Edith Cowan University

Research Online

Research outputs 2014 to 2021

2020

# A review on green caching strategies for next generation communication networks

M. Ishtiaque A. Zahed
*Edith Cowan University*

Iftekhar Ahmad
*Edith Cowan University*

Daryoush Habibi
*Edith Cowan University*

Quoc Viet Phung
*Edith Cowan University*

Md Munjure Mowla

*See next page for additional authors*

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

10.1109/ACCESS.2020.3040958
Zahed, M. I. A., Ahmad, I., Habibi, D., Phung, Q. V., Mowla, M. M., & Waqas, M. (2020). A review on green caching strategies for next generation communication networks. *IEEE Access, 8*, 212709-212737. https://doi.org/10.1109/ACCESS.2020.3040958
This Journal Article is posted at Research Online.
https://ro.ecu.edu.au/ecuworkspost2013/9355

## Authors

M. Ishtiaque A. Zahed, Iftekhar Ahmad, Daryoush Habibi, Quoc Viet Phung, Md Munjure Mowla, and Muhammad Waqas

This journal article is available at Research Online: https://ro.ecu.edu.au/ecuworkspost2013/9355

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 9, 2020, accepted November 24, 2020, date of publication November 26, 2020,
date of current version December 9, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3040958*

# A Review on Green Caching Strategies for Next Generation Communication Networks

**M. ISHTIAQUE A. ZAHED[1], IFTEKHAR AHMAD[1], (Member, IEEE), DARYOUSH HABIBI[1], (Senior Member, IEEE), QUOC VIET PHUNG[1], (Member, IEEE), MD. MUNJURE MOWLA[2], AND MUHAMMAD WAQAS[3,4], (Member, IEEE)**

[1]School of Engineering, Edith Cowan University, Perth, WA 6027, Australia
[2]Department of Electronics and Telecommunication Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[3]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[4]Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Science and Technology, Topi 23460, Pakistan

Corresponding author: M. Ishtiaque A. Zahed (mazizzah@our.ecu.edu.au)

**ABSTRACT** In recent years, the ever-increasing demand for networking resources and energy, fueled by the unprecedented upsurge in Internet traffic, has been a cause for concern for many service providers. Content caching, which serves user requests locally, is deemed to be an enabling technology in addressing the challenges offered by the phenomenal growth in Internet traffic. Conventionally, content caching is considered as a viable solution to alleviate the backhaul pressure. However, recently, many studies have reported energy cost reductions contributed by content caching in cache-equipped networks. The hypothesis is that caching shortens content delivery distance and eventually achieves significant reduction in transmission energy consumption. This has motivated us to conduct this study and in this article, a comprehensive survey of the state-of-the-art green caching techniques is provided. This review paper extensively discusses contributions of the existing studies on green caching. In addition, the study explores different cache-equipped network types, solution methods, and application scenarios. We categorically present that the optimal selection of the caching nodes, smart resource management, popular content selection, and renewable energy integration can substantially improve energy efficiency of the cache-equipped systems. In addition, based on the comprehensive analysis, we also highlight some potential research ideas relevant to green content caching.

**INDEX TERMS** Caching, energy, Internet traffic, IoT, video streaming.

## I. INTRODUCTION

In recent years, the advancements in smart hand-held devices, IoT applications, on-demand streaming services, and wireless data transfer [1]–[3] have resulted in a phenomenal growth in network traffic demand. By 2030, the number of communication devices, served by diverse network paradigms, is expected to reach 100 billion, and the mobile data traffic will be 10,000 times higher compared to the traffic in 2010 [4].

The future Internet is envisioned as a networked society with ubiquitous connectivity and information dissemination opportunities [3], [5]. In this aspect, the major role of wireless networks has evolved from pair-wise telephony conversations towards content distributions [6], [7]. This paradigm shift in communication technology is expected to support diverse applications, including online video streaming, smart homes, connected vehicles, smart cities, and machine to machine (M2M) communications [8]–[10]. These applications require high data rates, improved quality of experience (QoE), low latency, longer battery lifespan, robust security protections, and optimal bandwidth utilization [2], [11]. To fulfill these demands, operators are constantly deploying networking resources, where capital expenditure is increasing dramatically [8], [12].

The explosively growing Internet traffic is also increasing energy requirements at a staggering rate [3], [4]. Every single content object of the Internet traffic requires efficient processing, transmission, and dissemination resulting in excessive energy consumption. Energy cost makes up about 30% of the operational costs of a communication network [4]. This upsurge in energy consumption also leads to excessive carbon footprint and drastic environmental consequences [13], [14]. Furthermore, the carbon emission rate is recognized as a key performance indicator for the company's brand image [15]. Another prime concern is limited battery lifetime and the

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


This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

increased radiations of mobile devices. Per annum more than 25 thousand of batteries are disposed [7]. These batteries disposal not only increase costs, but also impact human health and ecological balances [4].

Consequently, energy-efficient network design and the smart management of networking resources are regarded as the key attributes of next generation communication networks [16]–[18]. The inclusion of green strategies in communication networks has attained considerable interest from both industry and academia [4], [19]. The stakeholders have already targeted to reduce the energy consumption by 90% in 5G networks [20]. A number of research initiatives, such as EARTH, GreenTouch, Green Radio, TREND, and OPERA-NET have been introduced to design energy-efficient solutions for communication networks [19], [21]. However, in spite of the remarkable success of these projects, energy-aware communication system design is still a prime research field in the telecommunication industry.

The major portion of the proliferating Internet traffic is mainly comprised of two key data sources [22], [23]. The first source is on-demand video (VoD) streaming and content sharing services (e.g., Netflix, youtube). In global Internet traffic, the contribution of video streaming is anticipated to be more than 80% by 2021 [24]. The streaming of high resolution 4k and 8k videos requires very high data rates and low latency [25]. Therefore, in near future, the exponentially growing popularity of video traffic will cause further increase in the data volume.

The second source is the IoT data generated from thousands of sensors, actuators, and applications in IoT networks [26], [27]. It is expected that more than 75 billion IoT devices will be connected to the Internet within 6 years [28]. These devices will also escalate the volume of the network traffic.

According to recent studies [11], [24], [29], a very small portion of the content objects makes up the major portion of overall Internet traffic. In addition, a common feature of both the aforementioned data types is that users repeatedly requests for them [22], [26], [30]. However, these data are not suitable for broadcasting as requests are generated in an asynchronous manner [11]. Hence, broadcasting can cause significant delay to earlier requests and degrade user experience. On the other hand, a conventional host based network serves the subscribers from the remote servers of the content providers. As a result, the network causes duplicate content transmission and creates backhaul bottlenecks [31].

To combat these technical limitations, caching popular content at the network nodes is envisioned as a prospective countermeasure [2], [25]. In caching, frequently requested data items are stored in an intermediate network node between primary content server and user device [6], [32]. These popular items are cached before the arrival of content requests and utilized to serve the users during peak-hours. Caching mainly replaces additional bandwidth requirements by cheaper data storage units [25]. Therefore, cache-equipped networks reduce content delivery distance and avoid duplicate transmissions [33]. As a result, not only user-perceived latency is decreased, but also the backhaul bottleneck is alleviated. In addition, caching improves both spectrum efficiency and network throughput [34].

Caching nodes are frequently updated because of the changing popularity of cached content objects [35], [36]. Therefore, content replacement techniques require accurate identification of less popular items to accommodate the popular files. In addition to content popularity, an efficient caching strategy considers content size, delivery path, and transmission time [22], [37], [38].

Although caching has been introduced to achieve faster content delivery [39], it has shown remarkable potential in reducing energy demand [40], [41]. This is because shortening content delivery distance reduces both content transmission energy and network resource utilization. However, the operational procedure of caching units requires additional energy. Hence, the design of innovative caching techniques, which not only reduces energy consumptions but also maintains expected QoE, is of widespread interest. Such innovative techniques are categorized as green caching strategies. These strategies investigate the energy aspects of caching and improve energy efficiency of the networking systems. The fundamental consideration of green caching networks is to minimize the difference between utilized and offered caching resources. To address this, a number of studies [2], [11], [42]–[44] have investigated and proposed diverse optimal caching resource allocation, network management, and renewable energy utilization techniques for latency-critical and bandwidth-greedy applications.

Energy-aware cache node selection and optimal network planning substantially improve energy savings [2], [42]. In this process, the nodes serve maximum number of user requests by deploying minimum number of caching units in the network. According to the existing studies [43], [44], the smart selection of content delivery paths and the activation of caching nodes based on network traffic demand exhibit considerable improvement in energy efficiency. Furthermore, the utilization of renewable energy to power up caching nodes reduces not only the dependence on grid energy but also carbon emission [14]. In addition, the joint utilization of these techniques exhibits significant potential in limiting peak-to-average disparity of the energy load and improving network load adaptation [45]. However, the exploitation of the green approaches in cache-equipped networks may impact network throughput and transmission delay [29], [46]. The inclusion of the additional services at the caching nodes also causes additional energy consumption [47]. Therefore, some studies [48]–[50] have investigated the trade-off analysis of the caching schemes based on energy consumption and network performance.

As reported in the existing literature [45], [51], [52], a number of green caching strategies have significantly improved both network performance and energy efficiency. However, the above mentioned challenges need to be addressed in future initiatives. Motivated by these considerations,

**FIGURE 1.** The detailed taxonomy of the survey on green caching strategies.

we provide a detailed survey of the state-of-the-art caching strategies, which have been exploited to investigate the energy aspects of caching. The major contributions of this study are as follows:

- In this article, a comprehensive review of the existing survey papers on existing caching strategies is provided. The key focus of this study is to analyze the energy-related concerns of caching systems. Therefore, we highlight the existing strategies on green caching and identify the research gaps for further study.
- This article also identifies the distinctive role of individual aspects of caching techniques towards energy savings. These aspects, include cache node placement, content selection, and caching strategies.
- We summarize the existing green caching strategies into three categories, which are energy-aware resource allocation, green resource management, and renewable energy utilization. In this article, the major contributions, solution methods, different networking paradigms, and limitations of existing articles, related to the above mentioned categories, are identified.
- The study also provides a comprehensive summary of the utilized performance metrics and diverse applications of the green caching techniques.
- Finally, this article categorically explains different future opportunities and related technical challenges.

The remaining sections of this survey paper are organized as follows. Section II reviews existing survey articles and identifies the research gap. In Section III, we investigate the major issues involving proposed caching strategies. Section IV to Section VI analyze existing studies and classify them into energy-aware network planning, green resource management, and utilization of renewable energy, respectively. In addition, section VII discusses different performance metrics, which have been considered in the literature. Section VIII highlights different use cases and applications of

caching systems. In Section IX, we discuss the future research directions and challenges. Section X concludes the article. For the convenience of the readers, the detailed taxonomy of the article is drawn in Fig. 1.

## II. RELATED SURVEY PAPERS

### A. SURVEY ARTICLES ON GREEN COMMUNICATIONS

Energy-efficient network panning, deployment, and operation have become recent concerns for the telecommunication industry [7], [19]. In the literature, a number of survey papers have reviewed green solutions for communication networks.

In [53], the advancements and importance of energy-aware technologies for core network operation were explicitly discussed. Energy-efficient initiatives for edge computing systems and relevant research challenges were outlined in [54]. Buzzi *et al.* [8] identified the key aspects of energy-aware network design. In addition, the study also analyzed the importance of holistic approach, confronting randomness in 5G networks, and innovative energy models involving caching and computing. For 5G networks, the conventional approaches of energy cost reduction and the modern concepts of renewable energy utilization were studied in [12].

In [17], a 5G multi-tier architecture was designed for energy-efficient network operation. To address the energy related concerns, the study also analyzed cell zooming, device to device communication, massive multiple input multiple output (MIMO) techniques, and hardware solutions. Performance metrics related to energy-efficient 5G systems were summarized in [4]. In addition, a novel spectrum sharing technique was introduced for the enhancement of battery lifetime.

Energy-efficient system design initiates trade-off considerations related to user experience and system performance. These trade-offs in wireless networks were explicitly explained in [16], [55].

### B. SURVEY ARTICLES ON CACHING

Caching has emerged as a paradigm shifting technology to explore the in-network storage capability of the networking nodes and improve QoE through low dissemination latency [56]. In-network caching is used in information-centric networks (ICN) [57]–[59] to serve user requests from network nodes, which are located closer to the subscribers.

ICN is based on named content objects, and subscribers send their requests as interest packets containing content names [60]. The network itself finds the most suitable caching node to serve the requests and maps content delivery routes. A detailed description of the ICN operating procedure was provided in [37], [59]–[61]. The authors in [38], [62] illustrated some ICN frameworks, related features, and described key influencing factors for the prompt delivery of the cached content. Ioannou *et al.* [37] not only investigated on-path caching strategies, but also reviewed diverse forwarding techniques in ICNs.

A comprehensive review of the cache management strategies in ICN, including their advantages and limitations, is illustrated in [61]. Lal *et al.* [63] summarized existing research initiatives addressing efficient cache placement and content placement techniques for ICNs. Qiao *et al.* [64] presented an in-depth study of the upper-layer applications and services in content centric networks (CCN). In addition, ICN functionality and performance analysis were discussed in [60].

On the other hand, probable security threats and privacy concerns for information centric networks (ICN) were presented in [57], [65]. AbdAllah *et al.* [65] studied privacy, integrity, and confidentiality requirements of the ICNs. Tourani *et al.* [57] also identified probable access control measures to allocate stringent security services.

Arshad *et al.* [66] explained the prospects of incorporating ICN features into IoT applications. Furthermore, Nour *et al.* [67] presented existing efforts and identified ICN as a key enabling technology for communications in IoT systems.

Exploiting caching in mobile networks also showed remarkable performance improvement [2], [22], [68] Wang *et al.* [22] reviewed the existing strategies on mobile edge networks. They also analyzed the cooperative role of communication, caching, and computing at the network edge. Yao *et al.* [68] summarized different studies on mobile edge caching and addressed associated technical challenges.

Wireless caching techniques for diverse cellular networks were reviewed in [34]. The study compared different caching algorithms based on mathematical tools, performance metrics, and network types. For IoT systems, a detailed study on the existing edge caching techniques in radio access networks (RAN) was provided in [69].

Parvez *et al.* [70] studied probable solutions to achieve real-time transmission in a 5G network. They summarized the strategies to achieve low latency into three parts, which are access network, core network, and caching solutions. Prerna *et al.* [71] outlined existing D2D technologies for content caching and sharing in 5G networks. The authors also pointed out associated research challenges and limitations in implementing D2D caching.

The integration of information-centric features into mobile networks creates opportunities for faster content download and mobility support [52], [72]. In addition, the design of an efficient information-centric mobile (ICM) system and networking resource optimization were explicitly discussed in [59]. A detailed survey on the integration of ICN networks with mobile edge computing (MEC) systems was provided in [72]. The authors also proposed future research directions on ICN over MEC for IoT applications.

Content popularity also plays a pivotal role in designing an efficient caching technique [73], [74]. Goian *et al.* [75] discussed the features of video files and summarized existing techniques adopted in content caching for the prediction of content popularity.

Table 1 briefly summarizes the contributions of the aforementioned survey articles.

### C. NOVELTY OF THIS WORK

Caching has been established as an acceptable solution for the design of an energy-efficient communication network [24], [43]. However, the green aspects of caching and associated challenges did not achieve significant attention in the literature. Most of the existing survey papers analyzed caching techniques from the perspectives of faster content delivery and alleviating backhaul bottlenecks. To the best of our knowledge, only Fang *et al.* [77] reviewed energy related issues of a cache-equipped network. The authors analyzed energy-efficiency related metrics, network design, and resource management. However, the article was published way back in 2015. Since then a number of modern caching techniques have been introduced to optimize energy-efficiency and system performance.

In this article, we summarize the recent advancements in energy-aware network design and operation, including D2D caching systems, IoT data caching, software defined networking (SDN), and MEC technologies. In addition, these studies analyse diverse techniques involving high resolution video streaming, popular content selection, and social relationships in communication networks. The role of caching has evolved from wire-based networks to wireless networks. To address this transformation, our proposed study also reviews recent articles on heterogeneous networks (HetNets), cloud radio access networks (C-RAN), and fog radio access networks (F-RAN). Moreover, our study also highlights the utilization of state-of-the-art technologies in next generation communication networks. These technologies include, deep reinforcement learning, neural network, dynamic decision making procedures, queuing theory, and heuristic approaches. The existing survey on green caching [77] did not consider any of these modern approaches. Moreover, the utilization of renewable energy in cache-equipped systems was overlooked in the existing article [77], which we have included in this study. Future research challenges and their prospective solutions are also outlined to overcome these technical concerns.

## III. AN OVERVIEW OF CACHING

The use of caching units was first introduced in the central processing unit (CPU) of the computers [69]. This cache worked as a memory unit to store frequently accessed data by the CPU. The process remarkably improved data rate and reduced delay. Web caching strategies, based on content recentness and frequency, also achieved significant performance gain [78].

Currently, this concept of caching is explored in the field of communication networks [34], [61]. Fig. 2a shows a conventional host-based network, which delivers content to subscribers' from the network server. Therefore, the content travels a long-distance through the core network. On the other hand, Fig. 2b presents a cache-equipped network. In this configuration, if the requested content is cached at the core node, subscribers can download from that. Therefore, the content delivery distance shortens, which reduces not only content delivery latency but also core network traffic.

Caching mainly investigates three key aspects: cache storage allocation, popular content selection, and caching strategies [39], [59], [69]. These points are briefly discussed as follows:

### A. CACHE STORAGE PLACEMENT

The caching units can be placed at any of the network nodes based on their position and allocated resources. However, this placement varies in wired and wireless networks.

