A Survey on Multi-Access Edge Computing Applied to Video Streaming: Some Research Issues and Challenges

Xiantao Jiang, F. Richard Yu, *Fellow, IEEE,* Tian Song, and Victor C.M. Leung, *Life Fellow, IEEE*

***Abstract*—Driven by the quality of experience (QoE) requirement of video streaming applications in the smart city, smart education, immersive service, and connected vehicle scenarios, the existing network poses significant challenges, including ultra-high bandwidth, ultra-large storage, and ultra-low latency requirements, etc. Multi-access edge computing (MEC) is a potential technology, which can provide computation-intensive and caching-intensive services for video streaming applications to satisfy the requirement of QoE. Thus, focusing on video streaming schemes, a comprehensive summary of the state of the art applying MEC to video streaming is surveyed. Firstly, the related overview and background knowledge are reviewed. Secondly, resource allocation issues have been discussed. Thirdly, the enabling technologies for video streaming are summarized by taking account of caching, computing, and networking. Then, a taxonomy of MEC enabled video streaming applications is classified. Finally, challenges and future research directions are given.**

***Index Terms*—Multi-Access Edge Computing, Video Streaming, Resource Allocation, QoE, Caching, Blockchain.**

I. Introduction

With the rapid development of mobile Internet technology, mobile network data traffic has experienced an explosive growth trend. According to the 2019 Cisco VNI report [1], global mobile data traffic will reach 61 EB (exabyte) per month by 2022, when mobile video traffic will account for nearly four-fifths (79%) of global mobile data traffic. The explosive growth of data traffic, especially video traffic, poses a huge challenge to mobile networks. Currently, immersive videos such as virtual reality (VR) and augmented reality (AR) have very good potential to be the next most important application market. Globally, VR/AR traffic will account for up to 254 PB (petabytes) per month in 2022. Together with the development of the 5th generation (5G) technology and smart wearable devices, the growth of VR/AR applications is rapid. However, VR/AR applications place higher demands on the quality and performance of future networks [2], [3].

Moreover, 5G is expected to be an enabling technology for smart city, smart education, AR/VR service, and connected vehicle [4]. With the development of wireless communication technology, 5G and smart cities are inextricably linked: (1) smart surveillance needs to improve the efficiency of object recognition/detection. (2) connected vehicles need to analyze and adapt to their environment by using artificial intelligence (AI) technology. Finally, 5G promises a more friendly ecosystem for the Internet of Thing (IoT) applications, where data collected by IoT devices need to be analyzed at the network edge. Therefore, with regards to video streaming applications, mobile networks need to provide higher data transfer rates and lower network latency, which will provide users with a better quality of experience (QoE).

To address these network challenges, based on the development of capture and display devices dedicated to video coding and transmission technologies have become an important task of standardization [5]. In terms of video coding, High Efficiency Video Coding (HEVC) which has been standardized by the Joint Collaborative Team on Video Coding (JCT-VC) can be used to encode VR/360° video. As the successor of HEVC, the future video coding standard named Versatile Video Coding (VVC) has been developed by the Joint Video Experts Team (JVET) [6], which takes into account camera-view video, high dynamic range (HDR) video and VR/360° video. In terms of video transmission, the content delivery network (CDN) is adopted to distribute in cloud-manner content, enhancing media availability and distribution performance [7]. CDN nodes are usually deployed in multiple locations, often over multiple Internet backbones. Benefits include reducing bandwidth costs, improving page load times, or increasing global availability of content. However, as a continuous increase in the number of video users further triggers the quick growth of video traffic, the CDN-based video systems need to increase the number of replica servers to achieve the trade-off between bandwidth and the massive traffic demand of users, which increases the cost of video systems and affects negatively on QoE.

Multi-access edge computing (MEC) architecture promoted mainly by ETSI [8], can offer low latency, high bandwidth, location, and context awareness service for video streaming. The MEC resources are placed at the network edge, such as

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61701297, in part by the China Postdoctoral Science Foundation under Grant No. 2018M641981, in part by the Mitacs, and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). (Corresponding author: Xiantao Jiang)

Xiantao Jiang is with the Department of Information Engineering, Shanghai Maritime University, NO.1550, Haigang Ave., Shanghai 201306, China (e-mail: xtjiang@shmtu.edu.cn).

F. Richard Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: richard.yu@carleton.ca).

Tian Song is with Department of Electrical and Electronics Engineering, Tokushima University, 2-24, Shinkura-cho, Tokushima, Japan (e-mail: tiansong@ee.tokushima-u.ac.jp).

V. C. M. Leung is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vleung@ece.ubc.ca).

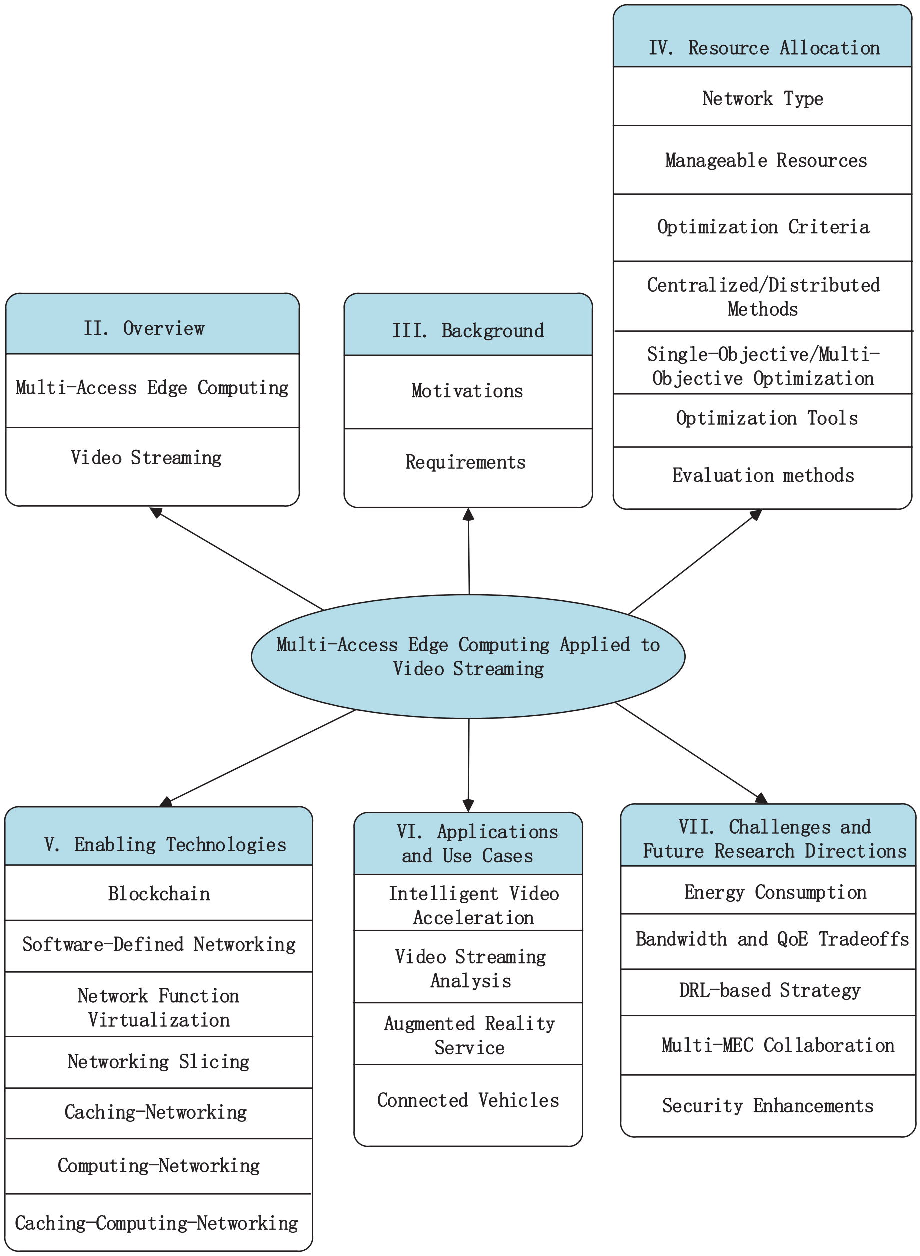

Fig. 1: Road map of multi-access edge computing applied to video streaming.

base stations (BSs), access points (APs), and radio access network (RAN). MEC is integrated into RAN and provides computing, storage, and communication resources at the edge of the mobile network. The MEC reference architecture consists of the mobile edge host and its management [9], where the mobile edge host includes a virtualization infrastructure and the mobile edge platform. Fog computing (FC) introduced by Cisco [10], refers to extending cloud computing to the edge of an enterprise's network. Fog standards are driven by OpenFog consortium that defines FC as a system-level horizontal architecture [11]. The node devices of FC include routers, switches, APs, and gateways. Compared to the traditional cloud/fog network, MEC has some obvious advantages, including low delay and high energy efficiency. Firstly, delay-saving characteristics of MEC are particularly evident in delay-sensitive applications, such as 360° video and VR applications [12]. By using the storage resources of MEC that caches the content on the MEC server, the user can directly obtain content from the MEC server, which can greatly decrease the waiting time between the request and the response, thereby improving the user's QoE [13]–[15]. Secondly, MEC computing and storage resources can perform partial computational tasks to reduce the computational energy consumption of the core network [16], [17]. Moreover, the cost of storage resources is lower than the cost of bandwidth resources. The deployment of MEC is also a way of storing bandwidth for storage. The local storage of content can greatly reduce the necessity of remote transmission, thereby reducing transmission energy

TABLE I: Summary of existing surveys on MEC for video streaming.

Theme	Reference	Major Contribution	Related Content	Differences of Related Content in this Paper
MEC	[19]	A review of MEC definition, application scenarios, challenges and opportunities, as well as security and privacy issues.	Sections II and VI	A discussion about resource allocation for MEC-based video streaming.
	[20]	An overview of state-of-art MEC works by considering resource management of communication/computation in different applications including single-user, multi-user, and multi-server MEC systems.	Sections IV, VI	Focusing on the state of art and difficulties of resource allocation for MEC-based video streaming.
	[21]	A comprehensive survey of MEC's use cases, advancement, and the previous works about computation offloading, as well as open research challenges.	Sections II and V	A thorough review of the MEC for video streaming and converging of computing offloading, communication, and caching.
	[22]	An exhaustive review of previous works about MEC including the convergence and integration of communication, computation, and caching.	Sections V and VI	Providing a taxonomy for the resource allocation for MEC-enabled video streaming.
	[23]	A complete summary of MEC architectures, enabling technologies, and MEC-enabled IoT applications, as well as the selected use case.	Sections V and VI	Focusing mainly on MEC-enabled video applications.
	[24]	A summary of MEC definition, architectures, advantages, applications, and state-of-art related researches, as well as security and privacy issues.	Sections II, III, and VI	Addressing the main challenges and enabling technologies in MEC-based video streaming.
	[25]	A comprehensive of state-of-art MEC researches, the integration of MEC with some emerging technologies, and a summary of lessons learned.	Section II	Pinpointing on the challenges of resource allocation in MEC-based video applications.
MEC for Video Streaming	[26]	A comprehensive survey of applications of edge caching including smart city, vehicular content networks, virtual reality and healthcare.	Section V	A more extensive summary of MEC-enabled video applications and enabling communication, computing, and caching.
	[27]	A systematic review of edge video analytics applications, algorithms and platforms for public safety, as well as open issues and future works.	Section VI	Giving a more detailed description of resource allocation and applications for MEC-based video streaming.
	[28]	An overview of MEC feature capabilities including edge caching, computing and multimedia Internet of Things, as well as research challenges.	Sections II and VI	Not only resource allocation but also service scenarios for MEC-based video streaming are presented.

consumption [18]. MEC-based video streaming schemes are largely ignored in most previous works [19]–[25]. In addition, the enabling technologies and resource allocation issues need to be investigated in future work.

A. Survey Novelty and Contributions

Some existing works have studied MEC and video streaming, which are listed in Table I with a brief description of their related topics and the differences with this survey. To the best of our knowledge, different from the previous works, this survey is the first to discuss methodologies and approaches to solve the impediments when performing the resource allocation for MEC-based video streaming. Most previous works [26]–[28] only mentioned video streaming applications in MEC systems, but the possible solutions of conducting resource allocation in MEC systems are crucial for the success of video streaming application in MEC systems, which need further research.

To fill this gap, this survey focuses on the use of MEC for video streaming applications. Different from the previous works, this work provides a comprehensive survey on resource allocation, the enabling technologies, and the applications and use cases for MEC-based video streaming. The main contributions of this article are summarized as follows.

- Overviews of MEC and video streaming are conducted, and the benefits of MEC are stated.
- The resource allocation issues for MEC-based video streaming are discussed. In this context, the network architecture, manageable resources, optimization criteria, control structure, and evaluation methods are reviewed. Moreover, the main optimization tools for solving the problem are summarized.
- The enabling technologies for MEC-based video streaming are surveyed, such as blockchain, SDN, NFV, and networking slice. Moreover, caching, computing and networking are considered to support video streaming, where previous works are summarized logically. Moreover, the system abstraction has been described in optimization alternatives.
- The existing solutions to video streaming applications by using MEC are classified, including intelligent video acceleration, video streaming analysis, augmented reality service, and connected vehicles.
- Challenges and future research directions of video streaming are stated based on MEC.

B. Survey Structure

The remainder of this article is organized as follows.

- Section II: We provide basic background knowledge of multi-access edge computing and video streaming. In addition, the comparison of the concepts of cloud, fog, and edge computing is given.
- Section III: We describe the motivation and requirements for applying multi-access edge computing to video streaming.
- Section IV: We survey solution techniques that solve the resource allocation problem of MEC-based video streaming. We discuss the pros and cons of each solution technique.
- Section V: We survey the enabling technologies for MEC-based video streaming, including communication, caching, and computing.
- Section VI: We classify existing video applications in the MEC system.

TABLE II: The main abbreviations.

Term	Definition	Term	Definition
ABR	Adaptive Bitrate	MEC	Multi-Access Edge Computing
ADMM	Alternating Direction Method of Multipliers	MSN	Mobile Social Network
AP	Access Point	NDN	Named-Data Networking
AR	Augmented Reality	NFV	Network Functions Virtualization
AVC	Advanced Video Coding	OTT	Over-the-Top
CDN	Content Delivery Network	PVRV	Panoramic Virtual Reality Video
D2D	Device to Device	QoE	Quality of Experience
DDoS	Distributed Denial of Service	QoS	Quality of Service
DASH	Dynamic Adaptive Streaming over HTTP	RAN	Radio Access Network
DRL	Deep Reinforcement Learning	RA	Resource Allocation
FoV	Field of View	SDN	Software-Defined Networking
HetNet	Heterogeneous Network	SVC	Scalable Video Coding
HEVC	High Efficiency Video Coding	VR	Virtual Reality
IoT	Internet of Things	VVC	Versatile Video Coding
ICN	Information-Centric Networks	UE	User Equipment

TABLE III: Cloud, Fog, and Edge Computing Concepts Comparison.

	Cloud Computing	Fog Computing	Edge Computing
Node Devices	Data Center	Routers, Switches, Access Points, Gateways	Base Stations
Location of data processing and storage	Network Core	Near-Edge, Edge	Network Edge
Service	Virtualization	Virtualization	Virtualization
Network Architecture	Centralized	Decentralized, Distributed	Decentralized, Distributed
Location Awareness	Low	Medium	High
Mobility	Limited	Supported	Supported
Latency	High	Medium	Low

- Section VII: We suggest future research directions based on the overall trends observed from the survey results.
- Section VIII: We summarize the key ideas from this survey.

Fig. 1 illustrates a road map of our approach, and the main abbreviations in this paper are summarized in Table II.

II. Overview

In this section, different variations of the computing paradigm, including cloud computing, fog computing, and multi-access edge computing are described. Then, video streaming technology is introduced.

A. Multi-Access Edge Computing

1) Cloud, Fog, and Edge Computing:

Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. The motivation of cloud computing is to provide IT services on-demand. More comprehensively, initial steps toward decentralization of cloud computing are realized through the emergence of fog and edge computing [29]. Fog computing extends the concept of cloud computing to the network edge, making it ideal for IoT and other applications that require real-time interactions. Fog provides data, storage, computing, and application services to end-users thanks to the intermediation of a local proxy, often called Smart Gateway (SG) [30]. Moreover, fog computing is closer to end-users and has a wider geographical distribution. Edge computing is a distributed computing pattern that can bring computation and data storage closer to the location, and improve response times and save bandwidth [19]. At the edge, the devices can consume or produce data. Moreover, the devices can request or perform tasks from the cloud. Similarly, the edge can offload the computing tasks, process data caching/storage, and service delivery.

Table III details conceptual differences and similarities in their current approaches. Cloud computing paradigm is based on the data centers which are capable of handling storage and processing of large scales of data. For cloud computing, the location of data processing and storage is located close to the network core. Fog computing implementation is a decentralized computing infrastructure based on routers, switches, access points, and gateways. For fog computing, the location of data processing and storage is located close to the near-edge and edge. Edge computing is to bring computational and storage capacities to the edge of the network within the radio access network. The edge computing nodes are usually co-located with the macro base-station. For edge computing, the location of data processing and storage is located close to the network edge. Moreover, cloud computing lacks of location awareness, while the location awareness of fog computing and edge computing is medium and high. The mobility of cloud computing is limited, while fog and edge devices support mobility.

2) MEC Standardization:

Mobile edge computing provides an information service environment and cloud-computing capabilities at the edge of the mobile network, within the RAN and near mobile subscribers [24]. In 2017, mobile edge computing was renamed to multi-access edge computing (MEC) by the European Telecommunications Standards Institute (ETSI) Industry Specification Group (ISG). In February 2018, the document about the deployment of MEC in the network function virtualization (NFV) environment was approved by ETSI [31]. Moreover, the use cases and requirements specification on MEC were published by ETSI in October 2018 [32], which include an annex describing example use cases and their technical benefits, for deriving requirements. Most recently, the framework

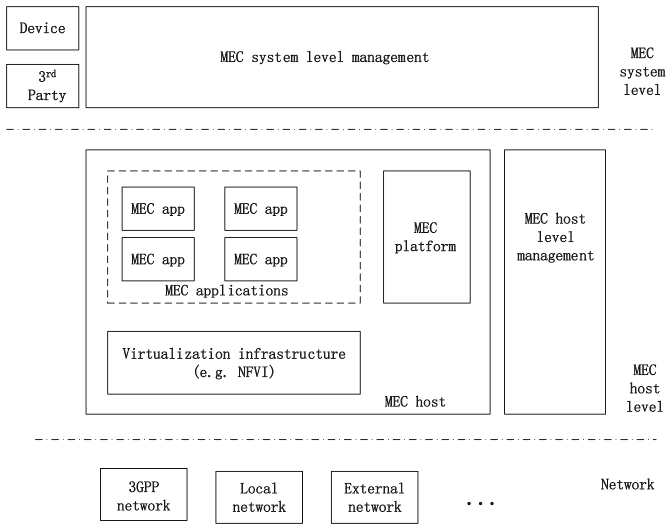

Fig. 2: ETSI MEC framework [8].

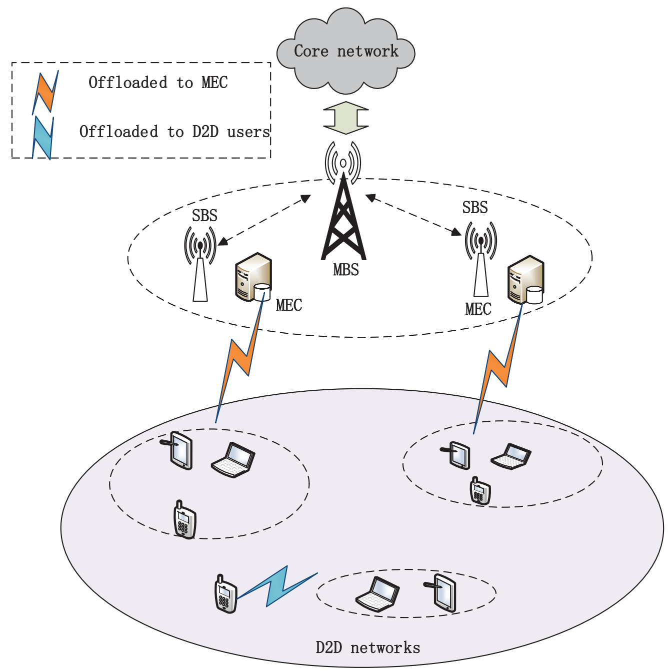

Fig. 3: The D2D-enabled MEC.

and reference architecture specification of MEC were released by ETSI [8], which describe the functional elements and the reference points between them, and some MEC services. In July 2019, ETSI has announced the Proof of Concept (PoC) framework that was accepted by ETSI ISG MEC [33]. Most recently, the document about MEC 5G integration was produced by ETSI [34].

MEC enables the implementation of applications as software-only entities that run on top of a virtualization infrastructure that is located close to the network edge. The framework of ETSI MEC is shown in Fig. 2, which includes system-level, host-level and network-level entities. The networks-level offers connectivity to a variety of accesses, and the host-level provides the virtualization infrastructure and the MEC platform, facilitating the execution of MEC applications. The system-level management provides an abstraction of the underlying MEC system facilitating access for user equipments (UEs) and third parties.

