow the gap between the Internet's original design and the current applications, such as high-definition video on-demand streams, 3D gaming, as well as augmented and virtual reality, with ever increasing traffic volumes. In order to optimize caching and content distribution, ICN proposes to redesign the Internet architecture as a content-centric network which adopts two design concepts, namely networking named contents (rather than hosts) and in-network caching, e.g., at MEC servers, to relieve the pressure on bandwidth as well as to improve data delivery [105]–[108].

C. DEVICE-TO-DEVICE (D2D) COMMUNICATION

The exponential growth of mobile data traffic and context-aware applications require innovative approaches to utilize the bandwidth more efficiently and to increase coverage, while lowering delay and energy consumption. The star topology of cellular networks with a centralized control point, e.g., a BS or AP, suffers inefficiencies as all communication has to be relayed by the centralized control point. In contrast, D2D communication is a radio technology that enables direct data exchanges between two adjacent UEs without the involvement of the central control point or core network of the cellular network, i.e., without traversing the BS or AP [12]–[15]. This direct D2D communication brings several benefits, such as improved spectral efficiency, increased data rates between devices, reduced power consumption, and reduced end-to-end delay. D2D communication has been employed in several studies for computation offloading to other nearby UEs (while not utilizing any MEC resources), e.g., [109]–[115]. Also, accessing caches at nearby UEs (while not utilizing MEC caches) has been considered in prior studies, e.g., [116]–[123], whereby specifically video file caching at other UEs has been considered in [124], [125].

However, D2D communication also poses some implementation challenges. One challenge is the need to collect precise channel information, e.g., for estimating the channel and controlling the communication, which adds overhead. Security is another important challenge in D2D communication. Since a UE's data passes through other UEs, D2D communication is inherently susceptible to security attacks. Selfish exploitative UE behavior is another obstacle for collaborative multi-device D2D communication, as some UEs may use the communication resources of other UEs, e.g., for multi-hop D2D communication via intermediate relay UEs, without contributing their own resources to aid others. Interference and mobility management are also key challenges. Therefore, these D2D communication challenges need to be carefully considered when designing device-enhanced MEC systems that involve D2D communication.

Despite these challenges, D2D communication holds significant promise for a wide range of practical use case scenarios in future communication systems. We proceed to briefly outline a few example use-case scenarios.

- National security and public safety:
  The reliance of cellular wireless communication on the availability of the cellular network infrastructure gives rise to severe problems in emergency and disaster scenarios, such as earthquakes and floods. Such disasters often damage the cellular network infrastructure, disrupting cellular wireless communication. In contrast, D2D communication does not require a fixed installed infrastructure and thus can continue to operate when the cellular network infrastructure is damaged. This advantage has made direct D2D communication a key component in projects proposed for next-generation national security and public safety networks by the U.S. National Public Safety Telecommunications Council as well as the European Conference of Postal and Telecommunications Administrations [126].
- Proximity and local-based services:
  The growing interest in multiplayer gaming, advertising, and social network services (e.g., Facebook and Instagram) has increased the need for efficient short-range communications to support interactions between nearby people with low latency and battery consumption while supporting high levels of user privacy [127]. D2D communication can facilitate such connections between different machines in close proximity, such as a mobile phone connecting to a PC or other mobile phones to store and share video files and images [128].
- Vehicle-to-Vehicle (V2V) communication:
  Vehicular or V2X communications is another important use case of D2D communication which is divided into three categories such as vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-network (V2N) communication [129]–[132]. Recent significant enhancements in computing and communication platforms as well as sensing capabilities of vehicles have shifted attention towards V2X communication to improve public safety and intelligent transportation system [133], collision avoidance systems [134], as well as the charging of electric vehicles [135].

We note that these outlined use cases and a wide range of other D2D communication use cases have the potential to significantly benefit from jointly exploiting installed MEC computing and caching resources as well as the resources of other nearby mobile end devices, i.e., from device-enhanced MEC. In order to facilitate the further advancement of exploiting device-enhanced MEC through D2D communication, we comprehensively survey in the following two sections the existing research literature on device-enhanced MEC.

III. ENHANCING MEC COMPUTATION OFFLOADING WITH END DEVICES

A. OVERVIEW

With device-enhanced MEC, end devices, such as UEs, can offload tasks that require heavy computations to powerful MEC servers or to nearby UEs in order to fulfill the low-latency demands of applications and extend their battery life time [158]. The offloading to nearby UEs is conducted over D2D communication links, which reduces the load on the cellular network infrastructure and frees up some cellular bandwidth for other usages.

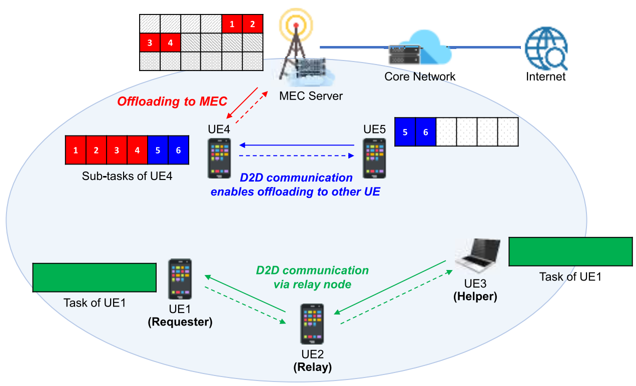

FIGURE 2. Illustration of device-enhanced MEC computation offloading via D2D communication: Partial offloading from UE4 to MEC server and UE5; Binary (full) offloading from UE1 via relay UE2 to helper UE3.

Given the widespread consideration of UEs as end devices in the existing device-enhanced MEC studies, we consider the terms "end device" and "UE" as interchangeable in this article. There are two categories for offloading, depending on whether the tasks can be partitioned or not, namely binary offloading and partial offloading. Binary offloading is employed for tasks that cannot be partitioned. Binary offloading either executes the entire task locally or offloads the entire task to an MEC server or another nearby UE, as illustrated for the offloading from UE1 via UE2 to UE3 in the bottom part of Figure 2. Partial offloading is employed for tasks that can be partitioned into independent parts (sub-tasks) and executed in parallel, either locally or at MEC servers or other nearby UEs, as illustrated in the top part of Figure 2, where UE4 offloads its sub-tasks to an MEC server and UE5. The offloading to

TABLE 1. Summary of device-enhanced MEC computation offloading studies: We categorize the studies according to their main objective, each objective is treated in a subsection of Section III.

Objective	Study	Year	D2D Access Technology	Mobility of UEs	Incentive Mechanism	Evaluation Environment
Latency minimization, Sec. III-B	[136]	2017	Cellular resource	No	No	Simulation using Python and Matlab
	[137]	2018	Cellular resource, TDMA	No	No	Simulation
	[114]	2019	Cellular resource, TDMA	No	No	Simulation
	[138]	2019	–	No	Bandwidth incentive	Simulation
Energy consumption minimization, Sec. III-C	[139]	2018	–	No	No	Simulation
	[140]	2017	Cellular resource	Yes	No	Simulation
	[141]	2018	Cellular resource	No	No	Simulation
	[142]	2019	Cellular resource	No	No	Simulation
	[143]	2016	Cellular resource, OFDMA	Working day movement model	Tit-for-tat	Simulation
	[144]	2019	Cellular resource	No	No	Simulation
	[145]	2019	Cellular resource, FDMA	No	No	Simulation
	[146]	2019	Cellular resource	No	No	Simulation
Joint minimization of latency + energy consumption, Sec. III-D	[147]	2018	Cellular resource	No	No	Simulation
	[148]	2017	Cellular resource	Working day movement model	No	Simulation
	[149]	2017	Cellular resource, OFDMA downlink, SCFDMA uplink	Random direction movement model	No	Simulation using Matlab
	[150]	2019	Cellular resource	No	No	Simulation
	[151]	2017	Near Field Communication	No	No	Android
	[152]	2017	–	Yes	Credit and reputation	Real and lab tests with Android OS
Capacity enhancement, Sec. III-E	[153]	2019	Cellular resource, OFDMA	No	No	Simulation
Security enhancement, Sec. III-F	[154]	2015	WiFi and Bluetooth	Real world mobility traces	No	Android
	[155]	2018	Cellular or WiFi resource	Random waypoint model	Reward	Simulation
	[156]	2018	Cellular resource	No	No	Simulation
	[157]	2019	Cellular resource	Real user mobility trace	No	Simulation

other UEs exploits the idle resources of nearby UEs via D2D communication, which can significantly improve the service to UEs [115], [139]. End devices can generally play three distinct roles in device-enhanced MEC:

- Helper node: A helper node computes offloaded tasks on behalf of UEs that require computation services.
- Relay node: A relay node helps other UEs through communication in order to offload their computation tasks to nearby devices or an MEC server for remote execution.
- Helper and relay node: A device can act as both helper and relay in order to execute and communicate offloaded tasks.

The main objectives of the existing device-enhanced MEC computation offloading studies have been the minimization of the latency and the energy consumption through the optimization of communication and computation resources. We organize our survey according to the main objective of the existing device-enhanced MEC computation offloading studies, as summarized in Table 1. As Table 1 indicates, several studies have considered the joint minimization of latency and consumed energy, while some studies have focused on enhancing the security aspects of device-enhanced MEC computation offloading. The D2D access technology column in Table 1 gives the type of frequency resources considered for the D2D communication links in the studies, as well as the channel access method if a specific channel access method is considered in a study. The dash sign '–' indicates that no specific D2D access technology is considered in the study.

B. LATENCY MINIMIZATION

MEC system failures diminish the quality of the service provided to the UEs. MEC server downtimes can incur enormous costs for businesses that rely on MEC server computations. The study [136] proposed two recovery schemes for an MEC server that is overloaded from serving too many computation tasks or for an MEC server that failed. The first scheme offloads the tasks of the overloaded or failed MEC server to available MEC servers within a transfer range. However for situations when there is no available neighboring MEC server, the proposed second scheme uses the UEs that are adjacent to an MEC server as ad-hoc relay nodes in order to provide a connection between the failed MEC server and a new MEC server. The study [136] assumes that an ad-hoc relay node can relay up to three LTE Frequency Division

Duplex (FDD) resource blocks (RBs) at a time. It is shown that the proposed method works well in dense areas. However, the study [136] has only considered the data downlink from the recovery MEC server, while the data uplink to the recovery MEC server has been neglected. The availability of neighboring resources is also not guaranteed by the protection strategies.

Importantly, the study [136] has only considered the UEs as relay nodes towards the new MEC server and ignored the usage of their computation resources. However, it is beneficial in terms of delay to use the available resources in the vicinity. Considering this fact, a joint task assignment and resource allocation for device-enhanced MEC computing has been proposed in [137]. In this study, a UE can offload its computational-heavy tasks to several nearby end devices, such as smart wearable devices, cell phones, tablets, laptops, as well as infrastructure nodes, such as WiFi APs and cellular BSs, as helper nodes. The task assignment is optimized to minimize the latency, subject to UE and helper energy constraints. Each UE can compute a task locally, or offload the task to a helper node for remote execution. The tasks are considered non-partitionable, however parallel execution of independent UE tasks is possible. A time-slotted communication protocol with three phases is developed. Within the three phases of a time slot, the task is offloaded to one of the helper nodes and the computation results are sent back to the UE. The resulting mixed-integer non-linear minimization problem is solved by relaxing the integer task assignment variables, which results in an efficient, albeit suboptimal solution.

The follow-up study [114] reduced the overall latency by considering controllable computation frequencies instead of fixed processing capacities. Nevertheless, there are still some limitations. The UEs and channel condition are considered static; however, in reality UEs are mobile and channels are dynamic. Therefore, UE mobility and dynamic channels should be addressed through adaptive mechanisms in future research. In addition, only TDMA is used due to its ease of implementation; other orthogonal multiple access methods for D2D communications should be examined in future research to improve the system performance.

The design of an incentive mechanism to motivate UEs to share their computation resources is a key factor in device-enhanced MEC computation offloading and has been neglected in the studies surveyed so far. The study [138] presented bandwidth incentives for UEs. The considered system contains one BS and numerous UEs. The UEs are either computing UEs (CUEs), which have computationally intensive tasks, or helper UEs (HUEs), which help by taking over some of the computation sub-tasks. CUEs motivate HUEs to take over some computation sub-tasks as follows. CUEs give some of their available communication bandwidth to the HUEs in exchange for the help with computations. Thus, HUEs essentially trade in some of their computation resources in order to increase their overall communication bandwidths. A CUE can either offload a task to an MEC server using its full available bandwidth or offload a part of the task to the MEC server and the rest to an HUE; thereby lending some of its bandwidth to that HUE. An optimization problem has been formulated to model the decisions for pairing a CUE to a suitable HUE, the task offloading, and the partitioning of the MEC server resources among UEs. The study [138] assumed that each HUE can only assist a single CUE. Also, the specifics of the bandwidth lending process were neglected and UE mobility was not considered.

C. ENERGY CONSUMPTION MINIMIZATION

In order to improve the MEC performance, a joint computation and communication cooperation method has been presented in [139]. The study [139] considers a basic three-node MEC system with two UEs, whereby one UE needs computation resources and the other UE is the helper/relay. Moreover, one AP node is attached to an MEC server. A four-slot protocol is proposed to enable energy-efficient device-enhanced MEC that minimizes the total energy consumption at both UEs, but also considers the UE's latency-constrained computation requirements. UE computation tasks are assumed to be partitionable; thus, a computations task can be partitioned and the different partitions can be executed locally, offloaded to a helper, or offloaded to the MEC server. However, the examined approach does not fully exploit the capacity of the multiple access channel from the multiple UEs to the MEC server. This limits the performance of multi-user MEC systems [159]. Another drawback of this study is the simple evaluation topology, which included only two UEs.

A cellular D2D framework with a massive crowd of devices at the network edge for joint computation and communication resource sharing has been proposed in [140]. The UE energy consumption is minimized by optimizing the task assignment with a graph matching policy, which can achieve good D2D task assignments. However, the energy-efficiency of the D2D clusters is not considered in the study [140], since it mainly deals with the D2D crowd task assignment problem [160]. In addition, in order to make the proposed framework practical, scenarios with changing D2D connections need to be considered in future research. Moreover, to prevent UEs from over-exploiting other UEs and from free-riding behaviors, an incentive mechanism should be added in future research.

The minimization of the energy consumption of computation task offloading in device-enhanced MEC with a large number of UEs poses significant modeling and computational challenges. The two studies [141] and [142] have investigated game-theoretic models for device-enhanced MEC offloading with large UE numbers. More specifically, the study [141] has formulated the offloading decision problem as a sequential game and examined the stable Nash equilibrium for the system. The study [142] has formulated the problem as a non-cooperative strategic potential game [161]. Both studies found that the game-theory based device-enhanced MEC computation offloading reduces the consumed energy compared to computation on only the MEC servers or computation on only the local UEs.

To take the long-term UE incentive constraints into account and avoid free riding behaviors of UEs which may deter other UEs from collaborating, a D2D framework is presented in [143] to minimize the time-average energy consumption with a Lyapunov optimization based online task offloading. UEs can dynamically share their resources, whereby the sharing is controlled by the BS. The BS establishes in-band LTE-direct Orthogonal Frequency-Division Multiple Access (OFDMA) D2D links between UEs (out-of-band links, e.g., Bluetooth, cannot be controlled by the BS). The working day movement model has been used to characterize the UE mobility patterns. This model is based on people's daily life activities, including commuting from home to work, spending time at the work place, and commuting back from work to home. The working day movement model has shown close similarity to real-world mobility measurements [162]. Three types of tasks have been considered, namely pure computations tasks, such as image processing, pure communications tasks, such as file downloading, and hybrid tasks requiring both computation and communication resources, such as video streaming. The evaluation model generates the UE application layer tasks according to a Poisson process to represent the stochastic nature of real-life task generation. Tasks are admitted based on a best-effort first-come-first-serve admission policy. The task admission policy is independent from the scheduling of the task offloading and only operates at the start of a time frame. Interactions between task admission and task offloading should be examined in future research.

The rapid growth of the IoT and fog computing have brought computing devices, which are referred to as fog computing devices, with idle resources close to the UEs. Accessing both the MEC and the fog computing devices can improve energy savings. Towards this goal, an energy-efficient joint computation offloading via cellular networks to the MEC server and via D2D communications to fog computing devices in a 5G network has been presented in [144]. Some UEs are deployed around one MEC server in the considered system. The access technology between UEs and the MEC server is an LTE radio access network. Fog computing devices with idle computing resources near the UEs functions as helpers. In particular, each UE has a fixed fog computing helper device and communicates with its helper through D2D links. Since the helpers have also limited computing resources, three computation task execution models are considered depending on the UE demands for computation resources: local, fog computing device, and MEC server execution. The computation offloading framework has two parts, namely a control plane and a data plane. The control plane includes the controller, which is responsible for offloading decisions according to the network status. The data plane includes the task queue buffer and the task data transmission parts in the UEs. Simulation results have demonstrated that the proposed method is effective; however, several issues, such as communication overhead, synchronization, data recovery overhead, security, and incentive mechanisms, are neglected in the framework.