An ICN is regarded as the most widely accepted cache-equipped wired network and caches content at the intermediate routers [6], [61]. Fig. 3 presents a detailed overview of caching in an ICN, which explores receiver-driven policy for efficient content download. Every node in the network maintains communication using interest packets and data packets. Subscribers send their interest packets to the connected ICN routers [57]. An interest packet contains content name and searches content store (CS). If the packet matches to the stored content, the node serves request in the reverse path. Otherwise, the pending interest table (PIT) records the packet and forwards to the other nodes. In this process, a forwarding information base (FIB) table is utilized to find the next node for interest transmission, where every node checks the CS [6], [61]. If any of the CS contains the same interest packet, data packet uses the reversed path to serve user requests. When none of these intermediate nodes contains the interest packet, content server serves the request. If there is empty spaces in the cache storage units, the intermediate nodes cache the requested content. Later, the cached content is utilized to provide faster content delivery to other subscribers [57].

In caching, the in-network storage capability of the routers is exploited to provide a robust connection and improved QoE through low dissemination latency. In-network caching can be categorized into off-path caching and on-path caching [6], [37].

In off-path caching, caches work as an auxiliary information publisher and contents are replicated within the network without considering the delivery path [6]. The main goal of off-path caching is to make content objects available at the edge nodes for long-term access. Off-path caching requires sufficient prior knowledge and the decision is taken on the basis of optimization techniques [37].

Conversely, on-path caching is an opportunistic technique which caches content when it moves from the source through the delivery path [6]. It maintains a strict forwarding mechanism but requires minimum prior knowledge. This dynamic strategy is suitable for short-term caching and mostly depends on low-complexity online algorithms [37].

Caching in cellular networks also achieves remarkable performance gain [39]. In a wireless cellular network, the cache storage can be placed either at the evolved packet core or at the access network [30], [39]. Caching units can also

**TABLE 1. Summary of the related survey articles.**

| Research Area | Sl. No. | Authors | Key Focus | Publication Year |
|---|---|---|---|---|
| Green Communication Strategies | 1 | Idzikowski *et al.* [53] | Energy-efficient techniques for core network operations | 2016 |
| | 2 | Buzzi *et al.* [8] | State-of-the-art strategies to reduce energy consumption in 5G networks | 2016 |
| | 3 | Zhang *et al.* [12] | Resource management and renewable energy utilization strategies for sustainable 5G | 2018 |
| | 4 | Gandotra *et al.* [4] | Energy-efficient 5G technologies and associated performance metrics | 2017 |
| | 5 | Sofi *et al.* [17] | Energy efficiency of a multi-tier 5G architecture | 2018 |
| | 6 | Mahapatra *et al.* [16] | Trade-offs related to energy-efficient wireless networks | 2016 |
| | 7 | Jiang *et al.* [54] | Energy-efficient initiatives for edge computing systems | 2020 |
| Caching in ICNs | 8 | Zhang *et al.* [62] | Increasing content availability and associated challenges in ICN | 2013 |
| | 9 | Abdullahi *et al.* [38] | Existing ICN architectures and caching methods to improve network performance | 2015 |
| | 10 | Ioannou *et al.* [37] | Content caching and forwarding techniques in ICN | 2016 |
| | 11 | Din *et al.* [61] | Cache management techniques in ICN | 2018 |
| | 12 | Lal *et al.* [63] | Caching methodologies for efficient cache placement and content allocation in ICN | 2019 |
| | 13 | Qiao *et al.* [64] | Applications and services provisioning of CCN | 2019 |
| | 14 | Khandaker *et al.* [60] | ICN functionality with qualitative and quantitative analysis | 2019 |
| | 15 | AbdAllah *et al.* [65] | Probable security attacks and security requirements in ICNs | 2015 |
| | 16 | Tourani *et al.* [57] | Security concerns and mitigation approaches for access control in ICNs | 2018 |
| Incorporation of ICN and IoT | 17 | Arshad *et al.* [66] | The integration of ICN and IoT for ubiquitous connectivity | 2018 |
| | 18 | Nour *et al.* [67] | State-of-the-art ICN enabled solutions for IoT applications | 2019 |
| Content Popularity | 19 | Goian *et al.* [75] | Content prediction approaches and popularity measurement for video caching | 2019 |
| Caching in Mobile Networks | 20 | Wang *et al.* [22] | The prospects and role of mobile edge networks | 2017 |
| | 21 | Wang *et al.* [76] | Integrated approach combining computing, caching, and networking for wireless systems | 2018 |
| | 22 | Yao *et al.* [68] | State-of-the-art mobile edge caching mechanisms and technical challenges | 2019 |
| | 23 | Li *et al.* [34] | Caching strategies for diverse cellular networks, and comparisons in terms of mathematical tools and performance metrics | 2018 |
| | 24 | Piao *et al.* [69] | State-of-the-art edge caching techniques in RAN for IoT | 2019 |
| | 25 | Parvez *et al.* [70] | Low latency solutions for 5G networks | 2018 |
| | 26 | Prerna *et al.* [71] | State-of-the-art D2D caching strategies and challenges for 5G networks | 2020 |
| Mobile ICN | 27 | Jin *et al.* [59] | Caching architectures and optimization in ICM networks | 2017 |
| | 28 | Ullah *et al.* [72] | The prospects and challenges of ICN over MEC for IoT systems | 2018 |
| | 29 | Fang *et al.* [58] | Mobile ICNs and associated technical challenges | 2018 |
| Green caching | 30 | Fang *et al.* [77] | Green perspectives of ICNs | 2015 |

be placed at the subscribers' devices [79]. In addition, the placement of caching units at the network core can decrease up to two-thirds of the data traffic [22]. Caching content at base stations (BSs) provides an opportunity to address user requests locally within the coverage area [33], [56]. When a mobile user sends a request to the connected BS, the edge server connected to the BS detects whether the content is cached or not. The domain name system (DNS) enables the user to download content from the closest BS that caches the content [14], [80].

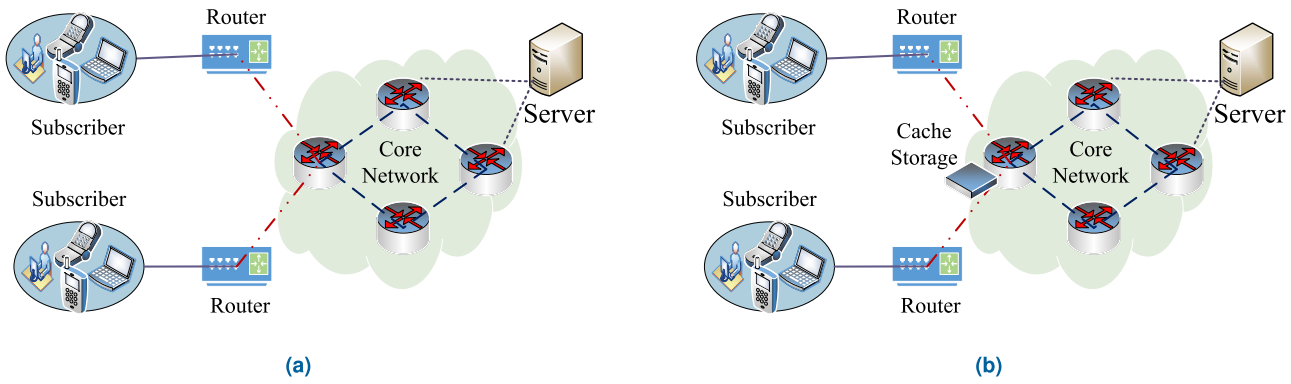


**FIGURE 2.** (a) A conventional host-based network. (b) A cache-equipped network.

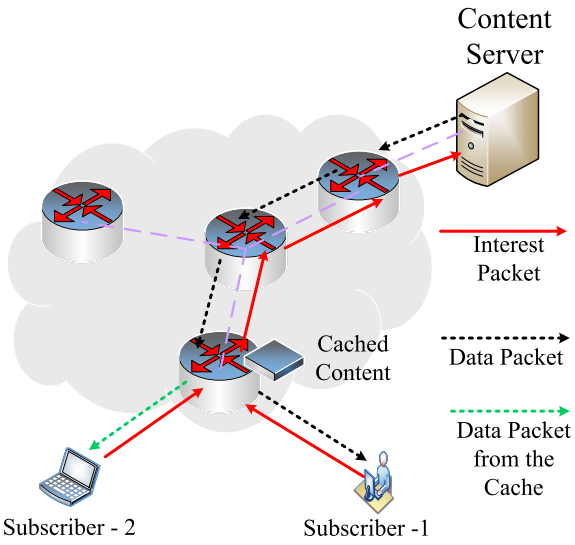


**FIGURE 3.** Caching in an ICN.

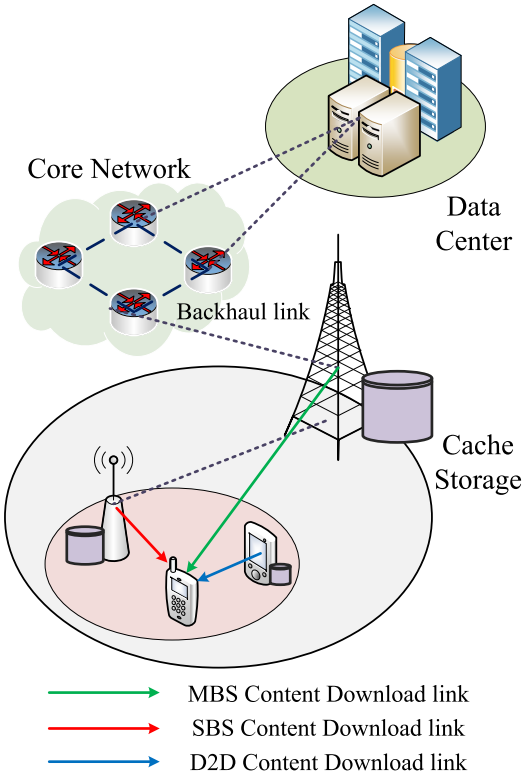


**FIGURE 4.** Caching in a HetNet.

In HetNet, the deployment of cache storages at the small cell BSs (SBS) brings content closer to the subscribers [41], [81]. Therefore, network data rates improve and content delivery latencies reduce. On the other hand, the D2D caching technique exploits the storage capacity of the user devices [71], [82]. The users have the option to store already downloaded content in their devices and utilize them later to serve adjacent users. D2D caching has been proven as an efficient offloading scheme to release network resources [71], [83].

In a cellular network with D2D caching capability, users first check the requested content at the nearby user devices [71], [84]. The device immediately serves the request if the content is cached. Otherwise, the content request is passed to the SBS. Again, the search is conducted at the SBS, macro BS (MBS), and the core network, respectively. If any of these entities cache the content, the request is not transferred to the upper level, and the caching node delivers the requested content [41], [84]. The content caching framework for a HetNet is illustrated in Fig. 4. Successful content delivery from the SBSs and MBSs significantly reduces backhaul traffic and alleviates bottleneck [43], [70].

Cooperative caching has also been utilized in communication networks with spectrum sharing capabilities [85], [86]. In this type of networks, secondary BSs cached popular content objects to serve the requests received at the primary BS. During content transmission from secondary BSs, if the primary users are given priority compared to secondary users then perceived delay can be reduced [85]. Furthermore, for a cognitive network, the satisfaction rate of secondary requests can be maximized by utilizing cache-split method and secondary user partitioning [86].

### B. POPULAR CONTENT SELECTION

Usually most of the user requests are for a small number of popular content objects. According to a recent study, 83% of the total requests for Facebook videos originate from only 1% of the uploaded content [87]. Therefore, the identification as well as the placement of popular content objects in close proximity to the users is crucial to achieve maximum performance gain.

Content popularity is defined in terms of the number of requests for a specific content with respect to the total number of requests for all the files [68], [88]. The popularity changes over time and is difficult to predict accurately [75]. Furthermore, content providers and subscribers add new content every minute to the Internet [89]. Spatial disparity in content popularity is also crucial for efficient caching. This is because a limited number of caching resources are deployed in the coverage area, and subscribers exhibit similar interest [75]. After certain time, the popularity of previously cached content shows a declining trend, and other content objects become popular. Multimedia videos show constant popularity for a week whereas short-form videos remain popular for only few hours [35], [80]. Therefore, the eviction and the replacement of less popular files with currently popular files are crucial.

Video files usually come with bigger data sizes compared to documents and images [14], [90]. The requirement of higher data rates and resolutions makes the situation even more critical [68]. Generally, static popularity is modeled using an independent Poisson process according to power law [91]. In a number of studies [30], [88], [92], the popularity of video content is modeled as a Zipf distribution, characterized by a number of content objects and the skewness parameter.

User preference for a specific content and social relationship also influence the popularity [5], [73]. A learning-based method [93] explored instantaneous content requests and estimated content popularity to improve cache performance. Previously cached content can also be exploited in evaluating content request probability [94].

For content with dynamic popularity, the shot noise model (SNM) is a good option [95]. This model depends on the life span and the instantaneous popularity of the requested content [22], [95]. In addition, dynamic popularity-based caching models consider multiple factors to evaluate content popularity. These factors include delivery distance, cache capacity, node centrality, and so on [74], [96].

Currently, big data analytics and machine learning techniques are widely used to evaluate content popularity [32], [97]. In cooperative strategies, caching nodes also consider the popularity of the cached contents at the neighboring nodes [2], [5], [9]. Smart traffic management, temperature monitoring, and industrial automation applications require frequent content update for seamless operations. Therefore, content update strategy also plays a pivotal role along with popularity in IoT content caching [66].

### C. CACHING STRATEGIES

The storage capacity of the caching units is limited, and it is impractical to cache all the popular content in a caching node. Therefore, the smart utilization of networking and caching resources is a prerequisite to improve system performance.

Caching strategies can be categorized into uncoded techniques [98]–[100] and coded techniques [101]–[103] based on the stored content type. In the uncoded technique, the whole file or a fractional portion of the file is stored in the caching node. On the other hand, the coded technique divides every content object into multiple segments and utilizes coding methods to store the content [34]. Although a coding process causes some overheads, it saves cache spaces [104].

The controlling procedure of the caching units can be either centralized or distributed. In a centralized process, all the caching nodes are controlled by a central entity, which gathers required information and instructs the nodes to cache and serve specific content [105]–[107]. In a distributed technique, the networking nodes identify the most suitable content for caching and serve user requests locally [45], [108]. Centralized strategy reduces caching redundancy and ensures global optimal solution whereas distributed strategy reduces system complexity.

The content objects are cached in both reactive and proactive manner [80]. In reactive caching, the caching decision is executed after the arrival of the content requests. This technique is more suitable for caching during peak hours when the intermediate nodes in the content delivery path stores the file. Proactive caching estimates content popularity exploiting user behavior and request patterns [35], [45]. Then the popular content objects are stored before the request arrivals, usually in the off-peak hours.

As mentioned earlier, content popularity changes over time and previously cached files are replaced by newly popular content objects [73], [75], [80]. Therefore, cache replacement strategies are also important for efficient system design. Recent request patterns and the frequency of requests are the key factors in cache replacement techniques.

The first in first out (FIFO) technique is the simplest method in which, the oldest content is replaced by the most recently requested content [109]. This technique often replaces popular content and causes cache redundancy. The least recently used (LRU) technique eliminates the oldest content in terms of content request by the most recently requested content [110], [111]. A least frequently used (LFU) technique replaces the content which has received minimum number of requests within a certain time [111], [112]. These two techniques are most popular and widely deployed in communication networks. The time aware least recently used (TLRU) and the least frequent recently used (LFRU) strategies are designed combining the features of these techniques [109]. TLRU counts the requests for a content object within stipulated time span and caches the item with more requests [10], [109]. If the cache is full, it utilizes LRU technique [61]. In LFRU, the cache space is partitioned into

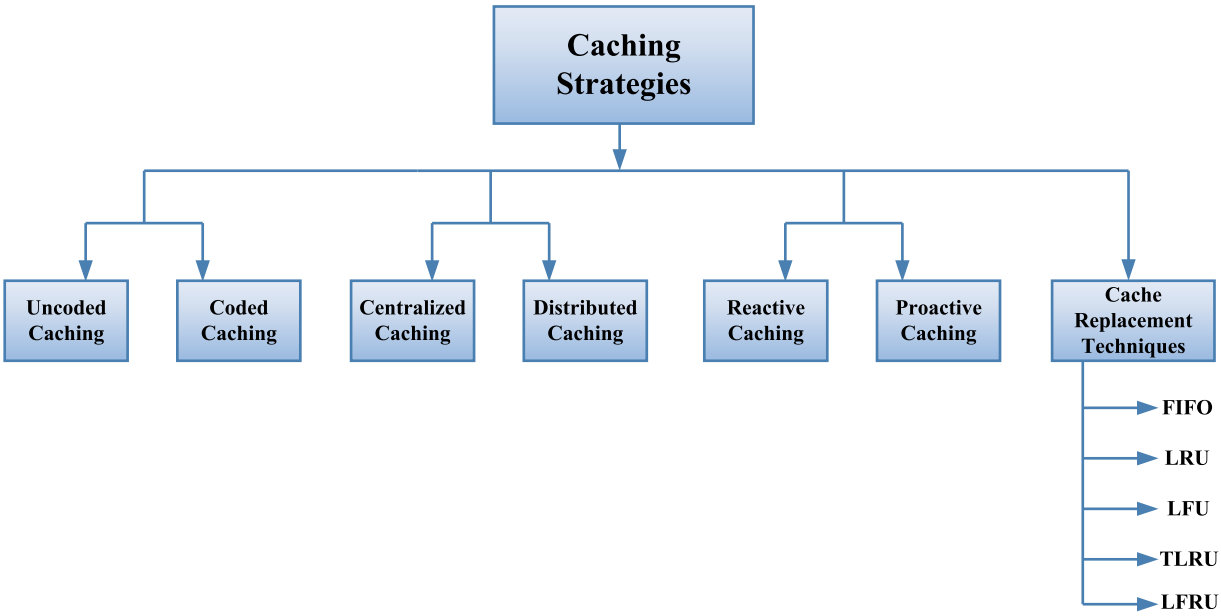


**FIGURE 5.** Classification of diverse caching strategies.

two parts. The privileged part adopts LRU strategy for content objects with dynamic popularity. In contrast, the unprivileged part gathers request history of the content for a limited time period and sets a benchmark for content replacement based on number of requests [10].