Moreover, D2D communication is direct communication between two mobile users without traversing the Base Station (BS) or the core network [35]. D2D-enabled MEC has been investigated to support end device resource sharing [36]. The integration of D2D and MEC can further improve the computation capacity of the cellular networks, where end-devices with computing and storage capacities collaborate with the conventional MEC infrastructure. As shown in Fig. 3, the D2D-enabled MEC system consists of a macro base station (MBS), small cell-based stations (SBSs), and end devices. In this case, the end device can offload the task to either the nearby MEC or a group of D2D nodes.

B. Video Streaming

1) HD/UHD Video:

With the development of display technology, the video definition is higher. The high definition (HD) video refers to higher quality video than standard definition video, which has between 720 and 1080 lines of vertical resolution. The ultra high definition (UHD) video refers to the higher quality video that has a higher resolution than standard definition ones, which has between 2160 and 3840 lines of vertical resolution. Due to the increase in video streaming traffic over the Internet, users' quality of experience is significantly affected.

Video streaming is a type of media streaming, which is continuously transmitted over the Internet to users. The size of each streaming is determined by the compressed video size, bandwidth, and latency. The widely-used video coding formats include H.264/AVC [37], H.265/HEVC [38], AVS2 [39], VP6 [40], VP9 [41], VC-1 [42], and AV1 [43]. In addition, H.265/HEVC was designed by the Joint Collaborative Team on Video Coding (JCT-VC[1]). As a successor to H.264/AVC, H.265/HEVC achieves to increase the coding efficiency by 50%, while at the same time video quality is not reduced. H.265/HEVC is friendly for real-time encoding, which supports HD, UHD, and 8K UHD video. Moreover, the next-generation video standard Versatile Video Coding (VVC) will be released around 2020 by the Joint Video Experts Team (JVET[2]). AV1 is an open, royalty-free video codec designed for video delivery over the Internet, while the hardware design of AV1 codec will be implemented at least in 2020.

To adapt to different network environments and varying user requirement in video transmission and storage systems, Scalable video coding (SVC) and Scalable High efficiency Video Coding (SHVC) that are the scalable extensions of the H.264/AVC and H.265/HEVC were presented to offer the scalability [44], [45]. In lossy transmission environments, the bit rate and format can be adjusted to adapt to various terminal capabilities or network conditions. Furthermore,

[1]https://hevc.hhi.fraunhofer.de
[2]https://jvet.hhi.fraunhofer.de

TABLE IV: HD/UHD Video Streaming Applications.

Applications	Video Codec	Highest Resolution	Bandwidth
Amazon Video	VC-1	1280 × 720 (HD)	2.5-6 Mbit/s
Hulu	On2 VP6	1280 × 720 (HD)	2.5 Mbit/s
Apple TV	H.264/AVC	1920 × 1080 (HD)	
Vimeo	H.264/AVC	1920 × 1080 (HD)	
Netflix	VC-1	3840 × 2160 (UHD)	25 Mbit/s
Youtube	H.264/AVC, VP9	7680 × 4320 (8K UHD)	

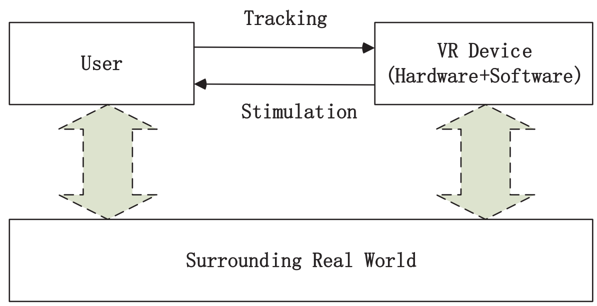

Fig. 4: The VR system.

video traffic has caused a huge impact on mobile networks. Mobile networks are required to provide a higher transmission rate and lower network latency capability. To address these challenges, adaptive bitrate (ABR) technology is adopted in video distribution [46]–[48]. Firstly, in the mobile network, the video is encoded into a variety of bit rate streaming versions. Secondly, each streaming version is cut into multiple segments. Therefore, according to terminal capabilities and network conditions, the suitable streaming version will be provided to the user dynamically. The benefit of ABR is that the probability of choppy video can be reduced, while the user's QoE can be enhanced. Recently, ABR is used to streaming transcoding and caching with different scenarios [49].

Encoded video streams are assembled in a container "bitstream", and the bitstream is transmitted by using a transport protocol. Popular streaming protocols include MEPG-DASH [50], Apple HTTP Live Streaming (HLS) [51], Adobe HTTP Dynamic Streaming (HDS) [52], and Microsoft Smooth Streaming [53]. Table IV shows HD/UHD video streaming applications.

2) AR/VR Video:

Immersive technologies such as augmented reality (AR) and virtual reality (VR) bring new opportunities and challenges for content storage, delivery, and display in the multimedia network industry. By 2025, VR and AR ecosystem are expected to be an $80 billion market. Virtual reality is a simulated experience that can be similar to or completely different from the real world. VR system has the following three basic characteristics: immersion, interaction, and imagination, and it emphasizes the dominant role of a user in virtual systems [54]. Fig. 4 emphasizes the whole VR system including person, VR device, and the real world. The VR device contains at least one screen, a set of sensors, and a set of computing components, which can generate the stimuli and track user's motion by using various sensors. The user can respond to the stimuli from the real world. VR applications include two categories: 360° VR video and computer-generated (CG) VR, which have a broad application prospect. In the short term, the most promising applications include video games, entertainment, medical care, manufacturing [55], and education [56].

Augmented reality is an interactive experience of a real-world environment where the objects that reside in the real-world are enhanced by computer-generated perceptual information, sometimes across multiple sensory modalities [57]. AR system is closely bound up to hardware, software, and applications. For hardware, the combination of processor, display, sensor, and input device can be suitable for the AR platform. In terms of software, having the ability to combine augmented objects with the real world is very critical. Therefore, the AR algorithm must obtain the coordinates of the real world from the images in the equipment, and then superimpose the augmented objects onto the coordinates. AR technology was used in various industries, and typical applications include medicine, assembly, maintenance and repair, entertainment, sport and marking, collaborative visualization space, tourism, architecture and construction, teaching, education, and training [58].

III. Background

In this section, motivations for applying MEC to video streaming are introduced. Moreover, the requirements for video streaming are surveyed.

A. Motivation

MEC is a key enabling technology for 5G networks. MEC-enabled video streaming applications can accelerate the commercialization of enabling technologies. The core characterizations of MEC can be summarized as flowing.

1) Computational Capacity: MEC can provide ultra-low latency and ultra-high bandwidth, which is one of the ambitious need of MEC to obtain real-time network analysis. MEC is suitable for computation-intensive services such as AR and VR. Moreover, MEC is geographically close to users or information sources, which can significantly reduce the network latency of service response and reduce the possibility of network congestion in the backpass network and core network [59]. Finally, MEC can perceive network data in real-time, including the condition of wireless links, users' behavior information, and location.

2) Storage Capacity:

Different from traditional networks, MEC has many significant advantages, which can effectively solve the problems of high delay and low efficiency in traditional networks. For video streaming applications, MEC's computing/storage resources and network perception ability can effectively support ABR technology. On one hand, video content can be cached in distributed edge networks and offloaded to the local server, where the transmission latency can be reduced and the energy efficiency can be improved significantly. On the other hand, by using ABR technology, the bitrate of the video version can be adjusted adaptively based on mobile devices and network conditions. Then, multiple bitrate versions can be cached, which will cause a large cache cost. The possible solution is that a higher bitrate version to be cached in the MEC server. When the bitrate version is not hit, the cached bitrate version can be transcoded to the requested bitrate version in MEC server, which can improve the efficiency of caching resources. Therefore, the popular bitrate version is cached and transcoded between different bitrate versions to respond to user requests on the edge of the network, which has been considered as an important approach to solve content distribution in adaptive video streaming applications.

B. Requirements

With the development of the acquisition and transmission technologies of immersive media, great challenges are foreseen in the network system for supporting a variety of video streaming applications, which can be outlined in the following.

1) Ultra-High Bandwidth:

As users' expectations for high resolution video quality increase, the disadvantage of "occupying too much bandwidth" is highlighted, and the delivery of video streaming in real-time faces new challenges. On one hand, the pixel of 4K (3820×2160) video is four times that of the 1080p HD video, which can consume a lot of bandwidth. Moreover, compared with 4K video, 8K (7640×4320) video needs a significant more amount bandwidth to transmit.

On the other hand, as the 360° video and VR applications such as virtual classrooms and virtual galleries are becoming popular, people are also increasingly eager for a real mobile experience. Providing 360° video and edge-based VR applications require ultra-high bandwidth and ultra-low latency, which is difficult to achieve in mobile networks. Furthermore, transmitting a high bandwidth will bring high operating costs and potentially large delays. The Field of View (FoV) is used to reduce the streaming bandwidth. However, extracting and transmitting FoVs that respond to user head movements can increase high latency and negatively impact the user experience.

2) Ultra-Large Storage:

The digital video is rapidly evolved to 8K and beyond, but their adoption and consumption have created immense challenges for storage domains. Firstly, transcoding is one of the critical media services that is required to produce content. It is a conversion of data from one format to another. Transcoded content must support many different formats, devices, and platforms. Therefore, a lot of video files are proceed concurrently, requiring high throughput, massive storage and CPU intensive resources.

Moreover, for 360° video and AR/VR applications, cloud storage is not suitable, because there is a constant need to retrieve, edit, and enhance the content. Although upfront costs of cloud storage are minimal, the cost of continuing storing video data in the cloud is expensive as data grows.

3) Ultra-Low Latency:

Low latency is very vital for real-time streaming applications. When distributing videos using online streaming platforms, latency issues are especially noticeable. The latency over the network is proportional to bitrate and inversely proportional to bandwidth. Some factors are impacting the network latency: file upload, CDN propagation, and file delivery.

TABLE V: Typical bandwidth and latency for VR.

Standard	Previous VR	Entry VR	Advance VR	Ultimate VR
Resolution	960×960	1920×1920	3840×3840	7680×7680
Ratio	165:1	165:1	215:1	350:1
Frame rate	30	30	60	60
Bandwidth	25Mbps	100Mbps	418Mbps	4.95Gbps
Latency	40ms	30ms	20ms	10ms

Moreover, the bandwidth is the bottleneck for the VR network latency. Transmission time thus dominates the overall delay. There are two categories of latency for VR applications: the transmission latency and the user's perceived latency. VR immersive experience brings shorter latency requirements to the network. Table V shows the typical bandwidth requirement and latency for various VR applications. The tolerable network latency is 10ms for the ultimate VR, while the tolerable network latency is 40ms for the previous VR. Therefore, the processing of the VR video should be either on local devices or on the MEC server.

4) QoE:

QoE has become a prominent problem for real-time video streaming applications. The growth of online video streaming can lead to network congestion, which brings a poor QoE for users. The average bitrate is the most accurate indicator of video quality. To increase the user's QoE, ABR streaming technology can be used to adapt video bit-rate to network congestion. However, other streaming metrics including re-buffering ratio and start buffering time have a major effect on the QoE of the real-time streaming.

Furthermore, immersive video streaming is extremely welcomed by users. However, it is hard to understand the QoE of VR streaming. Several factors, including presence, usability, and cyber-sickness have a great correlation with the QoE of VR streaming [60]. To offer satisfying end-user QoE, edge computing has the potential to be the solution to transmit VR video streaming [61].

5) Security and Privacy:

Currently, the security and privacy issues in smart city and smart compus environments have attracted the academic and industry. In general, security is associated with the robustness of VR/AR devices against some attacks, and privacy is associated with the protection and secrecy of sensitive data [62]. For one thing, VR/AR video technologies have been employed increasingly to increase efficacy and outcomes by the education community. For instance, VR devices can be used to visit a virtual campus, which enhances student engagement. However, in the public environment, the data are not secretive, and the data integrity can be compromised. For another, distance learning and training can be enhanced by VR technology. However, the distributed denial of service (DDoS) attacks would happen in an on-line classroom more frequently. Therefore, it is very vital to provide decentralized security and privacy for VR/AR applications in IoT environments.

IV. Resource Allocation

In this section, we study resource allocation issues for video streaming. Firstly, different network architectures are introduced, and the manageable resources in video communication are reviewed. Secondly, the optimization criteria are described. Thirdly, the control structures of resource allocation are discussed. Next, the single-objective and multi-objective optimization problems are introduced. Then, optimization tools for solving the resource allocation problem are described. Finally, evaluation methods are surveyed.

A. Network Architectures

We consider the following four typical network architectures: Macro-Cellular Networks, D2D Networks, Heterogeneous networks (HetNets), and C-RANs.

- Macro-Cellular Networks: Macro-cellular networks [63]–[65] provide radio coverage served by a high power cell site (tower, antenna or mast). Generally, macro-cellular networks provide coverage larger than micro-cellular networks. Macro-cellular networks base stations have power outputs of typically tens of watts. Macro-cellular networks performance can be increased by increasing the efficiency of the transceiver.
- D2D Networks: D2D networks [66], [67] are cellular networks that allow direct communication between two mobile users without traversing the Base Station (BS) or core network. D2D communication is generally non-transparent to the cellular network and it can occur on the cellular frequencies or unlicensed spectrum.
- HetNets: HetNets [68]–[70] consist of macro-cells, small cells (femtocells, picocells), and relay nodes. In HetNets, the small cell BSs and relay nodes transmit at low power levels.
- C-RANs: A radio access network (RAN) is part of a mobile telecommunication system, which implements a radio access technology. RAN resides between a mobile device and provides the connection with its core network (CN). The typical RANs include cloud-based radio access network (C-RAN) [71], [72]. Another similar concept is called fog-based radio access network (F-RAN).

B. Manageable Resources

In the wireless cellular networks, the joint framework considering computation offloading, spectrum resource allocation, and content caching are proposed to improve the performance of video streaming with MEC. For wireless communication, manageable resources are spectrum/bandwidth, power/battery, time, and space resources [73]. The computation resource is evaluated by computational ability. The caching resource is evaluated by storage availability.

C. Optimization Criteria

In this work, the general criteria have been proposed for resource optimization. We present the networking, caching and computing related metrics, respectively. All the objective metrics are classified and briefly described in Table VI.

1) Energy Consumption:

The energy consumption metric refers to the energy consumed by network nodes, either user devices or servers, for the execution of the computation task. The energy consumption metric is in close relation to the execution time metric, since energy consumption depends on the time during which the network nodes work. In the MEC system, the energy consumed by user devices for computation task offloading, and the energy consumed by the MEC server for computation results feedback, can also be included in this energy consumption metric.

In the previous works, the approaches for energy consumption minimization have been widely studied in [74], [75], [76], [77]. In general, energy efficiency consists of signal transmission power consumption and equipment energy consumption. Furthermore, in caching-computing-networking architecture, the cache energy consumption and the computation energy consumption can be considered in [78], [79].

2) Throughput:

The throughput metric refers to the data units delivered through the network per unit time interval. This metric is often used in the MEC system to serve as a joint indicator of the network transmission capability and the network node's computation capability. Throughput maximization is an important topic for optimal video experience. The criterion could be used to maximize the total base stations (BSs) throughput [80], the overall throughput of the networks [81]–[83], and the video capacity [84].

3) QoE:

The characteristics of video streaming services are high real-time requirements and high bandwidth resources. Therefore, the impact of network parameter changes is very significant. The quality of video streaming services is especially sensitive to time-varying network characteristics. From the user's point of view, QoE analyzes the user's experience quality of the video streaming service and it is an intuitive feeling of the user's network video quality.

The QoE-driven approaches may be used to maximize the aggregate average video distortion reduction [85]. Moreover, video stalling, video quality, bitrate switching, and initial/startup delay are related to QoE in [86], [64], [87]. In addition, the subjective evaluation approaches of QoE can be considered in [68], [88], [89].

4) Service Latency:

The service latency metric refers to the delay induced by preparation and propagation of data packets in the system, including the propagation delay, serialization delay and queuing delay.

The MEC-based video streaming applications may be used to minimize the average delay of its mobile users in [90], [7], [91], [92], [93], maintain high transmission and low latency in [65], [94], [95].

5) Cost:

The cost metric refers to the investment paid by infrastructure providers and system operators for infrastructure construction and system operation, including Capital expenditures (CAPEX) and operating expenses (OPEX). In MEC systems, CapEX is composed of the expenses of constructing

TABLE VI: Optimization Metrics.

Metrics	Types	Main Findings
Energy Consumption	Computing metric	Energy consumed for the execution of the computation task.
Throughput	Networking and computing metrics	Data units delivered through the network per unit time interval.
QoE	Networking and caching metrics	Quality of video experienced by end users.
Service Latency	Networking metric	Packet delay.
Cost	General metric	Including CAPEX and OPEX.
Fairness	General metric	The resource shared by each AP/user fairly.
Cache Hit Ratio	Caching metric	Number of content requests responded to by one network node.
Revenue	General metric	Earned revenue.

base stations, radio network controller equipment, backhaul transmission equipment, core network equipment, caching and computing equipment installation in network nodes [96]. OpEx consists of energy charge, equipment lease and operation and maintenance fees.

6) Fairness:

During the resource allocation for video streaming applications, fairness is a very vital optimization criterion. The fairness in resource allocation can ensure that the scheduling policy provides services to the access points (APs)/users fairly. At present, the common fairness strategies include max-min fairness [86], QoE-fairness [87], [97], [98] and proportional fairness. Moreover, the work in [80] presents the fairness measure approach using Jain's fairness index.

7) Cache Hit Ratio:

The cache hits metric indicates the load of a node by measuring the number of content requests responded to by the node. When intended to reflect the load savings of a node due to conducting caching, the node hits metric is determined as a fraction called cache hit ratio, which is normalized by the total number of content requests generated in the whole network. In order to improve the cache capacity of video streaming, the approaches of the cache-hit ratio maximization have been studied in [82], [99], [70], [96], [71].

8) Revenue:

During a smart contract, blockchain technology provides a solution to enforce a business strategy through robust distributed authentication avoiding repudiability. For video streaming applications, the revenues include the profit of caching miners and the profit of transcoder. In communication-computation-caching architecture, this criterion could refer to maximizing the total profit of caching miners [67], the average reward for transcoders [100], [101], the total revenue of resource cost and the price for the transcoding and delivery market [102], the aggregate reward for caching/computing resources [103], and the total revenue, including communication/caching/computing revenues [69].

D. Centralized/Distributed Methods

RA can be controlled by one entity in a centralized structure, or by multiple entities in a distributed structure. In this subsection, the two control structures are introduced.

1) Centralized:

Based on the centralized approach, a centralized entity, including BS and eNodeB is responsible for the resource allocation. There are some advantages for the centralized-based RA method: (1) It is easy to implement and the network environment is stable and predictable. (2) It is easy to obtain the optimal solution [63], [68], [104]. However, the complexity of the centralized scheme is high.

2) Distributed:

Compared with the centralized approach, the distributed approach is a low-complexity solution the flexibility and robustness. With the distributed-based approach, each node/user can make the decision individually, and the optimization problem can be divided into some subproblems. Several distributed approaches are adopted to solve the resource optimization problem, such as the alternating direction method of the multipliers (ADMM) [100], multiuser-based game model (MGM), and blockchain-based approaches [100], [105].

E. Single-Objective/Multi-Objective Optimization

In MEC systems, the resource allocation can be divided into the single-objective and multi-objective optimization. Some details about single-objective and multi-objective optimization are described as follows.

1) Single-Objective Optimization:

Optimization problems are normally stated in a single-objective optimization (SOO) way, where the process must optimize a single-objective function complying with a series of constraints. A single-objective optimization problem may be stated as follows [106]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize/maximize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p, \end{aligned} \tag{1}$$

where the function to be optimized is $f(\mathbf{x})$, where $\mathbf{x}$ is the optimization variable. Functions $g_i(\mathbf{x})$ and $h_i(\mathbf{x})$ are the constraints of the model.

For the SOO problem, three sets of solutions include the universal set, set of feasible solutions, and set of optimal solutions. The traditional optimization tools can be used to find a minimum or maximum of objective function, whether local or global, which will be introduced in the Subsection E.

2) Multi-Objective Optimization:

Multi-objective optimization (MOO) is an area of multiple criteria decision making that is concerned with mathematical optimization problems involving more than one objective

TABLE VII: Multi-Objective Optimization Methods.

Category	Technique	Advantage	Disadvantage
Mathematical Methods	Weighted Sum	Computationally efficient in generating a non-dominated solution.	Depending on weight coefficients.
	ε-Constraint	Can identify trade-off points between multi-objective functions.	Time-consuming process.
	Utility Function Method	This idea is simple and ideal if adequate value function information is available.	The obtained solution entirely depends on the chosen value function.
Metaheuristics Methods	Evolutionary Algorithm	Provides heuristic but close-to-optimal solutions.	Computationally intensive.
	Simulated Annealing	Good approximation solution for a large size solution search space.	No guarantee for the global optimum.

function to be optimized simultaneously. A multi-objective optimization model may be stated as follows [107]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize/maximize}} \quad & (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_k(\mathbf{x})) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p, \end{aligned} \tag{2}$$

where the functions to be optimized (whether minimize or maximize) are the set of functions $f_k(\mathbf{x})$, where $\mathbf{x}$ is the set of independent variables. Functions $g_i(\mathbf{x})$ and $h_i(\mathbf{x})$ are the constraints of the model.