Advanced energy harvesting techniques to power mobile devices with renewable energy, such as solar and wind energy, can extend the battery life time of devices. A new device-enhanced MEC computing and networking framework called D2D Edge Computing and Networking (D2D-ECN) has been proposed in [145] toward designing a green computation MEC system that exploits advanced energy harvesting techniques. The examined D2D-ECN system includes a BS and some UEs, whereby one UE is called the master and the rest are secondary devices. The master device is the UE with a computation-intensive task. The master device is equipped with energy harvesting elements. The offloading process is divided into successive time slots of the same length. The task assignment decision, CPU frequency adjustment, and power control are accomplished at the beginning of each time slot. The task transmission and computation at the master and secondary UEs fill the total task execution time in each time slot. The communication setting between UEs is based on the LTE-D2D standard with the FDMA protocol for dedicated D2D transmissions. The energy cost model for each time slot includes the energy consumed for task transmission and processing at the master and secondary devices. A system operation cost is defined to give a reward or penalty to the D2D-ECN system. The reward or penalty is a function of the energy consumption and cost for a unit of energy. The joint optimization of the computation offloading and the resource management to reach a good tradeoff between low system operation cost and short task execution time is formulated as a constrained Markov Decision Process (MDP). In order to execute this joint optimization problem, a Q-learning algorithm is employed, which helps to address the stochastic features of harvesting energy and network information. In addition, a low-complexity online Lyapunov optimization based algorithm is developed to tackle the challenges of high dimensionality of the D2D-ECN offloading framework. However, in the D2D-ECN study, the system status is considered static in each time interval, which may not be a realistic assumption for scenarios with high UE mobility. The simple system model with only one BS and one master UE device is another drawback of this study.

Based on recent advances in antenna design, the study [146] has proposed an energy efficient offloading scheme using full duplex (FD) relays. The network consists of one BS and several UEs forming multiple clusters. One UE with FD antennas is selected as the cluster head, referred to as FD-DCH, in each cluster. This FD-DCH acts as a relay between normal UEs in the cluster (DUEs) and the BS. When DUEs send a proportion of their tasks to their associated FD-DCHs, the tasks will be relayed simultaneously to the BS on the same frequency band used for D2D communication. To avoid interference, it is assumed that DUEs and FD-DCHs work on orthogonal spectrums in both uplink and downlink. The cluster head selection algorithm is based on the Chinese Restaurant Process (CRP) [163] and the weighted sum method considers several metrics, such as UEs' social behaviors, energy and storage resources, and the transfer rate from the BS to the UEs. The mobility of UEs, which can change the social attributes and consequently the cluster head selection procedure is neglected in this study.

D. JOINT MINIMIZATION OF LATENCY AND ENERGY CONSUMPTION

A simple scenario to minimize the task execution cost which can jointly consider latency and energy consumption minimization for a system with one BS has been proposed in [147]. The problem is transformed into a computation offloading subproblem and a resource allocation subproblem which are solved by the Kuhn-Munkres algorithm [164] and the Lagrangian dual method, respectively. In [147], UE tasks are considered partitionable and parallel execution at the requesting UE and at an MEC server or helper UE is possible.

The total task execution cost problem is further investigated in [148] with the consideration of users movements using a hybrid offloading framework called HyFog. The cost problem has been defined as the weighted sum of the UE computational time and the UE energy consumption. HyFog chooses between UE task offloading to the MEC or to nearby end devices using D2D communication (cellular D2D or WiFi-direct). The working day movement model has been used as UE mobility pattern. A novel three-layer graph matching algorithm has been developed to represent the choice space consisting of local (UE) task execution, D2D task offloading to nearby UEs, and task offloading to the MEC. The total task execution cost is minimized through problem mapping to a minimum weight matching problem in the three-layer graph and the Edmonds' Blossom algorithm [165]. The study [148] has only focused on spectrum allocation problems. However, the development of mechanisms that overcome the instinctive selfishness of the UEs remains a key challenge. Instinctively, each IoT user typically optimizes its own quality of experience (QoE) individually without following the strategies for optimizing the overall system performance [166].

Some IoT applications require ultra-low latency computation services. However, poor channel conditions between end devices and the MEC server may impede latency-constrained IoT applications. To address this problem, the study [149] proposed a forwarding scheme to improve resource sharing for mission-critical IoT devices which fall under the coverage of neighboring end devices. A greedy example heuristic has been proposed to solve the optimization problem for task allocations [149]. In particular, the tasks are allocated according to two main criteria: the proximity of the devices and the number of tasks that have already been allocated to a given device. The evaluations in [149] demonstrated through simulations that by using D2D communication in this way, lower latency, energy consumption, and traffic load through the network can be achieved and improvements in the cooperation of IoT devices at the edge of the network are possible.

LTE-Direct with OFDMA and Single-Carrier FDMA have been employed for the downlink and uplink D2D communications, respectively. A round robin scheduler divided the RBs equally between the candidate D2D transmissions (with 6 RBs for D2D). This RB division avoided interference. The random direction movement model, which is a variant of the widely used random waypoint model [167], is considered as the mobility model. An interference coordination scheme that reuses parts of the available frequencies could achieve additional performance gains.

The study [150] has proposed an offloading method with frequency reuse for IoT applications. In the studied architecture, UEs send their computation requests to the MEC server. The MEC server determines the offloading destination according to a two-step algorithm. The first step processes delay-sensitive tasks, while the second step processes tasks of UEs with energy restrictions. The offloading problem for delay-sensitive tasks is modeled as a delay-aware adjacency graph, which is solved for a maximum matching with minimum cost with Edmonds' Blossom method [165]. The result specifies whether the computation requests are offloaded through D2D communication to nearby UEs or to the MEC server. The MEC server then conducts an analogous graph-based solution procedure for the remaining requests from UEs with energy limitations and allocates the computation resources of the remaining idle nearby UEs and its own resources. If the MEC server becomes overloaded, it offloads the computation tasks of energy-limited UEs to the central cloud.

Common drawbacks of the preceding studies on the joint minimization of latency and energy is their use of conventional cellular and WiFi technologies for D2D communication as well as their simulation based evaluation. It is important to examine novel D2D communication technologies as well as to examine the effectiveness of an offloading algorithm through real implementations. The study [151] addressed these drawbacks by proposing the first task offloading framework with near field communication (NFC) based D2D communication and a real implementation evaluation. NFC has several advantages over the longer-range Bluetooth and WiFi technologies due to its short communication range, including lower interference, lower energy consumption, and intrinsic security. The proposed framework circumvents some of the limitations of default Android NFC protocols: The NFC-based task offloading enables bidirectional communications between two UEs and makes the task offloading smoother. The performance evaluation in [151] demonstrated that the NFC interface reduces the UE energy consumption and reduces the execution time of the offloaded task, especially for powerful helper devices. Nevertheless, the NFC-based task offloading in [151] has several limitations. First, the data transfer rate of NFC based communications is only 53 kB/s, because the used hardware can transfer only one message per connection; therefore, the framework is not suitable for data-demanding application scenarios. Moreover, the device heterogeneity and the potential of parallel connections using Bluetooth and/or WiFi-direct as well as user mobility should be examined in future research.

Although the study [151] is based on a practical implementation, the study [151] as well as all prior studies on joint latency and energy minimization lack an incentive mechanism. An incentive mechanism is generally required to make the offloading attractive for users in real D2D systems. A generalized offloading scheme with an incentive approach based on credit and reputation to increase the cooperation among UEs via D2D communication has been proposed in [152]. The proposed task offloading system enhances the accessibility of UEs to offloading support and improves their Quality of Service (QoS). The social-characteristics of the UEs [168], [169] are exploited to form offloading communities. An offloading community is formed by a group of UEs that trust each other with offloading tasks. A UE gains points when it shares computation resources with other UEs, stays in a certain location for a longer time, or pre-caches some tasks. A UE loses points when utilizing the community resource pool. In [152], the community assignment is based on the frequencies and durations with which the UEs are detected. This assignment approach requires the activation of the UE discovery interfaces. A learning method for predicting communities can improve the discovery process and save energy [170]. Also, new task process acceleration techniques that exploit multiple devices are an important direction for future research.

E. CAPACITY ENHANCEMENT

The study [153] has examined the maximization of the total computing capacity of a device-enhanced MEC system. In particular, the maximization of the supported number of UEs subject to communication and computation resource constraints is formulated as a mixed integer non-linear program. The program is decoupled into a sub-problem that minimizes the required MEC resources and a sub-problem that maximizes the UE D2D pairings. The simulation results indicate that the developed optimization approach significantly increases the number of supported UEs compared to a pure MEC system when the MEC resources are limited and when the number of UEs is high. The main limitations of the computing capacity maximization study [153] are the lack of consideration of UE energy consumption, UE incentives, and latencies.

F. SECURITY ENHANCEMENT

The HoneyBot security scheme for collaborative offloading using D2D communication in MEC platforms has been presented in [154]. HoneyBot is a novel defense technique for malicious D2D communication. HoneyBot consists of some nodes for detecting and tracking malicious activities in a D2D network. When HoneyBot nodes identify an insider attacker, they isolate the attacker from the network to protect the network. The evaluations in [154] are based on real mobility trace data collected at the Infocom 2006 and Sigcomm 2009 conferences from people who carried experimental devices with Bluetooth contact discovery logs and data communications. The detection, tracking, and isolation phases were evaluated. In the detection phase, the time required for identifying a malicious D2D communication was evaluated as a function of the number and placement of HoneyBot nodes in the network. The results indicated how the number and placement of HoneyBot nodes impact the detection speed. Once a malicious D2D communication has been detected, the tracking phase commences. The tracking overhead, and accuracy were evaluated and found to depend on the position and number of the HoneyBots as well as the number of attacker nodes. The isolation phase was evaluated through the localization accuracy of the malicious node(s).

Security and cooperation incentives are often discussed as separate topics. However, security and cooperation incentives are tightly interwoven in device-enhanced MEC computation offloading. On the one hand, strong cooperation incentives may lead to more cooperation, which may facilitate the propagation of malicious attacks. On the other hand, increased risks of malicious attacks may deter users from cooperating. In order to address this issue, the study [155] has presented a novel mathematical framework which jointly investigates the user incentives and interdependent security risks in D2D offloading, since an attack can be the result of the users' collective decisions on the cooperation for offloading. In this framework, UEs set their participation levels according to a Stackelberg game [171] to maximize their utility. The game model determines the operator's optimum incentive mechanism based on the users' incentives under infection risks. In the system, D2D offloading is employed either when the MEC server computation load is very high or when the network is congested and WiFi-Direct or LTE-Direct can be used for D2D communications. In addition, the well-known Susceptible-Infected-Susceptible (SIS) epidemic model [172]–[175] is used to model the attacks. The network security state, i.e. the fraction of normal UEs, has an effect on the operator's objective function and the UE participation incentives. Therefore, the fraction of compromised UEs, the UE participation levels, and the operator's optimal utility have been investigated through simulations.

To comprehensively enhance security in D2D task offloading frameworks, social networking characteristics and UEs heterogeneity should also be considered. A socially-motivated cooperative approach has been presented in [156] to improve the security level of task offloading by leveraging the social tie structure among UEs. The cooperative approach incorporates the social tie structure with the UE computation and network resource sharing processes. To minimize the overall system task execution overhead, a socially-aware bipartite matching based algorithm is then proposed. The matching based algorithm exploits the similarity of the structure of the worker assignment in the matching algorithm and the examined problem. Since the social community can provide a structure of UEs with stable relationships, it can be further exploited in future research to offload large-size tasks.

The main drawback of the approach in [156] is that the UE mobility has not been considered. However, UE mobility has a very strong impact on the social graph of the UEs, i.e., the UEs' social ties and relationships. To tackle this issue, the study [157] has proposed a dynamic social-aware computation offloading method that jointly minimizes the task computation latency and the UE energy consumption. A dynamic offloading mode selection, which includes local computation, offloading to the MEC, and D2D offloading, has been formulated into an infinite-horizon time-average renewal-reward problem. A Lyapunov optimization based method and a drift-plus-penalty algorithm are used to solve this problem. The evaluations in [157] considered the delay and energy consumption, but did not specifically evaluate security metrics.

G. SUMMARY OF DEVICE-ENHANCED MEC COMPUTATION OFFLOADING

The main objectives of the existing device-enhanced MEC computation offloading studies have been the minimization of the latency and energy consumption of the UEs as well as the enhancement of security.

Most existing studies have considered partitionable and non-partitionable computation tasks, depending on the application scenario. Individual sub-tasks of partitionable tasks or complete non-partitionable tasks can be executed locally (if the UE has sufficient computation resources and the latency of local UE execution can be tolerated), offloaded to adjacent UEs directly via D2D communication or via relays, or offloaded to an MEC server.

The wireless channel characteristics and the UE resource availabilities are generally stochastic and change with time due to UE mobility. Therefore, offloading decisions should be based on the latest status of the system and be computed online. Overall, Lyapunov optimization based algorithms have so far been the predominant optimization tools for tackling the challenges of the high dimensionality of the offloading frameworks. Lyapunov optimization based algorithms can solve the offloading optimization problems with low-complexity online computations based on the current state of the system, as well as the drift-plus-penalty function for stabilizing the queues.

The examined computation offloading decisions, i.e., the task assignments to other UEs or an MEC server, have typically been based on various aspects of the UE and MEC server resources as well as the UE computation demands. Despite the considerable amount of research devoted to task assignment, the proposed approaches are generally oversimplified. In particular, they did typically not consider the dynamics of wireless communication links. Also, the heterogeneous computational capabilities and time-varying availabilities of the computational resources of the end devices and MEC servers have typically only been partially considered. Future research needs to develop practical approaches that optimize the computation offloading (task assignment) while comprehensively considering the wireless network dynamics as well as heterogeneity and dynamic availabilities of the end devices and MEC servers.

The task assignment, i.e., the decision on where to execute a computation task or sub-task, can generally either be made in a distributed manner or a centralized manner (at the BSs). While the centralized control approach is appropriate for small network sizes, a purely centralized approach may become infeasible or inefficient for large-scale networks. This is because the adaptation to the network dynamics requires frequent data collection from the entire network domain and subsequent centralized processing. This centralized processing translates into long signalling delays, large control signaling overhead, and high computational complexity in large-scale networks [176]. The existing research studies that considered distributed task assignments, neglected the network dynamics; thus they cannot be readily applied to device-enhanced MEC systems [137]. Future research needs to explore hybrid decision approaches that delegate some scope of the decision making to local nodes, while slow-timescale global decisions can still be made at a central controller. While such hybrid approaches have begun to be explored for general wireless resource allocation problems [177]–[182], they remain an open research area for device-enhanced MEC computation task offloading.

In order to reach the main goal of efficient device-enhanced MEC computation offloading for real world applications and scenarios, future research needs to further examine the interactions between task admission policies and the scheduling of task offloading as well as effective ways to continuously maintain the offloading service when UEs are mobile. In addition, relying only on orthogonal multiple access technologies, such as TDMA and FDMA, may limit the performance of multiuser MEC systems [159]; hence, there should be more focus on using new channel access technologies that exploit the particular network architecture. Generally, WiFi appears to be the most practical medium access technology for D2D communication between UEs. Nevertheless, emerging physical layer technologies should be evaluated for providing D2D UE communications. Despite a wide range of studies on the design of incentive mechanisms, there is still a pronounced lack of systematic research on participation incentives that consider the interdependent security risks.

The evaluation methodology in most of the existing computation offloading studies is simulation and only few studies have considered practical scenarios. Future research needs to broaden the evaluation to consider mathematical analysis when appropriate to obtain relevant insights through tractable analysis. Also, prototypes of the proposed device-enhanced MEC computation offloading systems should be developed and evaluated through measurements for representative work loads and mobility patterns.

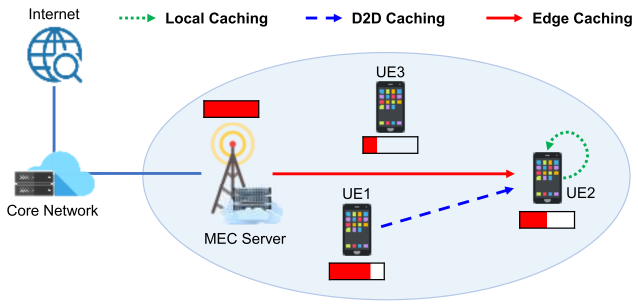

FIGURE 3. Illustration of device-enhanced MEC content caching via D2D communication: The content request at UE2 on the right is served through a combination of locally cached content (dotted green arrow), content cached at the nearby UE1 via D2D communication (dashed blue arrow), and content from the MEC cache server (red arrow).

IV. ENHANCING MEC CONTENT CACHING WITH END DEVICES

A. OVERVIEW

Mobile video streaming and related social networking already account for a large traffic proportion in wireless networks. The forecast continuous growth of this data-intensive traffic will likely overwhelm installed MEC caching resources or require substantial additional investments by wireless operators (or lead to service degradations). Device-enhanced MEC caching exploits the extensive storage capacities in modern wireless end devices to supplement the MEC cache infrastructure. UE requests for data-intensive video streams, web pages, and related social networking applications can be *collaboratively* served by MEC cache servers, the local UE cache, and the caches of other nearby UEs (see Figure 3), which are reached via D2D communication [206], [207]. The caching contributions from the UE caches reduce duplicate content transmissions by the BS, which would result when popular content items are requested by the UEs in the range of a BS at different times. In particular, for social networking applications, exploiting the social relationships among UEs and their common interests using local D2D communication can be a key enabler for pre-caching popular content items in the caches of UEs with rich social ties [203].

This section comprehensively surveys the existing research studies on device-enhanced MEC caching. We have organized the survey according to the main study objectives and then the examined caching aspects, as summarized in Table 2. Generally, there are two main aspects of caching, namely content placement and content delivery. Caching placement studies strive to design methods for optimally storing (placing) the content item files in caches at BSs and UEs. In contrast, content delivery studies focus on the transmission of the requested files to the end devices. There are also a few studies that have jointly examined content placement and content delivery, as indicated in the caching aspect column in Table 2.