Fig. 5 illustrates a detailed summary of the classification of diverse caching strategies based on different features of the operational procedures.

## IV. ENERGY-AWARE PLANNING AND RESOURCE ALLOCATION

In this section, we discuss the energy-aware allocation of caching units and resources in a communication network.

### A. NETWORK PLANNING AND CACHING NODE SELECTION

A network comprises hundreds of nodes, and their geographical positions have a significant influence on network performance [36]. Compared to the hardly accessed nodes, caching popular content at a node, which is frequently accessed by the subscribers, is much more beneficial to offload network traffic. Therefore, the selection of the most suitable caching nodes is crucial for efficient network design.

Kitsuwan *et al.* [113] showed that the position of caching units in a CCN has a direct influence on the number of hops. They developed a betweenness-centrality based technique to optimally determine the cache node positions and minimized overall hop counts. The number of hop counts is directly proportional to the required transmission energy. Thus, the study achieved significant reduction in energy consumption.

From the user perspective, the best location for cache placement was selected in [114]. The introduced technique estimated the distance from the node to the users and placed caching units considering minimum energy requirements. Another study [115] utilized Poisson point process and Poisson hole process to model the positions of the BSs and caching nodes. The collaborative strategy not only reduced power consumption but also improved spectral efficiency.

In [36], the authors considered node importance and content popularity for energy-efficient in-network caching. They optimized cache node placement and introduced a neighborhood-based cache replacement policy. The study achieved both energy savings and higher cache hit rate. In a CCN, another energy-efficient technique [116], which considered QoS requirement, optimally allocated caching nodes and delivered requested content objects.

Cluster-based selection of the caching nodes also reduces overall energy requirements [117], [118]. A framework for MEC clustering and cache space allotment was introduced in [2]. The collaborative approach partitioned the cache space into three logical groups and optimized storage allocation. For a HetNet, another clustering framework was developed by Hajri *et al.* [117]. The proposed strategy optimally activated SBSs and allocated them to different clusters. In [118], the authors clustered a number of SBSs in an ultra-dense network (UDN) and exploited user mobility to reduce energy consumption.

In [119], a three-layer framework was designed to exploit temporal and spatial properties of wireless node mobility. The proposed model optimally selected edge clusters and proactively cached data at the edge servers. Another clustering technique [51] was proposed to reduce energy consumption and improve QoS of a C-RAN. The technique cached content in the local cluster and efficiently designed remote radio head (RRH) association.

For network planning, a number of novel strategies were also developed to achieve energy savings [120]–[122]. An energy-aware strategy introduced a set of cloudlets

as an intermediate layer in between the users and cloud servers [120]. These cloudlets cached offloading-data and executed user tasks to reduce energy consumption. The authors in [121] designed communication links for both edge and fronthaul connections to achieve energy savings in a C-RAN.

For the efficient allotment of cache storage units in 5G networks, Jia *et al.* [122] explored network slicing and increased overall revenue of the infrastructure providers. In addition, the authors in [123] utilized moving vehicles as caching nodes and exploited user mobility to reduce energy consumption of a CCN.

### B. ENERGY-EFFICIENT RESOURCE ALLOCATION

The energy-aware allocation of the networking resources also achieved significant attention in the designing of caching strategies [42], [124]–[126]. Huo *et al.* [124] explored the impact of the joint utilization of networking, caching, and computing techniques on the energy-saving rate and average-delay experience of network subscribers. Chen *et al.* [42] utilized SDN for the allocation of computing, caching, and networking resources in an energy-efficient manner. In addition, an information-centric architecture explored SDN and artificial intelligence for efficient resource allocation in IoT networks [27]. Another resource allocation strategy combined computing, caching, and communication for a F-RAN [46]. The proposed method optimized tolerant delay subject to energy requirements.

In [52], a caching architecture was designed to incorporate information-centric features into a 5G network. The study investigated the trade-off in energy consumption for content storage and transmission. Another strategy [1] optimized user QoE, cache allocation, and transmission rate for green resource allocation and content delivery.

In [125], transmission power, bandwidth, and caching resources were optimally allocated in a F-RAN. The proposed technique explored the economic perspectives of energy-aware caching and resource allocation. Similarly, Yin *et al.* [126] jointly optimized resource assignment, caching, and user connection to achieve energy savings.

Cheng *et al.* [128] optimally allocated transmission power to the subscribers and achieved not only maximum user satisfaction but also minimum outage probability. A context-aware technique selected user-BS pairs and allotted transmission power considering the limitations in backhaul and caching capability. In [130], a wireless-ICN architecture was introduced for dynamic power allocation at the network nodes based on network traffic and content popularity. This strategy also served requests from the network edge and reduced not only traffic load but also energy consumption.

For a two-tier HetNet, a caching resource and spectrum allocation scheme was proposed to determine the optimal SBS density and improve energy efficiency [131]. Another strategy optimally allocated constrained spectrum and storage resources at the cache-quipped BSs to minimize energy consumption [132]. A centralized framework for MEC and caching was proposed in [133], which jointly optimized computation and spectrum resources to reduce the overall latency of the system.

D2D-assisted systems also explored energy-efficient approaches [134], [135], [137]. In [134], a 3-tier MEC system was developed for optimal resource allocation and traffic offloading. The proposed model cached content not only at the BSs but also in user devices. Another study [135] optimally allocated downlink resources of a D2D-assisted caching system for simultaneous content distribution from the BSs and nearby users. In addition, a D2D-assisted edge caching strategy was designed for IoT systems. The novel tecnique not only reduced transmission energy consumption, but also offloaded traffic to SBSs and user devices [136]. Another framework optimally allotted transmission power and activated D2D links in a single-hop D2D network [137].

Virtual resource allocation in cache-equipped networks was studied in [138], [139]. Zhou *et al.* [138] developed an information-centric framework for a HetNet, which jointly optimized virtualization, computing, and caching. Moreover, Tan *et al.* [139] investigated both power control and user association for heterogeneous services. The study optimally allocated caching and computing resources for substantial performance gain. These studies considered energy consumption as one of the components for cost function while calculating overall revenue.

Table 2 summarizes the above mentioned studies related to energy-aware network planning and resource allocation for caching systems.

## V. OPTIMAL CONTENT PLACEMENT AND RESOURCE MANAGEMENT

In cache-equipped networks, the introduction of optimal resource management and content placement strategies exhibits significant potential in reducing energy requirements. A flurry of research [3], [43], [140]–[146] has already developed diverse green techniques for content management and energy savings.

### A. OPTIMAL SCHEDULING OF CACHING NODES

Optimal scheduling and the activation of caching nodes enable network operators to switch off redundant resources subject to user demand [24], [44], [147], [148]. In [149], the authors switched off some of the content routers and connected links of a CCN. This study reduced the energy cost while guaranteeing QoS requirements.

For a HetNet, Xie *et al.* [44] jointly optimized BS activation and caching to achieve maximum energy savings. Another proactive and energy-efficient caching scheme [147] cooperatively activated required SBSs and cached popular items accordingly. Optimal activation and the placement of caching nodes at different geographical segments of a network was investigated in [148]. For video files, the authors developed a content forwarding strategy, which considered both static and dynamic network load, to reduce energy consumption.

**TABLE 2. Summary of the literature review on energy-aware planning and resource allocation.**

| Key focus | Articles | Major contribution | Network type | Solution approach |
|---|---|---|---|---|
| Network planning and caching node selection. | [113]–[115] | Selected the most suitable nodes to place caching units and reduce energy consumption | CCN [113] ICN [114], HetNet [115] | Solved ILP optimization using Betweenness centrality based heuristic [113]. Utilized optimal stopping rule to reduce energy consumption [114]. Solved mixed integer optimization problem using Poisson cluster approximations [115]. |
| | [36], [116] | Content placement, replacement, and forwarding technique considering energy efficiency. | CCN | Exploited probability, centrality, and caching nodes' neighborhood for content delivery [36]. Domain-based aggregation and Ant colony optimization [116]. |
| | [2], [51], [117]–[119] | Optimized caching and selected networks nodes for content allocation based on clustering approach. | HetNet [2], [117], C-RAN [51], Mobile network [118], [119] | Solved optimization problem using heuristics and approximation algorithms [2], [118], Statistical modeling, stochastic geometry, and sub-modular function optimization [117], Coverage segmentation and distributed Hungarian algorithms [119], Stochastic geometry and Coalition game [51]. |
| | [120] | Energy-efficient cloudlet placement technique for Mobile Cloud Computing | Cloud computing network | Dynamic offloading and decision making algorithm. |
| | [121] | Energy-aware link design for CRAN | CRAN | Algorithm based on Concave-convex procedure. |
| | [122] | Exploited network slicing and optimally allocated caching units. | 5G mobile network | Solved ILP optimization using chemical reaction optimization algorithm. |
| | [123] | Introduced moving vehicles as caching nodes to improve energy-efficiency. | vehicular CCN | Nonlinear fractional programming and Lyapunov optimization. |
| Energy efficient resource allocation. | [42], [46], [124] | Caching, computing, and communication resource allocation. | Wireless network [124], SDN [42], F-RAN [46]. | SDN-based programmable control mechanism [124]. Solved optimization problems by using alternating direction method of multipliers (ADMM) [42], Lagrangian relaxed method and dual decomposition [46]. |
| | [1], [27], [52] | Green resource allocation and optimization exploiting ICN features. | Content centric IoT [1], [27], 5G network [52] | Deep Q-Learning [1], [27], Shortest path tree algorithm [1], and Dynamic self-optimization [52]. |
| | [125], [126] | Caching resource allocation to achieve energy savings in a communication network. | F-RAN [125], HetNet [126], Content-centric wireless network [127] | Fractional programming, Greedy technique, and Adaptive transmission [125]. Many-to-many matching game and MILP optimization using KKT conditions [126]. MDP and generalized monotonicity [127]. |
| | [128]–[130] | Optimal transmission power allocation and user association. | HetNet [128], [129], Wireless ICN [130]. | Lagrange method and Karush-Kuhn-Tucker (KKT) conditions [128], One-to-one matching game, Gale Shapley algorithm [129], Context aware and rate adaptive scheme [130]. |
| | [131]–[133] | Optimal energy-aware allocation of spectrum and caching resources. | HetNet [131], Mobile network [132], MEC network [133] | Stochastic geometry and numerical analysis [131], Decoupling of optimization problem and utilization of block nested procedures [132], Solved optimization problem using Branch and Bound method and Benders decomposition [133]. |
| | [134]–[136] | Energy-aware resource allocation techniques for the D2D-assisted systems. | MEC system [134], Cellular network [135], HetNet [136] | Solved non-convex optimization by using problem decomposition, iterative algorithm, and game theoretic approach [134]. Dinkelbach method, Modified branch and bound method, Lagrange dual decomposition [135]. MDP, Q-Learning and deep Q-network algorithm [136]. |
| | [137] | Optimal transmission power allotment and D2D link activation. | Wireless network | Solved non-convex optimization by using problem decomposition and iterative algorithm |
| | [138], [139] | Energy-aware virtual resource allocation | HetNet [138], [139], | Solution of optimization problems using distributed algorithm based on alternating direction method of multipliers (ADMM). |

Zhang *et al.* [3] investigated optimal cache allocation and transmission scheduling of a social-aware D2D network. A deep reinforcement learning (DRL) method also analyzed C-RAN mode and D2D mode for user device communication in [150]. In this study, the controller intelligently decided the operating mode and

on-off state of the processors to minimize transmission power.

To maximize energy efficiency and spectral efficiency, a caching strategy [140] activated RRHs and connected users. The proposed approach also explored unicasting and multicasting process for content delivery subject to constrained transmission power and data rate. Moreover, in [151], a caching technique was designed for CCN to achieve substantial gain in energy savings without violating delay requirements. To utilize limited caching and computing resources, Yang *et al.* [152] minimized content-transmission latency subject to constrained energy consumption.

### B. GREEN CACHING TECHNIQUES FOR CONTENT PLACEMENT

The optimal allocation of popular content objects at the caching nodes significantly reduces energy costs [153]–[155]. Fang *et al.* [141] introduced a distributed caching strategy for content placement and replacement in a CCN. They utilized game theoretic approach and achieved substantial energy savings.

A content allocation, caching, and replacement technique utilized big data analytics to evaluate energy cost of a 5G-RAN [153]. For energy-efficient network operation, another strategy, involving content caching and replacement at the mobile edge, was developed in [156].

Xu *et al.* [154] designed a three-layer framework to reduce energy consumption and latency in a HetNet. The framework investigated both content allocation problem and cache replacement technique for tactile Internet services. Another energy-efficient caching strategy optimized content allocation and access in a content-centric 5G network by exploiting D2D communication and content popularity [155].

In [157], the authors developed a collaborative strategy to cache content at the SBSs and MBSs of a HetNet. Another collaborative approach [158] optimally cached content at the core network and BSs. Both these studies substantially reduced energy costs.

Dynamic content popularity also plays a crucial rule in content replacement and reducing energy consumption [75]. For dynamic content popularity, the authors in [142] optimized cache size and hit-rate. The study also investigated the tradeoffs between cache size and energy requirements. In addition, an online caching strategy considered dynamic content popularity for content placement and exploited DRL to improve energy efficiency [159]. Another reinforcement-learning (RL) technique proactively cached and replaced content at the BSs considering dynamic content inclusion and user behavior [160].

In [161], neural network and DRL technique were utilized to design energy-aware optimum content allocation and forwarding for a network with D2D caching capability. The authors in [162] optimized transmission zone for content delivery and minimized energy consumption. The proposed architecture utilized multicasting for optimal D2D communication and power allocation. For a content distribution network, a multi-layer caching technique with D2D communication and wireless backhaul was developed in [163].

Coding based techniques can also reduce the energy demand of a network. Gabry *et al.* [13] introduced a maximum-distance separable encoding technique and achieved further energy efficiency compared to conventional content fragmentation approach. In [164], scalable coding based approach was explored in designing an energy-efficient caching technique for VoD services in HetNets.

### C. ENERGY-AWARE CACHING RESOURCE MANAGEMENT

Intelligent management of caching resources is important for the improvement of improve network performance without exceding energy budget. In [43], the authors showed that a number of factors play key roles in achieving maximum energy efficiency for BS caching. They optimally designed cache capacity of a BS in a HetNet. The proposed model considered backhaul capacity, BS density, content popularity, power consumption of caching units, and interference level.

In a HetNet, caching popular content objects and serving identical content requests, which are generated within a small time difference, using multicast approach can significantly reduce energy demand [165], [166]. A randomized-rounding approach in [165] optimally allocated caching space and scheduled multicast transmission based on energy consumption and constrained delay. In contrast to the optimal content caching, Zhou *et. al.* [166] optimized content delivery using multicast approach. They proposed a stochastic model on the basis of MDP, and scheduled multicasting using a low-complexity heuristic.

Another study [167] jointly investigated uncoordinated placement and coded delivery techniques. Based on finite content size, the proposed model cached content at the mobile devices and achieved lower bound of the required content length exploiting optimal multicast transmissions. In addition, an iterative method was introduced for multicast scheduling technique [168]. The backhaul-aware model optimally allocated caching resources and minimized content fetching costs.

D2D caching enables BSs to offload traffic to the connected devices and reduce network congestion [169]. However, the mobile subscribers hesitate to participate in D2D caching because of limited battery capacity and incurred energy consumption [170], [171]. Therefore, in recent studies [169]–[172], incentive designs for the participating mobile users have been introduced.

In [169], a Stackelberg game-based incentive scheme rewarded the participating devices and minimized costs at the BSs. On the other hand, the goal of these participating devices was to achieve maximum utility. This conflicting approach was resolved by utilizing an iterative mechanism, and the BS achieved substantial cost reduction. Another dynamic strategy [170] exploited MDP and designed incentives not only for diverse network statuses but also for successful deliveries. A recent study [171] has utilized stochastic geometry along with a Stackelberg game to reach the equilibrium between

the network operators and participating users. The proposed scheme maximized profits of both entities and allocated caching resources and incentives using a low-complexity algorithm.

Wang *et al.* [172] developed an incentive scheme combining D2D caching, user cooperation, and edge resource utilization. The study considered both uncoded and coded caching policies for energy consumption minimization and the fulfillment of QoS requirements. On the other hand, a collaborative technique [173] exploited storage capacity and the participation willingness of connected devices to achieve energy savings. Yi *et. al.* [174] reduced signaling heartbeat overheads and energy demand using a D2D system. In this model, the voluntary devices gathered signals from nearby devices and transmitted the aggregated signal to the BSs. In a sensor network, Zafari *et al.* [175] optimally compressed data and allocated content. The study reduced energy consumption without degrading quality-of-information.