In the MOO, objective functions can be divided into three types: minimize all objective, maximize all objective functions, and minimize some and maximize others [64]. Solving multi-objective problem can be complex, which requires optimizers that instead of providing a single optimal solution. For MOO problems, the Pareto optimal solution can be considered to reach with the aid of specific parameters of the scalarization [108]. There are various classes of methods for generating the Pareto set of MOO problems, including mathematical programming and metaheuristics methods. The MOO methods are classified and briefly listed in Table VII.

Mathematical programming methods can be adopted to solve multi-objective problems by using single-objective approximations, including weighted sum, ε-constraint, and utility function method [109]. These methods are described as follow:

- Weighted Sum: This method is to create a single-objective function by weighing the n objective functions. The multi-objective function (2) above can be restated as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize/maximize}} \quad & \sum_{i=1}^{n} r_i \times f_i(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p, \\ & \sum_{i=1}^{n} r_i = 1, \quad 0 \leq r_i \leq 1 \end{aligned} \tag{3}$$

  where r_i is the weight of each objective function. This weighted sum method is the simplest way to solve the MOO problem, which guarantees to find solutions on the entire Pareto-optimal set. However, in handling mixed optimization problems, all objectives have to be converted into one type.
- ε-Constraint: This method consists of creating a single-objective model, where only one of the functions will be optimized and the remaining function becomes constraints. Therefore, the MOO function shown above (2) can be rewritten as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize/maximize}} \quad & f_i(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p, \\ & f_k(\mathbf{x}) \leq \varepsilon_k, \quad k = 1, \dots, n, k \neq n \end{aligned} \tag{4}$$

  where $f_i(\mathbf{x})$ is the only one optimized and the remaining functions act as constraints. Different Pareto-optimal solutions can be found by using different ε_k. This method can also be used for problems having convex or non-convex objective spaces.
- Utility Function Method: This method assumed that the decision maker's utility function $\mathbf{U}$ is available over the entire feasible search space. The utility function provides interactions among different objectives, and the MOO problem is then to maximize the utility function as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{U}(f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_k(\mathbf{x})) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p, \end{aligned} \tag{5}$$

  The utility function methods are mainly used in practice to multi-attribute decision analysis problems with a discrete set of feasible solutions [85]. Wang $etal.$ [110] apply a utility function method to solve resource allocation in MEC systems. However, the utility function method requires users to come up with a value function which is globally applicable over the entire search space. Thus, there is a danger of using an over-simplified value function.

Metaheuristics are considered high-level algorithmic strategies that are used to guide other heuristics in search of the space of feasible solutions [111]. Metaheuristics are friendly techniques to solve MOO problems, which can accomplish the following functional features: (1) convergence toward optimal; (2) optimal solutions not withstanding convexity; (3) avoiding local optimal. Some typical metaheuristics include evolutionary algorithm and simulated annealing [112].

- Evolutionary Algorithm (EA): Evolutionary algorithm is a heuristic-based approach to solving problems that cannot be easily solved in polynomial time, which contains four overall steps: initialization, selection, genetic operators, and termination. EAs can be adopted to solving MOO problems, which can find the globally near-optimal

TABLE VIII: Summary of Optimization tools.

Technique	Description	Examples of applicable problems	Advantages	Disadvantages
Convex Optimization	The RA problems can be modeled into a maximization or minimization problem with constraints.	The optimization problem is suitable for the convex problem or can be transfered into convex problem, such as [114], [103], [69], [96], [68], [100] [115], [72], [95].	Any local optimal solution is the global optimal solution. Moreover, the distributed method such as ADMM can reduce the complexity of the RA problem.	It is hard to solve dynamic, time-varying optimization problems.
Game Theory	The optimization model consists of a set of makers who chooses the strategy to maximize the utility.	1) Nash equilibrium game is a concept of non-cooperative game solution involving two or more players.(eg. [105]) 2) Stackelberg game is a price leadership model. (eg. [90], [82] [102], [116])	The game theory optimization model has broad applicability and preferable accuracy.	The assumption that players have the knowledge about their own pay-offs and pay-offs of others is not practical.
Stochastic Optimization	The stochastic optimization problem consists uncertain operating conditions, which is a novel methodology including probability and randomness.	1) Queuing theory is constructed so that queue lengths and waiting time can be predicted, when making decisions about the resources needed to provide a service. (eg. [117]) 2) Lyapunov theory is used to make conclusions about the trajectories of a system. (eg. [118]–[120])	It is the potential for uncertain environments, where the optimization approach is hard to solve.	No guarantee for global optimization solution.
Reinforcement Learning	The RA problems are formalized as a discrete-time stochastic.	1) Q-learning is an value-based reinforcement learning algorithm, which can solve problems with stochastic transitions and rewards. (eg. [98], [101], [121]) 2) Actor-Critics method is temporal-difference methods that have a separate memory structure to explicitly represent the policy independent of the value function (eg. [122]).	RL can address the RA problem by maximizing a numerical reward signal while interacting with the unknown environment.	Reinforcement learning is not preferable to use for solving simple problems.

solutions by repeatedly evaluating the objective functions or fitness functions using exploration and exploitation methods [113]. Liu $et al.$ [76] use EA to maximize the utilization of the wireless bandwidth for panoramic VR video streaming. However, the solutions of EA may converge towards local optima.

- (SA): It is a probabilistic technique for identifying the global minimum of a cost function that may have a number of local minima. Li $et al.$ [85] use SA to solve a resource allocation problem in MEC systems. SA is useful when the search space is large, while it does not guarantee the global optimum.

F. Optimization Tools

In the following, some possible problem solutions of resource allocation in video streaming applications have been highlighted, including convex optimization, game theory, stochastic optimization and reinforcement learning. Moreover, we give a summary of the resource optimization tools in Table VIII.

1) Convex Optimization:

Convex optimization deals with the minimization of convex functions defined on convex sets. In some sense, convex optimization is simpler than general mathematical optimization. Moreover, in convex optimization, the local optimal value must be the global optimal value. The optimization problems can be divided into convex/non-convex, linear/non-linear, and the optimization variables include continuous and discrete. These optimization problems, including least squares method (LSM), linear programming (LP), Semi-definite programming (SDP), second-order cone program (SOCP), etc. are convex, or can be converted to convex problems by changing variables in [114], [103], [115], [95]. A convex optimization problem is in standard form if it is written as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p, \end{aligned} \tag{6}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, the function $f : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex, $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$, are convex, and $h_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, p$, are affine.

The convex optimization problem can be solved by the following methods: bundle method, sub-gradient method, and interior point method [69], [96]. The authors in [68], [72] propose the dual-decomposition method to solve the convex optimization problem. Furthermore, the work in [100], [123] presents the ADMM to solve the convex optimization problem, where ADMM is with superior convergence. However, non-convex optimization problems can be transformed into convex approaches.

2) Game Theory:

Game theory is a useful mathematical tool that considers the predicted behavior and actual behavior of individuals in games, and studies their optimization strategies. In the game model, all players must consider various possible courses of action of their opponents and try to choose the most favorable or reasonable one for themselves. Games theory can be divided into cooperative games and non-cooperative games. Cooperative game theory is more complex than non-cooperative game theory, therefore, the game theory generally refers to non-cooperative games in practical applications. Moreover, the Nash equilibrium (NE) is a proposed solution of a non-cooperative game involving two or more players in which each player is assumed to know the equilibrium strategies of the other players. In a NE, each player chooses a strategy that maximizes their expected payoff given the strategies employed by other players [105], [124]. Other optimization problems can be solved by using Stackelberg game [82], [90], where Stackelberg model is a price leadership model. Furthermore, the authors propose the three-stage Stackelberg game to solve the RA problem in [102], [116].

3) Stochastic Optimization:

Stochastic optimization refers to the optimization problem with stochastic factors, which is solved by using probability statistics, stochastic process, stochastic analysis, and other tools. Stochastic optimization solutions can be used to solve the optimization problem for video streaming applications, including queuing theory [117] and Lyapunov theory [118]–[120].

- Queuing Theory: Queuing optimization model is to minimize the total cost of waiting cost and service cost. The queuing systems consist of input processes, queuing and queuing rules, service agencies, and service rules. In the queuing model based RA problem, the main optimization objectives include: (1) to determine the optimum service level to minimize the total cost; (2) to maximize the difference between service income and service cost. The most important aspect of queuing optimization is how you value your customers compare to how you value the cost of your servers.
- Lyapunov Theory: Lyapunov function is defined to become larger when the system enters an undesired state [125]. By taking control measures to make the Lyapunov function drift to zero in the negative direction, the stability of the system is achieved. Suppose there are N queues in the network, and define the vector of queue backlogs at time t by:

$$Q(t) = (Q_1(t), \ldots, Q_N(t)) \tag{7}$$

The function $L(t)$ is defined as the sum of the squares of all queue sizes at time t, and is called a Lyapunov function:

$$L(t) = \frac{1}{2}\sum_{i=1}^{N} Q_i(t)^2 \tag{8}$$

Then, the Lyapunov drift is defined:

$$\Delta L(t) = L(t+1) - L(t) \tag{9}$$

Every slot t, the current queue state is observed and control actions are taken to greedily minimize a bound on the following drift-plus-penalty expression:

$$\Delta L(t) + Vp(t) \tag{10}$$

where $p(t)$ is the penalty function and V is a non-negative weight.

4) Reinforcement Learning:

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward. The general RL problem is formalized as a discrete time stochastic control process, such as Markov decision process (MDP). A Markov decision process is a 4-tuple: (S, A, P_a, R_a), where S is a finite set of states, A is a finite set of actions, $P_a(s, s')$ is the probability that action a in state s at time t will lead to state s' at time $t+1$, $R_a(s, s')$ is the immediate reward received after transitioning from state s to state s', due to action a. The core problem of MDPs is to find a "policy" for the decision-maker: a function π that specifies the action $\pi(s)$ that the decision-maker will choose when in state s. The goal is to choose a policy π that will maximize some cumulative function of the random rewards, typically the expected discounted sum over a potentially infinite horizon [126]:

$$E[\sum_{t=0}^{+\infty} \gamma^t R_{a_t}(s, s_{t+1})] \tag{11}$$

where a_t=$\pi(s_t)$, and γ is the discount factor satisfying 0$\leq \gamma \leq$ 1.

In general, there are two main types of RL methods, including value-based and policy-based. The value-based RL method try to find or approximate the optimal value function, which is a mapping between an action and a value, and the most famous algorithm is Q-learning. The policy-based RL method tries to find the optimal policy directly without the Q-value as a middleman. Moreover, some typical RL-based approaches are proposed to optimize the resources in video streaming applications, including Q-learning [121] and actor-critic methods [122].

- Q-learning: This is a value-based reinforcement learning algorithm, which can solve problems with stochastic transitions and rewards. Moreover, Q-learning can identify an optimal action-selection for a given Markov decision process. Q-learning algorithm has a function that calculates the quality of a state-action combination:

$$Q : S \times A \rightarrow \mathbb{R} \tag{12}$$

The core of Q-learning algorithm as a Bellman equation as a simple value iteration update, using the weighted average of the old value and the new information:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q_t(s_{t+1}, a) - Q(s_t, a_t)] \tag{13}$$

where r is the reward received when moving from the state s_t to the state s_{t+1}, and α is the learning rate ($0 \leq \alpha \leq 1$). The enhancements of Q-learning include deep Q network (DQN) and double dueling Q network (DDQN) [98], [101].
- Actor-Critics (AC): This methods are temporal-difference methods that have a separate memory structure to explicitly represent the policy independent of the value function [127]. Actor-Critics aims to take advantage of all the good stuff from both value-based and policy-based while eliminating all their drawbacks. There are two popular improvements of AC models, including advantage actor-critic (A2C) and asynchronous advantage actor-critic (A3C) [128].

G. Evaluation methods

The evaluation method is a tool to verify the performance of the optimization algorithm. The evaluation methods can be divided into the four groups:

- Numerical Analysis: Numerical analysis is the study of algorithms that use numerical approximation for the problems of mathematical analysis. The numerical analysis

can be adopted to solve the optimization problems. There is a trade-off between the computational complexity and the scalability of the numerical analysis.
- Simulation: Simulation methods have been widely applied to reflect the operation a RAN. The characteristics of simulation are simple and flexible. The discrete-event simulators are widely used to study the behavior of a RAN, including NS-3 [1] and openWNS [2].
- Testbed: A testbed is a platform for conducting experiment and is adopted to verify the practical feasibility of an on/off switching strategy in a controlled environment.
- Realistic Network: The performance of a on/off switching strategy can be evaluated by implementing the proposed algorithm into a real network. The most realistic evaluation can be performed in this manner.

H. Lessons Learned

Resource allocation in MEC systems is a fundamental problem that becomes increasingly complex with the consideration of system uncertainties. For video streaming applications, multiple resources should be considered to optimize resource utilization, including networking, computing, and caching. The computing related metrics include energy consumption and throughput, and the networking related metrics include QoE and latency. The caching related metric is cache hit ratio. It is noticed that the general metrics of entire integrated system include cost, fairness, and revenue.

Resource allocation problems are generally posed as optimization problems with one or more objective functions and some constraints. For the single-objective optimization problem, three sets of solutions include the universal set, set of feasible solutions, and set of optimal solutions. The traditional optimization tools can be used to find a minimum or maximum of objective function. The traditional optimization tools include convex optimization, game theory, stochastic optimization, and reinforcement learning. It is noticed that the solution of convex optimization is the global optimal solution, while stochastic optimization cannot guarantee that the solution is globally optimal. Moreover, reinforcement learning is a vital tools to address RA problems in the unknown environment.

Multi-objective optimization is an area of multiple criteria decision making that is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously. There are various classes of methods for generating the Pareto set of multi-objective optimization problems, including mathematical programming and metaheuristics methods. Moreover, mathematical programming methods can be adopted to solve multi-objective problems by using single-objective approximations, including weighted sum, ϵ-constraint, and utility function method. Some typical metaheuristics methods include evolutionary algorithm and simulated annealing. Evolutionary algorithms can identify feasible solutions tend to converge toward optimal or near-optimal solutions. Simulated annealing algorithm has a good approximation solution for a large size solution search space, while it can not guarantee that the solution is globally optimal.

[1]https://www.nsnam.org/
[2]https://launchpad.net/openwns

V. Enabling Technologies

In this section, enabling technologies of video streaming, including blockchain, software-defined networking, network function virtualization, networking slicing, caching-networking, computing-networking, and caching-computing-networking are surveyed.

A. Blockchain

Blockchain technology has aroused great interest in both academia and industry recently, which is a secured/distributed ledger, which does not require third part authorization [129]. In the peer-to-peer (P2P) based blockchain network, the transaction is recorded by the majority, which is transparent to all partners. Blockchain architecture combines cryptography, data management, networking and incentive mechanisms [129]. Moreover, the consensus algorithm is used to decide how the agreement is achieved to record the last block into the blockchain [130], [131]. Blockchain consensus algorithms are divided into two categories: (1) Proof-based consensus algorithm, (2) Byzantine Agreement based consensus algorithm. The core characteristics of blockchain are distribution, trust, transparency, and traceability. Firstly, blockchain is a distributed network system. Data storage, transmission, and accounting are based on distribution. Secondly, blockchain uses consensus machines to solve trust problems. Therefore, the ledger itself can be programmed to trigger transactions automatically. Thirdly, except the private transaction is encrypted in the blockchain network, and the data is transparent to all nodes. Finally, the data in the chain is traceable, and any subtle change is known and recorded by all the nodes.

Therefore, when edge computing meets blockchain, it makes their respective advantages complementary to each other [132]. Integrated edge computing and blockchain will bring many benefits [133]. On one hand, integrating blockchain into the MEC system can provide secure data protection. On the other hand, integrating MEC into the blockchain system can offer rich computation and storage spaces. Blockchain can provide a secure/distributed solution for solving many challenges in video streaming applications [134].

Ensuring security is one of the achievements for blockchain-enabled edge computing network. Rivera $et al.$ [135] propose a blockchain framework to provide a trusted collaboration mechanism among edge servers in a MEC environment. With the proposed framework, Hyperledger Fabric is adopted to add a security blanket to the task sharing/offloading processes among edge nodes. A Hyperledger Fabric network runs with M number of nodes, and node N requests additional resources to perform a specific task. Therefore, the proposed system will provide smart and distributed communication architecture between the edge and cloud levels.

Due to the separate storage locations in MEC systems, data integrity is criticized. The integration of blockchain and edge computing for data storage can guarantee decentralized data validity. Zyskind $et al.$ [136] develop a decentralized personal data management system that ensures users own and control their data. Based on a distributed hashtable (DHT), the integration of blockchain and off-blockchain storage is designed to provide privacy.

In the MEC systems, blockchain can work as the incentive as well as verification. the blockchain-enabled MEC systems, Feng $et al.$ [137] propose the blockchain-enabled MEC systems, where MEC servers process offloading and blockchain tasks simultaneously. The computation tasks can be offloaded to MEC, and the transactions are recorded by the blockchain. The benefit of this integrated system is to maximize the revenue of MEC computing and storage resources.

1) Lessons Learned: Key lessons learned from the integration of blockchain and MEC systems discussed above are summarized below.

- Multimedia streaming communications combined with the blockchain communications ensure the secure and reliable cooperation of network. Furthermore, the combination of blockchain and distributed multidomain networks can effectively accomplish the trusted cross-domain collaboration and topology privacy protection in 5G networks.
- In a fully decentralized network, blockchain-based framework can guarantee data integrity. However, due to the limited size of data storage provided by blockchain, offchain data storage is necessary for multimedia applications.
- In the MEC systems, blockchain can provide computing resources management. The computation offloading from the end device to the edge server enables the resource-limited end users to take part in the blockchain.

B. Software-Defined Networking

Software-defined networking (SDN) is an emerging network architecture with separate control and forwarding and direct programming. The core of SDN is to decouple the tightly coupled network architecture of traditional network equipment into a three-layer separation architecture of the application, control, and forwarding, and implement centralized control of the network and programmable of network applications through standardization [138]. SDN is based on OpenFlow protocol which is a communications protocol that gives access to the forwarding plane of a network switch or router over the network [139]. The SDN architecture is directly programmable, centrally managed, and programmatically configured.

The MEC system is composed of an array edge servers with three capabilities: storage, communication, and computing. MEC can push the services to the edge and closer to the end-user with ultra-low latency and high throughput. However, this approach comes with some challenges [140]: (1) The computational resources of MEC servers are limited; (2) The heterogeneity of requests from users needs to be handled; (3) It needs to achieve the load balancing among MEC servers. Therefore, the integration of SDN and MEC can bring some benefits, including effective control, service-centric implementation, adaptability, and interoperability [141]. SDN enables effective network management by presenting a global perspective of the network [142].

Salman $et al.$ [143] design an architecture that employs the SDN paradigm while extending the MEC concept, named SD-MEC. With the proposed system, the heterogeneity and the scalability of the IoT network can be addressed. Moreover, the software-defined gateways can provide interoperability and virtualized functions for the IoT nodes. A collection of network functions are performed by these gateways.

SDN can be adopted to improve the scalability and cost-efficiency of MEC servers. Jararweh $et al.$ [144] introduce software defined based framework to enable efficient mobile cloud computing (MCC) services. The SDN components consist of software-defined compute, software-defined storage, and software-defined security. The integration of SDN and MEC system can smoothly handle the global and local client requests.

Video over SDN can enhance users' QoS/QoE for real-time interactive video applications. Zhang $et al.$ [145] propose a video multicast orchestration approach by using SDN in ultra-dense networks (UDN) environment. The proposed framework of the hybrid and concurrent multicast (HCM) consists of group identification and management, hybrid multicast method, multicast orchestration, and packet loss recovery. Based on HCM, the bandwidth consumption and delivery latency can be reduced. The advantage of HCM is to optimize the network resource and limit the video traffic. However, the scalable video streaming approach is used considered.

Yang $et al.$ [120] focus on scalable video streaming over dense small-cell networks (DSCN), where segment-routing technology in SDN is proposed to enhance the network collaboration. Moreover, a joint optimization problem is modeled to maximize the time-averaged QoE subject. In this context, the Lyapunov optimization technique is used to solve this problem. The advantage of the proposed approach is that an intelligent scalable video transmission framework is designed to improve users' QoE. However, this imposes a challenge on the transmission schedule, because the SDN has to frequently configure forwarding paths for each flow.

In order to leverage the flexibility of SDN and features of forwarding device, Bentaleb $et al.$ [146], [147] focus on addressing HTTP adaptive streaming (HAS) scalability problem further, an intelligent streaming framework, named SDNHAS, is proposed to improve the users' QoE, where SDN is used to help HAS players make better decisions. Moreover, a novel algorithm is designed to group players into clusters to optimize the performance. In this context, the computational complexity of the proposed approach is low. However, SDNHAS should be extended to support multi-bottleneck shared networks.

In order to assure the best users' QoE, Liotou $et al.$ [148] develop a programmable QoE-SDN video application, where mobile network operators (MNOs) and video service providers (VSPs) are isolated by using SDN. Moreover, the Knapsack optimization method is used to model the video selection problem. The simulation results show that the proposed scheme can enhance QoE as well as saving bandwidth. The advantage of the proposed approach is that enabling network support feedback capabilities. However, the scalability need to be investigated.

1) Lessons Learned: Key lessons learned from the integration of SDN and MEC systems discussed above are summarized below.