B. THROUGHPUT MAXIMIZATION

1) CONTENT PLACEMENT

In order to increase the throughput of video files, two possible strategies for caching popular files in UEs with no additional infrastructure cost have been proposed in [183]. In the first method, the file placement is optimally controlled by the BS as the central controller, which knows the locations of each UE. In the second method, the caching is random without a centralized controller. When a UE demands a video file, the UE first sends the request to the UEs in close proximity, which are defined as a UE cluster. If the content is locally available in the cluster, then a D2D communication link will be established between the two UEs. The cluster size is a key system parameter and is controlled by the UE transmit power. The numerical simulation evaluations in [183] assumed a limited set of statistical models, e.g., for the user distribution and the storage capacities. The performance of the proposed schemes should be evaluated for a wider set of real-world scenarios in future research.

A similar content placement method with mobility-awareness has been studied in [184]. The study [184] has modeled the UE mobility through a contact rate random variable, which characterizes the probability that two UEs are in D2D communication contact. The simulation evaluations in [184] found that the proposed mobility aware caching placement achieves significantly higher throughput of content items served from caches than random caching and content-popularity based caching across a wide range of UE mobility levels. The general concept of a virtual money incentive for providing cached content to other UEs is mentioned, but not examined in detail.

The caching throughput specifically for video files has been further enhanced by a cooperative caching placement based on stochastic geometry in [185]. A main motivation for the study [185] is that in video streaming, a few highly popular video files typically dominate the overall video traffic load. The cooperative caching placement optimization is transformed to an equivalent biconvex optimization problem, which is solved with a block coordinate descent based algorithm [208]. In this system, users requesting content items, BSs, and D2D transmitters are placed according to Homogeneous Poisson Point Processes (HPPPs). The D2D and cellular links use the same spectrum resources. UEs can request a file either from the BS or a D2D transmitter. In the D2D case, files are obtained through one-hop D2D communication. The proposed approach is evaluated for range of Zipf parameters [209], [210] representing different levels of skewness of the video popularity distribution towards a few highly popular video files.

We note that a social trust scheme for video content distribution in a device-enhanced MEC caching system has been examined in [168]. The social trust scheme [168] validates the legitimacy and authenticity of users participating in the cache based video content delivery in systems such as [183], [185]. The trust evaluation is carried out in the MEC server and does not involve UE computations. The proposed social trust evaluation combines direct observations of the interactions with a particular UE that is being examined as well as indirect observations (i.e., observations of UEs that directly interact with the particular examined UE).

TABLE 2. Summary of device-enhanced MEC caching studies: The studies are categorized according to their main objective and each objective category is covered in a subsection of Section IV.

Objective	Study	Year	Caching Aspect	D2D Access Technology	UE Mobility	Incentive Mechanism	Evaluation Environment
Throughput maximization, Sec. IV-B	[183]	2014	Content placement	Cellular res. (reuse)	Random walk	No	Simulation
	[184]	2016	Content placement	Cellular res. (reuse)	Gamma distr. UE contact rate	Payment	Simulation
	[185]	2017	Content placement	Cellular res. (reuse)	No	No	Simulation
	[186]	2018	Content placement	Cellular res. (reuse)	No	No	Simulation
	[187]	2019	Content placement	Cellular res. (reuse)	Random jump	No	Simulation
	[188]	2017	Content delivery	Cellular res. (reuse)	No	No	Simulation in MatLab
	[189]	2019	Content delivery	Cellular res. (reuse)	No	No	Simulation
Energy consumption minimization, Sec. IV-C	[190]	2016	Content delivery	Cellular res. (reuse)	No	No	Simulation
	[191]	2017	Content delivery	Cellular res. (reuse)	No	No	Simulation
	[192]	2019	Content delivery	Cellular res. (reuse)	Yes	No	Simulation
	[146]	2018	Content delivery	Cellular res. (reuse)	No	No	Simulation
Joint optim. throughp. + energy, Sec. IV-D	[193]	2018	Content delivery	Cellular res. (reuse)	No	No	Simulation w. practical param.
Latency minimization, Sec. IV-E	[194]	2017	Content delivery	Cellular res. (dedicated)	No	No	Simulation
	[195]	2018	Content delivery	Cellular res.	No	No	Simulation in MaTLab
Utility maximization, Sec. IV-F	[196]	2019	Content placement	Cellular res. (reuse)	No	Reward or payment	Simulation
	[197]	2018	Content placement	–	No	Yes	Simulation
	[198]	2017	Content placement	Cellular res. (reuse)	Random walk	Reward or credit	Simulation
	[199]	2018	Content delivery	Cellular res.	No	Reward	Simulation
	[200]	2018	Joint content placement & delivery	WiFi res.	Yes	No	Simul. based on Xender experim. results
	[201]	2019	Joint content placement & delivery	–	No	Reward and penalty	Simulation
	[202]	2019	Task caching	Cellular res. (reuse)	No	No	Simulation
Other performance metrics, Sec. IV-G	[203]	2014	Content placement	Cellular res.	No	No	Simulation
	[118]	2019	Content placement	Cellular res. (reuse)	Yes	No	Simul. based on Xender experim. results
	[204]	2018	Content delivery	mmWave res.	Real world pedestrian trajectories	No	Simulation
	[205]	2018	Joint content placement & delivery	Cellular res. (reuse)	Random walk	No	Simulation

A cooperative caching method is presented in [186] by considering caching placement at both user nodes and relay nodes. The considered HetNet system contains macro BSs, low-power relays, and UEs. The BSs and the UEs communicate in half-duplex (HD) mode; however, the low-power relays operate in full-duplex (FD) mode. If the requested content cannot be found either at nearby UEs connected via D2D links or at nearby relays, then the requesting UE connects to the BS through a relay. Exploiting the FD communication, the relay receives the requested content from the BS and at the same time transmits the content to the requesting UE, which helps to reduce the latency as well as the UE power consumption. The joint caching and resource allocation optimization problem is formulated with the aim to maximize the throughput. This complex non-convex stochastic optimization problem is solved through decomposition into three suboptimal problems that separately address the placement optimization at the user and relay levels as well as the power control. However, considering the lower energy efficiency of FD relay communication compared to HD relay communication, it may be worthwhile to select the relay communication mode based on the task requirements and the communication gains [211].

The approach from [186] has been improved in [187] by considering the UE mobility which has been addressed by a mobility-aware coded caching method. The mobility-aware coded caching method [187] considers a random jump model, a discrete form of the random walk model, which is a variant of the random waypoint model, to characterize the UE mobility pattern. Furthermore, in this coded caching scheme, it is advantageous to retrieve a requested content file by receiving any subset of the content file segments which are cached in the local UE or at the BS. The BS will send the missing segments to the UE when a UE has not received enough encoded segments within a tolerable downloading time period. To maximize the throughput, two content assignment algorithms, namely non-overlapping and overlapping content assignment, are developed.

2) CONTENT DELIVERY

An edge caching technology for addressing the issue of asynchronism in multi-view video (MVV) in fog computing has been proposed in [188]. The proposed caching technology considers the social characteristics of the UEs, such as their similar interest in MVV streams. Streams are synchronized with the assistance of edge caching between UEs. First, the storage capacity of each UE is calculated based on a proposed spatial distribution model and then a greedy algorithm is proposed which chooses caching nodes for multicast groups with the aim to increase the total system throughput. The BS controls both the multicast group and the caching UEs. Interference between cellular and D2D links is avoided by using orthogonal frequencies.

Information-centric wireless networking (ICWN) is a wireless variant of the ICN paradigm which aims to distribute information by specifying a name for each data item [105]–[107], see Section II-B3. The study [189] proposed a novel resource allocation and power control method for improving the throughput of the content delivery in ICWNs by integrating MEC systems and D2D communication in order to maximize the spectrum efficiency and overall system capacity as well as reducing the traffic congestion. The system consists of several small cells connected to the Internet through the core network and MEC servers with content caching capability which are placed at the BSs. A spatial Poisson process is used to model the deployment of UEs in each cell. The UE broadcasts its request to determine whether the content is available in the caches of nearby UEs. Then, depending on the location of the cached content, the content is delivered in one of two communication modes, namely cellular communication or D2D communication. In the cellular communication mode, the UE communicates with another UE via the BS. In the D2D communication mode, the communication is through direct traffic [212]. In the D2D mode, the cellular users use the downlink resources of the small cell and D2D users reuse the resources non-orthogonally and the same resource block is shared between a cellular user and D2D UE pairs. It is assumed that the BS can allocate D2D users to different channels by using a resource scheduler; and to mitigate interference, the UE can adjusts its power level. The optimization problem of resource allocation decisions considering the quality of the channel between UEs and BSs as well as the interference between D2D users is modeled as an MDP to maximize the overall system capacity. A policy-gradient algorithm is proposed to solve this MDP. This policy-gradient algorithm is then divided into two subalgorithms. First, the communication mode selection mechanism is designed based on the cache matrix. Second, using deep reinforcement learning, D2D pairs are designed to be able to adaptively perform the channel and power selection strategies. The stochastic actions for power selection are created based on a Gaussian distribution. However, this study assumes that there is no interference between neighboring small cells, i.e., the cells are assumed to use channel resources of different bandwidths.

C. ENERGY CONSUMPTION MINIMIZATION

The studies on energy consumption minimization strive to minimize the UE transmission power and therefore, are mainly focused on content delivery strategies. Social networks are a key predictor for content caching, since they are highly representative of human activities [213]. Exploring social relationships, the spatial network structure, and D2D communication, the study [190] proposed a caching method to reduce the energy consumption of the end devices. Nodes are selected for caching the contents according to the social centrality of the users, which is based on their contacts and location information. The BS detects user locations and channel state information and keeps this information for a period of time. The BS then caches content files in end devices during off-peak periods according to a genetic algorithm which minimizes the energy consumption of the end devices. An end device first sends a content request to its adjacent UEs by D2D links. If the content cannot be found at the adjacent UEs, then the request will be sent to the BS. A frequency band has been dedicated to the downlink channel and cellular spectrum resources can be used by D2D users. This gives rise to interference problems when the cellular and D2D users share the same downlink spectrum resources. In this study, only one BS is considered and the inter-cell interference as well as device mobility are neglected.

The study [191] investigated the energy cost of a helper node and the continuity of its battery lifetime to maintain the D2D connection in the context of a static (non-mobile) network with different D2D communication modes. The proposed proactive user-centric caching and transmission method introduces a collaboration distance, whereby only UEs within the collaboration distance can act as helper nodes. This proactive caching policy is optimized so that the maximum amount of traffic can be offloaded through the D2D links. Furthermore, to increase the possibility of finding a requested file in adjacent UEs and its complete transmission with a data rate higher than a threshold, the transmit power at each helper node is optimized. The BSs and UEs are modeled as two separate HPPPs, whereby the BS knows the placement of content items at each UE and coordinates the D2D communications. There are two modes for D2D communications, namely underlay and overlay modes based on whether the D2D and cellular links reuse the same frequency band or not. In the underlay mode, the cellular frequency can be reused by D2D communication; however only the downlink reuse is considered in [191]. In the overlay mode, the cellular and D2D communications operate in orthogonal frequency bands. The preceding two energy consumption minimization studies neglected user mobility. However, user mobility is a critical aspect of real-life usage scenarios and will bring more challenges, as the social relationships among mobile devices and their physical distances need to be constantly updated.

Taking the user mobility and the asynchronous demands of the UEs into account, a distributed content delivery method has been proposed in [192] to minimize the average energy consumption for each delivery process. The proposed system model contains a wireless cell with one BS and UEs which can act as both requester and relay. The content requester makes the requests based on its preference indicators of content items, which are estimated using some learning methods or recommendation techniques. The content delivery optimization problem is formulated as a fractional program which is NP-hard. In order to facilitate the best content helper selection for the requester, a distributed energy-balancing algorithm based on a belief propagation framework [214] is then proposed. It is assumed that the UEs exchange information by actively sending and receiving D2D broadcast signals. The sizes of the content items as well as the content helpers' caching capacities are assumed equal in this study, which is not realistic for practical scenarios.

The study [146] has proposed a caching scheme that takes advantage of D2D multicast. The system model consists of one BS and several UEs forming multiple clusters according to their geographical locations. One UE is selected as the head in each cluster; this head UE is called DCH and acts as a relay. The DCHs have the ability to cache some contents proactively as well as to distribute content through multicast. Also, all regular UEs have the caching ability. The cluster head selection algorithm is based on the CRP [163] and the weighted sum methods in the clustering considers several factors such as UEs' social behaviors, energy and storage resources, and the transfer rate from the BS to the UEs. When UEs have requests, they first send the requests to the associated DCH. If the contents cannot be found in the associated DCH, then the associated DCH tries to obtain the content from the closest DCH (with the content) using D2D communication and then the contents are sent to the requesting UEs via D2D multicasting. If the contents have not been cached in any of the DCHs, then the BS multicasts the contents to the UEs. The energy consumption optimization of the multicast content delivery method is formulated and solved using a cooperation-based greedy caching algorithm. The study [146] has examined both a computation offloading scheme, which we covered in Section III-C, and the caching scheme covered here. However, these two schemes are independent and do not synergistically interact with each other. Collaborative computation offloading and caching in device-enhanced MEC is an important direction for future research as elaborated in Section V-B1.

D. JOINT OPTIMIZATION OF THROUGHPUT AND ENERGY CONSUMPTION

Based on the optimization of cooperation distance, an optimal caching policy is proposed in [193]. The proposed caching policy enhances both the throughput and the energy efficiency by comprehensively accounting for the caching in the local end device (that requires a content item), in other end devices reached via D2D communication, and in the BS. A cluster based D2D network architecture [183], [215] with a specific power control policy and a frequency reuse scheme has been adopted. D2D links only exist between users within a given cluster. Two types of UEs are defined, namely active UEs and inactive UEs. An active UE participates in sending requests and D2D communications, while an inactive UE does not place requests but still participates in the D2D collaborations. Both active and inactive UEs are independently distributed according to HPPPs. Two different network structures, namely random-push and prioritized-push are considered for analyzing network throughput. Pareto-optimality in multi-objective optimization [216] is exploited to solve the tradeoff design problem.

E. LATENCY MINIMIZATION

The study [194] has shown that the expansion of the cooperation possibilities among caching nodes can significantly improve the content delivery delay and cache hit rate. Three types of cooperation have been considered, namely inter-BS, inter-device, and cross-tier. The system model consists of BSs with limited cache capacities under the centralized control of a service gateway and UEs within the cell of each BS. The D2D communication is assumed to be interference-free and to be well integrated with the cellular network and the file popularity follows a static Zipf-like distribution.

However, human factors can result in dynamic file popularity variations. A learning based cooperative caching method that accounts for such file popularity variations has been presented in [195]. The system includes a server, which is normally placed far from the UEs and contains all contents, a service gateway to control the caching distribution decisions, and cache nodes (BSs and UEs). A requesting UE can obtain the content either from an adjacent UE through D2D communication, a nearby BS, or from the content server. The total content delivery delay problem is formulated considering the content sharing cost between UEs, the cooperative caching cost among BSs, the limited storage resources of UEs and BSs, as well as content popularity variations. This study assumes that content is cached by a UE if there is enough storage available; otherwise, the least popular content will be replaced by the most popular content. Then, a two-step multi-armed bandit game learning based algorithm is proposed to first estimate the content popularity and then solve the caching strategy based on a relaxation based approach [217]. UE mobility and UE incentives are neglected in this study.

F. UTILITY MAXIMIZATION

1) CONTENT PLACEMENT

To remove the need for advance knowledge of the accurate channel state information in designing the incentive mechanism, a method based on the statistical channel state information is proposed in [196]. The proposed method encourages UEs with available (yet limited) cache storage to participate in D2D communication. The problem of different interests between the operator, which is in general responsible for designing the incentive mechanism, and the D2D transmitters, is modelled as a Stackelberg game. The operator designs an incentive price such that the maximum profit can be achieved by considering the D2D transmitters willing to maximize their utility. The profit maximization builds on a mathematical analysis of the achievable cellular and D2D link throughputs. The D2D communication operates in the underlay mode. The interference resulting from the underlay D2D communication is considered in the analysis of the profits of the D2D transmitters and the operator.

Leveraging game theory as a powerful modeling tool for the cooperation among several players [218], a hybrid caching strategy has been proposed in [197] based on content awareness in the D2D network. There are two types of nodes in the proposed caching strategy, namely active nodes and silent nodes. Active nodes download the file from the BS and share the file with adjacent UEs. Silent nodes cache the file directly from nearby active nodes. The cache cost problem includes data cost and sharing cost, whereby the data cost includes the cost for downloading a file from the BS and the basic cost to connect with the BS. The cooperative caching problem is formulated as a local cooperative game and then a log-linear learning algorithm is modified to speed up convergence and to improve the performance in large strategy spaces. While the hybrid caching strategy [197] generally considered the cooperation among UEs, it ignored the dynamics of the communication links between UEs as well as UE mobility.

Unlike the preceding two utility maximization studies, the study [198] focused on a mobility-aware caching scheme. The presented mobility-aware caching scheme considers a portion of each content item for caching and the selfish behavior of devices, to maximize user utility (energy consumption and delay) and to minimize BS cost. There are three ways for requesting the content items, namely local, D2D, and BS caching. In this method, users are rewarded by the BS for sharing their contents. Rewards can be a virtual currency or credit. The user mobility pattern is represented by a probabilistic version of the random walk [219]. A gradient projection algorithm is deployed to solve the optimization problem. A simplistic scenario considering only one BS is a shortcoming of this study.