To support transcoded VoD services, Xie *et al.* [176] developed an energy-efficient caching technique and addressed user requests dynamically in a 5G-MEC system. Moreover, a DRL method [143] investigated adaption and transcoding of video streaming services in a cache-enabled mobile network. The proposed method utilized SDN and jointly optimized both QoE and energy. Zhu *et al.* [177] developed another DRL based policy for IoT content caching. They also investigated the trade-off between content freshness and associated content fetching cost.

Some of the existing studies [145], [178]–[181] exploited diverse networking features and reduced the energy costs of cache-equipped systems. In [178], the authors incorporated SDN and divided user plane and control plane in a HetNet. The study investigated the impacts of SBSs density on caching and evaluated energy efficiency, coverage probability, and throughput. In addition, a user-centric protocol was also designed to maximize traffic offloading gain and regulate power consumption based on content delivery distance of a cellular network [179]. For a hierarchical content delivery network (CDN), Bianco *et al.* [180] proposed an energy consumption model and illustrated the impact of network topology and server synchronization on energy consumption. An offline technique [181] optimized both caching and data rate to reduce the offline circuit power of a HetNet. For industrial IoT systems, Duan *et al.* [145] partitioned the storage space of the BSs for proactive and reactive caching. They achieved substantial energy savings without degrading QoS.

Cached content objects at the network nodes gradually become old, and the number of requests for these items decreases sharply [182]. Hence, content freshness plays a 1pivotal role in caching, and cache nodes require timely update to serve currently popular items. However, frequent upgrade causes additional overheads [183]. To accommodate the trade-off between data freshness and network performance, the authors in [182] switched cache status based on the load at the specific node. Banerjee and Biswash [183] considered the associated signal overheads and designed a mechanism to maintain the data freshness of an ICN. The proposed model determined specific lifetime for different content objects and cached content based on remaining lifetime.

IoT content objects are transient in nature and frequently updated [184]. In addition, storage capacity of IoT devices is limited, and IoT applications also have specific data freshness requirements. However, these freshness requirements lead to excessive energy demand. A recent study [185] has introduced Age of Information (AoI) as a representation of content freshness and updated caching nodes using MDP for the reduction of average cost. Moreover, to address the changing user behavior and content popularity, a DRL approach has been introduced for evolving caching resource management. Yu *et. al.* [186] developed a crowd-sourcing framework to collect photos from IoT and mobile devices for 3D model generation. They introduced a pricing method based on content freshness, size, and resolution and dynamically chose photos for the successful reconstruction of the model. The selection of photos was decided considering target coverage and associated cost.

An energy-efficient caching system [96] investigated the dynamic movement of mobile nodes. The proposed technique reallocated content to another node when the device was not within the coverage area of the old caching node. Tan *et al.* [187] also introduced a mobility-aware framework combining MEC and caching based on deep Q-learning. This technique not only optimally allocated network resources but also decreased overall system cost. Another study [188], developed a location prediction method for user requests and cached multimedia content at the network edge.

### D. TRADE-OFF ANALYSIS IN GREEN CACHING SYSTEMS

Although caching at the network edge achieves substantial energy savings, this technique can cause performance related concerns and security threats. Therefore, some studies investigated associated trade-offs.

In [146], the authors jointly investigated channel state information, backhaul load, and cache status for video transmission. The study improved physical layer security and minimized power consumption in a cellular network. Another strategy [189] introduced the concept of trusted caching nodes (TCNs), which stored IIoT contents and provided surveillance facility for the communication links. The proposed technique optimally placed TCNs and selected content delivery routes based on energy demands and probable security threats.

A packet update caching (PUC) technique studied energy-aware and secure operation of the IoT systems [26]. On the other hand, in [47], the authors exploited MEC and assigned security services to the offloaded IoT tasks. The optimal strategy analyzed trade-offs related to energy and security concerns of these latency-critical tasks.

Content caching and distribution in presence of eavesdroppers is investigated in [190], [191]. Zheng *et al.* [190] maximized both secrecy energy efficiency and throughput

considering content diversity and popularity. Xiang *et al.* [191] explored multiple-input-multiple-output content transport and scalable encoding of videos. They minimized associated transport energy and ensured QoS standard and secure transmission.

Li *et al.* [49] investigated energy-delay trade-off by exploiting cooperative caching and sleeping mechanism for SBSs. Furthermore, Luo *et al.* [50] studied the trade-off between delay and energy for cache-enabled MEC systems with constrained backhaul capacity.

Lee and Molisch [29] determined the cooperation distance for a cellular network with D2D cooperation. To design an optimal caching strategy, they studied the trade-offs between energy savings and network throughput. In [48], the trade-off between goodput and energy of a D2D-assisted HetNet was also investigated by using markovian model.

### E. GREEN TECHNIQUES COMBINING MOBILE EDGE COMPUTING AND CACHING

The proliferation of latency-sensitive and resource-constrained applications has heightened the requirements of storing popular data items and faster data processing. To address this demand, mobile cloud computing (MCC) emerged as a promising solution [192]. In MCC, a mobile device can offload requested tasks to the remote servers and achieve faster task execution [192], [193]. Although the introduction of cloud servers gained significant performance improvement, the operation of these cloud data centers causes excessive energy demand and high carbon footprint [120], [194]. Therefore, the network operators and service provides are investigating the prospect of renewable energy in powering up the cloud servers [194]. However, the intermittent nature of renewable energy makes the process complicated. In [194], the authors addressed diverse aspects and associated challenges of renewable energy powered cloud servers. For the cloud data centers, the study also suggested innovative techniques to address unexpected energy demand and grid-integrated renewable energy supply.

The design of energy-efficient cloud computing techniques has also achieved significant attention. A stochastic optimization based approach [193] introduced a transmission protocol for energy-aware communication between the cloud servers and user devices. The proposed strategy addressed the problems associated with unstable connectivity and fluctuating bandwidths of wireless systems. This adaptive framework addressed energy-delay trade-off and fetched popular content to confront poor connectivity conditions.

In MEC, edge clouds/servers are deployed at the network edge for the faster execution of computing tasks [2], [72]. MEC not only helps to overcome the limitations of user equipment, but also improves user satisfaction [195]. An edge cloud also provides an opportunity to cache popular content objects and reduce redundant content transmission [47], [72].

The joint approach combining caching and MEC categorizes computing tasks based on their popularity [47], [196]. First, the results of the most popular tasks are cached at the edge nodes, and user requests are addressed immediately. This technique does not require task computing for every request. Second, the tasks with a decent popularity are executed at the edge nodes. However, the edge clouds usually cache the input data items of these tasks. Third, the edge cloud caches neither the input data items nor the executed results of the tasks with little popularity. In this process, the tasks are executed at the edge on demand.

In [197], caching and computing resources were assigned to delay-critical offloaded tasks from the perspective of energy cost minimization. Tang *et al.* [144] also developed a model for the migration and caching of offloaded tasks to the edge servers. For a MEC system with constrained storage and computation resources, Hao *et al.* [198] proposed a task offloading and caching mechanism, which reduced energy consumption at the user devices.

Service-specific model for the energy consumption of communication networks was introduced in [199]. Additionally, Huang *et al.* [200] proposed a caching and routing mechanism for the service flows of an edge computing framework and reduced data transmission.

Although MEC reduces content transmission delay and energy requirements, the limited storage and computing capacity at the network edge calls for the smart utilization of networking resources [54], [119]. In addition, the changing user demand influences the requirements for MEC and caching services. The authors in [201] introduced an optimal approach for the selection of access network and the allocation of delay-sensitive services at the network edge. The study also addressed concerns over indirect user access, waiting queue, and dynamic switching of MEC applications. The proposed iterative algorithm achieved a near optimal solution for the long-term optimization problem and reduced overall delay.

In recent years, minority game (MG) has been introduced not only for distributed resource allocation but also for reducing associated signal overheads [202], [203]. Ranadheera *et. al.* [202] developed a MG-based strategy for the offloading of computation-intensive tasks and proposed a framework for mobile applications in future small cell networks. The objective of this strategy was to improve user satisfaction by ensuring optimum resource assignment and achieving reduced volatility. In [203], the authors also explored MG and reinforcement learning. To reduce the energy requirements of edge clouds, the study designed an optimal approach for the activation of these distributed servers.

Apostolopoulos *et. al.* [195] also activated MEC servers using both MG and distributed learning techniques. Instead of maximizing the QoS, the proposed technique focused on achieving QoS satisfaction. The proposed non-cooperative game obtained efficient-satisfaction-equilibrium and determined the transmission power for connected IoT devices. In a recent study [204], the theory of satisfaction games has been utilized for surveillance systems equipped with MEC capabilities. The novel technique also incorporated

fully autonomous aerial systems, and satisfied the service requirements of involved entities. According to the proposed framework, by achieving satisfaction equilibrium, connected entities fulfilled their individual QoS demand and reduced overall energy consumption.

In [205], an auction strategy was developed to address the emergency demand at the network edge. The online method introduced split incentives for distributed workload distribution based on market conditions. This dynamically adapted technique also switched off edge clouds to achieve energy cost reduction considering the acceptance level of QoS degradation tolerance.

The authors in [206] quantified the impact of different tasks and developed a machine-learning based framework for the allocation of the most important tasks at the edge servers.

In Table 3, the state-of-the-art green resource management and content placement techniques are briefly presented.

## VI. UTILIZATION OF RENEWABLE ENERGY IN CACHE-EQUIPPED NETWORKS

To address the ever-increasing energy consumption and carbon emission of communication networks, powering up network devices with green energy have become a prime consideration [14], [45], [209].

### A. CACHING RESOURCE ALLOCATION AND CONTENT DELIVERY USING RENEWABLE ENERGY

Several studies [11], [209]–[211] have proven that traffic-offloading at the green BSs not only improves energy efficiency but also ensures faster content delivery. In this regard, Zhou *et al.* [209] reduced the number of requests handled by the MBSs. This green technique proactively pushed and cached content based on the random arrival of green energy and user requests. In [11], a coded caching technique was proposed to offload multiple requests simultaneously from the MBSs to the green energy-powered SBSs.

A green framework, which incorporated caching and harvested energy sharing among the SBSs, optimally allocated network resources and minimized brown energy consumption [208]. In addition, a user association mechanism [40] reduced grid energy requirements for cache-equipped HetNets. In the HetNets, SBSs utilized renewable energy and optimally allocated network resources to the subscribers subject to limited backhaul and storage capacity.

Wu *et al.* [210] developed a coordinated approach for the energy harvesting SBSs in a HetNet. By utilizing context-aware caching and content delivery, these SBSs reduced the burden on the MBSs and improved network coverage probability. Another cooperative technique [211] jointly exploited intermittent green energy utilization and MEC. The proposed model optimized both backhaul traffic and bit rate for the dynamic adaptation of VoD services. Furthermore, the authors in [212] developed a relay selection strategy and optimized throughput of the links subject to harvested energy. This cache-aided technology not only transmitted wireless data, but also delivered power to the connected relays.

In [213], an energy-harvesting technology was utilized to charge the IoT sensors and popular items were cached at the connected gateways. The proposed model offered incentives for the harvesting technology providers and optimized both QoS and system utility. On the other hand, an optimal strategy [14] exploited green energy and combined content forwarding and caching to reduce the utilization of on-grid energy.

### B. ACTIVATION OF CACHING UNITS BASED ON RENEWABLE ENERGY SUPPLY

Network traffic demand shows substantial variation during different hours of a day [215], [216]. Therefore, the smart scheduling of network resources considering green energy supply has been proven as an effective strategy. This technique not only reduces energy requirements and operational costs, but also ensures the optimal utilization of green energy.

In [215], self-sustaining SBS with wireless backhaul was introduced as a roadside caching node to serve vehicular traffic requests. The study also modeled traffic-steering on the basis of temporal changes in network traffic and energy arrival. Another study developed a traffic steering model based on both temporal and spatial disparities in user requests and green energy availability [216]. The proposed technique proactively cached and pushed content for dynamic traffic adaptation.

The authors in [214] studied the trade-off between on-grid energy consumption and traffic delivery latency. For a HetNet with limited backhaul capacity, this study utilized hybrid energy supply and balanced the traffic load. For a D2D network, Chen *et al.* [217] introduced the concept of self-sustaining mobile helpers with caching capability. The mobile helpers served requests of the nearby users subject to battery status and achieved substantial reduction in energy consumption. A joint optimization technique [218] comprised both sleep-active scheduling and caching. This optimal approach activated BSs considering energy availability.

In [45], the authors introduced time as a management tool and determined the optimum time for content caching. The proposed model exploited solar energy to reduce not only energy costs but also peak to average disparity of on-grid energy consumption. Another strategy [24] scheduled and activated caching nodes based on user demand and intermittent solar energy supply. This green technique substantially reduced the dependence on grid energy and maintained QoE standards.

Table 4 summarizes the contributions and associated issues related to the renewable energy powered cache-equipped networks.

## VII. PERFORMANCE METRICS FOR GREEN CONTENT CACHING

This section briefly discusses major performance metrics utilized in green caching schemes. Content allocation and delivery strategies evaluate network performance and associated costs in terms of these metrics, which are utilized for caching

**TABLE 3. Summary of the literature review on optimal content placement and resource management for green content caching.**

| Key focus | Articles | Major contribution | Network type | Solution approach |
|---|---|---|---|---|
| Optimal scheduling of caching nodes. | [44], [147], [149], [207] | Energy-optimal and cooperative activation of the caching nodes. | CCN [149], HetNet [44], [147], UDN [207] | Solved optimization problems using heuristic approach [149], quantum-inspired evolutionary (QEA) algorithm [44], MDP and value iteration [147], and iterative method and KKT conditions [207], respectively. |
| | [148] | Energy-efficient switching of the CCN nodes and links. | Metro area network | Solved MILP optimization using centralized spanning tree algorithm and distributed dual decomposition. |
| | [3], [150] | Optimal transmission scheduling for D2D offloading and energy savings. | Cellular Network [3], F-RAN [150] | Stochastic MINLP optimization, Lyapunov drift-plus-penalty theory, and virtual queue model [3]. Markov decision process (MDP) and DRL [150]. |
| | [140] | RRH activation and user association with constrained power and transmission rate. | C-RAN | Solved MINLP optimization using iterative convex quadratic programming. |
| | [151], [152] | Delay-aware and energy-efficient strategies for caching systems. | CCN [151], MCC/MEC systems [152] | Solved optimization problems using Hopfield neural network [151] and McCornick envelops, binary variables relaxation, and ADMM [152], respectively. |
| Green caching techniques for content placement. | [141], [153]–[156] | Energy-efficient and distributed caching strategies for content allocation and replacement | CCN [141], 5G-RAN [153], Wireless network [156], HetNet [154], [155] | Solved non-cooperative game through the investigation of Nash equilibrium (NE) [141], Algorithm based on big data analysis [153]. Solved ILP optimization using Greedy heuristic [156]. Online algorithm for content allocation and replacement [154]. Knapsack problem mapping [155]. |
| | [157], [158] | Collaborative techniques for content placement and energy cost minimization. | HetNet [157], Mobile network [158] | Solved ILP optimization problem using Genetic algorithm [157] and QEA [158], respectively. |
| | [142], [159], [160] | The design of optimal caching strategies exploiting dynamic content popularity. | Cellular network [142], UDN [159], Wireless network [160] | Mathematical analysis and simulation [142], Reinforcement learning [159], [160]. |
| | [161]–[163] | Optimal content allocation for D2D caching. | D2D network | Recurrent neural network (RNN) and DRL [161]. Lagrange multipliers method and Mode selection algorithm [162]. Solved optimization problem using water-filling based power allocation and zero-forcing beam forming [163]. |
| | [13], [164] | Coding based content allocation at the caching nodes to reduce energy consumption. | HetNet | Maximum-distance separable (MDS) coding, Low-complexity solution tools for convex optimization [13], Logarithmic approximations, Standard gradient projection [164]. |
| Energy-aware caching resource management. | [43] | Optimal backhaul-aware caching for energy cost minimization. | HetNet [43] | Poisson process and Zero-forcing beamforming [43]. |
| | [165]–[168] | Multicast scheduling for cached content delivery. | HetNet [165], [166], Wireless network [167], [168] | Randomized-rounding technique [165]. Infinite horizon average cost MDP [166], [168] and Low-complexity value iterative method [168]. Random placement and clique cover delivery [167]. |
| | [169]–[173] | Incentive and reward based D2D caching resource management. | Wireless Network [169], [170], Cellular network [171], [173], Mobile edge network [172], | Stackelberg game and iterative gradient method [169]. MDP [170]. Stochastic geometry and Stackelberg game [171]. Gradient Projection Method and Contract theory [172]. Willingness-aware D2D cache modeling and validation through simulation [173]. |
| | [174] | The minimization of signaling overheads and energy consumption. | D2D network | D2D peer matching and message scheduling. |
| | [175], [176] | Optimal data compression, video transcoding, and content allocation for energy savings. | Wireless sensor network [175], 5G network [176] | Solved optimization problem using simulated annealing algorithm [176] and spatial brunch-and-bound method [175]. |