- SDN is the key enabling technology of 5G networks, which improve the performance of video streaming. SDN provides MEC system with decision making for computational offloading in 5G networks, including the channel conditions, server load.
- The integration SDN and MEC system brings easy pluggability of edge nodes, where new devices are easily detected and traffic information can be easily updated.
- The integrated framework can compute the load balancing by using a collection of OpenFlow-based statistics.

C. Network Function Virtualization

Network function virtualization (NFV) is a network architecture concept that uses the technologies of IT virtualization to virtualize entire classes of network node functions into building blocks. NFV can decouple software from hardware, offer flexible network function deployment, and support dynamic operation [149]. The whole NFV architecture can be divided into three parts: NFV infrastructure (NFVI), virtual network function and services (VNFs), and NFV management and orchestration (NFV MANO). NFV has the potential to lead to significant reductions in OPEX and CAPEX [150].

The MEC server includes a mobile edge host and associated management components, and the mobile edge host consists of the mobile edge platform and the virtualization infrastructure. The MEC and NFV have similar architectures and both require a network infrastructure virtualization platform. Therefore, the management function of MEC/NFV has components that can be combined and re-divided [151]. The MEC platform is deployed as a VNF, and the MEC applications behave as VNFs for the rest of NFV MANO components. The virtualization infrastructure is deployed as the NFVI [152].

NFV decouples the software and hardware to reduce the cost of communication systems. In Li's work [153], an architecture of NFV-based MEC platform and analyzes is designed to enhance QoE. Virtualized network functions can help MEC to diminish the negative effects of hardware, features, and platform. Moreover, NFV can provide flexible and scalable services for MEC. In this context, the open platform for NFV (OPNFV) is used for QoE analysis of HTTP video, and the mean opinion score (MOS) is adopted to evaluate QoE. The advantage of NFV-based MEC approach is that the high bandwidth and ultra-low latency are provided to improve video quality. However, virtualization's impact on cellular networks has not been considered.

Kourtis $etal.$ [154] focus on developing an appropriate QoE assessment solution in 5G small cell networks (SCNs) by taking account of virtualization's impact. The video quality assessment (VQA) is proposed to address the quality degradation based on the structural similarity (SSIM) index, where SSIM index can be as a VNF. The VQA task is performed at an edge server, which can reduce the power consumption of the user equipment (UE). Moreover, a long term evolution (LTE) infrastructure is adopted to implement, test, and evaluate the VQA approach. In addition, the proposed method can be offered as a service to mobile network operators. However, the dynamic deployment of NFV has not been studied.

To address network latency and stalling of video applications, Ma $etal.$ [155] proposes an NFV-based MEC system to improve ultra-high quality video streaming in 5G networks. The proposed NFV-based streaming system consists of the streaming server, MEC server, and UEs, where MEC server can be as a VNF. By the dynamic deployment of NFV, the power consumption of the system can be reduced. Based on DASH protocol, the adaptive video streaming technology is used to support streaming seamlessly with bandwidth fluctuation. An important advantage of the method lies in the fact that it provided $4K$ and $8K$ video streaming efficiently with lower latency. However, the context-awareness component is not investigated.

Furthermore, the flexibility of 5G networks is not exploited well in the state of the art. By integrating SDN and NFV, Alvarez $etal.$ [156] proposes a service virtualization platform (SVP) to facilitate the development, deployment, and operation of media services. The proposed system architecture is comprised of the development and service preparation and evaluation layer, the SVP layer, and the physical layer, which can support the cognitive management environment for the provisioning of network services (NSs) and media-related applications. Moreover, the integration of edge computing and multimedia service can simplify development and deployment time. The main advantage of the proposed platform is it can flexibly adapt service operations for dynamic conditions. However, the cost effectiveness of the operation is high.

1) Lessons Learned: Key lessons learned from the integration of NFV and MEC systems discussed above are summarized below.

- The integration of the 5G network with NFV can provide some benefits for reducing the infrastructure cost and network latency for video streaming applications.
- In 5G networks, NFV optimizes resource provisioning of the VNFs for cost and energy efficiency and ensures the coexistence of VNFs with non-virtualized network functions.
- NFV ensures performance guarantees of VNFs operations, including failure rate and latency.

D. Networking Slicing

5G network slicing is a network architecture that enables the multiplexing of virtualized and independent logical networks on the same physical network infrastructure [157]. Several techniques are employed, including network functions, virtualization, and orchestration. The main architecture of networking slicing consists of service layer, network function layer, infrastructure layer, and network slice controller [158]. Network slicing can optimize radio resources to provide performance guarantees and isolation for 5G use cases, including ultra-reliable low-latency communications (URLLC), massive machine-type communications (mMTC), and enhanced mobile broadband (eMBB) [159].

MEC can support virtual network function for core network and radio access network for latency sensitive services, and MEC server can be considered as a pool of virtualized resources [160]. MEC provides networking, computing, and storage resources, which is a key component of network slicing. Moreover, MEC plays different roles for different 5G network slicing [161]: (1) For URLLC applications, MEC can provide low latency and high reliability services. (2) For mMTC applications, MEC can ensure high performance and scalability. (3) For eMBB applications, edge caching can increase overall capacity. Therefore, the integration of MEC and network slicing can reduce network management complexity [162].

To achieve the extremely low latency for multimedia streaming in 5G networks, Dong $et al.$ [163] proposes a mobile multimedia streaming slice model. The mobile multimedia streaming slice consists of storage allocation slice, network bandwidth allocation slice, context tracking slice, route calculation slice, content provisioning and delivery slice. The base station (BS) mapping algorithm is proposed to optimize storage and bandwidth resources. Then, the orchestration procedures for the multimedia streaming service is designed to orchestrate the necessary services. In addition, the proposed approach is contextually aware. However, QoS's impact on networking slicing is not discussed.

Wang $et al.$ [164] focuses on studying the migration of multimedia services to 5G network challenges in the QoS, and the SliceNet framework is proposed to improve the performance of eHealth telemedicine services based on networking slicing. The SliceNet consists of the application programming interface (API), the RAN and core slicing, the low-complexity platform, and the control framework. Moreover, the proposed framework can be applied to the eHealth system with a high degree of QoS guarantee. An important advantage of the method lies in the fact that managing the QoS is considered for sliced services. However, the QoE of the multimedia system needs to be improved.

In order to improve user QoE of video streaming, Yan $et al.$ [119] propose a customized transmission protocol in sliced networks by using SDN and scalable video coding (SVC) technologies. The proposed protocol is composed of bottleneck queue management, selective caching, and enhanced transmission (ET). According to the network condition, ET can be devised to improve users' QoE. Therefore, the proposed approach can alleviate network congestion and provide better QoE. However, the reliability of the system is not studied.

Unlike Yan's work [119], Khan $et al.$ [165] propose a joint video quality selection and resource allocation method to enhance the vehicular devices' QoE in sliced vehicular networks. The optimization problem is formulated to maximize the QoE, which can be divided into two subparts. Moreover, a network slicing-enabled clustering method is adopted to partition the network. Then, the Lyapunov method is used to solve the vehicle scheduling and quality selection problem. The main advantage of the proposed approach is that it achieves low latency and high-reliability communications. However, the computational complexity of the proposed algorithm is high.

1) Lessons Learned: Key lessons learned from the integration of networking slicing and MEC systems discussed above are summarized below.

- Networking slicing is a key enabler of 5G networks to address the cost, efficiency, and flexibility requirements of video streaming applications.
- 5G RAN slicing can relieve the pressure on the core network, reduce the transmission delay for multimedia applications.
- The integration of MEC and network slicing can address architectural, security, and isolation-related challenges in 5G networks.

E. Caching-Networking

In this section, the state-of-the-art MEC caching categories include video content placement and delivery. Therefore, some typical caching schemes are introduced and discussed here. Finally, the summary of edge caching according to their structures is listed in Table IX.

1) Placement:

To improve QoE for video streaming in information-centric networks (ICNs), a centralized edge caching scheme is proposed in [63]. The controller collects the network conditions, and the video pre-fetching decision is made. In this context, the proposed scheme is based on dynamic adaptive streaming over HTTP (DASH) protocol. Assuming a Poisson arrival process of video streams [171], the controller can formulate as a function of video initial request arrival rate. The tasks performed by the controller include residual capacity estimation and segment scheduling. Moreover, the resource allocation mechanism is designed to improve users' QoE. The tested includes a video server, a cache, a proxy, and three clients. In this scenario, the video streaming is based on a variable bit rate (VBR), and the testing video is trimmed from the DASH data set.

The authors of [64] propose a DASH-aware video caching approach to enhance users' QoE by using edge computing in the cellular network. The main caching strategy is to ensure that the bit rates of cached video representations match the network's downlink capacity. Based on the content context (e.g., popularity) and network condition (e.g., throughput), a two-dimensional user QoE-driven cache replacement method is designed to maximize video quality as well as minimizing playback quality fluctuation. The performance of the proposed approach is evaluated on a real LTE-A wireless testbed. 3 identical copies of the video Big Buck Bunny [1] are used for video traces. However, the caching manager acts as the centralized controller for video caching at the local cache.

Similar to [63], [64], the authors of [65] propose the radio network-aware edge caching and updating approach to enhance users' QoE based on DASH and radio network information service (RNIS). A cached video is determined by network condition and video popularity. Firstly, the information of network edge is obtained to help MEC cache. In the context, assume that some segments with many representations are

[1] https://peach.blender.org/download/

TABLE IX: The summary of MEC caching for video streaming.

Ref.	Caching Strategy	Networking	Objective	Control	Optimization Tool	Streaming	Evaluation	Main Findings
[63]	Placement	Macro-cellular networks	Improving users' QoE	Centralized	N/A	VBR, HTTP-based	Tested	This approach is optimized for multi-rate DASH videos.
[64]			Maximizing video quality as well as minimizing playback quality fluctuation	Centralized	N/A	VBR, HTTP-based	A real LTE-A wireless testbed	This approach is driven by the popularity and network conditions.
[65]			Improving video quality as well as reducing buffering time based on video popularity	Centralized	N/A	VBR, HTTP-based	A real 4G LTE BS	A cached video is determined by network condition and video popularity.
[166]			Minimizing sum of content access delay	Centralized	Approximation algorithm	VBR	Simulation	Increasing caching capacity contributes to delay reduction and hit ratio increment.
[167]		D2D networks	Maximizing the system utility	Centralized	Approximation algorithm	VBR, HTTP-based	Simulation	The system utility achieved by every strategy decreases with the increase of the number of video quality levels.
[66]			Maximizing the QoE as well as minimizing the backhaul data traffic and edge data processing	Centralized	Online algorithm	VBR, HTTP-based	Simulation	The integration of D2D into the collaborative edge caching and processing can help to alleviate load from the backhaul network.
[85]		HetNets	Maximizing the aggregate average video distortion reduction as well as minimizing the additional cost of representation downloading	Centralized	Approximation algorithm	VBR	Simulation	The proposed approach is applied to VoD systems with larger settings.
[87]		RAN	Minimizing the rate of cache misses as well as maximizing the average video quality	Centralized	Approximation algorithm	VBR, HTTP-based	Simulation	Using the proposed approach yields noticeable reduction in the percentage of cache miss compared to both least-frequently-used (LFU) and least-recently-used (LRU).
[168]			Maximizing the QoE of individual client as well as minimizing the backhaul data traffic	Centralized	Approximation algorithm	VBR, HTTP-based	Simulation	The superiority of the proposed system can indeed act as guidelines for DASH video streaming in next generation of mobile networks.
[115]	Delivery	Macro-cellular networks	Maximizing the weighted average QoE of VoD users	Centralized	Convex optimization	VBR	Simulation	The proposed caching policy efficiently improves the average QoE of VoD users.
[67]		D2D networks	Maximizing the total profit	Distributed	Approximation algorithm	VBR	Simulation	The caching-enabled mobile device prefers to store the contents with larger popularity for the linear relationship.
[86]		HetNets	Maximizing the QoE, fairness as well as minimizing the overall data traffic	Centralized	Online algorithm	VBR, HTTP-based	Simulation	The collaborative edge caching is more effective in reducing backhaul data traffic than increasing video bitrate.
[169]			Maximizing the number of requests	Centralized	Greedy algorithm	VBR	Simulation	The traffic grows smoothly after applying the adaptive-rate approach when requests arrive.
[105]			Finding the best cache helper	Distributed	Game theory	VBR	Simulation	The reward pricing scheme with identical reward potential is able to achieve better fairness among the cache helpers
[99]			Maximizing the cache hit ratio	Distributed	N/A	VBR	A realistic network	The caching algorithm based on blockchain based CDN significantly reduces the delivery time.
[91]		RAN	Maximizing the revenue as well as minimizing the average delay	Distributed	Matching theory	VBR	Simulation	The average revenue obtained by the proposed optimal policy achieves the best performance.
[170]	Joint placement and delivery	HetNets	Maximizing the utility	Distributed	Greedy algorithm	VBR	Simulation	The caching price increases over the caching capacity of the edge node.
[90]		RAN	Maximizing the utility of the cache node	Distributed	Game theory	VBR	Simulation	The utility of mobile users in each social group decreases with the increase of price.

stored at MEC server. Secondly, the cache replacing algorithm is developed to improve video quality and reduce buffering time based on video popularity, including requested popularity and expected popularity. In addition, the testbed is based on a real 4G LTE BS, and the testing video is trimmed from the DASH data set and divided into 50 segments. However, the incentive mechanism of caching needs to be investigated in future research.

Existing caching schemes lack interaction of video caching and retrieving and collaboration among servers. The work in [166] proposes collaborative video caching and retrieving in multi-server MEC systems under two timescales. Moreover, a two-timescale optimization problem is formulated to minimize the sum of content access delay, including the long-term caching placement and short-term video retrieval mode selection. Finally, a sample average approximation based two-phase algorithm is adopted to solve this problem. In the simulation, the authors assume that video requests at SCceNBs follow a Poisson process, and the popularity of videos at different SCceNBs follows Zipf distribution [172]. However, trust management of collaboration servers should be investigated in future work.

To avoid the unnecessary caching cost and improve QoE, the authors of [167] propose an architecture for D2D-based cellular network, which integrates radio access network (RAN) traffic handling, UE caching, and video quality adapting abilities into MEC. In the proposed framework, multiple UEs can cache and share multi-bitrate videos, and the caching strategy is based on placement. In the system model, the authors assume that the video request arrival is a Poisson process, and the popularity of videos follows a Zipf distribution. Moreover, the proactive multi-bitrate video caching strategy problem is formulated to maximize the system utility, which is an integer linear programming problem. In this context, the approximate solutions are used to solve this problem. In future work, the cooperation between multiple MEC servers should be considered.

However, the impact of D2D on joint QoE-traffic optimization is still underexplored in previous works. The authors of [66] propose a system for joint QoE-traffic optimization and edge processing based on MEC and D2D communication. In the proposed system, the neighborhood edge servers can collaborate with the end devices, and neighborhood devices can share resources through D2D communication. The authors assume that the relation between video bitrate and the perceived quality by the client is linear, and the system design takes explicit measures to ensure fairness. Therefore, a joint optimization problem is formulated to maximize the QoE as well as minimizing the backhaul data traffic and edge data processing. Then, a low-complexity online algorithm is adopted to solve this problem. In future work, the switching mechanism of retention-based collaborative caching at the local cache need to be investigated.

Li $et al.$ [85] take account of the rate-distortion characteristics of the video encoder and the collaboration edge servers, and a QoE-driven edge caching placement approach is proposed to reduce the service load of BSs. In the system model, the authors assume a caching system where a representation of a video file is either cached fully or not cached at all in any edge server. Moreover, the rate-distortion (R-D) model is adopted to denote the distortion of the representation of the video [173]. In this context, the distributed caching placement problem is modeled to maximize the aggregate average video distortion reduction as well as minimizing the additional cost of representation downloading. The constraints include the storage capacity and the transmission and initial startup delay. Moreover, an approximation algorithm is used to solve this problem with low complexity. In the simulation, the authors assume that the neighboring edge servers are operating on the orthogonal bands and each edge server allocates its transmission resource in a fair and uniform way. Moreover, four test videos, including Crowd Run, Riverbed, Tractor, and Sunflower [1], are selected as the video files needed for caching. The proposed scheme should be extended to future network architectures, including SDNs and ICNs.

In order to achieve the trade-off between video traffic and QoE, the authors of [87] study the cache-aware QoE-traffic optimization. In this context, the impact of fixed video content caching on the optimal QoE of mobile clients has been investigated. The theoretical downlink throughput is based on Shannon upper bound approximation [174]. In order to avoid stall events, the authors assume that the player starts to play the video after the startup phase. Then, the joint optimization problem is formulated to minimize the rate of cache misses as well as maximizing the average video quality. Moreover, an approximation algorithm is adopted to solve this problem with low complexity. In the simulation, the adaptive streaming is transmitted by DASH, and the chunks of four videos with different popularities are available in ten different qualities. However, the collaborative edge caching scheme needs to be exploited in future work.

In previous work, the objective of jointly optimizing the trade-off between the QoE of mobile video streaming clients and created backhaul traffic has been overlooked. The authors of [168] propose the green mobile edge computing (GMEC) system to improve the QoE of HTTP adaptive streaming. In this context, the Gaussian-shape function is used to estimate the time-varying solar energy harvesting during one day time duration [175]. Firstly, the joint optimization problem is formulated to maximize the QoE of the individual clients as well as minimizing the backhaul data traffic. Then, a suboptimal algorithm using self-tuning parametrization mechanism is used to solve this problem. Moreover, a proactive edge caching heuristic algorithm is proposed to improve cache placement efficiency. For the performance verification, the adaptive streaming is based on DASH, and four videos with different popularities are divided into consecutive chunks. In addition, the energy consumption problem should be studied by using D2D communication among neighborhood nodes.

2) Delivery:

Furthermore, a QoE-oriented edge caching strategy for video on demand (VoD) applications taking into account video quality adaptation is investigated in [115]. The authors assume that the BSs are equipped with ideal backhaul links, and a VoD user can pull part of a video from the BS and part from the helper. Therefore, the instantaneous signal-to-interference-plus-noise ratio of the typical user is served by its nearest helper. Moreover, the caching policy optimization problem is formulated to maximize the weighted average QoE of VoD users, where the impact of video quality adaptation on QoE is considered. In this context, a low-complexity convex optimization algorithm is used to solve this problem. Results from the experiment show the proposed caching method can improve the QoE of users significantly. However, the incentive mechanism of edge caching should be studied in the future.

Taking the consensus mechanism and the content sharing of UEs into account, Cui $et al.$ [67] propose a blockchain-enabled framework for caching and video sharing by using MEC and D2D communication. In the network model, BSs offer the computing resource to UEs to perform caching tasks and receive the reward, while the cached content can be shared by the neighborhood UE over D2D. In the blockchain system, the process of solving problem can be modeled as a stochastic process following a Poisson process. Moreover, the optimal caching scheme is formulated to maximize the total profit for both the linear and nonlinear relationships between computing power and content sharing size. Finally, the difference of the convex (DC) program is adopted to address this problem. In the simulation, the popularity distribution is modeled as a Zipf distribution, and the results show that the caching-enabled mobile device prefers to store the contents with larger popularity for the linear relationship. However, the UE mobility management is ignored in this work.

Further research on [87], the authors of [86] propose a collaborative edge caching mechanism to improve users' QoE. In the proposed network framework, the neighborhood edge servers form the edge clusters in which they collaborate to serve the request of users. The impact of collaborative edge caching on joint QoE and backhaul data traffic is studied. In system model, the authors assume that there is no traffic incurred when the requested chunks are retrieved from the local caches. Moreover, joint QoE-traffic optimization is formulated to maximize the QoE, fairness as well as minimizing the overall data traffic. Finally, an online scheduling algorithm is adopted to solve this problem with low complexity. With the simulation, the collaborative edge caching is more effective in

[1] http://media.xiph.org/video/derf/

reducing backhaul data traffic than increasing video bit rate. However, the incentive mechanism needs to be investigated in future research.

In order to reduce network bandwidths and balance traffic loads, the authors of [169] propose an integrated video distribution approach, including in-network caching and collaboration distribution. In the proposed framework, a controller can collect fine-grained network and user information across various domains by using SDN. The authors assume that an edge server can be concretized as a BS that caches content and serves a number of user equipment. The request screening problem is formulated to maximize the number of requests based on graph theory, and a greedy algorithm is developed to solve this problem. Moreover, a joint source redirection and flow routing problem is formulated to maximize the minimum buffer time, and a linear program is adopted to solve this problem. In the simulation, 200 videos from Youtube dataset [1] are provided. The results show that the traffic grows smoothly after applying the adaptive-rate approach when requests arrive. However, the payment mechanism should be studied in future work.

To offer the financial incentive in the autonomous content-caching market, the authors of [105] propose a decentralized caching framework with blockchain. In the system model, Zipf distribution is adopted to model the probability of content. An automatically executed smart contract is constructed to manage the caching market. Then, a proof-of-stake (PoS)-based consensus mechanism is adopted to encourage the cache helpers to stay active in service. Moreover, the proposed caching system is mapped as a Chinese restaurant game theory, and a decentralized Nash Equilibrium (NE) searching method is used to find the best cache helper. The performance of the proposed caching algorithm significantly outperforms that of random content selection in both the cache helpers' average payoff and the total number of offloaded deliveries. However, the trust degree of end users needs to be quantified in future work.