2) CONTENT DELIVERY

Taking UE heterogeneity into account, a joint design of a D2D caching strategy and incentive scheme in MEC networks is presented in [199] to maximize the BS utility and to improve the social welfare of the cellular network. The optimization problem includes two subproblems: caching and contract design. The UE caching strategy and incentive method are designed based on UE context information which includes three parameters, namely UE desire for each content item, the degree of UE interest to share its contents, and the transmission delays among UEs. The content request pattern of each UE is heterogeneous and modelled with a Zipf distribution [209], [210], [220]. A coded caching method is used to cache the contents [221]. A content request is considered satisfied only if the requesting UE receives the complete requested content. Different types are considered for UEs based on their willingness to participate in D2D communications and share contents which are the result of each UE's privacy issues, remaining battery charge, and the desire to share contents. The BS rewards each UE according to its contributions and designs a contract based on the total content shared by a UE with the aim to maximize the utility. The UEs can either accept or reject any of these contracts. The difference between the amount of content transmissions saved for the BS and the reward which should be paid to the UE is defined as the BS utility function. To achieve the optimal caching strategy, a heuristic algorithm based on the gradient projection method is proposed. The type of UEs is considered to follow a uniform distribution in [199]; however, the UE type distribution may change based on network or device conditions in a real network environment. The simple scenario considering only one BS is also a shortcoming of this study.

3) JOINT OPTIMIZING CONTENT PLACEMENT AND DELIVERY

The study [200] investigated the caching cooperation between UEs and BSs by taking the UEs' social behaviors, preferences, and heterogeneous cache sizes, as well as the network infrastructure into account. The proposed hierarchical edge caching architecture includes a mobile network operator core connected to the Internet through backhaul links as well as BSs and UEs inside the RAN area. BSs are fully connected to each other via high-speed links, such as optical fibers, and UEs can communicate with each other through WiFi-Direct or Bluetooth D2D links. UEs are divided into four types according to their social relationships, namely self, close friends, normal friends, and strangers. A relationship factor is defined to simplify the effects that each type has on the social relationship among UEs. It is assumed that the user is only served by its local BS during the content delivery period and the local BS can satisfy the requesting UE by using its own storage, obtaining the requested content from other BSs, or downloading the content directly from the Internet through backhaul links. The caching method contains two phases, namely the content placement phase and the content delivery phase. In the content placement phase, in order to increase the content diversity and to enhance the cooperation among BSs for content delivery, maximum distance separable coding is used to code the content cached in BSs, while uncoded caching is employed in UEs to preserve the integrity of each content item. In the content delivery phase, the content is requested based on the UE's own preference; a user will be served by its local BS if the requested content cannot be found in its local cache or the caches of its adjacent UEs. Aiming to maximize the capacity of the network infrastructure for traffic offloading and to reduce system costs while satisfying UE requests, a hierarchical edge caching problem is formulated. This hierarchical edge caching problem is NP-hard. Towards developing a low-complexity and efficient caching scheme, the problem is decomposed into two subproblems. The two subproblems are the device caching problem and the BS caching problem with focus on different cooperation levels. This hierarchical edge caching scheme has been evaluated with the *Xender* application [228]. The D2D communication links in this application can be based on WiFi tethering, WiFi-Direct, or Bluethooth. However, the physical-layer transmissions have been neglected in [200] and this shortcoming is further addressed in [222].

The preceding video caching studies reviewed in Section IV-B have considered the caching of complete video files. However, complete video files are typically very large, while cache space at UEs is limited, allowing only very few complete videos to be cached on UEs. The study [201] has addressed this problem by proposing to cache only popular chunks, e.g., temporal segments, of video files. When a UE requests a video, the chunks forming the complete video can be obtained from nearby UEs that have some needed chunks cached, the MEC servers that have chunks cached, or the remote server. The decision problem for placing the chunks into the caches and for delivering the chunks from the caches or remote server is formulated as a utility maximization problem considering the costs for serving requests, the UE incentives paid, and the penalty costs for service dissatisfaction. The simulation evaluations compared the proposed chunk caching in the device-enhanced MEC with the caching of full video files. The results indicate that chunk caching achieves substantially higher cache hit ratios while significantly lowering the network cost and decreasing the average video access delay compared to the caching of full video files.

4) TASK CACHING

While device-enhanced MEC caching studies have generally examined the cache placement and delivery of abstract data files or video files, the study [202] has ventured into the new domain of task caching. Task caching was originally introduced in the pure MEC context in [223] and is conceptually similar to the MEC studies that jointly consider computation offloading and caching [32], [224]–[226]. In particular, task caching refers to the caching of the (potentially large) data set resulting from a complex computing task, e.g., a computation relating to augmented or virtual reality rendering. The study [202] has examined this task caching in the context of a device-enhanced MEC system with UEs, MEC computing nodes (referred to as fog nodes), and a remote cloud. Computed task data sets are cached on the UEs according to a proposed near-optimal task caching algorithm in an offline manner, e.g., over night. A UE task request can be served from the local UE task cache, the task caches of nearby UEs, or the task cache of the fog nodes. If a task cannot be served form a cache, then the task is computed locally, at a fog node, or at the remote cloud; task computation offloading to other UEs is not considered. The task caching problem is formulated as an optimization problem to maximize a utility defined based on delay and energy consumption. The utility maximization is solved through a genetic algorithm. The simulation evaluations in [202] indicate that the proposed task caching algorithm achieves higher utilities than various benchmark approaches, particularly for high task computing demands and dense nearby UE populations. A shortcoming of the task caching in [202] is that the caching is conducted in an offline manner, e.g., over night. A typical motivating example for task caching is the computation of augmented and virtual reality computations for visitors to a specific location, e.g., a museum, lecture hall, or laboratory for hands-on experiments. In such a use-case, the UEs of the visitors are only in the specific location during the active visiting hours (and not during the off-hours, when such facilities are closed). Accordingly, it will be important to develop real-time task caching approaches in future work that can share the task computation data sets among the UEs that are currently in the specific location. Also, the collaborative optimization of both the task computation offloading and the task caching while utilizing both the computation and caching resources of the nearby UEs and the fog (MEC) nodes is an important direction for future research, as elaborated in Section V-B1.

G. OTHER PERFORMANCE METRICS

1) CONTENT PLACEMENT

The benefits of proactive caching have been examined in [203] through two case studies by considering the spatial and social network structures. In the first case study, the proposed distributed caching method stores files in the

BS during off-peak periods by exploiting file popularity as well as correlations among UEs and file patterns, using supervised machine learning and collaborative filtering tools. This proactive caching procedure continues until reaching the maximum storage capacity. The social network structure and D2D communications are leveraged in the second case study. First, a set of influential UEs is determined by obtaining the social ties among UEs using a centrality metric [227]. The higher the node centrality of a UE, the more influential the UE is in its community. When a UE requests a file, the BS first searches for influential UEs with a cached copy of the requested file and in the case of availability, the BS directs the nodes to establish a D2D communication between the influential UE and the requester. If the BS cannot find any influential UE with a cached copy of the requested file, then the BS sends the file directly from the core network to the requester. However, the study [203] considers a static system; UE mobility can bring challenges, especially for determining influential UEs when social ties change dynamically.

A more general scenario considering UE mobility, UE social behaviors, and geographical distances has been presented in [118]. The system model includes a hierarchical architecture with cellular links between UEs and BSs, as well as backhaul links between BSs and the core network. The UEs first request content items from their adjacent devices; if the requested content cannot be found in the UE's own storage or the storage of the UE's neighbors, then the cellular links mode is activated. The requested content can be obtained from the neighbor BSs as well as through the links between BSs or the Internet via backhaul links. UEs' sharing activities have been measured using *Xender*, which is an application to trace D2D content sharing at a large-scale [228]. The social relationships are based on the user preferences (whereby user preference is characterized by the probability distribution of a user request for each content item) and content transmission rate condition. The cache replacement process is modeled as an MDP and the Double Deep Q-Learning method [229] is used to optimize the cache replacement. In this study, the content transmission process is assumed to be finished before the UE moves out of the coverage of its small cell, which is not a realistic assumption considering the complicated UE mobility patterns in practice. In addition, considering the limited UE computing resources, using learning algorithms, which require heavy computational resources, may be inefficient and increase delay.

2) CONTENT DELIVERY

The study [204] has focused on mobility-aware transmission scheduling for caching at end devices that are located near hotspots. Hotspots are places that are frequently visited by the UEs, such as cinemas, restaurants, and stadiums. Popular content items are cached at the end devices near hotspots through multi-hop D2D relaying from the BS and via high datarate millimeter wave (mmWave) communication [230]. UEs can download contents from the end devices near the hotspots instead of the BS while passing by the end devices near the hotspots. The scheduling problem is formulated as a stochastic nonlinear mixed integer program to maximize the expected amount of cached data. Then, a Multi-Hop D2D Relaying based Caching (MHRC) scheme is proposed. The MHRC scheme first establishes the multi-hop relay paths and then schedules the transmissions for caching at the end devices near hotspots by considering the statistical user mobility properties. The study considers only one BS which is responsible for scheduling the caching transmissions and for synchronizing the clocks of relay nodes. Also, the blockage problem for mmWave communications [231] is neglected.

3) JOINT OPTIMIZATION OF CONTENT PLACEMENT AND DELIVERY

The associations between content item popularity levels and user preferences have been mostly overlooked in the caching designs studies. Focusing on this issue, the study [205] proposed a model for synthesizing user preference from content popularity with the aim to maximize the offloading probability. The user preference can be learned using probabilistic latent semantic analysis [232] and a UE request behavior model. The caching policy can then be optimized for a given UE request behavior model. The overall system consists of a central processor (CP) in the core part, several BSs connected to the Internet through backhaul links and with knowledge of the cached files and the UE locations, and uniformly distributed UEs. The caching policy operates as follows in the placement phase. First, the user preferences are learned and then the CP informs the BSs about the cached files of the UEs. Subsequently, during off-peak time, the fetching process of the files from the server is performed, i.e., the files are unicast or multicast by the BS to be cached at the UEs. In the delivery phase, a UE sends its request to the BS if the content cannot be found in its local cache. Then, the BS informs the CP and the CP sends the file index to the BS and records the UE request. In the next step, the BS looks for the file in the local caches of the UEs that are adjacent to the requesting UE. If the BS can find the file, then the closest helper will be assigned to send the file to the requesting UE. The BS is also responsible for establishing D2D links between the UEs. Both overlay and underlay modes are considered in the analysis in [205] and bandwidth can be shared either with full reuse or FDMA. A low computational complexity algorithm is applied to solve this offloading gain problem. The impact of UE mobility on the offloading probability is analyzed based on a random walk mobility model. However, file sizes are assumed to be equal, which is not realistic. Also, the study [205] did not include a method for encouraging the helper nodes.

H. SUMMARY OF DEVICE-ENHANCED MEC CACHING

The main overall strategy of the existing device-enhanced MEC caching approaches has been to prioritize content placement at and delivery from other end devices via D2D communication links, followed by placement at and delivery from the MEC server, followed by delivery from the central cloud. The main objectives of the existing device-enhanced

MEC caching studies have been the minimization of the UE energy consumption, the maximization of the throughput, as well as cost reduction. Considering any of these objectives in isolation typically does not enhance the overall caching system performance, since finding an optimal solution for one metric can negatively affect other metrics. Therefore, designing a caching system with good overall performance requires the investigation of the tradeoffs between different objectives [193].

In order to achieve desired caching performance goals, the optimization problems have usually been solved with some mathematical tools, such as game theory, stochastic geometry, heuristic optimization, and machine learning. Game theory has generally been used when the players, i.e., BSs and UEs, compete with each other in order to maximize their benefits [233], [234]. Optimization algorithms have usually been employed to maximize a performance metric under some existing constraints, such as limited cache size, cache state, and remaining battery life times of UEs. Stochastic geometry approaches have been formulated for designing caching techniques in networks with random topologies [235]. An interesting future research direction will be to combine the various solution approaches and tools into hybrid approaches that comprehensively solve and trade off the various problem aspects. Also, the computational effort and signalling overheads required for the various approaches or their combinations need to be examined.

Most existing studies have considered the Homogeneous Poisson Point Process (HPPP) model for the distribution of the BSs and UEs. Future research should validate the spatial distribution models for a wide range of networking scenarios and environments, and then re-examine the performance of device-enhanced MEC caching mechanisms for a range of validated spatial distribution models and possibly spatial distribution traces of real wireless networks.

Device-enhanced MEC caching generally includes the placement phase and the delivery phase. The vast majority of existing studies have designed, optimized, and examined these two phases separately. Only relatively few studies have attempted to jointly improve these two phases. Future research should expand on the studies that have jointly pursued the placement and delivery and strive to extract gains from synergistically completing both phases.

Device-enhanced MEC caching strategies can generally be controlled in a distributed manner or in a centralized manner. The centralized control is typically implemented at the BS. Most existing studies have considered the centralized control since the BS typically tracks the required information, e.g., UE locations, preferences, and requests, as well as content popularities and channel states. Thus, the BS has the required information to provide an optimal solution for the entire network encompassing the UEs within the range of the BS [31].

However, there are still several open challenges in order to fully exploit the advantages of device-enhanced MEC caching. The channel interferences among UEs, the heterogeneous caching UE capacities, the social connections among UEs [169], and the effects of UE mobility on the social ties of the UEs are some important factors that need to be rigorously investigated in future research. Furthermore, future research needs to explore the tradeoffs between designing an effective incentive mechanism and protecting the privacy of users and content items.

Another drawback of most existing device-enhanced MEC caching studies is the consideration of very simple scenarios, typically consisting of only one BS. Cooperation opportunities among different BSs, which appear especially interesting for small cells, have only rarely been discussed. Cooperation among BSs can, for instance, be useful in dynamic scenarios where UEs move between different small cells. Also, all existing studies have based the evaluation on simulations, whereby only a few studies have considered practical parameter settings in the simulations. Future research should address the limitations of the existing studies by thoroughly investigating device-enhanced MEC caching mechanisms for large networks with multiple BSs. The multiple BSs should coordinate their caching placement and delivery. The performance evaluations should include simulations, but also mathematical analysis and testbed measurements.

V. OPEN PROBLEMS AND FUTURE RESEARCH DIRECTIONS

Device-enhanced MEC computation offloading and caching is still a nascent research area. There are still numerous open problems and challenges that need to be resolved in this area. This section outlines the main open problems and outlines future research directions to make device-enhanced MEC computation offloading and caching highly effective and efficient. We group the open problems and future research directions into four main clusters, namely the control and management of the device-enhanced MEC mechanisms, the improvement of the performance and scalability of the device-enhanced MEC mechanisms, security and privacy, as well as performance evaluation and testbeds.

A. CONTROL AND MANAGEMENT

1) DEVICE-ENHANCED MEC MANAGEMENT FRAMEWORK

Presently, no framework exists to exchange information about computational or caching capabilities of end devices in a real-time or near-real-time manner so as to inform offloading and caching decisions. Stand-alone solutions, such as connection sharing in smartphones, need to be set up per device and are not designed to operate automatically or at scale. Accordingly, there is an urgent need to develop and evaluate control and management frameworks for device-enhanced MEC.

One possible avenue for developing device-enhanced MEC control and management frameworks is to build on the recent success of Software Defined Networking (SDN) control [93], [94] for a wide range of general networking aspects, such as scalable control plane operation [236], [237], flow control [238], traffic engineering [239], [240], routing [241]–[243], and Internet of Things management [244], [245]. Moreover, SDN principles have been employed to control and manage general wireless networks, see e.g., [95], [246]–[250]. Building on these successful applications of the SDN control and management principles, future research could adapt autonomic SDN control and management mechanisms [251]–[253] to develop control and management frameworks for device-enhanced MEC. Importantly, these frameworks could build on the principles of hybrid SDN [254], [255] that allow for the control and management of hybrid systems that combine conventional devices that are not SDN-enabled as well as SDN-enabled devices.

Another important aspect for the framework development is the consideration of the timescales of the control and management. While real-time computational and caching capabilities can be signalled quickly over a localized network area [256], [257], large network areas may introduce substantial signalling delays. Thus, future control and management frameworks for device-enhanced MEC should incorporate some aspects of fast localized decision making with global coordination on a slower timescale, similar to recent multi-timescale wireless resource allocation studies [177]–[182].

Moreover, future control and management frameworks for device-enhanced MEC should accommodate heterogeneity across the gamut of system characteristics, such as heterogeneity of the wireless access technologies, end devices, and applications. For instance, the use of the various heterogeneous wireless medium access and transmission technologies needs to consider their implications for UE battery life time, link speed, and link reliability. Depending on the application needs and the communications scenario, the tradeoffs between UE battery lifetime, link speed, and link reliability may suggest to utilize a particular medium access technology, or a combination of medium access technologies, which in turn may imply heterogeneous achievable communications ranges and UE discoverability. Future device-enhanced MEC control and management frameworks should be able to account for these tradeoffs across the layers of the wireless networking protocol stack, from the physical layer, up to and including the application layer.

2) INTERFERENCE MANAGEMENT

If several UEs offload their tasks to MEC servers or adjacent end devices use the same resources (e.g., time slots and frequency channels), then interference among multiple ongoing D2D communication links and between D2D communication and cellular communication arises. This interference becomes worse with increasing numbers of UEs within a given cell coverage area [258]. Dedicating the resources exclusively to only D2D communications enables solutions to the interference problem [259], at the expense of reduced reuse efficiency. Therefore, multiple interference management techniques, such as power control, mode selection, and radio resource allocation, are generally used jointly to improve the network capacity as well as spectrum reuse efficiency.