Continued on next page

**TABLE 3.** *(Continued.)* **Summary of the literature review on optimal content placement and resource management for green content caching.**

| Continued from previous page | | | | |
|---|---|---|---|---|
| Key focus | Articles | Major contribution | Network type | Solution approach |
| Energy-aware caching resource management. | [143], [177] | DRL based caching resource management, | Mobile network [143], IoT network [177] | MDP, Lyapunov technique, and DRL [143], [177],. |
| | [178], [179] | User centric designs for caching networks. | HetNet [178], Cellular network [179] | Stochastic process and mathematical modeling [178]. Laplace Transform, Probability Theory, and KKT conditions for convex optimization [179]. |
| | [180] | The impact of network topology and server synchronization on energy consumption. | CDN | Mathematical modeling and validation through simulation. |
| | [181] | Offline circuit power estimation of a cache-equipped network. | HetNet | Solved non-convex optimization problem using KKT conditions and String Tautening Algorithm. |
| | [145] | Cache space partitioning to support both proactive and reactive caching. | HetNet | M/M/1 queueing theory and Gold selection algorithm. |
| | [182], [183] | Optimized data freshness considering the impact on network performance | ICN [182], [183] | Freshness-aware algorithm [182], [183]. |
| | [184]–[186] | Content freshness-aware transient IoT data caching | ICN-IoT [184], edge-IoT network [185], [186], | Freshness-aware algorithm exploiting traffic prediction and user location [184]. MDP and DRL [185]. Solved NP-hard optimization problem using greedy approximation [186]. |
| | [96], [187], [188], | Mobility-aware dynamically adapted content caching. | Mobile ad-hoc network [96], Vehicular network [187] | Dynamic weight adaptation and centrality based cache node selection [96]. MDP and deep Q learning [187]. Location prediction using historical traces, content prefetching, and optimized cache replacement algorithm [188]. |
| Trade-off analysis in green caching systems. | [26], [47], [146], [189]–[191], | Jointly investigated security and energy aspects of caching. | ICN-IoT network [26], [189], Cellular Network [146], Mobile edge network [47], HetNet [190], [191] | Cluster design, Circular buffer, Data Purge, Incapsula process [26]. Solved optimization problem using relaxation technique and greedy iterative algorithm [146], problem decomposition and heuristic algorithms [47], [189], zero-crossing of the derivative terms [190], Concave and quasi-concave functions [190], and Benders decomposition [191], respectively. |
| | [49], [50] | Energy-delay trade-off investigation for caching systems. | UDN [49], 5G network [50] | M/G/1 queueing and Lyapunov optimization [49]. Solved optimization problem using genetic algorithm [50]. |
| | [29], [48] | Investigated D2D cooperation and trade-off analysis between energy and throughput. | HetNet | Solved optimization problems using concave and quasi-concave programs [29]. Continuous time Markov chain [48]. |
| Green techniques combining MEC and Caching. | [193], [194] | The inclusion of cloud servers for resource-hungry applications. | Mobile-cloud network | Lyapunov optimization and stochastic control framework [193]. Mathematical modeling and optimization [194]. |
| | [144], [197]–[200] | Energy-efficient task offloading to the network edge | Mobile edge network [144], [197], [198], [200] , HetNet [199] | Solved optimization problem using genetic algorithm [144], KKT conditions [200], and String Tautening Algorithm [200], problem decomposition and iterative mechanism [197], branch and bound method [198]. Mathematical modeling and simulation [199]. |
| | [201]–[203] | Distribued resource allocation in MEC servers. | Mobile-edge network [201], [203], 5G network [202] | Solved long-term optimization converting into multiple one-shot problem and iterative algorithm [201]. MG and distributed decision making [202]. MG and RL [203]. |
| | [195], [204] | QoS satisfaction-aware MEC. | IoT network [195], Fully Autonomous Aerial System and MEC network [204]. | MG and distributed learning [195]. Distributed learning and RL [204]. |

**TABLE 4.** Summary of the literature review on utilization of renewable energy in cache-equipped communication networks.

| Key focus | Articles | Major contribution | Network type | Solution approach |
|---|---|---|---|---|
| Caching resource allocation and content delivery using renewable energy. | [14] | Used renewable energy for caching and content forwarding | Computer network | Solved MILP optimization by using relaxation and rounding, Sequential fixing, and Gradient-based algorithm. |
| | [40], [208] | Developed caching resource allocation schemes exploiting harvested energy | Wireless network [208], HetNet [40] | Solved MILP optimization by using decoupling and Water-filling method [208], solved weakly acyclic game by proving two-dimensional NE [40]. |
| | [11], [209], [210] | Content caching based traffic offloading from the MBSs. | HetNet | Optimal content transmission mechanism [209], MDP and Iterative algorithm [11], and Stochastic geometry, Laplace function, Gamma approximation [210]. |
| | [211] | Cooperative caching scheme utilizing renewable energy and MEC. | Mobile edge network | Solved optimization problem using self-tuning parametrization. |
| | [212] | Optimized relay selection technique for the cached content and power transfer. | Wireless network | Solved optimization problem by applying KKT conditions and Lambert function. |
| | [213] | Caching incentive scheme for harvested energy providers. | IoT network | Stackelberg game formulation. |
| Activation of caching units considering renewable energy supply. | [214]–[216] | Caching strategies for traffic steering and network load balancing. | HetNet | Solved optimization problem by using greedy technique [215], [216] and MDP [216], M/M/1 queuing and heuristic algorithms [214]. |
| | [217] | Introduced self-sustaining mobile helpers for caching subject to battery status. | D2D Mobile network | Mathematical modeling, numerical analysis and validation through simulation. |
| | [24], [45], [218] | Sleep-active scheduling of caching nodes using green energy. | HetNet [218], CCN [24], [45] | Solved optimaztion problem using ILP solver,network simulation [24], [45], and heuristic algorithm [24], [218]. |

decision making. In this study, we categorize the performance metrics as energy-related metrics, caching-related metrics, network-related metrics, and other metrics. These metrics are described as follows:

### A. ENERGY-RELATED METRICS

#### 1) ENERGY CONSUMPTION

In green caching systems, energy consumption is the most widely utilized performance metric. It represents the energy consumed by the whole system. In the existing studies [26], [49]–[51], [118], [119], [144], [145], [153]–[158], [161], [173], [176], [180], the authors utilized energy consumption to compare the performance gain of the proposed caching model with other systems. Wireless networks utilize a similar metric named area power consumption [115], [139]. This metric calculates the average power consumption of a network in per unit area. On the other hand, for caching strategies, energy savings are also evaluated as an indicator for performance improvement [36], [44], [114], [219].

#### 2) ON-GRID ENERGY CONSUMPTION

The on-grid energy consumption of a communication network denotes the energy demand from the power grid. Caching techniques utilizing renewable energy sources mostly consider this metric [24], [45], [215], [216]. The reduction in on-grid energy consumption portrays alleviated electricity load, which helps to reduce the operational expenditure of the network. This term is also referred as brown energy consumption [14].

#### 3) ENERGY EFFICIENCY

Energy efficiency is recognized as a prime indicator of 5G networking systems [40], [48], [121], [131], [142], [154], [159]. It denotes the achieved data rate over per unit energy consumption. Caching nodes serve more requests from the network edge and eliminate energy components related to backhaul traffic. Furthermore, network traffic-aware caching node activation also improves energy efficiency [148], [163]. This is because in peak hours the caching nodes utilize maximum content transmission capacity and in off-peak hours they remain switched off [24].

#### 4) ECONOMICAL ENERGY EFFICIENCY

Yan *et al.* [220] introduced the idea of economical energy efficiency. It jointly considered system costs, energy efficiency, and spectral efficiency. This parameter has also been utilized for not only energy-aware network management, but also network planning and deployment [125].

#### 5) BS TRANSMIT POWER

BS transmit power represents the minimum power requirement to maintain the QoE and security standards in caching

systems [146], [162], [191]. Degradation in QoE and security has catastrophic consequences. Therefore, the operators can not allow a lower BS transmit power compared to the threshold value.

### B. CACHING-RELATED METRICS

#### 1) CACHE HIT RATIO

Cache hit ratio (CHR) is calculated as a ratio of the number of requests served by the caching node over the total number of content requests. To measure the effectiveness of a caching system, a number of studies [36], [40], [153], [154], [161], [173], [217]. utilized this metric. One of the major goals of an efficient caching model is to achieve a higher CHR. When CHR is higher, more requests are served locally, which reduces both content delivery time and distance. Therefore, energy demand is reduced, and backhaul pressure is alleviated [52], [187].

#### 2) HOP COUNTS

Hop counts are calculated as the number of intermediate nodes in a content delivery path [36], [151], [180]. The fundamental concept of caching is to reduce content delivery path, which eventually reduces hop content. This reduction is also represented as a hop count ratio or a hop reduction ratio [130]. In these metrics, the hop count in a cache-equipped system is compared with the hop counts in absence of caching. Number of hop counts is directly proportional to the transmission energy consumption [113]. Therefore, this metric has been widely utilized in green caching.

#### 3) COOPERATION DISTANCE

Cooperation distance is mainly utilized in D2D caching [29], [137]. It represents the maximum allowable distance between caching device and user device for successful content transmission. Although a longer distance supports better cooperation, it has a negative impact on energy consumption and frequency reuse. Therefore, some of the green caching strategies optimized cooperation distance subject to energy consumption [29], [137].

#### 4) AVERAGE FRESHNESS

Data freshness expresses whether the data is up-to-date or not [177]. Although frequently updated caching nodes store recent data items, this process has higher communication cost. On the other hand, fresh contents are crucial for caching systems utilized in the industrial and health sectors. Average freshness helps to adjust the balance between these two conflicting conditions.

#### 5) SUCCESSFUL TRANSMISSION PROBABILITY

This metric defines the probability that a cached content will be successfully delivered from the caching node to the designated subscriber [162], [164]. Successful transmission probability significantly contributes in determining the acceptable backhaul load and achieving expected QoE. In this purpose, few studies [114], [187] have also utilized the term delivery success rate.

### C. NETWORK-RELATED METRICS

#### 1) DELAY

In a cache-equipped network, delay represents the total time required in sending the content request to the serving node and delivering the content [2], [46], [117], [152], [207], [219]. This term is also referred as latency [133], [154].

The prime objective of caching is faster content delivery. Therefore, content delivery delay has been considered as a key criteria in a number of green caching strategies, including caching node selection [119], resource allocation [124], [134], and network management [168], [173]. Furthermore, diverse latency-critical applications have specific delay requirements [47], [70], [197]. The utilization of energy-efficient caching approaches in these applications has substantial impact on the associated delay. As a result, energy and delay have been jointly investigated in existing literature [49], [50], [151].

#### 2) NETWORK THROUGHPUT

On demand streaming of the video files and improved QoE are key requirements for future generation networks [35], [75], [89]. Therefore, throughput is crucial for effective network design. Throughput defines the amount of network data transmitted per unit time [77], [118], [154], [199]. For optimal caching system design, average throughput achieved by every subscriber is investigated along with energy and costs of a network.

#### 3) SPECTRAL EFFICIENCY

Spectral efficiency (SE) is the amount of data rate that can be achieved utilizing per unit bandwidth [115], [121], [139], [140]. A higher SE is achieved when more users are served by the network. Caching improves SE by serving more users locally and alleviating backhaul traffic.

#### 4) COVERAGE PROBABILITY

Coverage probability is a key metric from the user perspective. It represents the probability that the signal-to-interference-plus-noise-ratio (SINR) will remain higher compared to a benchmark value [178], [210], [217]. User experience vastly depends on coverage probability because it also represents the probability of obtaining satisfactory service. Outage probability is reciprocal to the coverage probability and denotes the probability of achieving SINR level below threshold [128], [131].

#### 5) NETWORK TRAFFIC

Network traffic stands for the cumulative traffic flow through all the links of a network. A well-designed caching strategy substantially reduces traffic flow and reduces energy demands [11], [96], [209]. Backhaul traffic is also evaluated

to compare the network load and performance gain of the systems [211].

### D. OTHER METRICS

#### 1) COST

The costs related to caching techniques represent both capital expenditure (CapEx) and operational expenditure (OpEx) for cache-equipped networks [1], [50], [134], [197]. Caching and networking-resource deployment mainly contributes to CapEx. On the other hand, energy cost, licensed spectrum fees, and content subscription charges constitute OpEx.

#### 2) REVENUE

Revenue denotes the earnings of a network operator or service provider [122], [138]. The main source of revenue is the charges paid by the clients to avail network services (e.g., Netflix, Industrial automation). Some of the service providers earn from the advertisements in the user interface of different applications (e.g., Youtube).

#### 3) UTILITY/REWARD

A number of green caching strategies [40], [138], [172], [213] have utilized the term utility, which mainly refers to the difference between performance gain and energy cost. In the literature, a similar term have also been defined as reward [27], [143]. Game theoretic approaches find an equilibrium for the caching systems when every entity tries to achieve maximum utility.

#### 4) OFFLOADING EFFICIENCY

Caching offloads backhaul traffic to the network edge [25], [79]. Offloading efficiency is the percentage of the overall traffic that has been offloaded to the caching node [133], [179]. Higher offloading efficiency helps to overcome backhaul congestion and resource scarcity.

#### 5) EXECUTION/RUN TIME

Execution time is usually investigated for task caching [144], [175], [197]. A joint strategy incorporating caching and MEC not only caches data but also executes offloaded tasks at the edge nodes. The required time span for the execution of these offloaded tasks is known as execution time. This metric mainly depends on the task complexity and computation capacity of the server [221].

#### 6) SECURITY BREACH COST

In recent studies [47], [189], security breach cost has been introduced to measure the probable loss for caching data leakage. This metric has an important role in investigating the trade-off between energy and security.

Table 5 summarizes the performance metrics utilized in different green caching strategies.

## VIII. CACHING APPLICATIONS

The low-latency and high-data rate applications will constitute a significant portion of the future network traffic [1], [133], [154]. Caching has been widely accepted as a promising technology to overcome the limitations of existing host based networks [11], [105]. This section discusses the major applications where caching plays an instrumental role in achieving maximum system output.

### A. VIDEO STREAMING SERVICES

To ensure standard QoE, video streaming services aim to achieve low latency and high data [30], [73], [222]. For these services, edge caching has shown significant potential in maintaining expected QoE. In addition, to maximize network performance, proactive caching strategies have predicted user behavior and have optimized caching resource allocation [45], [148]. On the other hand, a number of caching techniques [26], [146], [191] have been introduced to tackle the energy and security related concerns of the video streaming services.

### B. VIRTUAL AND AUGMENTED REALITY APPLICATIONS

Virtual and augmented reality applications are anticipated to have a life-changing influence on the network subscribers [46], [223]. These applications gather data related to users' activity and require substantial computational and storage capacity. Moreover, latency-critical and context-aware data transmission are also primary concerns for these applications [224]. As a solution to these technical challenges, edge caching provides an opportunity to store processed data and transmit required content immediately.

### C. VEHICULAR COMMUNICATION

Recent advancements in connected and autonomous vehicles have been instrumental for the unprecedented network traffic growth [123], [215]. Although cloud servers were utilized for vehicular data processing, the system performance was not satisfactory because of the long content transmission path and associated costs. Mobility issues and the ever-increasing number of smart vehicles worsen the situation. To overcome these technical challenges and support latency-critical applications, road side units have been introduced [187], [225]. These units with caching and computing resources not only gather data from hundreds of vehicles and connected sensors, but also analyze them for vehicular traffic control, smart parking, and safety services.

### D. INDUSTRIAL IoT

Industry 4.0 services are expected to be vastly dependent on industrial IoT applications and ubiquitous connectivity [119], [189]. The IoT data for these applications require efficient acquisition process, intelligent management and processing, encrypted transmission, and energy-aware operation [145], [189].

**TABLE 5. Summary of the performance metrics utilized in green caching techniques.**

| Category | Performance Metrics | Relevant Articles |
|---|---|---|
| Energy related metrics. | Energy consumption | [13], [26], [47], [49]–[51], [118]–[120], [126], [127], [130], [132], [134], [136], [143]–[145], [147], [149]–[158], [160], [161], [173], [175], [176], [180], [181], [189], [197]–[200], [207], [211], [218]. |
| | Area power consumption | [115], [139] |
| | Energy savings | [2], [36], [42], [44], [114], [116], [117], [124], [141], [219] |
| | On-grid energy consumption | [14], [24], [45], [208], [215], [216] |
| | Energy efficiency | [29], [40], [43], [48], [121], [123], [129], [131], [135], [140], [142], [148], [154], [159], [163], [164], [178], [190] |
| | Economical energy efficiency | [125] |
| | BS transmit power | [146], [162], [191] |
| Caching related metrics. | Cache hit ratio | [24], [36], [40], [45], [52], [96], [123], [136], [153], [154], [161], [173], [177], [187], [217], [218] |
| | Hop counts | [36], [113], [148], [149], [151], [180]. |
| | Hop reduction ratio | [130]. |
| | Cooperation distance | [29], [137] |
| | Average freshness | [177] |
| | Successful transmission probability | [162], [164] |
| | Delivery success rate | [114], [187] |
| Network related metrics. | Delay | [2], [24], [46], [47], [49], [50], [117], [119], [124], [133], [134], [152], [154], [168], [173], [207], [219] |
| | Network throughput | [29], [43], [116], [119], [131], [135], [137], [142], [178], [190], [215] |
| | Spectral efficiency | [115], [121], [139], [140] |
| | Coverage probability | [178], [210], [217] |
| | Outage probability | [128], [131] |
| | Network traffic | [11], [96], [156], [209], [211] |
| Other metrics | Cost | [1], [3], [50], [134], [168], [177], [179], [187] |
| | Revenue | [122], [138] |
| | Utility / Reward | [27], [40], [138], [143], [172], [213] |
| | Offloading efficiency | [133], [179] |
| | Execution/ Run time | [144], [175], [197] |
| | Security breach cost | [47], [189] |

In edge caching techniques, edge servers store the frequently accessed IoT data and transmit them upon user requests [177]. However, unauthorized access at the edge nodes can disrupt overall industrial operation. Therefore, the adoption of stringent security protections and robust controlling mechanisms are also critical for edge nodes of an industrial IoT network. Furthermore, an edge node can also serve as a gateway for low-latency and reliable connectivity of the heterogeneous IoT entities.