To comprehensively offer user privacy and benefit to all partners, the authors of [99] propose a blockchain-enabled CDN architecture. The proposed system leverages users' information to different content provides and develops a public ledger for the demands of content providers. The cosine similarity [176] is adopted to measure the correlation between content and the global feature popularity. Moreover, feature-based edge caching algorithm is proposed to maximize the cache hit ratio. The video traces are based on MovieLens 20M Dataset [1]. The proposed caching algorithm can improve the performance in term of the cache hit ratio and quality of experience. However, UE mobility needs to be investigated in future work.

Existing caching schemes ignore the security of MEC. the authors of [91] propose a dynamic edge caching framework with blockchain. In the proposed system, edge servers can authorize legitimate users to predict and update their content demands based on blockchain and advanced machine learning technologies. Moreover, a caching resource allocation problem is modeled to maximize the revenue as well as minimizing the average delay. Finally, a matching theory is used to address this problem. The simulation shows that the average revenue obtained by the optimal policy achieves the best performance. However, collaborative caching and payment management in MEC are open research issues.

3) Joint Placement and Delivery:

When we jointly design the content placement and delivery, the traffic between BSs and UEs can be significantly offloaded. Therefore, in this subsection, the joint placement and delivery approaches are described as follows.

Securely caching content is still not discussed well in previous works. Xu $etal.$ [170] propose a blockchain-enabled trustworthy edge caching approach for the mobile cyber-physical system (MCPS). In the network model, the mobile cyber physical system consists of content provider, multiple edge nodes, and some social groups. In the content model, the authors assume that the popularity of contents follows the Zipf distribution. In the blockchain system, Proof-ofwork (PoW) consensus algorithm is utilized to reach consensus on a mined block. The caching transactions between BSs and UEs are recorded by the decentralized blockchain. Then, a trust management mechanism is designed to enjoy the trustworthy caching services, where the trust degree of BSs can be updated dynamically. Moreover, a max-min based fair caching resource allocation problem is formulated to maximize the utility. With the simulation, the proposed approach can reduce delivery delay to improve mobile users' QoE. However, the cooperative edge caching approach needs to be studied to improve the profits and QoE in future research.

Due to the limited caching capacity, how to allocate the caching capacity on a cache node with an optimal price is a great challenge. The authors of [90] propose an edge caching approach to cache layered videos in RAN. The proposed system consists of content servers, edge cache nodes, and multiple social groups. In the content model, the authors assume that one video content is stored for a social group on the cache node, and the content's popularity is characterized by a Zipf distribution. The framework is designed to cache the layered videos based on user requests. Then, the system and content models are formulated to maximize the utility of the cache node. Moreover, a Stackelberg game theory is utilized to solve this problem. However, a distributed content caching approach needs to be studied in future work.

4) Lessons Learned:

Mobile edge caching technology can not only improve the utilization of caching resource, but also reduce the network transmission cost. Generally, BSs usually cooperatively cache a set of video streaming. Then, a user can obtain the requested video content from either the local BS or from a set of neighboring BSs. In contrast to placement strategies that move the content to the edges of the cellular networks, e.g., BSs, delivery strategies serve the UEs using the cached content. D2D-enhanced MEC caching exploits the extensive storage capacities of end devices. Moreover, the blockchain can work as a distributed ledger and provide the cache sharing incentive mechanism.

[1] http://netsg.cs.sfu.ca/youtubedata/

[1] https://grouplens.org/datasets/movielens/20m/

The main purpose of the existing MEC caching approaches focus on maximizing users' QoE, the utility of the cache node, the number of requests, the revenue, as well as cache hit ratio, and minimizing the overall data traffic, the average delay, as well as the additional cost. However, to achieve the best overall performance, the trade-off between different purposes needs to be studied in future work.

For video streaming applications, the control structure of MEC caching can be divided into a centralized manner and a distributed manner [177]. For video streaming applications, the typical distributed structure refers to the blockchain-enabled MEC systems. In the resource allocation, an MEC with a centralized manner decides the caching placement or delivery strategies. The centralized approach can obtain the optimal solution in the cellular networks, while the computational complexity of solving-problem is high. In the distributed approach, each access point (AP) or user equipment (UE) can decide which file needs to be cached without a centralized controller. The distributed approach can not provide the globally optimal solution, while it has a lower complexity. Moreover, blockchain-based video streaming application is a typically decentralized structure, which can bring high profit for content producers and consumers, and the privacy of consumers can be enhanced.

To improve the performance of edge caching, the optimization problems have usually been solved with some mathematical tools, including convex optimization, game theory, stochastic optimization, and approximation algorithms. In the future, the various tools should be combined to solve and trade-off various problem aspects. In addition, most of the previous works are based on simulation, and some works are based on practical testbed. Therefore, the result evaluations should consist of simulation and testbed measurements in future research.

F. Computing-Networking

This subsection focuses on the MEC computation with the partial offloading or full offloading. The computation offloading is a very complex process affected by some factors, such as user preference, UE capability, and MEC capability. The related schemes are summarized in Table X.

1) Partial offloading:

With the rapid growth of immersive video technology, panoramic virtual reality video (PVRV) delivery over wireless networks can bring challenges for bandwidth and latency. The authors of [76] propose the tight of MEC with the multi-connectivity millimeter wave (mmWave) cellular network to accelerate PVRV streaming applications. With the proposed system, MEC is intermediate processing component to enable transcoding and execute viewport rendering computations, which can satisfy the high bandwidth and ultra-low latency demand of PVRV streaming. In the system model, the PVRV streaming system consists of the content provider, the MEC server and the mmWave/sub-6 GHz cellular network. The authors assume that chase combining with reduced half size retransmission is used and each re-transmission uses the same channel coding schemes mode. Moreover, the communication and computation resource allocation problem is formulated to minimize the energy consumption as well as maximizing the quality of the received viewport. Then, the non-dominated sorting genetic algorithms are adopted to address this problem. Finally, NS-3 simulator is used to verify the PVRV streaming performance of the proposed method. In addition, the payment mechanism needs to be studied in future work.

In the above works, the security enhancement issues are ignored in MEC-enabled video applications. Lin $etal.$ [114] propose the optimization of multimedia transmission delay and the computation delay of the MEC-based healthcare applications. In the proposed system, a blockchain is integrated into the MEC system, which includes the integrity chain and fraud chain. In the communication model, the authors assume that these mobile users employ OFDM schemes for communications, and the data rate of mobile user is based on Shannon theorem. The priority of a mobile user is recorded into the blockchain ledger and broadcast to all the users. The consensus mechanism is adopted to ensure authenticity. Moreover, a communication and computation resource allocation problem is formulated to minimize the overall delay with the authenticity of the user priorities. Finally, Lagrange-based approaches are used to solve these problems. In future work, blockchain-enabled incentive mechanisms need to be designed.

The main drawback in [74] is that it ignores the fairness and resource utilization for adaptive video streaming delivery. The authors of [97] propose an optimized approach for network assisted adaptation specifically targeted to mobile streaming in MEC systems. Moreover, a joint optimization problem of QoE and fairness is formulated to maximize the resource utility, which is an NP-hardness problem. Therefore, a near-optimal greedy-based online algorithm is adopted to address these problems, including the client to edge server mapping and the bitrate selection problem. The performance evaluation is based on a standard LTE simulator, called SimuLTE[1], and the results show that the proposed algorithm performs overall better in terms of QoE, fairness, and resource utilization. In future work, an optimized edge caching and bitrate adaptation approach should be studied to maximize QoE as well as minimizing the network traffic.

In order to reduce the delay of computation offloading, the authors of [104] study the joint computation partitioning and resource allocation based on MEC for an AR application. In the system model, the users share the bandwidth provided by the wireless network and the task offloading can be executed on both local devices and the edge servers. The joint optimization problem is formulated to minimize the average application delay. Then, an offline algorithm, namely multi-dimensional search and adjust, is used for theoretical performance analysis. Furthermore, an online algorithm, named by cooperative online scheduling, is adopted to solve this problem in practical deployment. In the simulation, a constant bit rate (CBR) is fixed, and the proposed approach can outperform the benchmark methods in terms of the overall application latency. However, the main limitations of this work are the lack of consideration of UE energy consumption and incentives.

[1] https://simulte.com/

TABLE X: The summary of MEC computation offloading for video streaming.

Ref.	Offloading Type	Networking	Objective	Control	Optimization Tool	Streaming	Evaluation	Main Findings
[76]	Partial offloading	Macro-cellular networks	Minimizing the energy consumption as well as maximizing the quality of the received viewport	Centralized	Genetic algorithm	VBR	NS-3 simulator	The proposed scheme improves the PVRV streaming performance in term of latency, energy efficiency, and QoE.
[114]			Minimizing the overall delay	Distributed	Convex optimization	VBR	Simulation	The network delay in the local/edge cloud computation model is much lower than that in both the local computation model and the edge cloud computation model.
[97]		HetNets	Maximizing the resource utility	Centralized	Greedy algorithm	VBR, HTTP-based	Simulation	The proposed approach performs overall better in terms of QoE, fairness, and resource utilization.
[104]		RAN	Minimizing the average application delay	Centralized	Online algorithm	CBR	Simulation	The average latency increases as the number of users increases.
[178]			Maximizing users' QoE	Centralized	Greedy algorithm	VBR, HTTP-based	Simulation	The proposed approach achieves high fairness value and efficiently uses the bandwidth.
[74]			Maximizing QoE while guaranteeing wireless communication energy efficiency	Centralized	The truncation-point-optimized algorithm	CBR	Simulation	The proposed approach has significant energy efficiency potentials for big data multimedia communications.
[179]			Maximizing the system performance	Centralized	Greedy algorithm	VBR	Simulation	The human detection accuracy increases with the average video coding ratio increasing.
[75]	Full offloading	Macro-cellular networks	Minimizing the mobile energy consumption	Centralized	Approximation algorithm	AR	Simulation	The proposed approach achieves minimal mobile energy expenditure.
[78]			Maximizing the streaming quality for users as well as minimizing the computation and bandwidth cost	Distributed	Heuristic algorithm	VBR, HTTP-based	Simulation	The startup delay is reduced with the improvement of download speeds.
[92]			Improving uers' QoE	Centralized	N/A	VBR	A real life testbed	The MOS value starts reducing with the degradation in network conditions.
[180]		D2D networks	Maximizing the transcoding revenue	Distributed	DRL	VBR	Simulation	The proposed scheme receives higher transcoding revenue.
[100]		HetNets	Maximizing the average transcoding profit	Distributed	ADMM	VBR	Simulation	The average transcoding profit of our proposed algorithm is very close to that achieved by the centralized algorithm.
[80]		RAN	Maximizing QoE of multiple video streams while satisfying the bandwidth and video transcoding capability	Centralized	Approximation algorithms	VBR	Simulation	The CloudEdge system performs in-network transcoding and resource management in a wireless access network.
[110]			Maximizing the bandwidth utility	Centralized	Approximation algorithm	VBR, HTTP-based	Simulation	The proposed system adapts to the diversity of users' channel conditions with lower transcoding costs.
[81]			Minimizing bandwidth and latency	Centralized	N/A	VBR	A 4G LTE tested	The proposed approach provides the immediate benefts in bandwidth saving.
[116]			Maximizing the revenue	Distributed	Game theory	VBR	Simulation	The proposed system achieves good performance in terms of average time to finality, average access delay, and network cost.

The authors of [178] propose an edge cloud-assisted rate adaptation solution to improve QoE of video streaming in the RAN. Firstly, a joint throughput estimation method is proposed to enhance video quality for HTTP adaptive streaming. Moreover, a rate adaptation optimization problem is modeled to maximize users' QoE, which is the non-deterministic polynomial-time NP-hard problem. In this context, a greedy algorithm is designed to solve this problem. The simulation results show the proposed algorithm is promising under slow-moving and fast-moving environments. In future work, the payment mechanism should be studied.

However, the above research focus on latency bounded multimedia delivery over RAN, while the energy efficiency considerations are ignored. The authors of [74] develop a multimedia sensing and communication framework, namely MSaaS, and propose a resource management strategy to improve the energy efficiency in MEC-enabled IoT networks. In this context, the optimization problem is formulated to maximize QoE while guaranteeing wireless communication energy efficiency. Moreover, a truncation-point-optimized algorithm is adopted to address this problem. The simulation results show the proposed approach has significant energy saving potentials. In addition, the payment mechanism should be studied in future work.

Most existing works about edge computing for processing video tasks mainly adopt the "partition and allocation" strategy without cooperative processing. The authors of [179] propose an edge computing framework to improve object detection accuracy. The proposed system consists of a camera node, edge node and server. In the system model, the authors assume that each edge node has the same computation rate, and the human detection accuracy is modeled as a function of the video coding rate. Moreover, jointly partition tasks, compress and allocate video sub-tasks problem is formulated to maximize the system performance, which can be divided into the group formation and video-group matching problems, respectively. In this context, a greedy algorithm is used to solve the group formation problem, and a low-complex heuristic algorithm is adopted to address the video-group matching problem. The simulation results show that the human detection accuracy increases and its speed slows down at the same time with the average video coding ratio increasing. However, the trade-off between energy consumption and the completion time should be considered in future research.

2) Full offloading:

The drawback of [104] is that the offloading decision does not take account of energy consumption. In [75], a task offloading framework is proposed for the AR application based on MEC. In the system model, the authors assume that the offloaded application has shared inputs, outputs and computational tasks, and the Channel State Information (CSI) is assumed to remain constant for the frame duration. Moreover, a joint optimization problem is formulated to minimize mobile energy consumption. In this context, the successive convex approximation (SCA) algorithm is adopted to address this problem. The simulation results show that the proposed approach can achieve minimal mobile energy expenditure. However, carrying out the non-collaborative components locally to reduce the energy consumption should be considered in future work.

The traditional adaptive video streaming approaches are oblivious to the user preferences of different peering servers. The authors of [78] propose a joint online transcoding and geo-distributed delivery strategy. In the system model, the authors assume the transcoding is performed from high-bitrate versions to low-bitrate versions, and segments of the highest versions are already replicated to the geo-distributed CDN regions. Moreover, a joint optimization problem is formulated to maximize the streaming quality for users and minimize the computation and bandwidth cost for transcoding and replicating the video segments, which is an NP-hardness problem. In this context, the heuristic and distributed algorithms are adopted to solve this problem with low complexity. Finally, the trace-driven experiments are used to evaluate the performance of the proposed algorithms in terms of users' experience and computation resource utilization. However, the incentive mechanism should be studied to encourage transcoding in future research.

Another idea aiming at improving users' QoE is introduced in [92]. Based on ABR technology, a QoE-aware transcoding approach is proposed based on the mobile edge. In the proposed system, the best configuration is used for transcoding, and the transcoding service can satisfy users' request based virtual network function (VNF). The performance verification is based on a real life testbed, and the simulation results demonstrate that it can ensure reduced latency and better QoE. However, the incentives need to be investigated in future research.

Liu $etal.$ [180] propose a transcoder selection framework for blockchain-based D2D networks, where the stakes, reputation of candidates, and communication and computation capability have been considered. In the system model, the download and upload rates are based on Shannon's Theorem. The DRL-based approach is proposed to determine a group of transcoders. Then, a joint task scheduling and resource allocation problem is formulated to maximize the transcoding revenue, and the constrained stochastic successive convex approximation (CSSCA)-based approach is utilized to solve this problem. The simulation results show the proposed approach can achieve high transcoding revenue as well as QoS requirement. In the future, the distributed transcoder selection with a light-weight DRL approach should be investigated.

Furthermore, the incentive mechanism to facilitate collaboration between video producers and consumers is ignored in the state of the art. The authors of [100] propose a blockchain-enabled video transcoding framework with MEC, while the block size can dynamically be changed to accommodate the time varying nature of video streaming and its low-latency requirements. Mobile users can offload their sub-tasks to MEC servers or nearby mobile nodes via D2D communication. In the system model, the authors assume that the arrival of video segments follows a Poisson distribution, and the video segments generated from the video stream source are maintained in a queue. Then, a joint video offloading and resource allocation problem is formulated to maximize the average transcoding profit. In this context, an ADMM-based approach is adopted to solve this problem with low-complexity. In future research, the transcoders selection approach considering the held stake, reputation value, and communication capability should be investigated.

Unlike [92], Smith $etal.$ [80] propose an adaptive video streaming framework to address the multi-resource management challenges. In the proposed system, the resource optimization problem maximizing QoE of multiple video streams while satisfying the bandwidth and video transcoding capability is proposed. Moreover, two simplified approximation algorithms are used to solve this problem. The first algorithm is the throughput maximization method that is a linear programming (LP) problem. The second algorithm is a heuristic iterative method. However, a prototype network should be built to validate the effectiveness of the algorithm in future work.

The resource utilization problem is further investigated in [110] with the consideration of adaptive bitrate streaming. An adaptive transcoding framework is proposed based on MEC. The source streaming is transmitted to the edge server and transcoded with various versions. In the system model, the user's channel capacity is approximated according to Shannon's Theorem. Moreover, a spectrum resource allocation problem is formulated to maximize the bandwidth utility, which is an NP-hardness problem and hard to solve. A constrained concave-convex procedure (CCCP) based algorithm is used to address this problem with low complexity. Simulations show that the proposed framework can significantly improve system performance. In addition, combining edge computing and caching schemes with DASH should be studied to enhance the performance of video transmission in future research.

Previous works ignore the requirements of 360° video delivery with ultra-low latency and ultra-high throughput, the authors of [81] propose a field of view (FOV) rendering approach at the edge of a mobile network for VR applications. In the proposed framework, the 360° video service can produce a 360° video stream at the video hub. Then, the edge server processes the 360° video and performs optional motion prediction, FOV extraction, and transcoding and optimization to minimize bandwidth and latency. Finally, FOV control indications are sent upstream from the user to the edge server. The experiment evaluation is based on a 4G LTE test lab environment. In future research, VR value-added services in the edge network should be investigated.

The consensus process, the latency of block generation, and incentive mechanisms are not well studied in [180]. The authors of [116] propose a blockchain-enabled MEC network framework for video transcoding and delivery in a decentralized manner, where blockchain maintains publicly auditable video transaction records. With the framework, PoS-based smart contracts are built to execute automatically the transcoding transactions among video producers, MEC servers and users, and the latency of the blockchain generation is considered. In the system model, the authors assume that the channel keeps unchanged during the transmission of a packet, and the transmission rate of user is approximated according to Shannon's Theorem. Moreover, a joint optimization problem is formulated to maximize revenue. In this context, a three-stage Stackelberg game technology is used to solve this problem. The results show that the proposed system could achieve good performance in terms of average time to finality, average access delay, and network cost. However, the security issues regarding the malicious consensus nodes in blockchain should be studied in future work.

3) Lessons Learned:

MEC can aid mobile end devices in providing compute intensive services with low latency. MEC computation offloading type is divided into full offloading and partial offloading. In the full offloading manner, the whole computation is offloaded and processed by the MEC. In the partial offloading manner, a part of the computation is processed locally and the rest is offloaded to the MEC. The partial offloading can save significantly more energy at the UE when compared to the full offloading. With D2D-enhanced MEC, end devices can offload tasks that require heavy computations to powerful MEC servers or nearby UEs in order to fulfill the low latency demands and extend the battery lifetime. Moreover, blockchain-enabled MEC can provide high profit for content creators and high privacy for consumers in the video streaming system.

For video streaming applications, the main purpose of the MEC computation schemes focuses on minimizing the network delay and energy consumption as well as maximizing QoE, the resource utility, bandwidth utility, and revenue. However, the joint optimization approach should be investigated to balance the different purposes in the future.

The structure of MEC computation offloading for video streaming applications is divided into a centralized and distributed manner. In the resource allocation, the centralized structure needs an entity such as BS or eNodeB to take charge of information collecting and computing resource decision. In general, the centralized structure can be implemented with simple connection, and it can obtain the optimal or near-optimal solution. Compared with the centralized approaches, the distributed approach do not need the central entity. Blockchain-based MEC system is a typical decentralized network, where each AP/UE can make the computing resource decision individually. In addition, the distributed approach is more flexible and robust.

For video offloading decisions, the optimization problems can be solved with some optimization tools, including convex optimization, game theory, DRL-based algorithm, and approximation algorithm. An interesting future research direction will be to combine the various solution approaches and tools into hybrid approaches that comprehensively solve and achieve the trade-off for various problem aspects. Most of the existing works are verified by the simulation. Therefore, the performance evaluation should include simulation and testbed environments.

Furthermore, the incentive mechanism of MEC computation offloading is ignored in the existing centralized approaches. In some previous works, the blockchain has been integrated into the MEC system to offer the revenue between video producers and consumers. In future work, trust management should be studied in the blockchain-based MEC networks.

G. Caching-Computing-Networking

For computation-intensive and time-consuming video streaming applications, the integration framework of caching, computing, and networking can provide rich computing and storage resources in the network edge, which can make users access to media through any platform and any device. The related approaches applying the integration framework to video streaming are summarized in Table XI.

Most existing studies focus on computation-constrained MEC, the authors of [118] propose a communications-constrained MEC framework for VR applications. The two computation modes are considered, including MEC computation and local computation. In the system model, the system is time-slotted with the time slot length, and the task request arrival process is modeled as a Bernoulli process. Moreover, a task scheduling algorithm based on the Lyapunov theory is proposed to minimize the average transmission data per task. Finally, the trade-offs between communications, computing, and caching are studied, and the closed-form expression about the average transmission data per task, the CPU frequency, and the caching size is derived. The results demonstrate that the proposed approach can achieve a significant reduction in the average transmission data consumption. However, in future work, the cooperative video caching and computing approach should be studied.