The interference management issues are particularly challenging in heterogeneous IoT networks which can greatly benefit from device-enhanced MEC due to the limited nodal resources of typical IoT nodes. The interference management issues are more challenging in IoT networks compared to conventional cellular networks due to the massive numbers of connected IoT devices. In addition, the heterogeneous transmit power levels of IoT devices result in heterogeneous interference levels. Thus, there are several open challenges to efficiently manage the interference arising from device-enhanced MEC in IoT networks. One possible approach to address this challenge is to link the decision algorithm for the transmission mode selection with the device-enhanced MEC computation offloading or caching mechanisms so as to make dynamic offloading and caching decisions in accordance with the interference in the network environments. The general concept of adaptive mode selection was introduced in [260]; future research needs to effectively couple the mode decision making with the offloading and caching decision making so as to arrive at overall optimized offloading and caching decisions that adapt according to the interference levels and are thus applicable in practical IoT networks.

3) MOBILITY MANAGEMENT

In the context of computation task offloading, UE movements, including movements of either requester, relay, or helper UEs, can break the D2D links. For instance, if the relay moves during the transmission of tasks, the link breakage between requester and helper will likely increase latencies and waste battery energy. Mobility can also influence the information derived from social graphs by changing social ties among UEs. Therefore, updating and predicting the availability and reliability of computation resources is a key prerequisite for enabling satisfactory user experiences and energy savings. Future research needs to develop and validate effective and efficient network management methods for assessing and predicting the availability of computation resources for a wide range of network scenarios and mobility levels.

In the caching context in dynamic scenarios, copies of content items should be cached in multiple BSs or UEs in order to maintain acceptable caching performance; however, the storage capacity limitations may make it hard to achieve this goal [31]. Another important aspect of proactive caching is the popularity profile of content items which is usually based on the Zipf model [209], [210], [220] or variants thereof. However, in real mobile scenarios, it may not be possible to effectively and reliably define popularity profiles in advance due to the UE movements [261]. Future research should develop and validate management mechanisms for determining the optimal number of content item copies as a function of the mobility level. Also, content item popularity profile models should be formulated and validated for a wide range of mobility levels.

Another future research direction is to adapt the responsiveness of device-enhanced MEC mechanisms to mobility according to the needs and characteristics of a given particular application. For instance, even delay-tolerant applications, such as downloading video files, may be very sensitive to UE movements if the entire file is to be downloaded over one D2D link [262]. Adapting the download strategies, e.g., partitioning large files into small file segments that are downloaded over multiple D2D links, can make applications more robust to UE mobility.

B. PERFORMANCE IMPROVEMENTS AND SCALABILITY

1) COLLABORATIVE COMPUTATION OFFLOADING AND CACHING

Our comprehensive survey of the device-enhanced MEC literature has indicated that the existing studies have either focused on computation offloading or on caching. To the best of our knowledge, there are no existing device-enhanced MEC studies that have examined collaborative computation offloading and caching. However, some applications, such as augmented or visual reality or some video applications may involve large data files that benefit from caching, while also requiring extensive computation resources. Indeed, collaborative computation offloading and caching methods have begun to be investigated in the MEC context (without enhancements through end device resources), see, e.g., [32], [223]–[226]. Thus, a future research direction is to develop and evaluate device-enhanced MEC methods that collaboratively address caching and computation offloading, while striving to extract efficiency gains from the joint consideration of the these two functionalities.

2) SOCIAL-AWARE D2D COOPERATIVE COMMUNICATION

Exploiting the social characteristics of UEs generally facilitates efficient data exchanges in D2D networks. In particular, the location of UEs in mobile D2D networks can indicate social communities [263]; therefore, changes in the locations of UEs can lead to social disconnections between UEs. Thus, future research needs to design mechanisms that can dynamically adjust the network according to the mobility changes and establish corresponding D2D links between nearby socially related devices.

Furthermore, social network discovery mechanisms must safeguard against dishonest artificially generated information about social relationships, which UEs may maliciously generate to improve their networking performance [264]. Another important issue is that maintaining the social awareness in D2D cooperative communications costs the UEs some energy; this energy expenditure needs to be small in order to keep with a main premise for D2D communication, namely low-energy short-distance communications. Hence, future research needs to investigate energy-efficient social awareness in D2D cooperative communication [265].

3) LEARNING ALGORITHMS

Machine Learning (ML) is widely considered as a promising solution to autonomously and optimally configure future wireless networks based on the information learned from network system behaviors [94], [266]–[270]. In fact, it has been speculated that most problems considered ''hard'' can be formulated as ML problems and solved by iteration and policy search [271]. For instance, the channel selection problem in D2D communication can be modeled as a multi-armed bandit game which falls into the category of reinforcement learning algorithms [272]. Similarly, wireless power control techniques based on distributed Q-learning have been developed in [273]. However, due to the required multiple iterations, ML approaches are often highly time-consuming. Future research should focus on time-efficient ML algorithms [274].

In the caching context, the Zipf distribution is widely used to model the UE request pattern [209], [210], [220]. With the emergence of the ''Big Data'' concept, the actual content popularity matrix for a particular network could be obtained from wireless big data analyses at the MEC using ML algorithms [199]. Thus, the obtained actual content popularity matrix could potentially improve the caching performance of device-enhanced MEC.

4) CODING

Device-enhanced MEC involves wireless communication both with MEC servers as well as other end devices. Wireless communication is generally error prone and requires safeguards at the physical, e.g., forward error correction coding, and link layers, e.g., automatic repeat request retransmissions, to ensure reliability. An emerging coding technique that appears well suited to be explored for the highly heterogeneous device-enhanced MEC in future research is network coding [275]–[280]. Network coding eliminates the coordination that is required between sender and receiver in many conventional coding techniques. Instead, network coding only requires the collection of a sufficiently large random set of coded packets for successful decoding and protection against wireless link errors. Network coding thus embraces the randomness that naturally occurs when communicating with and involving other end devices in providing services. Future research should explore network coding based transmission techniques for device-enhanced MEC so as to improve the communications with both MEC servers as well as other end devices. A related future direction is to combine innovative network coding based transmission with specialized transmission scheduling for the various types of application traffic. For instance, video streaming scheduling requires consideration of the video frame timing constraints, wireless link characteristics, and the related implications for video quality of experience [281]–[285]. Future research should explore such application-specialized scheduling in the context of device-enhanced MEC supporting the various application traffic types.

Computation offloading in the device-enhanced MEC context involves end devices that may generally be unreliable, e.g., may move or require their resources for their own computations. Recently, the paradigms of replicated computation [286] and coded distributed computing [283], [287]–[290] has emerged to support efficient distributed computing, especially if some of the distributed computing nodes are unreliable or are late in completing their assigned computing tasks (i.e., are so-called stragglers). Future research should explore the adaptation of replicated computation and coded distributed computing to computation offloading in device-enhanced MEC. Device-enhanced MEC poses particular challenges due to the asymmetric nature of typically offloading few large computing tasks to MEC servers and numerous small computing tasks to other end devices.

Similarly, distributed caching of content items can be improved through coding [291]–[293]. Coding can be particularly helpful for network systems with unreliable caching nodes. Future research should investigate how the coded caching concepts can be efficiently adapted to device-enhanced MEC caching, which involves large relatively reliable MEC cache servers and numerous small potentially unreliable end device caches.

C. SECURITY AND PRIVACY

Security in the context of offloading computation tasks to adjacent devices is a considerable problem. Side channel attacks [294] could allow the exploitation of UEs' personal information, thus violating their data security and privacy. Such data security breaches would likely deter users from adopting task offloading schemes. Moreover, such security breaches could counteract the positive effects of offloading incentive mechanisms. Consequently, users may lose interest in participating in cooperations. Another important problem is the user mobility which requires adaptive security mechanisms that account for the varying user locations.

Future research needs to comprehensively address the security and privacy aspects of device-enhanced MEC. One avenue could build on the social user communities. For instance, UEs could be divided into different groups based on their social relationships, interests, and locations. Depending on the security level of a UE group, a given user may or may not participate in task offloading. An immediate drawback of such grouping of UEs is the potential loss of collaboration opportunities, due to hesitation to engage in collaborations with a nearby stranger [295]. Throughout, the overheads of security methods for the network communication need to be carefully traded off against their benefits [264].

Security in the context of content caching is highly challenging, since caching or statistically processing encrypted content items needs to circumvent the encryption [296], [297]. Future research needs to carefully examine the tradeoffs between justifiable needs for content caching security versus the additional overheads.

D. PERFORMANCE EVALUATION AND TESTBEDS

1) EVALUATION FRAMEWORK AND BENCHMARK WORKLOADS

For the further advancement of the device-enhanced MEC area it will be critical to quantitatively compare various approaches and identify weaknesses that can then be addressed in future research. In order to facilitate quantitative performance comparisons, future research should develop comprehensive evaluation frameworks that specify the set of performance metrics as well as the performance evaluation methodologies that ensure rigorous replicable evaluations. The evaluation frameworks should include workload specifications, as well as wireless channel and mobility models, that the research community agrees on as being representative for common device-enhanced MEC scenarios.

2) TESTBEDS

While evaluations of novel research approaches typically employ a combination of mathematical analysis and discrete event simulations, comprehensive validation in real-life networking scenarios should include measurement evaluations in real physical testbeds. Real physical testbeds account for the various real-world issues that are often neglected in mathematical analyses and simulations. Future research should develop specifications for representative testbeds for device-enhanced MEC. Ideally, such testbeds should be built as a research community infrastructure that will be made accessible to the research community

VI. CONCLUSION

We have comprehensively surveyed the area of device-enhanced Multi-access Edge Computing (MEC). Device-enhanced MEC augments the MEC computing and storage (caching) resources with the computing and storage resources of the wireless end devices, e.g., User Equipment (UE) nodes. Device-enhanced MEC thus enlarges the resource pool that is available for providing services to end devices. This enlargement of the available resource pool is achieved without additional MEC infrastructure investments; albeit, device-enhanced MEC typically requires some incentives (e.g., payments) to the owners of the participating end devices. Nevertheless, with the ever-increasing computing and storage resources available in mobile end-devices, device-enhanced MEC is an attractive paradigm for improving the service quality without requiring large upfront capital investments in more MEC resources. Also, device-enhanced MEC works particularly well in dense networks, where each end-device has a large number of neighboring end devices within a short device-to-device (D2D) communication distance, e.g., in crowded stadiums. Such dense network scenarios pose scalability problems for conventional MEC with a fixed amount of installed resources. Generally, the possibilities for "recruiting" neighboring end devices to contribute computation and storage resources grow in dense networks, as there are more end devices near any given end device in dense networks. Thus, device-enhanced MEC holds a particular promise to mitigate MEC resource shortages in dense networking scenarios.

We have organized this survey into the two main categories of studies focusing on computation offloading to the device-enhanced MEC and studies focusing on caching in the device-enhanced MEC. Within each of these two main categories, we sub-categorized studies according to their main objective. The existing studies have strived to increase the throughput, to reduce the latency, and to reduce the energy consumption. Also, some studies have focused on enhancing security aspects, while others have focused on maximizing some utility measure. Overall, the device-enhanced MEC studies that have been conducted to date have made significant progress in advancing the protocol development and optimization for offloading computations and caching jointly to MEC resources and other end devices.

Nevertheless, device-enhanced MEC is a nascent research area; most studies have appeared within the past three years. Thus, the existing state-of-the-art research in the device-enhanced MEC area has severe limitations and requires extensive future research to address the numerous open challenges. Overall, only roughly half of the existing studies have accounted for end device mobility. Also, less than roughly a quarter of the existing studies has incorporated an incentive mechanism. Moreover, there is an overarching need to develop effective and efficient control and management frameworks for device-enhanced MEC that can cope with end device mobility and end device heterogeneity while scaling to large network sizes and device densities. Future research should also further improve device-enhanced MEC, e.g., by exploiting emerging machine learning techniques and improved models of the social relationships of end device users. Also, comprehensive performance evaluation frameworks and methodologies should be developed and agreed upon by researches to facilitate the comparison of different approaches to device-enhanced MEC.

ACKNOWLEDGMENT

The authors alone are responsible for the content of the paper.

REFERENCES

[1] J. Rodriguez, A. Radwan, C. Barbosa, F. H. P. Fitzek, R. A. Abd-Alhameed, J. M. Noras, S. M. R. Jones, I. Politis, P. Galiotos, G. Schulte, A. Rayit, M. Sousa, R. Alheiro, X. Gelabert, and G. P. Koudouridis, "SECRET—Secure network coding for reduced energy next generation mobile small cells: A European training network in wireless communications and networking for 5G," in *Proc. Internet Technol. Appl.*, Wrexham, U.K., Sep. 2017, pp. 329–333.

[2] X. Yang, X. Yu, H. Huang, and H. Zhu, "Energy efficiency based joint computation offloading and resource allocation in multi-access MEC systems," *IEEE Access*, vol. 7, pp. 117054–117062, 2019.

[3] X. Wang, Y. Ji, J. Zhang, L. Bai, and M. Zhang, "Low-latency oriented network planning for MEC-enabled WDM-PON based fiber-wireless access networks," *IEEE Access*, to be published.

[4] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.

[5] B. Shi, J. Yang, Z. Huang, and P. Hui, "Offloading guidelines for augmented reality applications on wearable devices," in *Proc. ACM Int. Conf. Multimedia*, Brisbane, QLD, Australia, 2015, pp. 1271–1274.

[6] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, "Enabling real-time context-aware collaboration through 5G and mobile edge computing," in *Proc. Int. Conf. Inf. Technol. New Gener.*, Las Vegas, NV, USA, Apr. 2015, pp. 601–605.

[7] Z. Xiang, F. Gabriel, E. Urbano, G. T. Nguyen, M. Reisslein, and F. H. P. Fitzek, "Reducing latency in virtual machines: Enabling tactile Internet for human-machine co-working," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1098–1116, May 2019.

[8] Cisco VNI Forecast, Cisco Inc. (Nov. 2018). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast 2017–2022*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[9] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[10] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[11] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[12] F. Al-Turjman, "5G-enabled devices and smart-spaces in social-IoT: An overview," *Future Gener. Comput. Syst.*, vol. 92, pp. 732–744, Mar. 2019.

[13] J. Iqbal, M. A. Iqbal, A. Ahmad, M. Khan, A. Qamar, and K. Han, "Comparison of spectral efficiency techniques in device-to-device communication for 5G," *IEEE Access*, vol. 7, pp. 57440–57449, 2019.

[14] L. Militano, G. Araniti, M. Condoluci, I. Farris, and A. Iera, "Device-to-device communications for 5G Internet of Things," *EAI Endorsed Trans. Internet Things*, vol. 1, no. 1, pp. e4-1–e4-15, 2015.

[15] R. Paul and Y. J. Choi, "Autonomous interface selection for multi-radio D2D communication," *IEEE Access*, vol. 7, pp. 108090–108100, 2019.

[16] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," in *Proc. Int. Conf. Intell. Syst. Control*, Coimbatore, India, Jan. 2016, pp. 1–8.

[17] M. T. Beck, M. Werner, S. Feld, and T. Schimper, "Mobile edge computing: A taxonomy," in *Proc. Int. Conf. Adv. Future Internet*, 2014, pp. 48–54.

[18] J. Moura and D. Hutchison, "Game theory for multi-access edge computing: Survey, use cases, and future trends," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 260–288, 2018.

[19] Q.-V. Pham, F. Fang, V. N. Ha, M. Le, Z. Ding, L. B. Le, and W.-J. Hwang, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," 2019, *arXiv:1906.08452*. [Online]. Available: https://arxiv.org/abs/1906.08452

[20] H. Tanaka, M. Yoshida, K. Mori, and N. Takahashi, "Multi-access edge computing: A survey," *J. Inf. Process.*, vol. 26, pp. 87–97, Feb. 2018.

[21] E. Ahmed and M. H. Rehmani, "Mobile edge computing: Opportunities, solutions, and challenges," *Future Gener. Comput. Syst.*, vol. 70, pp. 59–63, May 2017.

[22] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[23] B. Huang, Z. Li, P. Tang, S. Wang, J. Zhao, H. Hu, W. Li, and V. Chang, "Security modeling and efficient computation offloading for service workflow in mobile edge computing," *Future Gener. Comput. Syst.*, vol. 97, pp. 755–774, Aug. 2019.

[24] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges," *Future Gener. Comput. Syst.*, vol. 78, pp. 680–698, Jan. 2018.

[25] R. Yang, F. R. Yu, P. Si, Z. Yang, and Y. Zhang, "Integrated blockchain and edge computing systems: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1508–1532, 2nd Quart., 2019.

[26] T. Le Duc, R. G. Leiva, P. Casari, and P.-O. Östberg, "Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 94:1–94:39, Oct. 2019.

[27] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, 3rd Quart., 2017.

[28] O. Richard, O. Fadahunsi, M. Maheswaran, and M. F. Zhani, "Opportunistic edge computing: Concepts, opportunities and research challenges," *Future Gener. Comput. Syst.*, vol. 89, pp. 633–645, Dec. 2018.

[29] X. Shan, H. Zhi, P. Li, and Z. Han, "A survey on computation offloading for mobile edge computing information," in *Proc. IEEE 4th Int. Conf. Big Data Secur. Cloud, IEEE Int. Conf. High Perform. Smart Comput., IEEE Int. Conf. Intell. Data Secur.*, Omaha, NE, USA, May 2018, pp. 248–251.

[30] X. Zhao, K. Yang, Q. Chen, D. Peng, H. Jiang, X. Xu, and X. Shuang, "Deep learning based mobile data offloading in mobile edge computing systems," *Future Gener. Comput. Syst.*, vol. 99, pp. 346–355, Oct. 2019.

[31] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1710–1732, 3rd Quart., 2018.

[32] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[33] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[34] Y. Zhao, W. Wang, Y. Li, C. C. Meixner, M. Tornatore, and J. Zhang, "Edge computing and networking: A survey on infrastructures and applications," *IEEE Access*, vol. 7, pp. 101213–101230, 2019.