### *E. HEALTHCARE APPLICATIONS*

The proliferation in wearable devices and wireless sensor communications have introduced diverse smart healthcare services [226], [227]. The wearable devices are capable of monitoring physiological signals, which helps to assess patient's condition. Moreover, the emerging solutions are anticipated to support mobile and real-time monitoring of cricitcal patients. These solutions require high data rate, dedicated channel bandwidth, and faster data acquisition and processing [228]. Therefore, enormous data will be generated to ensure smooth and remote services irrespective of patient's address.

If these applications utilize caching, health data can be stored at the edge nodes and redundant transmissions can be avoided. In addition, caching also helps to improve service availability, reduce communication overheads, and alleviate network load.

### *F. SMART CITY APPLICATIONS*

To fulfill the diverse requirements of modern lifestyle, smart city has emerged as a key solution [219], [229]. The smart city applications include smart energy utilization, supply chain

management, eco-friendly waste disposal technique, online education services, e-banking, and so on [229]. These applications are expected to transfer latency-sensitive and safety-concern information to thousands of connected devices [230]. Cache-equipped systems can be exploited to address these demands and reduce both traffic congestion and access delay [219]. Moreover, caching can also address mobility, scalability, and connectivity related concerns of a smart city.

### G. OTHER APPLICATIONS

In addition, caching has been utilized in diverse real-life applications. Smart home system has designed a centralized framework for the operation of entertainment, lighting, air conditioning, and security services in an apartment [23]. Furthermore, intelligent transportation systems [231] and smart grid operations [232] have exploited caching for optimal system management.

## IX. FUTURE RESEARCH CHALLENGES

Although recent studies [13], [14], [117] have explicitly investigated diverse caching strategies and have reduced energy consumption, a number of open issues still remain to be addressed.

### A. SYNERGY BETWEEN CACHING AND MEC FOR IoT APPLICATIONS

Edge servers provide both computation and caching resources for traffic offloading and content forwarding [133], [198]. In future, these edge nodes will disseminate information not only to the cellular phones but also to hundreds of connected sensors and smart devices, which will form IoT network [47], [72]. Different applications, running on these devices, require a considerable share of both data storage and computing resources to serve user demands. Some of these applications require very precised data and latency standards. In addition, content freshness requirements ask for frequent cache update. Dynamic approaches involving deep reinforcement learning and markov decision process can help in designing optimal distribution and management policy for the edge servers.

Although some of the existing studies [47], [119], [138] have studies this problem, still the designing of low complexity resource allocation models combining both MEC and caching for IoT systems require further investigations.

### B. ADAPTIVE AND INTELLIGENT MANAGEMENT OF CACHING NODES

Next generation communication networks are expected to handle thousands of applications and user requests [64], [68]. Cache-equipped network access nodes will store information, which vary in terms of volume, value, data type, and complexity. The intelligent strategies are expected to predict content popularity, assess social relationship, and user mobility [5], [73], [84], [233]. Therefore, edge nodes should cache and update popular contents for faster content delivery. In future, caching resource management techniques should incorporate dynamic and intelligent techniques, which will automatically address individual requirements of diverse services.

### C. GREEN APPROACHES FOR SECURE D2D CACHING

Content popularity [75] and mobility prediction [123], [234] are indispensable parts of caching strategies. Sometimes these lead to the acquisition of confidential and sensitive personal data. Furthermore, D2D caching is also prone to unauthorized access to the devices and content breach [22], [235]. Therefore, robust security protection is essential for the caching nodes. However, energy requirements and transmission latency also increase with the strength of the security services [47]. Hence, future efforts should investigate green content caching strategies while ensuring expected security protections and QoE.

### D. GREEN CACHING FOR 5G APPLICATIONS

5G networks aim to support real-time systems and low-latency services, including industrial automation, tactile Internet, vehicular traffic management, and virtual reality applications [17], [21], [70]. Existing host-based networks are not suitable for the fulfilment of these requirements. ICN networks with caching capability and heterogeneous interfaces are anticipated to meet the conditions of low-latency and reduced bandwidth utilization [59], [67]. Furthermore, the edge nodes can provide flexibility to the industrial operation and control process subject to guaranteed security [68], [69].

In future, adaptive bandwidth allocation and social-relationship aware content transmission can support the latency-critical 5G applications. To mitigate the challenge of excessive energy requirements, dynamic network link formation is expected to allocate the under-utilized resources to the active users and serve the require content. This method will not only substantially reduce delay and energy consumption, but also ensure optimal resource utilization.

### E. INTEGRATION OF CACHING AND OTHER RECENT TECHNOLOGIES FOR ENERGY-EFFICIENT SYSTEM DESIGN

In recent years, a number of green approaches have been introduced for next generation communication networks [4], [54]. Among these approaches, caching has attained remarkable attention of the research community [8], [34]. In an integrated system, the utilization of caching and other modern approaches can achieve further energy savings and improved network performance. Some of these potential green solutions are as follows.

#### 1) SOFTWARE-DEFINED NETWORKING (SDN)

SDN introduces a logical and centralized controlling mechanism for communication networks [124], [229]. In this process, the control plane and the user plane are operated in a separate manner. Exploiting centralized administration, SDN provides an opportunity for dynamic adaptation of the changing network attributes [236].

The integration of SDN along with caching can significantly contribute to the optimal planning and management of future networks. Moreover, the cumulative contribution of these technologies can centrally control and monitor content distribution at the network edges. Therefore, further improvement can be achieved in energy savings, content transmission, and content diversity.

#### 2) NETWORK FUNCTION VIRTUALIZATION (NFV)

NFV technique decouples different networking services from conventional hardware and utilize virtual machines to provide same service using software [224], [237]. A server can operate multiple virtual machines and ensure optimal utilization of resources. As a result, the deployment and configuration process of these services become not only less time consuming but also efficient and cost-effective [237]. Even the changes in user demands can be addressed instantly.

NFV and caching can be jointly utilized for the activation of edge nodes considering user traffic demand. Therefore, the system will be able to operate network resources optimally and reduce carbon footprint of the network.

#### 3) mmWave COMMUNICATION

For data transfer in future networks, mmWave communication is envisaged as a key technology [238]–[240]. This is because the originated narrow beam provides better connectivity between the access nodes and the subscribers. However, these connections have very small connectivity duration, which cause repeated handoffs for moving subscribers. Furthermore, mmWave communication has high sensitivity [239], which often leads to pathloss because of the random changes in the communication channel.

To overcome these shortcomings, optimal and energy-efficient content allocation method can be introduced for cache-equipped mmWave communication [238]. Qiao *et al.* [241] have already utilized caching at the small cells to serve the handoff-subscribers using mmWave spectrum. Future studies should address the energy consumption issues and reduce content redundancy in such systems.

## X. CONCLUSION

This article presented a comprehensive survey and literature review on the state-of-the-art research efforts related to green content caching. Initially, we categorically analyzed existing survey articles on caching and green communications. On the basis of the analysis, we identified the research gaps, and discussed the importance of a survey article on the energy aspects of existing caching technologies. Caching-related major considerations, including caching storage allocation, popular content selection, and caching policy design were also explicitly explained.

The strategies for green content caching were subdivided based on resource allocation, resource management, and renewable energy utilization. A detailed review on these subdivisions highlighted major contributions, associated network types, performance metrics, and solution methods of individual studies. This article also discussed the advancements and utilization of caching technologies in next generation communication networks. However, green caching solutions still confront different challenges while maintaining the balance between network performance and energy consumption. These challenges and future research directions were also precisely identified for optimal caching strategy design.

## REFERENCES


[1] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, and S. Guo, "Green resource allocation based on deep reinforcement learning in content-centric IoT," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 3, pp. 781–796, Jul. 2020, doi: 10.1109/TETC.2018.2805718.

[2] D. Ren, X. Gui, K. Zhang, and J. Wu, "Hybrid collaborative caching in mobile edge networks: An analytical approach," *Comput. Netw.*, vol. 158, pp. 1–16, Jul. 2019.

[3] X. Zhang, P. Huang, L. Guo, and Y. Fang, "Social-aware energy-efficient data offloading with strong stability," *IEEE/ACM Trans. Netw.*, vol. 27, no. 4, pp. 1515–1528, Aug. 2019.

[4] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.

[5] D. Wu, B. Liu, Q. Yang, and R. Wang, "Social-aware cooperative caching mechanism in mobile social networks," *J. Netw. Comput. Appl.*, vol. 149, Jan. 2020, Art. no. 102457.

[6] C. Fang, F. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of energy-efficient caching in information-centric networking," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 122–129, Nov. 2014.

[7] M. Ismail, W. Zhuang, E. Serpedin, and K. Qaraqe, "A survey on green mobile networking: From the perspectives of network operators and mobile users," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1535–1556, 3rd Quart., 2015.

[8] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[9] W.-C. Chien, H.-Y. Weng, and C.-F. Lai, "Q-learning based collaborative cache allocation in mobile edge computing," *Future Gener. Comput. Syst.*, vol. 102, pp. 603–610, Jan. 2020.

[10] M. Bilal and S.-G. Kang, "A cache management scheme for efficient content eviction and replication in cache networks," *IEEE Access*, vol. 5, pp. 1692–1701, 2017.

[11] T. Li, M. Ashraphijuo, X. Wang, and P. Fan, "Traffic off-loading with energy-harvesting small cells and coded content caching," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 906–917, Feb. 2017.

[12] S. Zhang, X. Cai, W. Zhou, and Y. Wang, "Green 5G enabling technologies: An overview," *IET Commun.*, vol. 13, no. 2, pp. 135–143, Jan. 2019.

[13] F. Gabry, V. Bioglio, and I. Land, "On energy-efficient edge caching in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3288–3298, Dec. 2016.

[14] A. Khreishah, H. Bany Salameh, I. Khalil, and A. Gharaibeh, "Renewable energy-aware joint caching and routing for green communication networks," *IEEE Syst. J.*, vol. 12, no. 1, pp. 768–777, Mar. 2018.

[15] L. Budzisz, F. Ganji, G. Rizzo, M. Ajmone Marsan, M. Meo, Y. Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, M. Pickavet, A. Conte, I. Haratcherev, and A. Wolisz, "Dynamic resource provisioning for energy efficiency in wireless access networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2259–2285, 4th Quart., 2014.

[16] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. Ul Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 686–705, 1st Quart., 2016.

[17] I. B. Sofi and A. Gupta, "A survey on energy efficient 5G green network with a planned multi-tier architecture," *J. Netw. Comput. Appl.*, vol. 118, pp. 1–28, Sep. 2018.

[18] M. M. Mowla, I. Ahmad, D. Habibi, and Q. V. Phung, "Energy efficient backhauling for 5G small cell networks," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 3, pp. 279–292, Jul. 2019.

[19] H. Al Haj Hassan, A. Pelov, and L. Nuaymi, "Integrating cellular networks, smart grid, and renewable energy: Analysis, architecture, and challenges," *IEEE Access*, vol. 3, pp. 2755–2770, 2015.
[20] M. M. Mowla, I. Ahmad, D. Habibi, and Q. V. Phung, "A green communication model for 5G systems," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 264–280, Sep. 2017.
[21] A. Abrol and R. K. Jha, "Power optimization in 5G networks: A step towards GrEEn communication," *IEEE Access*, vol. 4, pp. 1355–1374, 2016.
[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, 2017.
[23] K. Xu, Y. Wan, and G. Xue, "Powering smart homes with information-centric networking," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 40–46, Jun. 2019.
[24] M. I. A. Zahed, I. Ahmad, D. Habibi, Q. V. Phung, and L. Zhang, "A cooperative green content caching technique for next generation communication networks," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 1, pp. 375–388, Mar. 2020.
[25] M. I. A. Zahed, I. Ahmad, D. Habibi, Q. V. Phung, L. Zhang, and A. Mathew, "Security aware content caching for next generation communication networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.
[26] I. U. Din, S. Hassan, A. Almogren, F. Ayub, and M. Guizani, "PUC: Packet update caching for energy efficient IoT-based information-centric networking," *Future Gener. Comput. Syst.*, vol. 111, pp. 634–643, Oct. 2020.
[27] F. Xu, F. Yang, S. Bao, and C. Zhao, "DQN inspired joint computing and caching resource allocation approach for software defined information-centric Internet of Things network," *IEEE Access*, vol. 7, pp. 61987–61996, 2019.
[28] Statista. *Internet Things (IoT) Connected Devices Installed Base Worldwide From 2015 to 2025 (in Billions)*. Accessed: Jun. 3, 2020. [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
[29] M.-C. Lee and A. F. Molisch, "Caching policy and cooperation distance design for base station-assisted wireless D2D caching networks: Throughput and energy efficiency optimization and trade-off," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7500–7514, Nov. 2018.
[30] E. Baccour, A. Erbad, K. Bilal, A. Mohamed, and M. Guizani, "PCCP: Proactive video chunks caching and processing in edge networks," *Future Gener. Comput. Syst.*, vol. 105, pp. 44–60, Apr. 2020.
[31] W. Chu, M. Dehghan, J. C. S. Lui, D. Towsley, and Z.-L. Zhang, "Joint cache resource allocation and request routing for in-network caching services," *Comput. Netw.*, vol. 131, pp. 1–14, Feb. 2018.
[32] E. Bastug, M. Bennis, E. Zeydan, M. A. Kader, I. A. Karatepe, A. S. Er, and M. Debbah, "Big data meets telcos: A proactive caching perspective," *J. Commun. Netw.*, vol. 17, no. 6, pp. 549–557, Dec. 2015.
[33] G. Sun, D. Ayepah-Mensah, L. Lu, W. Jiang, and G. Liu, "Delay-aware content distribution via cell clustering and content placement for multiple tenants," *J. Netw. Comput. Appl.*, vol. 137, pp. 112–126, Jul. 2019.
[34] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1710–1732, 3rd Quart., 2018.
[35] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.
[36] Y. An and X. Luo, "An in-network caching scheme based on energy efficiency for content-centric networks," *IEEE Access*, vol. 6, pp. 20184–20194, 2018.
[37] A. Ioannou and S. Weber, "A survey of caching policies and forwarding mechanisms in information-centric networking," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2847–2886, 4th Quart., 2016.
[38] I. Abdullahi, S. Arif, and S. Hassan, "Survey on caching approaches in information centric networking," *J. Netw. Comput. Appl.*, vol. 56, pp. 48–59, Oct. 2015.
[39] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.
[40] F. Guo, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Joint optimization of caching and association in Energy-Harvesting-Powered small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6469–6480, Jul. 2018.
[41] F. H. Panahi, F. H. Panahi, and T. Ohtsuki, "Energy efficiency analysis in cache-enabled D2D-aided heterogeneous cellular networks," *IEEE Access*, vol. 8, pp. 19540–19554, 2020.
[42] Q. Chen, F. R. Yu, T. Huang, R. Xie, J. Liu, and Y. Liu, "Joint resource allocation for software-defined networking, caching, and computing," *IEEE/ACM Trans. Netw.*, vol. 26, no. 1, pp. 274–287, Feb. 2018.
[43] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 907–922, Apr. 2016.
[44] R. Xie, Z. Li, T. Huang, and Y. Liu, "Energy-efficient joint content caching and small base station activation mechanism design in heterogeneous cellular networks," *China Commun.*, vol. 14, no. 10, pp. 70–83, Oct. 2017.
[45] M. I. A. Zahed, I. Ahmad, D. Habibi, Q. V. Phung, and M. M. Mowla, "Proactive content caching using surplus renewable energy: A win–win solution for both network service and energy providers," *Future Gener. Comput. Syst.*, vol. 105, pp. 210–221, Apr. 2020.
[46] T. Dang and M. Peng, "Joint radio communication, caching, and computing design for mobile virtual reality delivery in fog radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 7, pp. 1594–1607, Jul. 2019.
[47] M. I. A. Zahed, I. Ahmad, D. Habibi, and Q. V. Phung, "Green and secure computation offloading for cache-enabled IoT networks," *IEEE Access*, vol. 8, pp. 63840–63855, 2020.
[48] S. S. Kafiloğlu, G. Gür, and F. Alagöz, "A Markovian model for satellite integrated cognitive and D2D HetNets," *Comput. Netw.*, vol. 169, Mar. 2020, Art. no. 107083.
[49] P. Li, S. Gong, S. Gao, Y. Hu, Z. Pan, and X. You, "Delay-constrained sleeping mechanism for energy saving in cache-aided ultra-dense network," *Sci. China Inf. Sci.*, vol. 62, no. 8, pp. 131–144, Aug. 2019.
[50] Z. Luo, M. LiWang, Z. Lin, L. Huang, X. Du, and M. Guizani, "Energy-efficient caching for mobile edge computing in 5G networks," *Appl. Sci.*, vol. 7, no. 6, p. 557, May 2017.
[51] Z. Zhao, M. Peng, Z. Ding, W. Wang, and H. V. Poor, "Cluster content caching: An energy-efficient approach to improve quality of service in cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1207–1221, May 2016.
[52] S. Yang, X. Qiu, H. Xie, J. Guan, Y. Liu, and C. Xu, "GDSoC: Green dynamic self-optimizing content caching in ICN-based 5G network," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 1, p. e3221, Jan. 2018.
[53] F. Idzikowski, L. Chiaraviglio, A. Cianfrani, J. Lopez Vizcaino, M. Polverini, and Y. Ye, "A survey on energy-aware design and operation of core networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1453–1499, 2nd Quart., 2016.
[54] C. Jiang, T. Fan, H. Gao, W. Shi, L. Liu, C. Cérin, and J. Wan, "Energy aware edge computing: A survey," *Comput. Commun.*, vol. 151, pp. 556–580, Feb. 2020.
[55] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 33–56, 1st Quart., 2017.
[56] Z. Zhang, C.-H. Lung, I. Lambadaris, and M. St-Hilaire, "When 5G meets ICN: An ICN-based caching approach for mobile video in 5G networks," *Comput. Commun.*, vol. 118, pp. 81–92, Mar. 2018.
[57] R. Tourani, S. Misra, T. Mick, and G. Panwar, "Security, privacy, and access control in information-centric networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 566–600, 1st Quart., 2018.
[58] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2353–2371, 3rd Quart., 2018.
[59] H. Jin, D. Xu, C. Zhao, and D. Liang, "Information-centric mobile caching network frameworks and caching optimization: A survey," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, p. 33, Dec. 2017.
[60] F. Khandaker, S. Oteafy, H. S. Hassanein, and H. Farahat, "A functional taxonomy of caching schemes: Towards guided designs in information-centric networks," *Comput. Netw.*, vol. 165, Dec. 2019, Art. no. 106937.
[61] I. U. Din, S. Hassan, M. K. Khan, M. Guizani, O. Ghazali, and A. Habbal, "Caching in information-centric networking: Strategies, challenges, and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1443–1474, 2nd Quart., 2018.