Software-defined mobile network (SDMN), in-network cache and MEC are the important next-generation technologies of mobile networks, which are important for enhancing the quality of video services. The authors of [68] study the video rate adaptation problem in an MEC-SDMN, where the Unified Video Mean Opinion Score (U-vMOS) is used to evaluate video performance [181]. U-vMOS considers three elements: video quality, viewing experience, and interactive experience. Rated on a 1 to 5 scale, sub-categories include video definition, the quantity of video sources, screen size, usage experience, and streaming smoothness [182]. In the system model, the authors assume that each video streaming is served by one data flow, and the spectrum efficiency of wireless link is approximated according to Shannon's bound. Then, a joint optimization problem is formulated to maximize the mean U-vMOS of a HetNet. Then, the dual-decomposition algorithm is used to decouple variables, including video data rate, computing resources, and traffic engineering. Finally, the ADMM-based algorithm is adopted for solving this problem. However, a communications-constrained MEC approach should be studied in the future.

The RA problem of massive video content delivery is ignored in previous works. The authors of [69] propose a virtualized heterogeneous network framework to support video transcoding, caching, and multicast, where information-centric networking (ICN) is integrated into the MEC system. In the system model, each content can be transcoded into versions with different bitrates, and the minimum achievable spectrum efficiency is approximated according to Shannon's theory.

TABLE XI: The summary of MEC caching and computation offloading for video streaming.

Ref.	Networking	Caching Strategy	Offloading Type	Objective	Control	Optimization Tool	Streaming	Evaluation	Main Findings
[118]	Macro-cellular networks	Delivery	Partial offloading	Minimizing the communication-resource consumption	Centralized	Lyapunov theory	VR	Simulation	The proposed approach achieves a significant reduction in the average transmission data consumption.
[68]	HetNet	Placement and delivery	Full offloading	Maximizing the mean U-vMOS	Distributed	ADMM	VBR, HTTP-based	Simulation	Increasing the capability of MEC servers and the capacity of caches can further improve the performance.
[69]		Delivery	Full offloading	Maximizing the aggregate revenue	Centralized	Convex optimization	VBR	Simulation	The proposed approach promotes the alleviation of backhaul and bring a higher caching revenue.
[70]	RAN	Placement	Partial offloading	Minimizing the expected delay cost of video retrieval	Centralized	Greedy algorithm	VBR, RTSP-based	Simulation	The proposed approach provides significant performance improvement in terms of cache hit ratio, backhaul traffic load, and average access delay.
[96]		Placement and delivery	Full offloading	Minimizing the backhaul network cost	Centralized	Alternative optimization algorithm	VBR	Simulation	Improving the users access data rate leads to increasing the provisioning cost of slices.
[77]		Placement and delivery	Full offloading	Minimizing the system energy consumption	Centralized	Heuristic algorithm	VBR, HTTP-based	Simulation	The proposed approach has obvious performance improvement when the bitrate version interval grows.
[79]		Delivery	Partial offloading	Minimizing the average required transmission rate under a given latency constraint	Centralized	Approximation algorithm	VR	Simulation	The approach saves communication bandwidth while meeting low latency requirement.
[103]		Delivery	Partial offloading	Maximizing the aggregate reward	Centralized	Approximation algorithm	VR	Simulation	The proposed framework takes advantage of the higher number of base stations available to pool and utilize their resources.
[71]		Placement	Partial offloading	Maximizing the edge caching	Centralized	Greedy algorithm	VBR	Simulation	The proposed approach improves the performance in terms of cache hit ratio, access delay and cost.
[72]		Placement and delivery	Partial offloading	Maximizing the average tolerant delay	Centralized	Convex optimization	VR	Simulation	The proposed scheme improves spectral efficiency while meeting high transmission rate requirements.

Moreover, a joint optimization problem is formulated to maximize the aggregate revenue, including communication revenue, caching revenue, and computing revenue. Finally, the primal-dual interior point method is used to solve this problem. Simulations results show the effectiveness of the proposed approach.

The existing joint optimization scheme is mainly based on the scenario of single server independent caching and transcoding task decision, without considering the cooperation among servers. The authors of [70] propose a joint collaborative caching and processing framework for adaptive video streaming over Real-time Transport Protocol (RTSP). In this context, there are three ways to get the requested video contents: (1) obtaining the cache from the local server and transcoding, (2) obtaining the cache from the collaboration server, transcoding and returning and (3) obtaining the cache from the collaboration server and transcoding locally after returning from the collaboration server. In the system model, each server is provisioned with a storage capacity. Moreover, a joint optimization problem is formulated to minimize the expected delay cost of video retrieval, which is NP-complete. Then, the original problem is decomposed into a cache placement problem and a request scheduling problem. Finally, a greedy algorithm is used to solve the proactive cache placement problem, and an online algorithm is adopted to solve video request scheduling problem. The results show that the proposed approach provides significant performance improvement in terms of cache hit ratio, backhaul traffic load, and average access delay. However, the incentive mechanism should be investigated in future work.

Furthermore, the authors of [96] study MEC-enabled cooperative video transcoding and caching approach in heterogeneous virtualized MEC networks. In the system model, each user is subscribed to only one slice. Then, a joint optimization problem is formulated to maximize slice revenues. Moreover, the problem is decomposed into a cache placement problem and a delivery problem, which is an NP-hard problem. Finally, the alternate optimization algorithm is used to solve this problem with low complexity. Simulation results show the proposed scheme has a better convergence performance. However, the trust management of mobile nodes should be considered in future work.

In previous works, the problem of energy efficiency is largely ignored. The authors of [77] propose a joint caching and transcoding schedule strategy for HTTP adaptive video streaming delivery in RAN. In the system model, each segment can be encoded to different bitrates. Moreover, a joint optimization problem is modeled to minimize the system energy consumption, which takes the caching and transcoding into consideration. In this context, a heuristic algorithm is adopted to solve this problem. Simulation results show that the proposed strategy presents an excellent performance. However, the payment mechanism should be studied in the future.

This challenge of VR over wireless networks has not been well addressed in previous works, including ultra-low latency and ultra-high transmission rates. The authors of [79] propose an MEC-based mobile VR delivery framework to enable VR caching and post-processing at the mobile device. In the system model, the request stream at the mobile VR device conforms to the independent reference model, and both 2D and 3D FOVs of all the viewpoints are cached at the MEC server. Then, a joint caching and computing decision problem is formulated to minimize the average required transmission rate under a given latency constraint. When FoVs of VR streaming is homogeneous, a closed-form expression is obtained for the optimization problem, which reveals interesting communications-caching-computing trade-offs. When FOVs are heterogeneous, the concave convex procedure (CCCP) based algorithm is used to solve this problem. However, the incentive mechanism between VR produces and consumers should be studied in future research.

Furthermore, in order to achieve the trade-offs among caching, computing, and communication for VR/AR applications, the authors of [103] propose the 360° video delivery framework, where only the scenic viewpoint of present interest needs to be transmitted to the user. In the system model, the video access rates at each base station are modeled as independent Zipf distributions. Moreover, an optimization problem is formulated to maximize the aggregate reward, taking account of caching, computing, and streaming. In this context, an efficient fully-polynomial time approximation algorithm is used to solve this problem. Simulation results demonstrate that the proposed scheme achieves better efficiency gains compared with the state of the art. However, collaborative caching and computing should be investigated in future work.

Further improvement of [103], the authors of in [71] proposes a collaborative caching and transcoding model for VoD in MEC networks, where neighboring MEC servers can collaborate to store and transcode different video chunks. In the system model, the authors assume that these servers can share resources, and all requested videos have different bitrate versions. Firstly, a probabilistic model is adopted to study the popularity of different chunks of videos. Then, two popularity-aware caching policies are proposed to cache only the highest probably requested chunks, including proactive caching policy (PcP) and cache replacement Policy (CrP). Finally, a greedy algorithm is designed to schedule collaborative caching and transcoding. In the future, the payment mechanism should be studied.

In order to balance latency, throughput, and complexity in an economic way, the authors of [72] propose a fog radio access networks (F-RANs) enabled mobile VR delivery framework, where the projection component from the two-dimensional monocular video (MV) to three-dimensional stereoscopic video (SV) is allowed to be computed at fog access points (F-APs) and mobile VR devices. In the system model, the VR delivery system consists of one mobile VR device, one F-AP and several remote radio heads. The service requests of the mobile VR device conform to the independent reference model. Moreover, a joint radio communication, caching, and computing decision problem is modeled to maximize the average tolerant delay under a given transmission rate constraint. The optimization problem can be transformed into a multiple knapsack problem, where the Lagrangian dual decomposition approach is used to solve this problem. Simulation results show the proposed scheme achieves an intuitive communication-caching-computing trade-off.

1) Lessons Learned:

Caching and computing at the edge is a powerful tool for multimedia streaming applications. Implementation framework by utilizing edge caching and computing capabilities of end devices can alleviate the traffic burden over wireless networks. For AR/VR application, the energy-efficient video delivery can be achieved in the combination MEC system.

This subsection jointly considers video offloading and content caching to improve the performance of cellular networks with MEC. The caching strategy can be divided into placement and delivery, and the offloading type includes full and partial offloading. Some collaborative caching and transcoding approaches are proposed to improve the system performance. However, blockchain-based joint optimization approach is rarely studied in previous works.

In order to optimize spectrum, computation, and storage resources, the purpose of the joint approaches is to maximize the QoE, the aggregate revenue and the average tolerant delay cost, or minimize the energy consumption, the expected delay, and the average required transmission rate.

For joint video caching and computation offloading approaches, the optimization problems can be solved with some optimization tools, including convex optimization, the Lyapunov theory, and approximation algorithm. Moreover, the Lyapunov algorithm can solve the offloading optimization problems with low-complexity online computations.

In addition, some payment mechanisms are proposed to encourage caching and offloading services with revenue/reward. However, the fairness of the payment mechanism needs to be studied in future work.

VI. Applications and Use Cases

In this section, we focus on how video streaming can leverage MEC in different application scenarios. The application-oriented video streaming classification is provided in this work, which includes intelligent video acceleration, video streaming analysis, augmented reality service, and connected vehicles. Therefore, MEC based video streaming applications are summarized as follows.

A. Intelligent Video Acceleration

Due to rapidly varying bandwidth for end users, the transmission control protocol (TCP)-based video delivery approach can lead to the inefficient use of the network resources and degrade application performance and user's QoE. MEC can enhance the QoE by reducing the content's time-to-start as well as video-stall occurrences, and guarantee maximum utilization of the radio resources [183]. Therefore, the radio analytics application located at the RAN can estimate the throughput by the indication for video server, and the information can be used for TCP congestion control to match the estimated throughput.

The MEC-based intelligent video acceleration service scenario is shown in Fig. 5, where a radio analytics application can provide the video server with an indication of the throughput estimated. Based on the indication, the video server can select the initial window size, set the value of the congestion window, and adjust the size of the congestion window. Moreover, the indication can be adopted to ensure that the application level coding matches the estimated throughput. In this context, the time-to-start and video-stall occurrences can be reduced, and the maximum utilization of network resources can be achieved. In other words, MEC can ensure efficient utilization of the network resources and improve user's QoE for video applications [184].

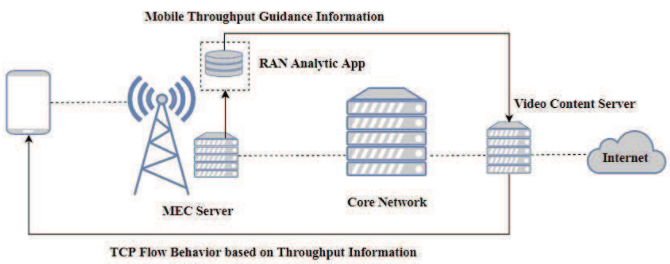

Fig. 5: MEC for video acceleration [185].

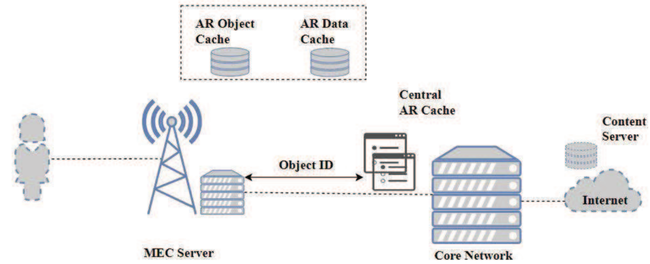

Fig. 7: MEC for AR service [190].

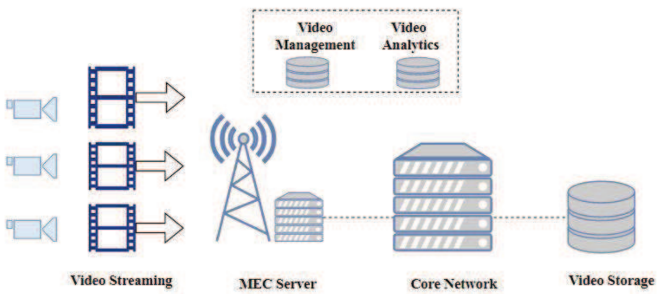

Fig. 6: MEC for video streaming analysis [185].

B. Video Streaming Analysis

Video streaming analysis has a range of applications such as license plate recognition, face recognition, and urban surveillance. The general tasks of video analysis include object detection, object tracking, and semantic segmentation. However, the computation complexity of video analysis algorithms is high. Therefore, moving the video analysis away from the video camera reduces the cost of cameras [186]. Compared to processing video analysis tasks in the central cloud, performing video analysis locally, i.e. close to edge nodes, can enjoy the benefits of low latency and avoid the network congestion [187]. The valuable data can be transmitted to the application server with the low data rate.

Fig. 6 shows the MEC-enabled video analysis architecture, where MEC server can execute the video analysis task with low complexity, and the extracted information is forwarded into the cloud server. Compared with a cloud-based video-analytics system, it saves from transporting high data video streaming to central cloud.

C. Augmented Reality Service

AR is a live direct or indirect view of physical, real-world environment whose elements are augmented by computer-generaged sensory inputs, including video, audio, and graphics. AR service requires an application to analyze the output from a device's camera. After analysing such information, the AR application can provide additional information in real-time [188]. For example, a smart museum AR application may provide additional information when a smart phone or a viewing device is pointed at a museum object, or a smart hard hat may project key information such as wiring diagrams onto components for field technicians. Augmented information pertaining to a point of interest is highly localized and thus hosting the information locally is advantageous compared with hosting in the cloud.

The MEC-enabled AR application is shown in Fig. 7, where the MEC server can analyze the output from the camera and transmit the AR data to the user. AR data requires low latency and a high rate of data processing. Therefore, The use of a MEC server is highly advantageous since augmented reality information is highly localized [189].

D. Connected Vehicles

The connected vehicle technology has the potential to enable diverse applications associated with traffic safety, traffic efficiency, and infotainment. The IEEE 802.11bd for the DSRC and NR V2X for C-V2X can supplement today's vehicular sensors in enabling autonomous driving [191]. Based on vehicular ad-hoc network (VANET), video streaming applications in connected vehicles can be divided into safety applications and non-safety applications. MEC is an enabling technology to reduce network delay and energy consumption in VANETs environment. Moreover, MEC-based video streaming can enhance safety for the driver's decision of overtaking or stopping. In intelligent transportation systems (ITS), different types of cameras can be used to detect the moving objects, where object detection tasks can be executed at the network edge [192]. When an emergency occurs, video broadcast messages can be transmitted to vehicles to warn drivers. MEC provides the hosting services for the application, enabling lower latency for the applications and a layer of abstraction from the host services running across the internet.

The MEC-based connected vehicle architecture is shown in Fig. 8, where MEC are deployed at the wireless base station site to provide the roadside functionality. The MEC applications can receive local messages directly from the applications in the vehicles and the roadside visual sensors, analyze them and then propagate hazard warnings and other latency sensitive messages to other vehicles [193].

VII. Challenges and Future Research Directions

In this section, some challenges are discussed, and potential research directions are proposed for future video streaming applications.

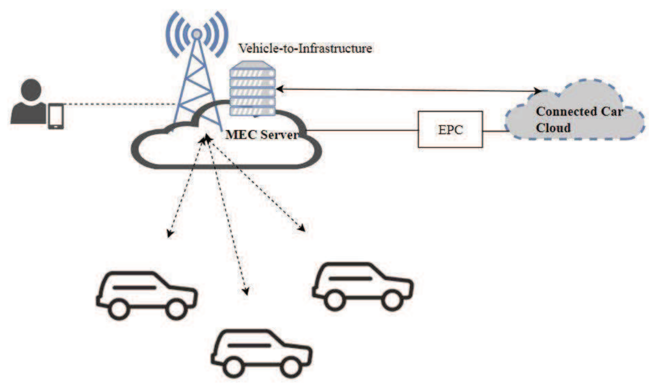

Fig. 8: MEC-based connected vehicle system [194].

A. Energy Consumption

The MEC deployment sinks storage and computing resources that were originally in the cloud to the edge of the network. On one hand, the network edge can respond to user requests, on the other hand, it can reduce the waste of returning resources. In terms of network optimization of MEC, energy efficiency is one of the key areas of concern. In the MEC deployment scenario, video caching, computing, and communication between MECs and users can result in significant energy consumption. Therefore, in order to schedule caching, computing, and communication resources reasonably, it is very vital to establish an efficient resource optimization mechanism to reduce the system energy consumption.

In addition, for the joint optimization of MEC-based video caching and transcoding, there are three kinds of energy consumption, including caching, transcoding, and transmission energy consumption. Therefore, how to jointly consider the caching, transcoding, and transmission, and optimize the total energy efficiency is a great challenge for future work.

B. Bandwidth and QoE Tradeoffs

QoE is a measure of the delight or annoyance of a customer's experiences with a service, which focuses on the entire service experience. The development of adaptive bitrate is a key driver for exploring an effective approach to improve QoE, thereby providing a differentiated service to users to enhance the user experience [85]. Moreover, the research on the tradeoff between bandwidth optimization and QoE optimization in MEC-based caching and transcoding approaches is a very meaningful research direction. From the perspective of video content providers, two important factors need to be considered for system optimization: (1) it is necessary to reduce the costs of caching and computing, (2) it is to ensure the users' QoE. Therefore, how to balance the cost of bandwidth resources of MEC-based caching and transcoding and the users' QoE is an important direction for future research.

C. DRL-based Strategy

Based on the deep reinforcement learning method, the ABR-based caching approach is an important research direction for video streaming applications [195]. In the adaptive bitrate streaming system, each video block has multiple bitrate versions. Owing to the capacity limitation of the MEC-based cache system, cache all bitrates can lead to a decrease in caching resource utilization and an increase in network costs. Therefore, by using MEC technology, network information can be perceived in real-time, including network link status and user behavior. Moreover, these information can be analyzed and processed by using the DRL-based method, and it is to predict the popularity of video content and bitrate version for user responses. Therefore, for video caching with the corresponding bitrate version, the resource allocation strategy can be decided in advance, which can improve the cache hit ratio and the utilization of caching resources.

D. Multi-MEC Collaboration

The deployment of MECs in edge networks is usually distributed, so that the MEC-based caching and computing resources are distributed in different locations on the network. The storage space and computing power of a single MEC are limited. Excessive caching and computing tasks will overload the MEC server. When these tasks are returned to the cloud data center, it will result in a higher cost. Therefore, based on multi-MEC with the distributed mode, adjacent MEC servers can collaborate to perform caching and computing tasks. When the current MEC server lacks of cached or computing resources, other idle MEC servers can be used to reduce network costs and improve network performance [196]. Moreover, sharing resources between different MEC nodes becomes an important research issue. For instance, when the target video content requested by the user is not cached by the local MEC server, how to select an optimal node among other MEC nodes that cache corresponding content. When the computing load of the local MEC server is overloaded, how to offload local computing tasks to other MEC nodes. The resource sharing mechanism based on distributed multi-MEC collaboration needs to be further studied to improve resource utilization and QoE in the future.

Furthermore, MEC node collaboration to fully support users' mobility should be investigated in future work. The Follow-Me Edge (FME) model is adopted to reduce core network traffic and ensure ultra-short latency [197]. In this context, the FME architecture includes two-tier, where cloud service providers use the cloud resource to deploy its application and mobile network operators manage its own set of edge server clusters. The FME ensures that the service constantly follows the user and that the user is always serviced from the closest edge.

E. Security Enhancements

Compared to traditional cloud computing, the benefits of MEC stem from the potential for location-awareness, bandwidth availability, real-time, privacy protection, and low-cost services to support emerging video streaming applications. However, the disadvantages of MEC are as follows: weak computational ability, attack unawareness, and communication protocol heterogeneities. Therefore, its emergence introduces more security issues, and the potential attack models include DDoS attacks, side-channel attacks, malware injection attacks, authentication, and authorization attacks [198] [199]. In order to defense different attacks, various types of security mechanisms of MEC need to design for identity and authentication, access control systems, trust management, and intrusion detection systems [200].

Moreover, the authors of [201] focus on the context-aware vehicular applications, and a blockchain-inspired approach is proposed to enhance the security of vehicular edge computing and networking. In this context, information and energy interactions are based on Proof of Work consensus in edge computing. The work in [202] presents a physical layer authentication enabled security mechanisms in MEC system, and the deep learning (DL)-based approach is designed to train the multi-user authentication model. In [203], a secure resource allocation approach is proposed to maximize the capacity of computation resources in the MEC system, where the the Lyapuonv theory is used to solve this problem.