[35] S. Wang, J. Xu, N. Zhang, and Y. Liu, "A survey on service migration in mobile edge computing," *IEEE Access*, vol. 6, pp. 23511–23528, 2018.

[36] M. J. Kaur, "A comprehensive survey on architecture for big data processing in mobile edge computing environments," in *Edge Computing*. Cham, Switzerland: Springer, 2019, pp. 33–49.

[37] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Gener. Comput. Syst.*, vol. 97, pp. 219–235, Aug. 2019.

[38] V. Nivethitha and G. Aghila, "Survey on architectural design principles for edge oriented computing systems," *J. Comput. Theor. Nanosci.*, vol. 16, no. 4, pp. 1617–1624, 2019.

[39] S. Raza, S. Wang, M. Ahmed, and M. R. Anwar, "A survey on vehicular edge computing: Architecture, applications, technical issues, and future directions," *Wireless Commun. Mobile Comput.*, vol. 2019, Feb. 2019, Art. no. 3159762.

[40] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.

[41] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[42] H. H. Elazhary, "Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions," *J. Netw. Comput. Appl.*, vol. 128, pp. 105–140, Feb. 2019.

[43] A. Pravin, P. Jacob, and G. Nagarajan, "A comprehensive survey on edge computing for the IoT," in *Edge Computing and Computational Intelligence Paradigms for the IoT*. Hershey, PA, USA: IGI Global, 2019, pp. 37–45.

[44] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for Internet of Things realization," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2961–2991, 4th Quart., 2018.

[45] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2018.

[46] Cisco. (2016). *Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are*. [Online]. Available: https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

[47] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, Hangzhou, China, 2015, pp. 37–42.

[48] R. Tandon and O. Simeone, "Harnessing cloud and edge synergies: Toward an information theory of fog radio access networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 44–50, Aug. 2016.

[49] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 416–464, 1st Quart., 2018.

[50] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, Sep. 2019.

[51] M. Mukherjee, L. Shu, and D. Wang, "Survey of fog computing: Fundamental, network applications, and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1826–1857, 3rd Quart., 2018.

[52] P. Zhang, J. K. Liu, F. R. Yu, M. Sookhak, M. H. Au, and X. Luo, "A survey on access control in fog computing," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 144–149, Feb. 2018.

[53] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.

[54] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2133–2168, 3rd Quart., 2018.

[55] P. Gandotra and R. K. Jha, "Device-to-device communication in cellular networks: A survey," *J. Netw. Comput. Appl.*, vol. 71, pp. 99–117, Aug. 2016.

[56] M. F. Hashim and N. I. A. Razak, "Ultra-dense networks: Integration with device to device (D2D) communication," *Wireless Pers. Commun.*, vol. 106, no. 2, pp. 911–925, May 2019.

[57] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1923–1940, 4th Quart., 2015.

[58] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in OFDMA cellular networks: A survey and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1885–1922, 4th Quart., 2015.

[59] S. T. Shah, S. F. Hasan, B.-C. Seet, P. H. J. Chong, and M. Y. Chung, "Device-to-device communications: A contemporary survey," *Wireless Pers. Commun.*, vol. 98, no. 1, pp. 1247–1284, Jan. 2018.

[60] P. Gandotra, R. K. Jha, and S. Jain, "A survey on device-to-device (D2D) communication: Architecture and security issues," *J. Netw. Comput. Appl.*, vol. 78, pp. 9–29, Jan. 2017.

[61] O. Hayat, R. Ngah, and Y. Zahedi, "In-band device to device (D2D) communication and device discovery: A survey," *Wireless Pers. Commun.*, vol. 106, no. 2, pp. 451–472, May 2019.

[62] O. A. Amodu, M. Othman, N. K. Noordin, and I. Ahmad, "Relay-assisted D2D underlay cellular network analysis using stochastic geometry: Overview and future directions," *IEEE Access*, vol. 7, pp. 115023–115051, 2019.

[63] M. Waqas, Y. Niu, Y. Li, M. Ahmed, D. Jin, S. Chen, and Z. Han, "Mobility-aware device-to-device communications: Principles, practice and challenges," *IEEE Commun. Surveys Tuts.*, to be published.

[64] M. Noura and R. Nordin, "A survey on interference management for device-to-device (D2D) communication and its challenges in 5G networks," *J. Netw. Comput. Appl.*, vol. 71, pp. 130–150, Aug. 2016.

[65] K. Zia, N. Javed, M. N. Sial, S. Ahmed, H. Iram, and A. A. Pirzada, "A survey of conventional and artificial intelligence/learning based resource allocation and interference mitigation schemes in D2D enabled networks," 2018, *arXiv:1809.08748*. [Online]. Available: https://arxiv.org/abs/1809.08748

[66] X. Cheng, Y. Li, B. Ai, X. Yin, and Q. Wang, "Device-to-device channel measurements and models: A survey," *IET Commun.*, vol. 9, no. 3, pp. 312–325, 2015.

[67] T. Wang, P. Li, X. Wang, Y. Wang, T. Guo, and Y. Cao, "A comprehensive survey on mobile data offloading in heterogeneous network," *Wireless Netw.*, vol. 25, no. 2, pp. 573–584, 2019.

[68] M. Haus, M. Waqas, A. Y. Ding, Y. Li, S. Tarkoma, and J. Ott, "Security and privacy in device-to-device (D2D) communication: A review," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1054–1079, 2nd Quart., 2017.

[69] M. Wang and Z. Yan, "A survey on security in D2D communications," *Mobile Netw. Appl.*, vol. 22, no. 2, pp. 195–208, Apr. 2017.

[70] M. Ahmed, Y. Li, M. Waqas, M. Sheraz, D. Jin, and Z. Han, "A survey on socially aware device-to-device communications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2169–2197, 3rd Quart., 2018.

[71] M. Nitti, G. A. Stelea, V. Popescu, and M. Fadda, "When social networks meet D2D communications: A survey," *Sensors*, vol. 19, no. 2, pp. 396-1–396-16, 2019.

[72] H. T. Mouftah, M. Erol-Kantarci, and M. H. Rehmani, "5G and D2D communications at the service of smart cities," in *Transportation and Power Grid in Smart Cities: Communication Networks and Services*. Hoboken, NJ, USA: Wiley, 2019, pp. 147–169.

[73] A. Boukerche and R. E. De Grande, "Vehicular cloud computing: Architectures, applications, and mobility," *Comput. Netw.*, vol. 135, pp. 171–189, Apr. 2017.

[74] N. Fernando, S. W. Loke, and W. Rahayu, "Computing with nearby mobile devices: A work sharing algorithm for mobile edge-clouds," *IEEE Trans. Cloud Comput.*, vol. 7, no. 2, pp. 329–343, Apr. 2019.

[75] I. Yaqoob, E. Ahmed, A. Gani, S. Mokhtar, M. Imran, and S. Guizani, "Mobile ad hoc cloud: A survey," *Wireless Commun. Mobile Comput.*, vol. 16, no. 16, pp. 2572–2589, Nov. 2016.

[76] G. A. McGilvary, A. Barker, and M. Atkinson, "Ad hoc cloud computing," in *Proc. IEEE Int. Conf. Cloud Comput.*, New York, NY, USA, Jun. 2015, pp. 1063–1068.

[77] W. Zhang, H. Flores, and P. Hui, "Towards collaborative multi-device computing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2018, pp. 22–27.

[78] A. J. Ferrer, J. M. Marquéés, and J. Jorba, "Towards the decentralised cloud: Survey on approaches and challenges for mobile, ad hoc, and edge computing," *ACM Comput. Surv.*, vol. 51, no. 6, pp. 111-1–111-36, 2019.

[79] H. Guo, J. Liu, and J. Zhang, "Computation offloading for multi-access mobile edge computing in ultra-dense networks," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 14–19, Aug. 2018.

[80] Y. Zhang, H. Liu, L. Jiao, and X. Fu, "To offload or not to offload: An efficient code partition algorithm for mobile cloud computing," in *Proc. IEEE 1st Int. Conf. Cloud Netw. (CLOUDNET)*, Paris, France, Nov. 2012, pp. 80–86.

[81] H. Ahlehagh and S. Dey, "Hierarchical video caching in wireless cloud: Approaches and algorithms," in *Proc. IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, Jun. 2012, pp. 7082–7087.

[82] N. Fernando, S. W. Loke, and W. Rahayu, "Mobile cloud computing: A survey," *Future Gener. Comput. Syst.*, vol. 29, no. 1, pp. 84–106, 2013.

[83] H. Qi and A. Gani, "Research on mobile cloud computing: Review, trend and perspectives," in *Proc. Int. Conf. Digit. Inf. Commun. Technol. Appl.*, Bangkok, Thailand, May 2012, pp. 195–202.

[84] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: State-of-the-art and research challenges," *J. Internet Services Appl.*, vol. 1, no. 1, pp. 7–18, May 2010.

[85] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," ETSI Sophia Antipolis, France, White Paper, 2015, no. 11, pp. 1–16.

[86] I. Morris. (Sep. 2016). *ETSI Drops Mobile From MEC*. [Online]. Available: https://www.lightreading.com/mobile/mec-(mobile-edge-computing)/etsi-drops-mobile-from-mec/d/d-id/726273

[87] L. Chen, S. Zhou, and J. Xu, "Computation peer offloading for energy-constrained mobile edge computing in small-cell networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 4, pp. 1619–1632, Aug. 2018.

[88] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo, and M. Zhou, "MEETS: Maximal energy efficient task scheduling in homogeneous fog networks," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4076–4087, Oct. 2018.

[89] S. Bi and Y. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4177–4190, Jun. 2018.

[90] X. Hu, K.-K. Wong, and K. Yang, "Wireless powered cooperation-assisted mobile edge computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2375–2388, Apr. 2018.

[91] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1784–1797, Mar. 2018.

[92] B. Blanco, J. O. Fajardo, I. Giannoulakis, E. Kafetzakis, S. Peng, J. Pérez-Romero, I. Trajkovska, P. S. Khodashenas, L. Goratti, and M. Paolino, E. Sfakianakis, F. Liberal, and G. Xilouris, "Technology pillars in the architecture of future 5G mobile networks: NFV, MEC and SDN," *Comput. Standards Int.*, vol. 54, pp. 216–228, Nov. 2017.

[93] J. H. Cox, J. Chung, S. Donovan, J. Ivey, R. J. Clark, G. Riley, and H. L. Owen, "Advancing software-defined networks: A survey," *IEEE Access*, vol. 5, pp. 25487–25526, 2017.

[94] W. Kellerer, P. Kalmbach, A. Blenk, A. Basta, M. Reisslein, and S. Schmid, "Adaptable and data-driven softwarized networks: Review, opportunities, and challenges," *Proc. IEEE*, vol. 107, no. 4, pp. 711–731, Apr. 2019.

[95] S. Jeon, C. Guimarães, and R. L. Aguiar, "SDN: Based mobile networking for cellular operators," in *Proc. ACM Workshop Mobility Evolving Internet Archit.*, Maui, HI, USA, 2014, pp. 13–18.

[96] C. Huang, M. Chiang, D. Dao, W. Su, S. Xu, and H. Zhou, "V2V data offloading for cellular network based on the software defined network (SDN) inside mobile edge computing (MEC) architecture," *IEEE Access*, vol. 6, pp. 17741–17755, 2018.

[97] E. Schiller, N. Nikaein, E. Kalogeiton, M. Gasparyan, and T. Braun, "CDS-MEC: NFV/SDN-based application management for MEC in 5G systems," *Comput. Netw.*, vol. 135, pp. 96–107, Apr. 2018.

[98] Y. Li and M. Chen, "Software-defined network function virtualization: A survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[99] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 236–262, 1st Quart., 2016.

[100] A. U. Rehman, R. L. Aguiar, and J. P. Barraca, "Network functions virtualization: The long road to commercial deployments," *IEEE Access*, vol. 7, pp. 60439–60464, 2019.

[101] H. Hantouti, N. Benamar, T. Taleb, and A. Laghrissi, "Traffic steering for service function chaining," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 487–507, 1st Quart., 2018.

[102] A. M. Medhat, T. Taleb, A. Elmangoush, G. A. Carella, S. Covaci, and T. Magedanz, "Service function chaining in next generation networks: State of the art and research challenges," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 216–223, Feb. 2017.

[103] G. Sun, Z. Xu, H. Yu, X. Chen, V. Chang, and A. V. Vasilakos, "Low-latency and resource-efficient service function chaining orchestration in network function virtualization," *IEEE Internet Things*, to be published.

[104] Z. Zhou, Q. Wu, and X. Chen, "Online orchestration of cross-edge service function chaining for cost-efficient edge computing," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1866–1880, Aug. 2019.

[105] D. Kutscher, *Information-Centric Networking (ICN) Research Challenges*, document RFC 7927, Internet Research Task Force, 2016, pp. 1–38.

[106] G. Piro, L. A. Grieco, G. Boggia, and P. Chatzimisios, "Information-centric networking and multimedia services: Present and future challenges," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 4, pp. 392–406, Apr. 2014.

[107] A. V. Vasilakos, Z. Li, G. Simon, and W. You, "Information centric network: Research challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 52, pp. 1–10, Jun. 2015.

[108] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 77–83, Aug. 2016.

[109] G. Ahani and D. Yuan, "BS-assisted task offloading for D2D networks with presence of user mobility," in *Proc. IEEE Veh. Techn. Conf. (VTC-Spring)*, May 2019, pp. 1–5.

[110] Q. Lin, F. Wang, and J. Xu, "Optimal task offloading scheduling for energy efficient D2D cooperative computing," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1816–1820, Oct. 2019.

[111] J. Liu, K. Luo, Z. Zhou, and X. Chen, "ERP: Edge resource pooling for data stream mobile computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4355–4368, Jun. 2019.

[112] W. Xiao, X. Zhu, W. Bao, L. Liu, and J. Yao, "Cooperative data sharing for mobile cloudlets under heterogeneous environments," *IEEE Trans. Netw. Service Manage.*, vol. 16, no. 2, pp. 430–444, Jun. 2019.

[113] J. Xie, Y. Jia, Z. Chen, Z. Nan, and L. Liang, "D2D computation offloading optimization for precedence-constrained tasks in information-centric IoT," *IEEE Access*, vol. 7, pp. 94888–94898, 2019.

[114] H. Xing, L. Liu, J. Xu, and A. Nallanathan, "Joint task assignment and resource allocation for D2D-enabled mobile-edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4193–4207, Jun. 2019.

[115] C. You and K. Huang, "Exploiting non-causal CPU-state information for energy-efficient mobile cooperative computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4104–4117, Jun. 2018.

[116] K. N. Doan, T. V. Nguyen, H. Shin, and T. Q. S. Quek, "Socially-aware caching in wireless networks with random D2D communications," *IEEE Access*, vol. 7, pp. 58394–58406, 2019.

[117] J. Huang, C. Huang, C. Xing, Z. Chang, Y. Zhao, and Q. Zhao, "An energy-efficient communication scheme for collaborative mobile clouds in content sharing: Design and optimization," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5700–5707, Oct. 2019.