[62] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking: A survey," *Comput. Netw.*, vol. 57, no. 16, pp. 3128–3141, Nov. 2013.

[63] N. Lal, S. Kumar, G. Kadian, and V. K. Chaurasiya, "Caching methodologies in content centric networking (CCN): A survey," *Comput. Sci. Rev.*, vol. 31, pp. 39–50, Feb. 2019.

[64] X. Qiao, H. Wang, W. Tan, A. V. Vasilakos, J. Chen, and M. B. Blake, "A survey of applications research on content-centric networking," *China Commun.*, vol. 16, no. 9, pp. 122–140, Sep. 2019.

[65] E. G. Abdallah, H. S. Hassanein, and M. Zulkernine, "A survey of security attacks in information-centric networking," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1441–1454, 3rd Quart., 2015.

[66] S. Arshad, M. A. Azam, M. H. Rehmani, and J. Loo, "Recent advances in information-centric networking-based Internet of Things (ICN-IoT)," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2128–2158, Apr. 2019.

[67] B. Nour, K. Sharif, F. Li, S. Biswas, H. Moungla, M. Guizani, and Y. Wang, "A survey of Internet of Things communication using ICN: A use case perspective," *Comput. Commun.*, vols. 142–143, pp. 95–123, Jun. 2019.

[68] J. Yao, T. Han, and N. Ansari, "On mobile edge caching," *IEEE Commun. Surveys Tuts.*, vol. 21, pp. 2525–2553, 3rd Quart., 2019.

[69] Z. Piao, M. Peng, Y. Liu, and M. Daneshmand, "Recent advances of edge cache in radio access networks for Internet of Things: Techniques, performances, and challenges," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 1010–1028, Feb. 2019.

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

[71] D. Prerna, R. Tekchandani, and N. Kumar, "Device-to-device content caching techniques in 5G: A taxonomy, solutions, and challenges," *Comput. Commun.*, vol. 153, pp. 48–84, Mar. 2020.

[72] R. Ullah, S. H. Ahmed, and B.-S. Kim, "Information-centric networking with edge computing for IoT: Research challenges and future directions," *IEEE Access*, vol. 6, pp. 73465–73488, 2018.

[73] Z. Wang, J. Liu, and W. Zhu, "Social-aware video delivery: Challenges, approaches, and directions," *IEEE Netw.*, vol. 30, no. 5, pp. 35–39, Sep. 2016.

[74] S. Tarnoi, W. Kumwilaisak, V. Suppakitpaisarn, K. Fukuda, and Y. Ji, "Adaptive probabilistic caching technique for caching networks with dynamic content popularity," *Comput. Commun.*, vol. 139, pp. 1–15, May 2019.

[75] H. S. Goian, O. Y. Al-Jarrah, S. Muhaidat, Y. Al-Hammadi, P. Yoo, and M. Dianati, "Popularity-based video caching techniques for cache-enabled networks: A survey," *IEEE Access*, vol. 7, pp. 27699–27719, 2019.

[76] C. Wang, Y. He, F. R. Yu, Q. Chen, and L. Tang, "Integration of networking, caching, and computing in wireless systems: A survey, some research issues, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 7–38, 1st Quart., 2018.

[77] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of green information-centric networking: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1455–1472, 3rd Quart., 2015.

[78] W. Ali, S. M. Shamsuddin, and A. S. Ismail, "A survey of Web caching and prefetching," *Int. J. Adv. Soft Comput. Appl.*, vol. 3, no. 1, pp. 18–44, Mar. 2011.

[79] B. Bai, L. Wang, Z. Han, W. Chen, and T. Svensson, "Caching based socially-aware D2D communications in wireless content delivery networks: A hypergraph framework," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 74–81, Aug. 2016.

[80] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 22–28, Sep. 2016.

[81] X. Li, X. Wang, Z. Sheng, H. Zhou, and V. C. M. Leung, "Resource allocation for cache-enabled cloud-based small cell networks," *Comput. Commun.*, vol. 127, pp. 20–29, Sep. 2018.

[82] A. S. Ali, K. R. Mahmoud, and K. M. Naguib, "Optimal caching policy for wireless content delivery in D2D networks," *J. Netw. Comput. Appl.*, vol. 150, Jan. 2020, Art. no. 102467.

[83] D. Malak, M. Al-Shalash, and J. G. Andrews, "Optimizing content caching to maximize the density of successful receptions in Device-to-Device networking," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4365–4380, Oct. 2016.

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

[85] D. Das and A. A. Abouzeid, "Co-operative caching in dynamic shared spectrum networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5060–5075, Jul. 2016.

[86] B. Zhou, S. Wang, E. Chen, and M. Tao, "Cooperative caching for spectrum access in cognitive radio networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[87] L. Tang, Q. Huang, A. Puntambekar, Y. Vigfusson, W. Lloyd, and K. Li, "Popularity prediction of Facebook videos for higher quality streaming," in *Proc. USENIX Annu. Tech. Conf.*, Santa Clara, CA, USA, Jul. 2017, pp. 111–123.

[88] M. Hajimirsadeghi, N. B. Mandayam, and A. Reznik, "Joint caching and pricing strategies for popular content in information centric networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 3, pp. 654–667, Mar. 2017.

[89] W. Hoiles, A. Aprem, and V. Krishnamurthy, "Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1426–1437, Jul. 2017.

[90] K. Bilal, E. Baccour, A. Erbad, A. Mohamed, and M. Guizani, "Collaborative joint caching and transcoding in mobile edge networks," *J. Netw. Comput. Appl.*, vol. 136, pp. 86–99, Jun. 2019.

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

[92] H. Ben-Ammar, Y. Hadjadj-Aoul, G. Rubino, and S. Ait-Chellouche, "On the performance analysis of distributed caching systems using a customizable Markov chain model," *J. Netw. Comput. Appl.*, vol. 130, pp. 39–51, Mar. 2019.

[93] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A learning-based approach to caching in heterogenous small cell networks," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1674–1686, Apr. 2016.

[94] P. Blasco and D. Gunduz, "Learning-based optimization of cache content in a small cell base station," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 1897–1903.

[95] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Temporal locality in today's content caching: Why it matters and how to model it," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 5, pp. 5–12, Nov. 2013.

[96] S. Naz, R. Naveed Bin Rais, P. A. Shah, S. Yasmin, A. Qayyum, S. Rho, and Y. Nam, "A dynamic caching strategy for CCN-based MANETs," *Comput. Netw.*, vol. 142, pp. 93–107, Sep. 2018.

[97] M. A. Kader, E. Bastug, M. Bennis, E. Zeydan, A. Karatepe, A. S. Er, and M. Debbah, "Leveraging big data analytics for cache-enabled wireless networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.

[98] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.

[99] C. Yapar, K. Wan, R. F. Schaefer, and G. Caire, "On the optimality of D2D coded caching with uncoded cache placement and one-shot delivery," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8179–8192, Dec. 2019.

[100] K. Wan, D. Tuninetti, and P. Piantanida, "An index coding approach to caching with uncoded cache placement," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1318–1332, Mar. 2020.

[101] P. Hassanzadeh, A. M. Tulino, J. Llorca, and E. Erkip, "On coding for cache-aided delivery of dynamic correlated content," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1666–1681, Aug. 2018.

[102] Q. Yang and D. Gunduz, "Coded caching and content delivery with heterogeneous distortion requirements," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4347–4364, Jun. 2018.

[103] M. Mohammadi Amiri and D. Gunduz, "Fundamental limits of coded caching: Improved delivery rate-cache capacity tradeoff," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806–815, Feb. 2017.

[104] T. Luo and B. Peleato, "The transfer load-I/O trade-off for coded caching," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1524–1527, Aug. 2018.

[105] Y. Li, H. Xie, Y. Wen, C.-Y. Chow, and Z.-L. Zhang, "How much to coordinate? Optimizing in-network caching in content-centric networks," *IEEE Trans. Netw. Service Manage.*, vol. 12, no. 3, pp. 420–434, Sep. 2015.

[106] Q. Li, W. Shi, X. Ge, and Z. Niu, "Cooperative edge caching in software-defined hyper-cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2596–2605, Nov. 2017.

[107] Q. Yang, P. Hassanzadeh, D. Gunduz, and E. Erkip, "Centralized caching and delivery of correlated contents over Gaussian broadcast channels," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 122–136, Jan. 2020.
[108] X. Lin, Y. Tang, X. Lei, J. Xia, Q. Zhou, H. Wu, and L. Fan, "MARL-based distributed cache placement for wireless networks," *IEEE Access*, vol. 7, pp. 62606–62615, 2019.
[109] M. Bilal and S.-G. Kang, "Time aware least recent used (TLRU) cache management policy in ICN," in *Proc. 16th Int. Conf. Adv. Commun. Technol.*, Pyeongchang, South Korea, Feb. 2014, pp. 528–532.
[110] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache 'less for more' in information-centric networks (extended version)," *Comput. Commun.*, vol. 36, no. 7, pp. 758–770, Apr. 2013.
[111] S. Podlipnig and L. Böszörmenyi, "A survey of Web cache replacement strategies," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 374–398, Dec. 2003.
[112] H. Che, Y. Tung, and Z. Wang, "Hierarchical Web caching systems: Modeling, design and experimental results," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 7, pp. 1305–1314, Sep. 2002.
[113] N. Kitsuwan, H. Tahara, and E. Oki, "Heuristic approach to determining cache node locations in content-centric networks," *IEEE Access*, vol. 6, pp. 1542–1549, 2018.
[114] X. Zheng, G. Wang, and Q. Zhao, "A cache placement strategy with energy consumption optimization in information-centric networking," *Future Internet*, vol. 11, no. 3, p. 64, Mar. 2019.
[115] S. Kuang and N. Liu, "Analysis and cache design in spatially correlated HetNets with base station cooperation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8754–8768, Sep. 2018.
[116] C. Li, W. Liu, L. Wang, M. Li, and K. Okamura, "Energy-efficient quality of service aware forwarding scheme for content-centric networking," *J. Netw. Comput. Appl.*, vol. 58, pp. 241–254, Dec. 2015.
[117] S. E. Hajri and M. Assaad, "Energy efficiency in cache-enabled small cell networks with adaptive user clustering," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 955–968, Feb. 2018.
[118] I. Keshavarzian, Z. Zeinalpour-Yazdi, and A. Tadaion, "Energy-efficient mobility-aware caching algorithms for clustered small cells in ultra-dense networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6833–6846, Jul. 2019.
[119] X. Li and J. Wan, "Proactive caching for edge computing-enabled industrial mobile wireless networks," *Future Gener. Comput. Syst.*, vol. 89, pp. 89–97, Dec. 2018.
[120] C. M. Sarathchandra Magurawalage, K. Yang, L. Hu, and J. Zhang, "Energy-efficient and network-aware offloading algorithm for mobile cloud computing," *Comput. Netw.*, vol. 74, pp. 22–33, Dec. 2014.
[121] J. Kim, D. Yu, S. E. Hong, and S. H. Park, "Energy-efficient joint design of fronthaul and edge links for cache-aided C-RAN systems with wireless fronthaul," *Entropy*, vol. 21, no. 9, p. 860, Sep. 2019.
[122] Q. Jia, R. Xie, T. Huang, J. Liu, and Y. Liu, "Efficient caching resource allocation for network slicing in 5G core network," *IET Commun.*, vol. 11, no. 18, pp. 2792–2799, Dec. 2017.
[123] Y. Zhang, C. Li, T. H. Luan, Y. Fu, W. Shi, and L. Zhu, "A mobility-aware vehicular caching scheme in content centric networks: Model and optimization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3100–3112, Apr. 2019.
[124] R. Huo, F. R. Yu, T. Huang, R. Xie, J. Liu, V. C. M. Leung, and Y. Liu, "Software defined networking, caching, and computing for green wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 185–193, Nov. 2016.
[125] Z. Yan, M. Peng, and M. Daneshmand, "Cost-aware resource allocation for optimization of energy efficiency in fog radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2581–2590, Nov. 2018.
[126] F. Yin, A. Wang, D. Liu, and Z. Zhang, "Energy-aware joint user association and resource allocation for coded cache-enabled HetNets," *IEEE Access*, vol. 7, pp. 94128–94142, 2019.
[127] H. Mohammed, Z. Chen, and W. Chen, "Energy-efficient joint pushing and caching based on Markov decision process," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 433–445, Jun. 2019.
[128] F. Cheng, Y. Yu, Z. Zhao, N. Zhao, Y. Chen, and H. Lin, "Power allocation for cache-aided small-cell networks with limited backhaul," *IEEE Access*, vol. 5, pp. 1272–1283, 2017.
[129] Z. Zhou, M. Dong, K. Ota, and Z. Chang, "Energy-efficient context-aware matching for resource allocation in ultra-dense small cells," *IEEE Access*, vol. 3, pp. 1849–1860, 2015.
[130] Q. N. Nguyen, M. Arifuzzaman, K. Yu, and T. Sato, "A context-aware green information-centric networking model for future wireless communications," *IEEE Access*, vol. 6, pp. 22804–22816, 2018.
[131] Z. Yan, S. Chen, Y. Ou, and H. Liu, "Energy efficiency analysis of cache-enabled two-tier HetNets under different spectrum deployment strategies," *IEEE Access*, vol. 5, pp. 6791–6800, 2017.
[132] F. Dong, T. Wang, and S. Wang, "Power consumption minimization in cache-enabled mobile networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6917–6925, Jul. 2019.
[133] J. Zhang, X. Hu, Z. Ning, E. C.-H. Ngai, L. Zhou, J. Wei, J. Cheng, B. Hu, and V. C. M. Leung, "Joint resource allocation for latency-sensitive services over mobile edge computing networks with caching," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4283–4294, Jun. 2019.
[134] Y. Lan, X. Wang, C. Wang, D. Wang, and Q. Li, "Collaborative computation offloading and resource allocation in cache-aided hierarchical edge-cloud systems," *Electronics*, vol. 8, no. 12, p. 1430, Nov. 2019.
[135] Y. Liu, Y. Wang, R. Sun, S. Meng, and R. Su, "Energy efficient downlink resource allocation for D2D-assisted cellular networks with mobile edge caching," *IEEE Access*, vol. 7, pp. 2053–2067, 2019.
[136] J. Tang, H. Tang, X. Zhang, K. Cumanan, G. Chen, K.-K. Wong, and J. A. Chambers, "Energy minimization in D2D-assisted cache-enabled Internet of Things: A deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5412–5423, Aug. 2020.
[137] L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for D2D-assisted wireless caching networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2438–2452, Jun. 2016.
[138] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, "Resource allocation for information-centric virtualized heterogeneous networks with in-network caching and mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11339–11351, Dec. 2017.
[139] Z. Tan, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Virtual resource allocation for heterogeneous services in full duplex-enabled SCNs with mobile edge computing and caching," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1794–1808, Feb. 2018.
[140] T. T. Vu, D. T. Ngo, M. N. Dao, S. Durrani, and R. H. Middleton, "Spectral and energy efficiency maximization for content-centric C-RANs with edge caching," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6628–6642, Dec. 2018.
[141] C. Fang, F. Richard Yu, T. Huang, J. Liu, and Y. Liu, "An energy-efficient distributed in-network caching scheme for green content-centric networks," *Comput. Netw.*, vol. 78, pp. 119–129, Feb. 2015.
[142] J. Ji, K. Zhu, R. Wang, B. Chen, and C. Dai, "Energy efficient caching in backhaul-aware cellular networks with dynamic content Popularity," *Wireless Commun. Mobile Comput.*, vol. 2018, Apr. 2018, Art. no. 7532049.
[143] J. Luo, F. R. Yu, Q. Chen, and L. Tang, "Adaptive video streaming with edge caching and video transcoding over software-defined mobile networks: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1577–1592, Mar. 2020.
[144] L. Tang, B. Tang, L. Kang, and L. Zhang, "A novel task caching and migration strategy in multi-access edge computing based on the genetic algorithm," *Future Internet*, vol. 11, no. 8, p. 181, Aug. 2019.
[145] P. Duan, Y. Jia, L. Liang, J. Rodriguez, K. M. S. Huq, and G. Li, "Space-reserved cooperative caching in 5G heterogeneous networks for industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2715–2724, Jun. 2018.
[146] L. Xiang, D. W. K. Ng, R. Schober, and V. W. S. Wong, "Cache-enabled physical layer security for video streaming in backhaul-limited cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 736–751, Feb. 2018.
[147] H. Ko, J. Lee, and S. Pack, "Energy-efficient proactive edge caching with sleep scheduling for green networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 190, Aug. 2018.
[148] O. Ayoub, F. Musumeci, M. Tornatore, and A. Pattavina, "Energy-efficient Video-On-Demand content caching and distribution in metro area networks," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 1, pp. 159–169, Mar. 2019.
[149] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "Distributed energy consumption management in green content-centric networks via dual decomposition," *IEEE Syst. J.*, vol. 11, no. 2, pp. 625–636, Jun. 2017.
[150] Y. Sun, M. Peng, and S. Mao, "Deep reinforcement learning-based mode selection and resource management for green fog radio access networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1960–1971, Apr. 2019.
[151] F. Dehghani and N. Movahhedinia, "CCN energy-delay aware cache management using quantized hopfield," *J. Netw. Syst. Manage.*, vol. 26, no. 4, pp. 1058–1078, Mar. 2018.