VIII. Conclusion

In this work, an exhaustive survey and research outlook of applying MEC to video streaming was presented. To this end, at first, the details about overview and background were introduced. At first, the categories of resource allocation schemes were presented, which included manageable resources, optimization criteria, and optimization approaches. Next, enabling technologies for video steaming applications ware summarized. Then, the taxonomy for MEC enabled video streaming applications was provided, from the perspectives of intelligent video acceleration, video streaming analysis, augmented reality service, and connected vehicles. Finally, some challenges were discussed.

In summary, research on applying MEC technology to video streaming is quite broad, and some challenges lay ahead. Nevertheless, it is favorable for the network community to address the challenges and go forward. This paper is trying to briefly summarize video streaming applications by using MEC, which may open a new avenue for the development of video streaming applications.

References

[1] C. M. VNI, "Cisco visual networking index: Global mobile data traffic forecast update, 2017 - 2022," Feb. 2019.

[2] J. Du, F. R. Yu, G. Lu, J. Wang, J. Jiang, and X. Chu, "Mec-assisted immersive VR video streaming over Terahertz wireless networks: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9517–9529, Oct. 2020.

[3] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile internet," *IEEE MultiMedia*, vol. 26, no. 1, pp. 21–30, Jan. 2019.

[4] D. Loghin, S. Cai, G. Chen, T. T. A. Dinh, F. Fan, Q. Lin, J. Ng, B. C. Ooi, X. Sun, Q.-T. Ta *et al.*, "The disruptions of 5g on data-driven technologies and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1179–1198, June 2020.

[5] M. Wien, J. M. Boyce, T. Stockhammer, and W. Peng, "Standardization status of immersive video coding," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 5–17, Mar. 2019.

[6] J. Ohm and G. Sullivan, "Versatile video coding-towards the next generation of video compression," in *Proc. Picture Coding Symposium*, San Francisco, USA, June 2018. [Online]. Available: http://mile-high.video/files/mhv2018/pdf/day1/1_05_Sullivan.pdf

[7] R. Viola, A. Martin, M. Zorrilla, and J. Montalbán, "MEC proxy for efficient cache and reliable multi-CDN video distribution," in *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Valencia, Spain, June 2018, pp. 1–7.

[8] "Multi-access edge computing (MEC); framework and reference architecture," *ETSI*, Jan. 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/02.01.01_60/gs_MEC003v020101p.pdf

[9] 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 Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[10] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proc. the first edition of the MCC workshop on Mobile cloud computing*, Helsinki, Finland, Aug. 2012, pp. 13–16.

[11] M. Yannuzzi, R. Irons-Mclean, F. van Lingen, S. Raghav, A. Somaraju, C. Byers, T. Zhang, A. Jain, J. Curado, D. Carrera *et al.*, "Toward a converged openfog and etsi mano architecture," in *Proc. IEEE Fog World Congress*, Oct. 2017, pp. 1–6.

[12] G. Ma, Z. Wang, M. Zhang, J. Ye, M. Chen, and W. Zhu, "Understanding performance of edge content caching for mobile video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1076–1089, May 2017.

[13] Y. Li, P. A. Frangoudis, Y. Hadjadj-Aoul, and P. Bertin, "A mobile edge computing-based architecture for improved adaptive HTTP video delivery," in *Proc. IEEE Conference on Standards for Communications and Networking*, Berlin, Germany, Oct. 2016, pp. 1–6.

[14] C. Ge, N. Wang, G. Foster, and M. Wilson, "Toward QoE-assured 4k video-on-demand delivery through mobile edge virtualization with adaptive prefetching," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2222–2237, Oct. 2017.

[15] C. Quadri, S. Gaito, R. Bruschi, F. Davoli, and G. P. Rossi, "A MEC approach to improve QoE of video delivery service in urban spaces," in *Proc. IEEE International Conference on Smart Computing*, Taormina, Italy, June 2018, pp. 25–32.

[16] X. Hou, Y. Lu, and S. Dey, "Wireless VR/AR with edge/cloud computing," in *Proc. 26th International Conference on Computer Communication and Networks*, Vancouver, Canada, July 2017, pp. 1–8.

[17] R. Schmoll, S. Pandi, P. J. Braun, and F. H. P. Fitzek, "Demonstration of VR/AR offloading to mobile edge cloud for low latency 5G gaming application," in *Proc. 15th IEEE Annual Consumer Communications Networking Conference*, Las Vegas, USA, Jan. 2018, pp. 1–3.

[18] S. Kumar, D. S. Vineeth, and A. F. A, "Edge assisted DASH video caching mechanism for multi-access edge computing," in *Proc. IEEE International Conference on Advanced Networks and Telecommunications Systems*, Indore, India, Dec. 2018, pp. 1–6.

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

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

[21] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1628–1656, thirdquarter 2017.

[22] 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, Mar. 2017.

[23] G. Premsankar, M. Di Francesco, and T. Taleb, "Edge computing for the internet of things: A case study," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1275–1284, Apr. 2018.

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

[25] Q. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W. Hwang, and Z. Ding, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116 974–117 017, June 2020.

[26] D. Antonogiorgakis, A. Britzolakis, P. Chatziadam, A. Dimitriadis, S. Gikas, E. Michalodimitrakis, M. Oikonomakis, N. Siganos, E. Tzagkarakis, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A view on edge caching applications," *arXiv e-prints*, p. arXiv:1907.12359, July 2019.

[27] Q. Zhang, H. Sun, X. Wu, and H. Zhong, "Edge video analytics for public safety: A review," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1675–1696, Aug. 2019.

[28] K. Kanai, K. Imagane, and J. Katto, "Overview of multimedia mobile edge computing," *ITE Transactions on Media Technology and Applications*, vol. 6, no. 1, pp. 46–52, Jan. 2018.

[29] A. J. Ferrer, J. M. Marques, and J. Jorba, "Towards the decentralised cloud: Survey on approaches and challenges for mobile, ad hoc, and edge computing," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–36, Jan. 2019.

[30] P. Bellavista, L. Foschini, and D. Scotece, "Converging mobile edge computing, fog computing, and IoT quality requirements," in *Proc. IEEE 5th International Conference on Future Internet of Things and Cloud*, Prague, Czech Republic, Aug. 2017, pp. 313–320.

[31] "Mobile edge computing (MEC); deployment of mobile edge computing in an NFV environment," *ETSI*, Feb. 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/MEC/001_099/017/01.01.01_60/gr_MEC017v010101p.pdf

[32] "Multi-access edge computing (MEC); phase 2: Use cases and requirements," *ETSI*, Otc. 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/02.01.01_60/gs_MEC002v020101p.pdf

[33] "Multi-access edge computing MEC); proof of concept framework," *ETSI*, July 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/005/02.01.01_60/gs_MEC005v020101p.pdf

[34] "Multi-access edge computing (MEC) MEC 5G integration," *ETSI*, Otc. 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/MEC/001_099/031/02.01.01_60/gr_MEC031v020101p.pdf

[35] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, Apr. 2014.

[36] M. Mehrabi, D. You, V. Latzko, H. Salah, M. Reisslein, and F. H. P. Fitzek, "Device-enhanced MEC: Multi-access edge computing (MEC) aided by end device computation and caching: A survey," *IEEE Access*, vol. 7, pp. 166 079–166 108, Nov. 2019.

[37] X. Jiang, T. Song, W. Shi, T. Katayama, T. Shimamoto, and L. Wang, "Fast coding unit size decision based on probabilistic graphical model in high efficiency video coding inter prediction," *IEICE Transactions on Information and Systems*, vol. 99, no. 11, pp. 2836–2839, Nov. 2016.

[38] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[39] X. Jiang, J. Feng, T. Song, and T. Katayama, "Low-complexity and hardware-friendly H. 265/HEVC encoder for vehicular ad-hoc networks," *Sensors*, vol. 19, no. 8, p. 1927, Apr. 2019.

[40] K. R. Rao, D. N. Kim, and J. J. Hwang, *VP6 Video Coding Standard*. Dordrecht: Springer Netherlands, Oct. 2014, pp. 159–197.

[41] D. Mukherjee, J. Han, J. Bankoski, R. Bultje, A. Grange, J. Koleszar, P. Wilkins, and Y. Xu, "A technical overview of VP9 -the latest open-source video codec," *SMPTE Motion Imaging Journal*, vol. 124, no. 1, pp. 44–54, Jan. 2015.

[42] H. Kalva and J. Lee, "The VC-1 video coding standard," *IEEE MultiMedia*, vol. 14, no. 4, pp. 88–91, Oct. 2007.

[43] X. Jiang, T. Song, D. Zhu, T. Katayama, and L. Wang, "Quality-oriented perceptual HEVC based on the spatiotemporal saliency detection model," *Entropy*, vol. 21, no. 2, p. 165, Feb. 2019.

[44] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[45] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, Jan. 2016.

[46] Y. Sani, A. Mauthe, and C. Edwards, "Adaptive bitrate selection: A survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2985–3014, Fourthquarter 2017.

[47] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann, "A survey on bitrate adaptation schemes for streaming media over HTTP," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 562–585, Firstquarter 2019.

[48] J. Yoon and S. Banerjee, "Hardware-assisted, low-cost video transcoding solution in wireless networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[49] H. Ahlehagh and S. Dey, "Adaptive bit rate capable video caching and scheduling," in *Proc. IEEE Wireless Communications and Networking Conference*, Shanghai, China, Apr. 2013, pp. 1357–1362.

[50] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[51] F.-L. Andrew, "A review of HTTP live streaming," Jan. 2010. [Online]. Available: http://files.andrewsblog.org/http_live_streaming.pdf

[52] X. Yin, V. Sekar, and B. Sinopoli, "Toward a principled framework to design dynamic adaptive streaming algorithms over HTTP," in *Proc. the 13th ACM Workshop on Hot Topics in Networks*, Los Angeles, USA, Oct. 2014, pp. 1–7.

[53] Z. Alex, "IIS smooth streaming technical overview," *Microsoft Corporation*, Mar. 2009. [Online]. Available: http://img.prodek.lt/documents/IIS_Smooth_Streaming_Technical_Overview.pdf

[54] M. A. Muhanna, "Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions," *Journal of King Saud University - Computer and Information Sciences*, vol. 27, no. 3, pp. 344–361, June 2015.

[55] L. P. Berg and J. M. Vance, "Industry use of virtual reality in product design and manufacturing: a survey," *Virtual Reality*, vol. 21, no. 1, pp. 1–17, Mar. 2017.

[56] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: State of the art and perspectives." in *Proc. The International Scientific Conference eLearning and Software for Education*, Bucharest, Rumania, Apr. 2015, p. 133.

[57] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," *Foundations and Trends in Human -Computer Interaction*, vol. 8, no. 2-3, pp. 73–272, Mar. 2015.

[58] F. Manuri and A. Sanna, "A survey on applications of augmented reality," *Advances in Computer Science: an International Journal*, vol. 5, no. 1, pp. 18–27, Jan. 2016.

[59] J. Feng, Q. Pei, F. R. Yu, X. Chu, and B. Shang, "Computation offloading and resource allocation for wireless powered mobile edge computing with latency constraint," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1320–1323, Oct. 2019.

[60] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *Proc. IEEE 19th International Workshop on Multimedia Signal Processing*, Luton, UK, Oct. 2017, pp. 1–6.

[61] D. He, C. Westphal, and J. J. Garcia-Luna-Aceves, "Joint rate and FoV adaptation in immersive video streaming," in *Proc. 2018 Morning Workshop on Virtual Reality and Augmented Reality Network*, Budapest, Hungary, Aug. 2018, pp. 27–32.

[62] A. Gulhane, A. Vyas, R. Mitra, R. Oruche, G. Hoefer, S. Valluripally, P. Calyam, and K. A. Hoque, "Security, privacy and safety risk assessment for virtual reality learning environment applications," in *Proc. 16th IEEE Annual Consumer Communications Networking Conference*, Las Vegas, USA, Jan. 2019, pp. 1–9.

[63] Yu-Ting Yu, F. Bronzino, R. Fan, C. Westphal, and M. Gerla, "Congestion-aware edge caching for adaptive video streaming in information-centric networks," in *Proc. 12th Annual IEEE Consumer Communications and Networking Conference*, Las Vegas, USA, Jan. 2015, pp. 588–596.

[64] C. Ge, N. Wang, S. Skillman, G. Foster, and Y. Cao, "QoE-driven DASH video caching and adaptation at 5G mobile edge," in *Proc. the 3rd ACM Conference on Information-Centric Networking*, Kyoto, Japan, Sep. 2016, pp. 237–242.

[65] Y. Tan, C. Han, M. Luo, X. Zhou, and X. Zhang, "Radio network-aware edge caching for video delivery in MEC-enabled cellular networks," in *Proc. IEEE Wireless Communications and Networking Conference Workshops*, Apr. 2018, pp. 179–184.

[66] A. Mehrabi, M. Siekkinen, G. Illahi, and A. Ylä-Jääski, "D2D-enabled collaborative edge caching and processing with adaptive mobile video streaming," in *Proc. IEEE 20th International Symposium on World of Wireless, Mobile and Multimedia Networks*, Washington, USA, Aug. 2019, pp. 1–10.

[67] H. Cui, Z. Chen, N. Liu, and B. Xia, "Blockchain-driven contents sharing strategy for wireless cache-enabled D2D networks," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, Shanghai, China, May 2019, pp. 1–5.

[68] C. Liang, Y. He, F. R. Yu, and N. Zhao, "Enhancing video rate adaptation with mobile edge computing and caching in software-defined mobile networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 7013–7026, Oct. 2018.

[69] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, "Video transcoding, caching, and multicast for heterogeneous networks over wireless network virtu-

alization," *IEEE Communications Letters*, vol. 22, no. 1, pp. 141–144, Jan. 2018.

[70] T. Tran and D. Pompili, "Adaptive bitrate video caching and processing in mobile-edge computing networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[71] E. Baccour, A. Erbad, A. Mohamed, K. Bilal, and M. Guizani, "Proactive video chunks caching and processing for latency and cost minimization in edge networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–7.

[72] T. Dang and M. Peng, "Joint radio communication, caching, and computing design for mobile virtual reality delivery in fog radio access networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 7, pp. 1594–1607, July 2019.

[73] J. Park, P. Popovski, and O. Simeone, "Minimizing latency to support VR social interactions over wireless cellular systems via bandwidth allocation," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 776–779, Oct. 2018.

[74] W. Wang, Q. Wang, and K. Sohraby, "Multimedia sensing as a service MSaaS: Exploring resource saving potentials of at cloud-edge IoT and fogs," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 487–495, Apr. 2017.

[75] A. Al-Shuwaili and O. Simeone, "Energy-efficient resource allocation for mobile edge computing-based augmented reality applications," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 398–401, June 2017.

[76] Y. Liu, J. Liu, A. Argyriou, and S. Ci, "MEC-assisted panoramic VR video streaming over millimeter wave mobile networks," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1302–1316, May 2019.

[77] Z. Li, R. Xie, Q. Jia, and T. Huang, "Energy-efficient joint caching and transcoding for HTTP adaptive streaming in 5G networks with mobile edge computing," in *Proc. IEEE International Conference on Communications Workshops*, Kansas City, USA, May 2018, pp. 1–6.

[78] Z. Wang, L. Sun, C. Wu, W. Zhu, Q. Zhuang, and S. Yang, "A joint online transcoding and delivery approach for dynamic adaptive streaming," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 867–879, June 2015.

[79] Y. Sun, Z. Chen, M. Tao, and H. Liu, "Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7573–7586, Nov. 2019.

[80] K. R. Smith, L. Hang, H. Li-Tse, X. de Foy, and R. Gazda, "Wireless adaptive video streaming with edge cloud," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–13, Dec. 2018.

[81] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "VR is on the edge: How to deliver 360° videos in mobile networks," in *Proc. the Workshop on Virtual Reality and Augmented Reality Network*, Los Angeles, USA, Aug. 2017, pp. 30–35.

[82] X. Xu, J. Liu, and X. Tao, "Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation," *IEEE Access*, vol. 5, pp. 16 406–16 415, 2017.

[83] Y. Qi, L. Tian, Y. Zhou, and J. Yuan, "Mobile edge computing-assisted admission control in vehicular networks: The convergence of communication and computation," *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 37–44, Mar. 2019.

[84] H. A. Pedersen and S. Dey, "Enhancing mobile video capacity and quality using rate adaptation, RAN caching and processing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 996–1010, Apr. 2016.

[85] C. Li, L. Toni, J. Zou, H. Xiong, and P. Frossard, "QoE-driven mobile edge caching placement for adaptive video streaming," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 965–984, Apr. 2018.

[86] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "QoE-traffic optimization through collaborative edge caching in adaptive mobile video streaming," *IEEE Access*, vol. 6, pp. 52 261–52 276, Sep. 2018.

[87] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "Cache-aware QoE-traffic optimization in mobile edge assisted adaptive video streaming," *arXiv e-prints*, p. arXiv:1805.09255, May 2018.

[88] C. Liang, Y. He, F. R. Yu, and N. Zhao, "Video rate adaptation and traffic engineering in mobile edge computing and caching-enabled wireless networks," in *Proc. IEEE 86th Vehicular Technology Conference*, Toronto, Canada, Sep. 2017, pp. 1–5.

[89] T. Lu, H. Dai, and B. Wang, "QoE-orientated resource allocation for wireless VR over small cell networks," in *Proc. 10th International Conference on Wireless Communications and Signal Processing*, Oct. 2018, pp. 1–6.

[90] Z. Su, Q. Xu, F. Hou, Q. Yang, and Q. Qi, "Edge caching for layered video contents in mobile social networks," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2210–2221, Oct. 2017.

[91] D. T. Hoang, D. Niyato, D. N. Nguyen, E. Dutkiewicz, P. Wang, and Z. Han, "A dynamic edge caching framework for mobile 5G networks," *IEEE Wireless Communications*, vol. 25, no. 5, pp. 95–103, Oct. 2018.

[92] S. Dutta, T. Taleb, P. A. Frangoudis, and A. Ksentini, "On-the-fly QoE-aware transcoding in the mobile edge," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Washington, USA, Dec. 2016, pp. 1–6.

[93] J. Chen, B. Balasubramanian, and Z. Huang, "Liv(e)-ing on the edge: User-uploaded live streams driven by "first-mile" edge decisions," in *Proc. IEEE International Conference on Edge Computing (EDGE)*, Milan, Italy, July 2019, pp. 41–50.

[94] S. Manariyo, A. Khakimov, D. Pyatkina, and A. Muthanna, "Optimization algorithm for IPTV video service delivery over SDN using MEC technology," in *Porc. Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. St. Petersburg, Russia: Springer International Publishing, Sep. 2018, pp. 419–427.

[95] T. Xu, Y. Sun, S. Xia, H. Li, L. Luo, and Z. Chen, "Optimal bandwidth allocation with edge computing for wireless VR delivery," in *Proc. IEEE/CIC International Conference on Communications in China*, Changchun, China, Aug. 2019, pp. 903–907.

[96] S. Rezvani, S. Parsaeefard, N. Mokari, M. R. Javan, and H. Yanikomeroglu, "Cooperative multi-bitrate video caching and transcoding in multicarrier noma-assisted heterogeneous virtualized mec networks," *IEEE Access*, vol. 7, pp. 93 511–93 536, July 2019.

[97] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "Edge computing assisted adaptive mobile video streaming," *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 787–800, Apr. 2019.

[98] Z. Chang, X. Zhou, Z. Wang, H. Li, and X. Zhang, "Edge-assisted adaptive video streaming with deep learning in mobile edge networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, April 2019, pp. 1–6.

[99] T. X. Vu, S. Chatzinotas, and B. Ottersten, "Blockchain-based content delivery networks: Content transparency meets user privacy," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.

[100] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, and M. Song, "Distributed resource allocation in blockchain-based video streaming systems with mobile edge computing," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 695–708, Jan. 2019.

[101] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung, and M. Song, "Deep reinforcement learning (DRL)-based transcoder selection for blockchain-enabled video streaming," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[102] Y. Liu, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Resource allocation for video transcoding and delivery based on mobile edge computing and blockchain," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[103] C. Jacob, "VR/AR immersive communication: Caching, edge computing, and transmission trade-offs," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network*, Los Angeles, USA, Aug. 2017, pp. 36–41.

[104] L. Yang, B. Liu, J. Cao, Y. Sahni, and Z. Wang, "Joint computation partitioning and resource allocation for latency sensitive applications in mobile edge clouds," *IEEE Transactions on Services Computing*, pp. 1–1, Jan. 2019.

[105] W. Wang, D. Niyato, P. Wang, and A. Leshem, "Decentralized caching for content delivery based on blockchain: A game theoretic perspective," in *Proc. IEEE International Conference on Communications*, Kansas City, USA, May 2018, pp. 1–6.

[106] Y. Donoso and R. Fabregat, *Multi-objective optimization in computer networks using metaheuristics*. CRC Press, 2016.

[107] J.-H. Cho, Y. Wang, R. Chen, K. S. Chan, and A. Swami, "A survey on modeling and optimizing multi-objective systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1867–1901, thirdquarter 2017.

[108] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001, vol. 16.