[118] W. Li, C. Wang, D. Li, B. Hu, X. Wang, and J. Ren, "Edge caching for D2D enabled hierarchical wireless networks with deep reinforcement learning," *Wireless Commun. Mobile Comput.*, vol. 2019, Feb. 2019, Art. no. 2561069.
[119] C. Ma, M. Ding, H. Chen, Z. Lin, G. Mao, Y.-C. Liang, and B. Vucetic, "Socially aware caching strategy in device-to-device communication networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4615–4629, May 2018.
[120] A. Said, S. W. H. Shah, H. Farooq, A. N. Mian, A. Imran, and J. Crowcroft, "Proactive caching at the edge leveraging influential user detection in cellular D2D networks," *Future Internet*, vol. 10, no. 10, pp. 93.1–93.17, Sep. 2018.
[121] S. Soleimani and X. Tao, "Caching and placement for in-network caching in device-to-device communications," *Wireless Commun. Mobile Comput.*, vol. 2018, Sep. 2018, Art. no. 9539502.
[122] N. Zhao, X. Liu, Y. Chen, Z. Li, B. Chen, and M.-S. Alouini, "Caching D2D connections in small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12326–12338, Dec. 2018.
[123] W. Zhang, D. Wu, W. Yang, and Y. Cai, "Caching on the move: A user interest-driven caching strategy for D2D content sharing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2958–2971, Mar. 2019.
[124] D. Wu, Q. Liu, H. Wang, D. Wu, and R. Wang, "Socially aware energy-efficient mobile edge collaboration for video distribution," *IEEE Trans. Multimedia*, vol. 19, no. 10, pp. 2197–2209, Oct. 2017.
[125] D. Wu, Q. Liu, H. Wang, Q. Yang, and R. Wang, "Cache less for more: Exploiting cooperative video caching and delivery in D2D communications," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1788–1798, Jul. 2019.
[126] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahmi, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, pp. 1510–1520, 2014.
[127] S. Mumtaz and J. Rodriguez, *Smart Device to Smart Device Communication*. New York, NY, USA: Springer, 2014.
[128] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 96–104, Jun. 2012.
[129] M. Höyhtyä, O. Apilo, and M. Lasanen, "Review of latest advances in 3GPP standardization: D2D communication in 5G systems and its energy consumption models," *Future Internet*, vol. 10, no. 1, p. 3, Jan. 2018.
[130] A. Masmoudi, S. Feki, K. Mnif, and F. Zarai, "Efficient scheduling and resource allocation for D2D-based LTE-V2X communications," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 496–501.
[131] G. Nardini, A. Virdis, C. Campolo, A. Molinaro, and G. Stea, "Cellular-V2X communications for platooning: Design and evaluation," *Sensors*, vol. 18, no. 5, pp. 1527.1–1527.22, 2018.
[132] S. Singh, J. Lianghai, D. Calabuig, D. Garcia-Roger, N. H. Mahmood, N. Pratas, T. Mach, and M. C. De Gennaro, "D2D and V2X communications," in *5G System Design: Architectural and Functional Considerations and Long Term Research*. Hoboken, NJ, USA: Wiley 2018, pp. 409–449.
[133] T. Bertram, *Fahrerassistenzsysteme 2018: Von der Assistenz zum Automatisierten Fahren 4. Internationale ATZ-Fachtagung Automatisiertes Fahren*. Wiesbaden, Germany: Springer, 2019.
[134] A. Feroz, N. Kavitha, M. Kasthuri, and R. Sudha, "Vehicle to vehicle communication for collision avoidance," *Int. J. Emerg. Technol. Innov. Eng.*, vol. 5, no. 7, pp. 544–549, 2019.
[135] Y. Cao, T. Jiang, O. Kaiwartya, H. Sun, H. Zhou, and R. Wang, "Toward pre-empted EV charging recommendation through V2V-based reservation system," *IEEE Trans. Syst., Man, Cybern. Syst.*, to be published.
[136] D. Satria, D. Park, and M. Jo, "Recovery for overloaded mobile edge computing," *Future Gener. Comput. Syst.*, vol. 70, pp. 138–147, May 2017.
[137] H. Xing, L. Liu, J. Xu, and A. Nallanathan, "Joint task assignment and wireless resource allocation for cooperative mobile-edge computing," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, May 2018. pp. 1–6.
[138] S. Gupta and A. Lozano, "Computation-bandwidth trading for mobile edge computing," in *Proc. IEEE Annu. Consum. Commun. Netw. Conf.*, Las Vegas, NV, USA, Jan. 2019, pp. 1–6.
[139] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for mobile edge computing," in *Proc. 16th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw.*, Shanghai, China, May 2018, pp. 1–6.
[140] X. Chen, L. Pu, L. Gao, W. Wu, and D. Wu, "Exploiting massive D2D collaboration for energy-efficient mobile edge computing," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 64–71, Aug. 2017.
[141] G. Hu, Y. Jia, and Z. Chen, "Multi-user computation offloading with D2D for mobile edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
[142] C. Wang, J. Qin, X. Yang, and W. Wen, "Energy-efficient offloading policy in D2D underlay communication integrated with MEC service," in *Proc. Int. Conf. High Perf. Compilation, Comput. Commun.*, 2019, pp. 159–164.
[143] L. Pu, X. Chen, J. Xu, and X. Fu, "D2D fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted D2D collaboration," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3887–3901, Dec. 2016.
[144] Q. Jia, R. Xie, Q. Tang, X. Li, T. Huang, J. Liu, and Y. Liu, "Energy-efficient computation offloading in 5G cellular networks with edge computing and D2D communications," *IET Commun.*, vol. 13, no. 8, pp. 1122–1130, May 2019.
[145] G. Qiao, S. Leng, and Y. Zhang, "Online learning and optimization for computation offloading in D2D edge computing and networks," *Mobile Netw. Appl.*, 2019, pp. 1–2.
[146] D. Wang, Y. Lan, T. Zhao, Z. Yin, and X. Wang, "On the design of computation offloading in cache-aided D2D multicast networks," *IEEE Access*, vol. 6, pp. 63426–63441, 2018.
[147] J. Lin, R. Chai, M. Chen, and Q. Chen, "Task execution cost minimization-based joint computation offloading and resource allocation for cellular D2D systems," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Bologna, Italy, Sep. 2018, pp. 1–5.
[148] X. Chen and J. Zhang, "When D2D meets cloud: Hybrid mobile task offloadings in fog computing," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.
[149] A. Orsino, I. Farris, L. Militano, G. Araniti, S. Andreev, I. Gudkova, Y. Koucheryavy, and A. Iera, "Exploiting D2D communications at the network edge for mission-critical IoT applications," in *Proc. Eur. Wireless Conf.*, 2017, pp. 68–73.
[150] R. Ranji, A. M. Mansoor, and A. A. Sani, "EEDOS: An energy-efficient and delay-aware offloading scheme based on device to device collaboration in mobile edge computing," *Telecommun. Syst.*, pp. 1–12, Aug. 2019.
[151] K. Sucipto, D. Chatzopoulos, S. Kostap, and P. Hui, "Keep your nice friends close, but your rich friends closer–computation offloading using NFC," in *Proc. IEEE Conf. Comput. Commun.*, Atlanta, GA, USA, May 2017, pp. 1–9.
[152] H. Flores, Rajesh Sharma, D. Ferreira, V. Kostakos, J. Manner, S. Tarkoma, P. Hui, and Y. Li, "Social-aware hybrid mobile offloading," *Pervasive Mobile Comput.*, vol. 36, pp. 25–43, Apr. 2017.
[153] Y. He, J. Ren, G. Yu, and Y. Cai, "D2D communications meet mobile edge computing for enhanced computation capacity in cellular networks," *IEEE Trans. Commun.*, vol. 18, no. 3, pp. 1750–1763, Feb. 2019.
[154] A. Mtibaa, K. Harras, and H. Alnuweiri, "Friend or foe? Detecting and isolating malicious nodes in mobile edge computing platforms," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.*, Vancouver, BC, Canada, Nov. 2015, pp. 42–49.
[155] J. Xu, L. Chen, K. Liu, and C. Shen, "Designing security-aware incentives for computation offloading via device-to-device communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6053–6066, Sep. 2018.
[156] X. Chen, Z. Zhou, W. Wu, D. Wu, and J. Zhang, "Socially-motivated cooperative mobile edge computing," *IEEE Netw.*, vol. 32, no. 6, pp. 177–183, Nov. 2018.
[157] Y. Gao, W. Tang, M. Wu, P. Yang, and L. Dan, "Dynamic social-aware computation offloading for low-latency communications in IoT," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7864–7877, Oct. 2019.
[158] A. A. Ateya, A. Muthanna, and A. Koucheryavy, "5G framework based on multi-level edge computing with D2D enabled communication," in *Proc. Int. Conf. Adv. Commun. Techn. (ICACT)*, Feb. 2018, pp. 507–512.
[159] F. Wang, J. Xu, and Z. Ding, "Optimized multiuser computation offloading with multi-antenna NOMA," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshop*, Singapore, Dec. 2017, pp. 1–7.
[160] F. Jiang, H. Wang, H. Ren, and S. Xu, "Energy-efficient resource and power allocation for underlay multicast device-to-device transmission," *Future Internet*, vol. 9, no. 4, p. 84, Nov. 2017.
[161] G. Scutari, S. Barbarossa, and D. P. Palomar, "Potential games: A framework for vector power control problems with coupled constraints," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 4, May 2006, pp. 241–244.

[162] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," in *Proc. ACM SIGMOBILE Workshop Mobility Models*, 2008, pp. 33–40.

[163] C.-Y. Wang, Y. Chen, and K. R. Liu, "Chinese restaurant game," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 898–901, Dec. 2012.

[164] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.

[165] S. Even, G. Even, and R. M. Karp, *Graph Algorithms*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[166] H. Shah-Mansouri and V. W. S. Wong, "Hierarchical fog-cloud computing for IoT systems: A computation offloading game," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3246–3257, Aug. 2018.

[167] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*. Boston, MA, USA: Springer, 1996, pp. 153–181.

[168] Y. He, F. R. Yu, N. Zhao, and H. Yin, "Secure social networks in 5G systems with mobile edge computing, caching, and device-to-device communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 103–109, Jun. 2018.

[169] H. Zhou, C. M. V. Leung, C. Zhu, S. Xu, and J. Fan, "Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10372–10383, Nov. 2017.

[170] H. Flores, P. Nurmi, and P. Hui, "AI on the move: From on-device to on-multi-device," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2019, pp. 310–315.

[171] Z. Han, D. Niyato, W. Saad, T. Ba ar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[172] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Computation: The Micro and the Macro View*. Hackensack, NJ, USA: World Scientific, 1992, pp. 71–102.

[173] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *Proc. Int. Symp. Reliable Distrib. Syst.*, Florence, Italy, 2003, pp. 25–34.

[174] N. T. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*. High Wycombe, U.K.: Griffin, 1975.

[175] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, Feb. 2009.

[176] B. Gu, Y. P. Chen, H. J. Liao, Z. Y. Zhou, and D. Zhang, "A distributed and context-aware task assignment mechanism for collaborative mobile edge computing," *Sensors*, vol. 18, no. 8, p. 2423, Jul. 2018.

[177] X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu, and Y. J. Guo, "Multi-timescale decentralized online orchestration of software-defined networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 12, pp. 2716–2730, Dec. 2018.

[178] N. Prasad, M. Arslan, and S. Rangarajan, "A two time scale approach for coordinated multi-point transmission and reception over practical backhaul," in *Proc. Int. Conf. Commun. Sys. Netw. (COMSNETS)*, 2014, pp. 1–8.

[179] P. Shantharama, A. Thyagaturu, N. Karakoc, L. Ferrari, M. Reisslein, and A. Scaglione, "LayBack: SDN management of multi-access edge computing (MEC) for network access services and radio resource sharing," *IEEE Access*, vol. 6, pp. 57545–57561, 2018.

[180] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.

[181] M. Wang, N. Karakoc, L. Ferrari, P. Shantharama, A. S. Thyagaturu, M. Reisslein, and A. Scaglione, "A multi-layer multi-timescale network utility maximization framework for the SDN-based LayBack architecture enabling wireless backhaul resource sharing," *Electronics*, vol. 8, no. 9, pp. 937-1–937-28, 2019.

[182] W. Xia, T. Q. S. Quek, J. Zhang, S. Jin, and H. Zhu, "Programmable hierarchical C-RAN: From task scheduling to resource allocation," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 2003–2016, Mar. 2019.

[183] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, Jul. 2014.

[184] M. Chen, Y. Hao, M. Qiu, J. Song, D. Wu, and I. Humar, "Mobility-aware caching and computation offloading in 5G ultra-dense cellular networks," *Sensors*, vol. 16, no. 7, p. 974, 2016.

[185] Y. Wang, X. Tao, X. Zhang, and Y. Gu, "Cooperative caching placement in cache-enabled D2D underlaid cellular network," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1151–1154, May 2017.

[186] T. Le Thanh, R. Q. Hu, and Y. Qian, "D2D communications in heterogeneous networks with full-duplex relays and edge caching," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4557–4567, Oct. 2018.

[187] L. T. Tan, R. Q. Hu, and L. Hanzo, "Heterogeneous networks relying on full-duplex relays and mobility-aware probabilistic caching," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5037–5052, Jul. 2019.

[188] X. Wang, S. Leng, X. Liu, Q. Zhao, K. Wangy, and K. Yang, "Fog computing aided multi-view video in mobile social networks," in *Proc. Int. Conf. Adv. Infocomm Technol.*, Chengdu, China, Nov. 2017, pp. 361–366.

[189] D. Wang, H. Qin, B. Song, X. Du, and M. Guizani, "Resource allocation in information-centric wireless networking with D2D-enabled MEC: A deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 114935–114944, 2019.

[190] S. Wang, X. Huang, Y. Liu, and R. Yu, "CachinMobile: An energy-efficient users caching scheme for fog computing," in *Proc. IEEE/CIC Int. Conf. Commun. China*, Chengdu, China, Jul. 2016, pp. 1–6.

[191] B. Chen, C. Yang, and A. F. Molisch, "Cache-enabled device-to-device communications: Offloading gain and energy cost," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4519–4536, Jul. 2017.

[192] J. Chuan, L. Wang, and J. Wu, "Belief propagation based distributed content delivery scheme in caching-enabled D2D networks," in *Proc. IEEE Int. Conf. Commun.*, Shanghai, China, May 2019, pp. 1–5.

[193] M.-C. Lee and A. F. Molisch, "Caching policy and cooperation distance design for base station-assisted wireless D2D caching networks: Throughput and energy efficiency optimization and tradeoff," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7500–7514, Nov. 2018.

[194] P. Lin, Q. Song, Y. Yu, and A. Jamalipour, "Extensive cooperative caching in D2D integrated cellular networks," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2101–2104, Sep. 2017.

[195] Y. Niu, X. Qin, and Z. Zhang, "A learning-based cooperative caching strategy in D2D assisted cellular networks," in *Proc. Asia–Pacific Conf. Commun.*, Ningbo, China, Nov. 2018, pp. 269–274.

[196] L. Shi, L. Zhao, G. Zheng, Z. Han, and Y. Ye, "Incentive design for cache-enabled D2D underlaid cellular networks using Stackelberg Game," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 765–779, Jan. 2019.

[197] Y. Zhang, Y. Xu, Q. Wu, X. Liu, K. Yao, and A. Anpalagan, "A game-theoretic approach for optimal distributed cooperative hybrid caching in D2D networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 324–327, Jun. 2018.

[198] D. Wang, H. Li, S. Wang, J. Yang, X. Liu, and X. Zhang, "Mobility aware caching incentive scheme for D2D cellular networks," in *Proc. IEEE Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[199] S. Wang, X. Zhang, L. Wang, J. Yang, and W. Wang, "Joint design of device to device caching strategy and incentive scheme in mobile edge networks," *IET Commun.*, vol. 12, no. 14, pp. 1728–1736, Aug. 2018.

[200] X. Li, X. Wang, P.-J. Wan, Z. Han, and V. C. M. Leung, "Hierarchical edge caching in device-to-device aided mobile networks: Modeling, optimization, and design," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1768–1785, Aug. 2018.

[201] E. Baccour, A. Erbad, A. Mohamed, and M. Guizani, "CE-D2D: Dual framework chunks caching and offloading in collaborative edge networks with D2D communication," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 1550–1556.

[202] Y. Lan, X. Wang, D. Wang, Z. Liu, and Y. Zhang, "Task caching, offloading, and resource allocation in D2D-aided fog computing networks," *IEEE Access*, vol. 7, pp. 104876–104891, 2019.

[203] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[204] Y. Niu, Y. Liu, Y. Li, Z. Zhong, B. Ai, and P. Hui, "Mobility-aware caching scheduling for fog computing in mmWave band," *IEEE Access*, vol. 6, pp. 69358–69370, 2018.

[205] B. Chen and C. Yang, "Caching policy for cache-enabled D2D communications by learning user preference," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6586–6601, Dec. 2018.