[109] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and multidisciplinary optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[110] D. Wang, Y. Peng, X. Ma, W. Ding, H. Jiang, F. Chen, and J. Liu, "Adaptive wireless video streaming based on edge computing: Oppor-

tunities and approaches," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 685–697, Sep. 2019.

[111] M. T. Emmerich and A. H. Deutz, "A tutorial on multiobjective optimization: fundamentals and evolutionary methods," *Natural computing*, vol. 17, no. 3, pp. 585–609, May 2018.

[112] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 550–586, Firstquarter 2016.

[113] E. Zitzler and L. Thiele, "An evolutionary algorithm for multiobjective optimization: The strength pareto approach," *TIK-report*, vol. 43, 1998.

[114] D. Lin, S. Hu, Y. Gao, and Y. Tang, "Optimizing MEC networks for healthcare applications in 5G communications with the authenticity of users' priorities," *IEEE Access*, vol. 7, pp. 88 592–88 600, June 2019.

[115] S. Han, H. Su, C. Yang, and A. F. Molisch, "Proactive edge caching for video on demand with quality adaptation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 218–234, Jan. 2020.

[116] Y. Liu, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Decentralized resource allocation for video transcoding and delivery in blockchain-based system with mobile edge computing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 169–11 185, Nov. 2019.

[117] L. Vigneri, T. Spyropoulos, and C. Barakat, "Low cost video streaming through mobile edge caching: Modelling and optimization," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1302–1315, June 2019.

[118] X. Yang, Z. Chen, K. Li, Y. Sun, N. Liu, W. Xie, and Y. Zhao, "Communication-constrained mobile edge computing systems for wireless virtual reality: Scheduling and tradeoff," *IEEE Access*, vol. 6, pp. 16 665–16 677, Mar. 2018.

[119] S. Yan, P. Yang, Q. Ye, W. Zhuang, X. Shen, X. Li, and J. Rao, "Transmission protocol customization for network slicing: A case study of video streaming," *IEEE Vehicular Technology Magazine*, vol. 14, no. 4, pp. 20–28, Otc. 2019.

[120] J. Yang, B. Yang, S. Chen, Y. Zhang, Y. Zhang, and L. Hanzo, "Dynamic resource allocation for streaming scalable videos in SDN-aided dense small-cell networks," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2114–2129, 2019.

[121] Y. He, N. Zhao, and H. Yin, "Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 44–55, Jan. 2018.

[122] X. Jiang, F. R. Yu, T. Song, and V. C. M. Leung, "Intelligent resource allocation for video analytics in blockchain-enabled internet of autonomous vehicles with edge computing," *IEEE Internet of Things Journal*, pp. 1–1, 2020, doi:10.1109/JIOT.2020.3026354.

[123] M. Liu, Y. Teng, F. R. Yu, V. C. M. Leung, and M. Song, "A mobile edge computing (MEC)-enabled transcoding framework for blockchain-based video streaming," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 81–87, Apr. 2020.

[124] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. ZHANG, "Mobile-edge computing for vehicular networks: A promising network paradigm with predictive off-loading," *IEEE Vehicular Technology Magazine*, vol. 12, no. 2, pp. 36–44, June 2017.

[125] V. Lakshmikantham, S. Leela, M. Sambandham *et al.*, "Lyapunov theory for fractional differential equations," *Communications in Applied Analysis*, vol. 12, no. 4, p. 365, 2008.

[126] C. C. White III and D. J. White, "Markov decision processes," *European Journal of Operational Research*, vol. 39, no. 1, pp. 1–16, Mar. 1989.

[127] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Proc. Advances in neural information processing systems*, 2000, pp. 1008–1014.

[128] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. International conference on machine learning*, 2016, pp. 1928–1937.

[129] F. R. Yu, J. Liu, Y. He, P. Si, and Y. Zhang, "Virtualization for distributed ledger technology (vDLT)," *IEEE Access*, vol. 6, pp. 25 019–25 028, Apr. 2018.

[130] G. T. Nguyen and K. Kim, "A survey about consensus algorithms used in blockchain," *Journal of Information Processing Systems*, vol. 14, no. 1, pp. 101–128, Feb. 2018.

[131] X. Yang, N. Zhang, W. Lou, and Y. T. Hou, "A survey of distributed consensus protocols for blockchain networks," *arXiv e-prints*, p. arXiv:1904.04098, Apr. 2019.

[132] Z. Xiong, Y. Zhang, D. Niyato, P. Wang, and Z. Han, "When mobile blockchain meets edge computing," *IEEE Communications Magazine*, vol. 56, pp. 33–39, Aug. 2018.

[133] 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 Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1508–1532, Secondquarter 2019.

[134] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: A state of the art survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 858–880, Firstquarter 2019.

[135] A. V. Rivera, A. Refaey, and E. Hossain, "A blockchain framework for secure task sharing in multi-access edge computing," *IEEE Network*, pp. 1–8, 2020, doi:10.1109/MNET.011.2000497.

[136] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *Proc. IEEE Security and Privacy Workshops*, San Jose, CA, USA, May 2015, pp. 180–184.

[137] J. Feng, F. R. Yu, Q. Pei, J. Du, and L. Zhu, "Joint optimization of radio and computational resources allocation in blockchain-enabled mobile edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4321–4334, June 2020.

[138] T. Huang, F. R. Yu, C. Zhang, J. Liu, J. Zhang, and Y. Liu, "A survey on large-scale software defined networking (SDN) testbeds: Approaches and challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 891–917, Secondquarter 2017.

[139] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, Mar. 2008.

[140] A. Wang, Z. Zha, Y. Guo, and S. Chen, "Software-defined networking enhanced edge computing: A network-centric survey," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1500–1519, Aug. 2019.

[141] A. C. Baktir, A. Ozgovde, and C. Ersoy, "How can edge computing benefit from software-defined networking: A survey, use cases, and future directions," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2359–2391, Fourthquarter 2017.

[142] W. Rafique, L. Qi, I. Yaqoob, M. Imran, R. U. Rasool, and W. Dou, "Complementing IoT services through software defined networking and edge computing: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1761–1804, thirdquarter 2020.

[143] O. Salman, I. Elhajj, A. Kayssi, and A. Chehab, "Edge computing enabling the internet of things," in *Proc. IEEE 2nd World Forum on Internet of Things*, Milan, Italy, Dec. 2015, pp. 603–608.

[144] Y. Jararweh, A. Doulat, A. Darabseh, M. Alsmirat, M. Al-Ayyoub, and E. Benkhelifa, "SDMEC: Software defined system for mobile edge computing," in *2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW)*, Berlin, Germany, Apr. 2016, pp. 88–93.

[145] X. Zhang, M. Yang, Y. Zhao, J. Zhang, and J. Ge, "An SDN-based video multicast orchestration scheme for 5G ultra-dense networks," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 77–83, 2017.

[146] A. Bentaleb, A. C. Begen, R. Zimmermann, and S. Harous, "SDNHAS: An SDN-enabled architecture to optimize QoE in HTTP adaptive streaming," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2136–2151, Oct. 2017.

[147] A. Bentaleb, A. C. Begen, and R. Zimmermann, "QoE-aware bandwidth broker for HTTP adaptive streaming flows in an SDN-enabled HFC network," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 575–589, June 2018.

[148] E. Liotou, K. Samdanis, E. Pateromichelakis, N. Passas, and L. Merakos, "QoE-SDN APP: A rate-guided QoE-aware SDN-APP for HTTP adaptive video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 598–615, Mar. 2018.

[149] J. Gil Herrera and J. F. Botero, "Resource allocation in NFV: A comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, Sep. 2016.

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

[151] L. Bing, Z. Yunyong, and X. Lei, "An MEC and NFV integrated network architecture," *ZTE Communications*, vol. 15, no. 2, pp. 19–25, Apr. 2017.

[152] K. Antevski, C. J. Bernardos, L. Cominardi, A. de la Oliva, and A. Mourad, "On the integration of NFV and MEC technologies: architecture analysis and benefits for edge robotics," *Computer Networks*, p. 107274, July 2020.

[153] S. Li, Z. Guo, G. Shou, Y. Hu, and H. Li, "QoE analysis of NFV-based mobile edge computing video application," in *Proc. IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, China, Sep. 2016, pp. 411–415.

[154] M. Kourtis, H. Koumaras, G. Xilouris, and F. Liberal, "An NFV-based video quality assessment method over 5G small cell networks," *IEEE MultiMedia*, vol. 24, no. 4, pp. 68–78, June 2017.

[155] L. Van Ma, V. Q. Nguyen, J. Park, and J. Kim, "NFV-based mobile edge computing for lowering latency of 4K video streaming," in *Proc. Tenth International Conference on Ubiquitous and Future Networks*, Prague, Czech, July 2018, pp. 670–673.

[156] F. Alvarez, D. Breitgand, D. Griffin, P. Andriani, S. Rizou, N. Zioulis, F. Moscatelli, J. Serrano, M. Keltsch, P. Trakadas, T. K. Phan, A. Weit, U. Acar, O. Prieto, F. Iadanza, G. Carrozzo, H. Koumaras, D. Zarpalas, and D. Jimenez, "An edge-to-cloud virtualized multimedia service platform for 5G networks," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 369–380, 2019.

[157] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, thirdquarter 2018.

[158] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.

[159] E. J. dos Santos, R. D. Souza, J. L. Rebelatto, and H. Alves, "Network slicing for URLLC and eMBB with max-matching diversity channel allocation," *IEEE Communications Letters*, vol. 24, no. 3, pp. 658–661, Mar. 2019.

[160] "Multi-access edge computing (MEC); support for network slicing," *ETSI*, Nov. 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/MEC/001_099/024/02.01.01_60/gr_mec024v020101p.pdf

[161] A. Ksentini and P. A. Frangoudis, "Toward slicing-enabled multi-access edge computing in 5G," *IEEE Network*, vol. 34, no. 2, pp. 99–105, March/April 2020.

[162] L. Cominardi, T. Deiss, M. Filippou, V. Sciancalepore, F. Giust, and D. Sabella, "MEC support for network slicing: Status and limitations from a standardization viewpoint," *IEEE Communications Standards Magazine*, vol. 4, no. 2, pp. 22–30, June 2020.

[163] L. Dong and R. Li, "Latency guarantee for multimedia streaming service to moving subscriber with 5G slicing," in *Proc. International Symposium on Networks, Computers and Communications*, Rome, Italy, Nov. 2018, pp. 1–7.

[164] Q. Wang, J. Alcaraz-Calero, R. Ricart-Sanchez, M. B. Weiss, A. Gavras, N. Nikaein, X. Vasilakos, B. Giacomo, G. Pietro, M. Roddy, M. Healy, P. Walsh, T. Truong, Z. Bozakov, K. Koutsopoulos, P. Neves, C. Patachia-Sultanoiu, M. Iordache, E. Oproiu, I. G. B. Yahia, C. Angelo, C. Zotti, G. Celozzi, D. Morris, R. Figueiredo, D. Lorenz, S. Spadaro, G. Agapiou, A. Aleixo, and C. Lomba, "Enable advanced QoS-aware network slicing in 5G networks for slice-based media use cases," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 444–453, Mar. 2019.

[165] H. Khan, S. Samarakoon, and M. Bennis, "Enhancing video streaming in vehicular networks via resource slicing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3513–3522, Feb. 2020.

[166] Y. Wang, Y. Zhang, M. Sheng, and K. Guo, "On the interaction of video caching and retrieving in multi-server mobile-edge computing systems," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1444–1447, Oct. 2019.

[167] H. Zhu, Y. Cao, Q. Hu, W. Wang, T. Jiang, and Q. Zhang, "Multi-bitrate video caching for D2D-enabled cellular networks," *IEEE MultiMedia*, vol. 26, no. 1, pp. 10–20, Jan. 2019.

[168] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "Energy-aware QoE and backhaul traffic optimization in green edge adaptive mobile video streaming," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 828–839, Sep. 2019.

[169] H. Xing and W. Song, "Collaborative content distribution in 5G mobile networks with edge caching," in *Proc. IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[170] Q. Xu, Z. Su, and Q. Yang, "Blockchain-based trustworthy edge caching scheme for mobile cyber-physical system," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1098–1110, 2020.

[171] J. Hsu, "Buffer behavior with poisson arrival and geometric output processes," *IEEE Transactions on Communications*, vol. 22, no. 12, pp. 1940–1941, 1974.

[172] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE INFOCOM'99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now*, vol. 1, New York, NY, USA, Mar. 1999, pp. 126–134.

[173] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on selected areas in communications*, vol. 18, no. 6, pp. 1012–1032, June 2000.

[174] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, Jan. 2001.

[175] T. D. Nguyen, J. Y. Khan, and D. T. Ngo, "A distributed energy-harvesting-aware routing algorithm for heterogeneous IoT networks," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 4, pp. 1115–1127, Dec. 2018.

[176] L. Lee, "Measures of distributional similarity," in *Proc. of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, College Park, Maryland, June 1999, pp. 25–32.

[177] 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 Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1710–1732, 2018.

[178] W. U. Rahman, C. S. Hong, and E. Huh, "Edge computing assisted joint quality adaptation for mobile video streaming," *IEEE Access*, vol. 7, pp. 129 082–129 094, Sep. 2019.

[179] C. Long, Y. Cao, T. Jiang, and Q. Zhang, "Edge computing framework for cooperative video processing in multimedia IoT systems," *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1126–1139, May 2018.

[180] M. Liu, Y. Teng, F. R. Yu, V. C. M. Leung, and M. Song, "A deep reinforcement learning-based transcoder selection framework for blockchain-enabled wireless D2D transcoding," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3426–3439, June 2020.

[181] J. Kunstler and A. Dobardziev, "Wireline video QoE using U-vMOS," *Ovum*, 2016. [Online]. Available: https://www.omdia.com/~/media/informa-shop-window/tmt/files/whitepapers/huawei_wireline_video_online.pdf

[182] T. Liufei, "The view's great with U-vMOS," *Industry Perspectives*, pp. 29–31, Mar. 2016.

[183] K. Zahoor, K. Bilal, A. Erbad, and A. Mohamed, "Service-less video multicast in 5G: Enablers and challenges," *IEEE Network*, vol. 34, no. 3, pp. 270–276, May/June 2020.

[184] M. F. Tuysuz and M. E. Aydin, "QoE-based mobility-aware collaborative video streaming on the edge of 5G," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7115–7125, Nov. 2020.

[185] "Mobile-edge computing (MEC); service scenarios," *ETSI*, Nov. 2015. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/mec-ieg/001_099/004/01.01.01_60/gs_mec-ieg004v010101p.pdf

[186] W. Kim and C. Youn, "Lightweight online profiling-based configuration adaptation for video analytics system in edge computing," *IEEE Access*, vol. 8, pp. 116 881–116 899, June 2020.

[187] X. Zhang, H. Lin, M. Chen, B. Kang, and L. Wang, "MEC-enabled video streaming in device-to-device networks," *IET Communications*, vol. 14, no. 15, pp. 2453–2461, Sep. 2020.

[188] J. Ahn, J. Lee, D. Niyato, and H. S. Park, "Novel QoS-guaranteed orchestration scheme for energy-efficient mobile augmented reality applications in multi-access edge computing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 631–13 645, Nov. 2020.

[189] Y. Cheng, "Edge caching and computing in 5G for mobile augmented reality and haptic internet," *Computer Communications*, vol. 158, pp. 24–31, Apr. 2020.

[190] J.-H. Huh and Y.-S. Seo, "Understanding edge computing: Engineering evolution with artificial intelligence," *IEEE Access*, vol. 7, pp. 164 229–164 245, Oct. 2019.

[191] G. Naik, B. Choudhury, and J. Park, "IEEE 802.11bd & 5G NR V2X: Evolution of radio access technologies for V2X communications," *IEEE Access*, vol. 7, pp. 70 169–70 184, May 2019.

[192] X. Jiang, F. R. Yu, T. Song, Z. Ma, Y. Song, and D. Zhu, "Blockchain-enabled cross-domain object detection for autonomous driving: A model sharing approach," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3681–3692, 2020.

[193] J. Pereira, M. Diaz-Cacho, S. Sargento, A. Zuquete, L. Guardalben, and M. Luis, "Vehicle-to-vehicle real-time video transmission through IEEE 802.11p for assisted-driving," in *Proc. IEEE 87th Vehicular Technology Conference*, Porto, Portugal, June 2018, pp. 1–6.

[194] B. H. Allah and I. Abdellah, "MEC towards 5G: A survey of concepts, use cases, location tradeoffs," *Transactions on Machine Learning and Artificial Intelligence*, vol. 5, no. 4, 2017.

[195] S. Sengupta, N. Ganguly, S. Chakraborty, and P. De, "Hotdash: Hotspot aware adaptive video streaming using deep reinforcement learning," in *Proc. IEEE 26th International Conference on Network Protocols (ICNP)*, Cambridge, UK, Sep. 2018, pp. 165–175.

[196] L. Lei, X. Xiong, L. Hou, and K. Zheng, "Collaborative edge caching through service function chaining: Architecture and challenges," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 94–102, June 2018.

[197] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, "Mobile edge computing potential in making cities smarter," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 38–43, 2017.

[198] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, "Edge computing security: State of the art and challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1608–1631, Aug. 2019.

[199] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, pp. 680 – 698, Jan. 2018.

[200] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu, "Data security and privacy-preserving in edge computing paradigm: Survey and open issues," *IEEE Access*, vol. 6, pp. 18 209–18 237, Mar. 2018.

[201] H. Liu, Y. Zhang, and T. Yang, "Blockchain-enabled security in electric vehicles cloud and edge computing," *IEEE Network*, vol. 32, no. 3, pp. 78–83, May 2018.

[202] R. Liao, H. Wen, J. Wu, F. Pan, A. Xu, H. Song, F. Xie, Y. Jiang, and M. Cao, "Security enhancement for mobile edge computing through physical layer authentication," *IEEE Access*, vol. 7, pp. 116 390–116 401, Aug. 2019.

[203] H. Hui, C. Zhou, X. An, and F. Lin, "A new resource allocation mechanism for security of mobile edge computing system," *IEEE Access*, vol. 7, pp. 116 886–116 899, Aug. 2019.

Xiantao Jiang (Member, IEEE) received an M.E. degree from Shanghai Maritime University (SMU) in 2012, and the Dual Ph.D. degrees from Tongji University of China and Tokushima University of Japan in 2016. After that, he joined the college of information engineering at Shanghai Maritime University as an Assistant Professor. He is currently a post-doc in Carleton University, Ottawa, Canada. His research interests include multimedia networks, multi-access edge computing, and blockchain. He has co-authored over 30 technical papers in archival journals and refereed conference proceedings.

F . Richard Yu (Fellow, IEEE) received the PhD degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2006, he was with Ericsson (in Lund, Sweden) and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. He received the IEEE TCGCC Best Journal Paper Award in 2019, Distinguished Service Awards in 2019 and 2016, Outstanding Leadership Award in 2013, Carleton Research Achievement Awards in 2012 and 2021, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom'10, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and the Best Paper Awards at IEEE ICNC'18, VTC'17 Spring, ICC'14, Globecom'12, IEEE/IFIP TrustCom'09 and Int'l Conference on Networking'05. His research interests include connected/autonomous vehicles, security, artificial intelligence, blockchain and wireless cyber-physical systems.

He serves on the editorial boards of several journals, including Co-Editor-in-Chief for Ad Hoc & Sensor Wireless Networks, Lead Series Editor for IEEE Transactions on Vehicular Technology, IEEE Communications Surveys & Tutorials, and IEEE Transactions on Green Communications and Networking. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. He has been named in the Clarivate Analytics list of "Highly Cited Researchers" in 2019 and 2020. He is an IEEE Distinguished Lecturer of both Vehicular Technology Society (VTS) and Comm. Society. He is an elected member of the Board of Governors of the IEEE VTS and Editor-in-Chief for IEEE VTS Mobile World newsletter. Dr. Yu is a registered Professional Engineer in the province of Ontario, Canada, and a Fellow of IEEE, IET, and Engineering Institute of Canada (EIC).

Tian Song (Member, IEEE) received his B.E. degree from Dalian University of Technology, China, in 1995, and his M.E. and Dr.E. degrees from Osaka University in 2001 and 2004, respectively. He joined Tokushima University in 2004 as an Assistant Professor. Presently, he is an Associate Professor of the Department of Electrical and Electronic Engineering, Graduate School of Advanced Technology and Science, Tokushima University. He is a member of IEICE and IEEE. His current research interests include video coding algorithms.

Victor C. M. Leung (Life Fellow, IEEE) is currently a Professor of electrical and computer engineering and holds the TELUS Mobility Research Chair, The University of British Columbia (UBC). His research is in the broad areas of wireless networks and mobile systems. He has co-authored over 1100 technical papers in archival journals and refereed conference proceedings, several of which has received the best paper awards. He is a fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. He received the IEEE Vancouver Section Centennial Award, the 2011 UBC Killam Research Prize, the 2017 Canadian Award for Telecommunications Research, and the 2018 IEEE ComSoc TGCC Distinguished Technical Achievement Recognition Award. He has co-authored papers that received the 2017 IEEE ComSoc Fred W. Ellersick Prize, the 2017 IEEE Systems Journal Best Paper Award, and the 2018 IEEE ComSoc CSIM Best Journal Paper Award. He is serving on the Editorial Boards of the IEEE Trans. Green Comm. Net., the IEEE Trans. Cloud Computing, the IEEE ACCESS, and several other journals.