[206] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[207] P. Zhang, X. Kang, Y. Liu, and H. Yang, "Cooperative willingness aware collaborative caching mechanism towards cellular D2D communication," *IEEE Access*, vol. 6, pp. 67046–67056, 2018.
[208] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, Dec. 2007.
[209] L. A. Adamic and B. A. Huberman, "Zipf's law and the Internet," *Glottometrics*, vol. 3, no. 1, pp. 143–150, 2002.
[210] G. K. Zipf, *Human Behavior Principle Least Effort: An Introduction to Human Ecology*. Boston, MA, USA: Addison-Wesley, 1949.
[211] M. Hu, B. Liu, H. Huang, W. Xiang, and Y. Tao, "Social-aware relay selection scheme for relay-based D2D communications," *IEEE Access*, vol. 6, pp. 73293–73304, 2018.
[212] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *Proc. Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2009, pp. 617–621.
[213] X. Cao, L. Liu, Y. Cheng, and X. S. Shen, "Towards energy-efficient wireless networking in the big data era: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 303–332, 1st Quart., 2018.
[214] M. Wainwright, T. Jaakkola, and A. Willsky, "MAP estimation via agreement on (hyper) trees: Message-passing and linear programming approaches," in *Proc. Allerton Conf. Commun. Control Comput.*, 2002, vol. 40, no. 3, pp. 1565–1575.
[215] M. Ji, G. Caire, and A. F. Molisch, "The throughput-outage tradeoff of wireless one-hop caching networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6833–6859, Oct. 2015.
[216] M. Ehrgott, *Multicriteria Optimization*, 2nd ed. Berlin, Germany: Springer, 2005.
[217] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
[218] Y. Song, C. Yang, J. Shen, and L. Chang, "Joint delay and energy management for cache-enabled ultra-dense cellular networks: A game-theoretic learning," in *Proc. IEEE Int. Conf. Commun. Syst.*, Chengdu, China, Dec. 2018, pp. 401–406.
[219] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 483–502, Sep. 2002.
[220] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE Conf. Comput. Commun.*, New York, NY, USA, vol. 1, 1999, pp. 126–134.
[221] Z. Chen, Y. Liu, B. Zhou, and M. Tao, "Caching incentive design in wireless D2D networks: A Stackelberg game approach," in *Proc. IEEE Int. Conf. Commun.*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[222] X. Wu, Q. Li, X. Li, V. Leung, and P. Ching, "Joint long-term cache allocation and short-term content delivery in green cloud small cell networks," 2019, *arXiv:1904.10882*. [Online]. Available: https://arxiv.org/abs/1904.10882
[223] Y. Hao, M. Chen, L. Hu, M. S. Hossain, and A. Ghoneim, "Energy efficient task caching and offloading for mobile edge computing," *IEEE Access*, vol. 6, pp. 11365–11373, 2018.
[224] A. Ndikumana, S. Ullah, T. LeAnh, N. H. Tran, and C. S. Hong, "Collaborative cache allocation and computation offloading in mobile edge computing," in *Proc. IEEE Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2017, pp. 366–369.
[225] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, and M. Song, "Joint computation offloading and content caching for wireless blockchain networks," in *Proc. IEEE INFOCOM Workshops*, Apr. 2018, pp. 517–522.
[226] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, "Computation offloading and resource allocation in wireless cellular networks with mobile edge computing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4924–4938, Aug. 2017.
[227] M. E. J. Newman, *Networks: An Introduction*. Oxford, U.K.: Oxford Univ. Press, 2016.
[228] S. Wang, Y. Zhang, H. Wang, Z. Huang, X. Wang, and T. Jiang, "Large scale measurement and analytics on social groups of device-to-device sharing in mobile social networks," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 203–215, Apr. 2018.
[229] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2094–2100.
[230] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.
[231] Y. Niu, C. Gao, Y. Li, L. Su, and D. Jin, "Exploiting multi-hop relaying to overcome blockage in directional mmwave small cells," *J. Commun. Netw.*, vol. 18, no. 3, pp. 364–374, Jun. 2016.
[232] T. Hofmann, "Probabilistic latent semantic analysis," in *Proc. 15th Conf. Uncertainty Artif. Intell.*, 1999, pp. 289–296.
[233] F. N. John, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, Sep. 1951.
[234] A. Antonopoulos, E. Kartsakli, and C. Verikoukis, "Game theoretic D2D content dissemination in 4G cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 125–132, Jun. 2014.
[235] H. Elsawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, Jun. 2013.
[236] F. Bannour, S. Souihi, and A. Mellouk, "Distributed SDN control: Survey, taxonomy, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 333–354, 1st Quart., 2018.
[237] M. Karakus and A. Durresi, "A survey: Control plane scalability issues and approaches in software-defined networking (SDN)," *Comput. Netw.*, vol. 112, pp. 279–293, Jan. 2017.
[238] M. Alsaeedi, M. M. Mohamad, and A. A. Al-Roubaiey, "Toward adaptive and scalable OpenFlow-SDN flow control: A survey," *IEEE Access*, vol. 7, pp. 107346–107379, 2019.
[239] I. F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "A roadmap for traffic engineering in SDN-OpenFlow networks," *Comput. Netw.*, vol. 71, pp. 1–30, Oct. 2014.
[240] A. Mendiola, J. Astorga, E. Jacob, and M. Higuero, "A survey on the contributions of software-defined networking to traffic engineering," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 918–953, 2nd Quart., 2016.
[241] Z. N. Abdullah, I. Ahmad, and I. Hussain, "Segment routing in software defined networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 464–486, 1st Quart., 2018.
[242] J. W. Guck, A. Van Bemten, M. Reisslein, and W. Kellerer, "Unicast QoS routing algorithms for SDN: A comprehensive survey and performance evaluation," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 388–415, 1st Quart., 2018.
[243] F. Y. Okay and S. Ozdemir, "Routing in fog-enabled IoT platforms: A survey and an SDN-based solution," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4871–4889, Dec. 2018.
[244] N. Bizanis and F. Kuipers, "SDN and virtualization solutions for the Internet of Things: A survey," *IEEE Access*, vol. 4, pp. 5591–5606, 2016.
[245] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging SDN and NFV security mechanisms for IoT systems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 812–837, 1st Quart., 2019.
[246] X. Costa-Perez, A. Garcia-Saavedra, X. Li, T. Deiss, A. de la Oliva, A. di Giglio, P. Iovanna, and A. Moored, "5G-Crosshaul: An SDN/NFV integrated fronthaul/backhaul transport network architecture," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 38–45, Feb. 2017.
[247] I. Elgendi, K. S. Munasinghe, D. Sharma, and A. Jamalipour, "Traffic offloading techniques for 5G cellular: A three-tiered SDN architecture," *Ann. Telecommun.*, vol. 71, nos. 11–12, pp. 583–593, 2016.
[248] D. King, A. Farrel, E. N. King, R. Casellas, L. Velasco, R. Nejabati, and A. Lord, "The dichotomy of distributed and centralized control: METRO-HAUL, when control planes collide for 5G networks," *Opt. Switching Netw.*, vol. 33, pp. 49–55, Jul. 2019.
[249] J. de C. Silva, J. J. P. C. Rodrigues, J. Al-Muhtadi, R. A. L. Rabêlo, and V. Furtado, "Management platforms and protocols for Internet of Things: A survey," *Sensors*, vol. 19, no. 3, p. 676, 2019. [Online]. Available: http://www.mdpi.com/1424-8220/19/3/676
[250] A. S. Thyagaturu, Y. Dashti, and M. Reisslein, "SDN-based smart gateways (Sm-GWs) for multi-operator small cell network management," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 4, pp. 740–753, Dec. 2016.
[251] A. A. Abbasi, A. Abbasi, S. Shamshirband, A. T. Chronopoulos, V. Persico, and A. Pescapè, "Software-defined cloud computing: A systematic review on latest trends and developments," *IEEE Access*, vol. 7, pp. 93294–93314, 2019.
[252] A. Binsahaq, T. R. Sheltami, and K. Salah, "A survey on autonomic provisioning and management of QoS in SDN networks," *IEEE Access*, vol. 7, pp. 73384–73435, 2019.

[253] A. A. Neghabi, N. J. Navimipour, M. Hosseinzadeh, and A. Rezaee, "Load balancing mechanisms in the software defined networks: A systematic and comprehensive review of the literature," *IEEE Access*, vol. 6, pp. 14159–14178, 2018.

[254] R. Amin, M. Reisslein, and N. Shah, "Hybrid SDN networks: A survey of existing approaches," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3259–3306, 4th Quart., 2018.

[255] X. Huang, S. Cheng, K. Cao, P. Cong, T. Wei, and S. Hu, "A survey of deployment solutions and optimization strategies for hybrid SDN networks," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1483–1507, 2nd Quart., 2018.

[256] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, 1st Quart., 2019.

[257] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3098–3130, 4th Quart. 2018.

[258] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.

[259] D. You and D. H. Kim, "Multi-device-to-multi-device communication in cellular network for efficient contents distribution," in *Proc. IEEE Int. Conf. Consum. Electron.*, Las Vegas, NV, USA, Jan. 2014, pp. 244–247.

[260] D. You, T. V. Doan, R. Torre, M. Mehrabi, A. Kropp, V. Nguyen, H. Salah, G. T. Nguyen, and F. H. P. Fitzek, "Fog computing as an enabler for immersive media: Service scenarios and research opportunities," *IEEE Access*, vol. 7, pp. 65797–65810, 2019.

[261] T. Hou, G. Feng, S. Qin, and W. Jiang, "Proactive content caching by exploiting transfer learning for mobile edge computing," *Int. J. Commun. Syst.*, vol. 31, no. 11, Jul. 2018, Art. no. e3706.

[262] A. Orsino, D. Moltchanov, M. Gapeyenko, A. Samuylov, S. Andreev, L. Militano, G. Araniti, and Y. Koucheryavy, "Direct connection on the move: Characterization of user mobility in cellular-assisted D2D systems," *IEEE Veh. Technol. Mag.*, vol. 11, no. 3, pp. 38–48, Sep. 2016.

[263] B. Zhang, Y. Li, D. Jin, P. Hui, and Z. Han, "Social-aware peer discovery for D2D communications underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2426–2439, May 2015.

[264] R. I. Ansari, C. Chrysostomou, S. A. Hassan, M. Guizani, S. Mumtaz, J. Rodriguez, and J. J. P. C. Rodrigues, "5G D2D networks: Techniques, challenges, and future prospects," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3970–3984, Dec. 2018.

[265] E. Datsika, A. Antonopoulos, N. Zorba, and C. Verikoukis, "Green cooperative device–to–device communication: A social–aware perspective," *IEEE Access*, vol. 4, pp. 3697–3707, 2016.

[266] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE Access*, vol. 6, pp. 32328–32338, 2018.

[267] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[268] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts.*, to be published.

[269] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 393–430, 1st Quart., 2019.

[270] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[271] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," *IEEE Access*, vol. 6, pp. 24411–24432, 2018.

[272] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[273] Z. Fan, X. Gu, S. Nie, and M. Chen, "D2D power control based on supervised and unsupervised learning," in *Proc. IEEE Int. Conf. Comput. Commun.*, Chengdu, China, Dec. 2017, pp. 558–563.

[274] J. Xu, X. Gu, and Z. Fan, "D2D power control based on hierarchical extreme learning machine," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Bologna, Italy, Sep. 2018, pp. 1–7.

[275] F. Gabriel, S. Wunderlich, S. Pandi, F. H. Fitzek, and M. Reisslein, "Caterpillar RLNC with feedback (CRLNC-FB): Reducing delay in selective repeat ARQ through coding," *IEEE Access*, vol. 6, pp. 44787–44802, 2018.

[276] J. Huang, Y. Liao, C.-C. Xing, and Z. Chang, "Multi-hop D2D communications with network coding: From a performance perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2270–2282, Mar. 2019.

[277] N. Ma and M. Diao, "CoFi: Coding-assisted file distribution over a wireless LAN," *Symmetry*, vol. 11, no. 1, pp. 71:1–71:19, 2019.

[278] S. Pandi, F. Gabriel, J. A. Cabrera, S. Wunderlich, M. Reisslein, and F. H. Fitzek, "PACE: Redundancy engineering in RLNC for low-latency communication," *IEEE Access*, vol. 5, pp. 20477–20493, 2017.

[279] R. Torrea-Duran, M. M. Céspedes, J. Plata-Chaves, L. Vandendorpe, and M. Moonen, "Topology-aware space-time network coding in cellular networks," *IEEE Access*, vol. 6, pp. 7565–7578, 2018.

[280] S. Wunderlich, F. H. Fitzek, and M. Reisslein, "Progressive multicore RLNC decoding with online DAG scheduling," *IEEE Access*, vol. 7, pp. 161184–161200, 2019.

[281] F. H. P. Fitzek and M. Reisslein, "A prefetching protocol for continuous media streaming in wireless environments," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2015–2028, Oct. 2001.

[282] J. Kim, G. Caire, and A. F. Molisch, "Quality-aware streaming and scheduling for device-to-device video delivery," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 2319–2331, Apr. 2016.

[283] S. Li, M. A. Maddah-Ali, and A. S. Avestimehr, "Coding for distributed fog computing," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 34–40, Apr. 2017.

[284] K. Miller, D. Bethanabhotla, G. Caire, and A. Wolisz, "A control-theoretic approach to adaptive video streaming in dense wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1309–1322, Aug. 2015.

[285] M. Reisslein and K. W. Ross, "High-performance prefetching protocols for VBR prerecorded video," *IEEE Netw.*, vol. 12, no. 6, pp. 46–55, Nov. 1998.

[286] K. Li, M. Tao, and Z. Chen, "Exploiting computation replication for mobile edge computing: A fundamental computation-communication tradeoff study," 2019, *arXiv:1903.10837*. [Online]. Available: https://arxiv.org/abs/1903.10837

[287] S. Li, M. A. Maddah-Ali, Q. Yu, and A. S. Avestimehr, "A fundamental tradeoff between computation and communication in distributed computing," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 109–128, Jan. 2018.

[288] A. Reisizadeh, S. Prakash, R. Pedarsani, and A. S. Avestimehr, "Coded computation over heterogeneous clusters," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4227–4242, Jul. 2019.

[289] Y. N. Shnaiwer, S. Sorour, T. Y. Al-Naffouri, and S. N. Al-Ghadhban, "Opportunistic network coding-assisted cloud offloading in heterogeneous fog radio access networks," *IEEE Access*, vol. 7, pp. 56147–56162, 2019.

[290] Q. Yu, N. Raviv, J. So, and A. S. Avestimehr, "Lagrange coded computing: Optimal design for resiliency, security and privacy," 2018, *arXiv:1806.00939*. [Online]. Available: https://arxiv.org/abs/1806.00939

[291] A. Ahmed, H. Shan, and A. Huang, "Modeling the delivery of coded packets in D2D mobile caching networks," *IEEE Access*, vol. 7, pp. 20091–20105, 2019.

[292] W. He, Y. Su, X. Xu, Z. Luo, L. Huang, and X. Du, "Cooperative content caching for mobile edge computing with network coding," *IEEE Access*, vol. 7, pp. 67695–67707, 2019.

[293] Z. Lin, Y. Wang, Y. Lin, L. Wu, and Z. Chen, "Analysis and optimization of RLNC-based cache placement in 5G D2D networks," *IEEE Access*, vol. 6, pp. 65179–65188, 2018.

[294] R. Spreitzer, V. Moonsamy, T. Korak, and S. Mangard, "Systematic classification of side-channel attacks: A case study for mobile devices," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 465–488, 1st Quart., 2017.

[295] M. Peng and K. Zhang, "Recent advances in fog radio access networks: Performance analysis and radio resource allocation," *IEEE Access*, vol. 4, pp. 5003–5009, 2016.

[296] D. Naylor, K. Schomp, M. Varvello, I. Leontiadis, J. Blackburn, D. R. López, K. Papagiannaki, P. Rodriguez Rodriguez, and P. Steenkiste, "Multi-context TLS (mcTLS): Enabling secure in-network functionality in TLS," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 199–212, 2015.

[297] G. Paschos, E. Bastug, I. Land, G. Caire, and M. Debbah, "Wireless caching: Technical misconceptions and business barriers," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 16–22, Aug. 2016.

MAHSHID MEHRABI received the B.Sc. and M.Sc. degrees in electronics from Zanjan and Amirkabir University of Technology, Iran, in 2011 and 2015, respectively. She is currently a Doctoral Researcher at the Deutsche Telekom Chair of Communication Networks, Technical University Dresden, Germany. Her current research interests include 5G technologies for energy-efficient and low latency communication. Her research focuses on mobile edge computing and computational offloading schemes.

DONGHO YOU received the B.S., M.S., and Ph.D. degrees from the Seoul National University of Science and Technology, South Korea, in 2012, 2014, and 2018, respectively, all in media IT engineering. He was a Visiting Researcher with the IoT Convergence Research Center, Korea Electronics Technology Institute, in 2014. He is currently a Postdoctoral Researcher at the Deutsche Telekom Chair of Communication Networks, Technical University Dresden, Germany, and also conducting research as a Research Director of the Fostering Next-generation Researchers Program of the National Research Foundation of Korea (2018–2019). His current research interests include 5G technologies for robust and reliable immersive media communications.

VINCENT LATZKO received the Diploma degree in electrical engineering from Technische Universität Darmstadt, Germany, in 2017. He is currently pursuing the Ph.D. degree with the Deutsche Telekom Chair of Communication Networks, Technical University Dresden, Germany. His current research interests include efficient compression, softwarized network infrastructure, and reliable low latency communication. His main research focus is on joining artificial intelligence in multi-access edge computing and optimized vehicular traffic flows.

HANI SALAH received the B.Sc. degree in computer systems engineering from Palestine Polytechnic University, the M.Sc. degree in internetworking from the Royal Institute of Technology (KTH), Sweden, and the Ph.D. degree in computer science from TU Darmstadt, in 2015, specialized in measurements, performance, and security of large-scale distributed systems, under the supervision of Prof. Dr. T. Strufe. He is currently a Senior Researcher with the Deutsche Telekom Chair of Communication Networks, TU Dresden, Dresden, Germany. From September 2016 and June 2018, he worked as a Lecturer and also a Researcher with the Palestine Polytechnic University, Hebron, Palestine. Previously, he has worked as the Research and Development Employee with TU Darmstadt, Darmstadt, Germany, and flexOptix GmbH, Dietzenbach, Germany, for 1.5 years, as a Research Assistant at TU Darmstadt, for four years, and as a Lecturer and Technical Assistant with the Palestine Polytechnic University, for about 5.5 years. His current research interests lie in the areas of distributed systems and modern networking technologies (SDN, NFV, ICN), with the emphasis on their performance, security, and monitoring.

MARTIN REISSLEIN (S'96–M'98–SM'03–F'14) received the Ph.D. degree in systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 1998. He is currently a Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University (ASU), Tempe, AZ, USA, and also an External Associate Investigator with the Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, Germany. He chairs the Steering Committee of the IEEE Transactions on Multimedia. He currently serves as an Associate Editor for the IEEE Transactions on Mobile Computing, the IEEE Transactions on Education, IEEE Access, as well as the *Computer Networks*. He is an Associate Editor-in-Chief of the IEEE Communications Surveys and Tutorials and a Co-Editor-in-Chief of the *Optical Switching and Networking*.

FRANK H. P. FITZEK received the Dipl.Ing. degree in electrical engineering from the University of Technology and Rheinisch-Westfälische Technische Hochschule (RWTH), Aachen, Germany, in 1997, the Ph.D. (Dr. Ing.) degree in electrical engineering from Technical University, Berlin, Germany, in 2002, and the Doctor Honoris Causa degree from the Budapest University of Technology and Economy, in 2015. He became an Adjunct Professor with the University of Ferrara, Italy, in 2002. In 2003, he joined Aalborg University as an Associate Professor and later became a Professor. He has co-founded several start-up companies starting with acticom GmbH, Berlin, in 1999. He is currently a Professor and the Head of the Deutsche Telekom Chair of Communication Networks, Technical University Dresden, Germany, coordinating the 5G Lab Germany. He is the Spokesman of the DFG Cluster of Excellence Centre for Tactile Internet with Human-in-the-Loop (CeTI). His current research interests are in the areas of wireless and mobile 5G communication networks, mobile phone programming, network coding, cross-layer as well as energy-efficient protocol design, and cooperative networking. He was selected to receive the NOKIA Champion Award several times in a row, from 2007 to 2011. He received the Nokia Achievement Award for his work on cooperative networks, in 2008, the SAPERE AUDE Research Grant from the Danish Government, in 2011, and the Vodafone Innovation prize, in 2012.