[152] X. Yang, Z. Fei, J. Zheng, N. Zhang, and A. Anpalagan, "Joint multi-user computation offloading and data caching for hybrid mobile Cloud/Edge computing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11018–11030, Nov. 2019.

[153] M. Yan, C. A. Chan, W. Li, L. Lei, A. F. Gygax, and C.-L. I, "Assessing the energy consumption of proactive mobile edge caching in wireless networks," *IEEE Access*, vol. 7, pp. 104394–104404, 2019.

[154] J. Xu, K. Ota, and M. Dong, "Energy efficient hybrid edge caching scheme for tactile Internet in 5G," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 483–493, Jun. 2019.

[155] M. Furqan, W. Yan, C. Zhang, S. Iqbal, Q. Jan, and Y. Huang, "An energy-efficient collaborative caching scheme for 5G wireless network," *IEEE Access*, vol. 7, pp. 156907–156916, 2019.

[156] G. Gür, "Energy-aware cache management at the wireless network edge for information-centric operation," *J. Netw. Comput. Appl.*, vol. 57, pp. 33–42, Nov. 2015.

[157] J. Ma, J. Wang, and P. Fan, "A cooperation-based caching scheme for heterogeneous networks," *IEEE Access*, vol. 5, pp. 15013–15020, 2017.

[158] R. Xie, Q. Tang, and T. Huang, "Energy-efficient hierarchical cooperative caching optimisation for 5G networks," *IET Commun.*, vol. 13, no. 6, pp. 687–695, Apr. 2019.

[159] W. Li, J. Wang, G. Zhang, L. Li, Z. Dang, and S. Li, "A reinforcement learning based smart cache strategy for cache-aided ultra-dense network," *IEEE Access*, vol. 7, pp. 39390–39401, 2019.

[160] S. O. Somuyiwa, A. Gyorgy, and D. Gunduz, "A reinforcement-learning approach to proactive caching in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1331–1344, Jun. 2018.

[161] L. Li, Y. Xu, J. Yin, W. Liang, X. Li, W. Chen, and Z. Han, "Deep reinforcement learning approaches for content caching in cache-enabled D2D networks," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 544–557, Jan. 2020.

[162] Y. Xu, S. Jiang, and J. Wu, "Towards energy efficient Device-to-Device content dissemination in cellular networks," *IEEE Access*, vol. 6, pp. 25816–25828, 2018.

[163] T. X. Vu, S. Chatzinotas, B. Ottersten, and T. Q. Duong, "Energy minimization for cache-assisted content delivery networks with wireless backhaul," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 332–335, Jun. 2018.

[164] X. Zhang, T. Lv, W. Ni, J. M. Cioffi, N. C. Beaulieu, and Y. J. Guo, "Energy-efficient caching for scalable videos in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1802–1815, Aug. 2018.

[165] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Exploiting caching and multicast for 5G wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2995–3007, Apr. 2016.

[166] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, Sep. 2016.

[167] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis., "Finite-length analysis of caching-aided coded multicasting," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5524–5537, Oct. 2016.

[168] B. Zhou, Y. Cui, and M. Tao, "Optimal dynamic multicast scheduling for cache-enabled content-centric wireless networks," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2956–2970, Jul. 2017.

[169] Z. Chen, Y. Liu, B. Zhou, and M. Tao, "Caching incentive design in wireless D2D networks: A stackelberg game approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[170] Z. Yang, H. Tian, S. Fan, and G. Chen, "Dynamic incentive design in content dissemination process through D2D communication," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1799–1802, Aug. 2017.

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

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

[173] P. Zhang, X. Kang, Y. Liu, and H. Yang, "Cooperative willingness aware collaborative caching mechanism towards cellular D2D communication," *IEEE Access*, vol. 6, pp. 67046–67056, 2018.

[174] X. Yi, L. Pan, Y. Jin, F. Liu, and M. Chen, "EDirect: Energy-efficient D2D-assisted relaying framework for cellular signaling reduction," *IEEE/ACM Trans. Netw.*, vol. 28, no. 2, pp. 860–873, Apr. 2020.

[175] F. Zafari, J. Li, K. K. Leung, D. Towsley, and A. Swami, "Optimal energy consumption for communication, computation, caching, and quality guarantee," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 1, pp. 151–162, Mar. 2020.

[176] R. Xie, Z. Li, J. Wu, Q. Jia, and T. Huang, "Energy-efficient joint caching and transcoding for HTTP adaptive streaming in 5G networks with mobile edge computing," *China Commun.*, vol. 16, no. 7, pp. 229–244, Jul. 2019.

[177] H. Zhu, Y. Cao, X. Wei, W. Wang, T. Jiang, and S. Jin, "Caching transient data for Internet of Things: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2074–2083, Apr. 2019.

[178] J. Zhang, X. Zhang, and W. Wang, "Cache-enabled software defined heterogeneous networks for green and flexible 5G networks," *IEEE Access*, vol. 4, pp. 3591–3604, 2016.

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

[180] A. Bianco, R. Mashayekhi, and M. Meo, "On the energy consumption computation in content delivery networks," *Sustain. Comput., Informat. Syst.*, vol. 16, pp. 56–65, Dec. 2017.

[181] C. Meng, G. Wang, X. Dai, S. Chen, and W. Ni, "An energy-efficient transmission strategy for cache-enabled wireless networks with non-negligible circuit power," *IEEE Access*, vol. 7, pp. 74811–74821, 2019.

[182] Y. Takatsuka, H. Nagao, T. Yaguchi, M. Hanai, and K. Shudo, "A caching mechanism based on data freshness," in *Proc. Int. Conf. Big Data Smart Comput. (BigComp)*, Jan. 2016, pp. 329–332.

[183] S. Banerjee and S. K. Biswash, "Freshness management of cache content in information-centric networking," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2019, pp. 1–6.

[184] M. Meddeb, A. Dhraief, A. Belghith, T. Monteil, K. Drira, and S. Alahmadi, "Cache freshness in named data networking for the Internet of Things," *Comput. J.*, vol. 61, no. 10, pp. 1496–1511, Oct. 2018.

[185] X. Wu, X. Li, J. Li, P. C. Ching, and H. Vincent Poor, "Deep reinforcement learning for IoT networks: Age of information and energy cost tradeoff," 2020, *arXiv:2010.12297*. [Online]. Available: http://arxiv.org/abs/2010.12297

[186] S. Yu, X. Chen, S. Wang, L. Pu, and D. Wu, "An edge computing-based photo crowdsourcing framework for real-time 3D reconstruction," *IEEE Trans. Mobile Comput.*, early access, Jul. 7, 2020, doi: 10.1109/TMC.2020.3007654.

[187] L. T. Tan and R. Q. Hu, "Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10190–10203, Nov. 2018.

[188] Y. Tang, K. Guo, J. Ma, Y. Shen, and T. Chi, "A smart caching mechanism for mobile multimedia in information centric networking with edge computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 590–600, Feb. 2019.

[189] M. I. Aziz Zahed, I. Ahmad, D. Habibi, and Q. V. Phung, "Content caching in industrial IoT: Security and energy considerations," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 491–504, Jan. 2020.

[190] T.-X. Zheng, H.-M. Wang, and J. Yuan, "Secure and energy-efficient transmissions in cache-enabled heterogeneous cellular networks: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5554–5567, Nov. 2018.

[191] L. Xiang, D. W. K. Ng, R. Schober, and V. W. S. Wong, "Secure video streaming in heterogeneous small cell networks with untrusted cache helpers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2645–2661, Apr. 2018.

[192] F. Liu, P. Shu, H. Jin, L. Ding, J. Yu, D. Niu, and B. Li, "Gearing resource-poor mobile devices with powerful clouds: Architectures, challenges, and applications," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 14–22, Jun. 2013.

[193] F. Liu, P. Shu, and J. C. S. Lui, "AppATP: An energy conserving adaptive mobile-cloud transmission protocol," *IEEE Trans. Comput.*, vol. 64, no. 11, pp. 3051–3063, Nov. 2015.

[194] W. Deng, F. Liu, H. Jin, B. Li, and D. Li, "Harnessing renewable energy in cloud datacenters: Opportunities and challenges," *IEEE Netw.*, vol. 28, no. 1, pp. 48–55, Jan. 2014.

[195] P. A. Apostolopoulos, E. E. Tsiropoulou, and S. Papavassiliou, "Game-theoretic learning-based QoS satisfaction in autonomous mobile edge computing," in *Proc. Global Inf. Infrastruct. Netw. Symp. (GIIS)*, Thessaloniki, Greece, Oct. 2018, pp. 1–5.

[196] M. Chen, Y. Hao, L. Hu, M. S. Hossain, and A. Ghoneim, "Edge-CoCaCo: Toward joint optimization of computation, caching, and communication on edge cloud," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 21–27, Jun. 2018.

[197] P. Liu, G. Xu, K. Yang, K. Wang, and X. Meng, "Jointly optimized energy-minimal resource allocation in cache-enhanced mobile edge computing systems," *IEEE Access*, vol. 7, pp. 3336–3347, 2019.
[198] Y. Hao, M. Chen, L. Hu, M. S. Hossain, and A. Ghoneim, "Energy efficient task caching and offloading for mobile edge computing," *IEEE Access*, vol. 6, pp. 11365–11373, 2018.
[199] M. Yan, C. A. Chan, A. F. Gygax, J. Yan, L. Campbell, A. Nirmalathas, and C. Leckie, "Modeling the total energy consumption of mobile network services and applications," *Energies*, vol. 12, no. 1, p. 184, Jan. 2019.
[200] M. Huang, Y. Liu, N. Zhang, N. N. Xiong, A. Liu, Z. Zeng, and H. Song, "A services routing based caching scheme for cloud assisted CRNs," *IEEE Access*, vol. 6, pp. 15787–15805, 2018.
[201] B. Gao, Z. Zhou, F. Liu, and F. Xu, "Winning at the starting line: Joint network selection and service placement for mobile edge computing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Paris, France, Apr. 2019, pp. 1459–1467.
[202] S. Ranadheera, S. Maghsudi, and E. Hossain, "Minority games with applications to distributed decision making and control in wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 184–192, Oct. 2017.
[203] S. Ranadheera, S. Maghsudi, and E. Hossain, "Mobile edge computation offloading using game theory and reinforcement learning," 2017, *arXiv:1711.09012*. [Online]. Available: http://arxiv.org/abs/1711.09012
[204] P. A. Apostolopoulos, M. Torres, and E. E. Tsiropoulou, "Satisfaction-aware data offloading in surveillance systems," in *Proc. 14th Workshop Challenged Netw. (CHANTS)*, Los Cabos, Mexico, Oct. 2019, pp. 21–26.
[205] S. Chen, L. Jiao, L. Wang, and F. Liu, "An online market mechanism for edge emergency demand response via cloudlet control," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Paris, France, Apr. 2019, pp. 2566–2574.
[206] Q. Chen, Z. Zheng, C. Hu, D. Wang, and F. Liu, "On-edge multi-task transfer learning: Model and practice with data-driven task allocation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 6, pp. 1357–1371, Jun. 2020.
[207] L. Pei, G. Shen, H. Yaoyue, P. Zhiwen, and Y. Xiaohu, "Joint caching and sleeping optimisation for D2D-aided ultra-dense network," *IET Commun.*, vol. 14, no. 1, pp. 1–10, Jan. 2020.
[208] X. Huang and N. Ansari, "Content caching and distribution in smart grid enabled wireless networks," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 513–520, Apr. 2017.
[209] S. Zhou, J. Gong, Z. Zhou, W. Chen, and Z. Niu, "GreenDelivery: Proactive content caching and push with energy-harvesting-based small cells," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 142–149, Apr. 2015.
[210] H. Wu, X. Tao, N. Zhang, D. Wang, S. Zhang, and X. Shen, "On base station coordination in cache-and energy harvesting-enabled HetNets: A stochastic geometry study," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3079–3091, Jul. 2018.
[211] A. Mehrabi, M. Siekkinen, and A. Yla-Jaaski, "Energy-aware QoE and backhaul traffic optimization in green edge adaptive mobile video streaming," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 828–839, Sep. 2019.
[212] S. Gautam, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Cache-aided simultaneous wireless information and power transfer (SWIPT) with relay selection," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 187–201, Jan. 2019.
[213] J. Yao and N. Ansari, "Caching in energy harvesting aided Internet of Things: A game-theoretic approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3194–3201, Apr. 2019.
[214] T. Han and N. Ansari, "Network utility aware traffic load balancing in backhaul-constrained cache-enabled small cell networks with hybrid power supplies," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2819–2832, Oct. 2017.
[215] S. Zhang, N. Zhang, X. Fang, P. Yang, and X. S. Shen, "Self-sustaining caching stations: Toward cost-effective 5G-enabled vehicular networks," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 202–208, Nov. 2017.
[216] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-sustainable traffic steering for 5G mobile networks," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 54–60, Nov. 2017.
[217] X. Chen, X. Song, C. Jia, and T. Cao, "Device to device networks with cache-enabled and self-sustained mobile helpers," *IEEE Access*, vol. 7, pp. 93820–93829, 2019.
[218] D. Xu, H. Jin, C. Zhao, and D. Liang, "Joint caching and sleep-active scheduling for energy-harvesting based small cells," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–6.
[219] X. Sun and N. Ansari, "Dynamic resource caching in the IoT application layer for smart cities," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 606–613, Apr. 2018.
[220] Z. Yan, M. Peng, and C. Wang, "Economical energy efficiency: An advanced performance metric for 5G systems," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 32–37, Feb. 2017.
[221] M. Zeng, Y. Li, K. Zhang, M. Waqas, and D. Jin, "Incentive mechanism design for computation offloading in heterogeneous fog computing: A contract-based approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[222] S. H. Mohamed, M. B. A. Halim, T. E. H. Elgorashi, and J. M. H. Elmirghani, "Fog-assisted caching employing solar renewable energy and energy storage devices for video on demand services," *IEEE Access*, vol. 8, pp. 115754–115766, 2020.
[223] Y. Cheng, "Edge caching and computing in 5G for mobile augmented reality and haptic Internet," *Comput. Commun.*, vol. 158, pp. 24–31, May 2020.
[224] J. Li, J. Wu, G. Xu, J. Li, X. Zheng, and A. Jolfaei, "Integrating NFV and ICN for advanced driver-assistance systems," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5861–5873, Jul. 2020.
[225] Z. Zhao, L. Guardalben, M. Karimzadeh, J. Silva, T. Braun, and S. Sargento, "Mobility prediction-assisted Over-the-Top edge prefetching for hierarchical VANETs," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1786–1801, Aug. 2018.
[226] M. Saleemi, M. Anjum, and M. Rehman, "The ubiquitous healthcare facility framework: A proposed system for managing rural antenatal care," *IEEE Access*, vol. 7, pp. 161264–161282, 2019.
[227] X. Wang and Z. Jin, "An overview of mobile cloud computing for pervasive healthcare," *IEEE Access*, vol. 7, pp. 66774–66791, 2019.
[228] M. Asif-Ur-Rahman, F. Afsana, M. Mahmud, M. S. Kaiser, M. R. Ahmed, O. Kaiwartya, and A. James-Taylor, "Toward a heterogeneous mist, fog, and cloud-based framework for the Internet of healthcare things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4049–4062, Jun. 2019.
[229] Y. He, F. R. Yu, N. Zhao, V. C. M. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 31–37, Dec. 2017.
[230] D. Wang, B. Bai, K. Lei, W. Zhao, Y. Yang, and Z. Han, "Enhancing information security via physical layer approaches in heterogeneous IoT with multiple access mobile edge computing in smart city," *IEEE Access*, vol. 7, pp. 54508–54521, 2019.
[231] E. H.-C. Lu, H.-S. Chen, and V. S. Tseng, "An efficient framework for multirequest route planning in urban environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 869–879, Apr. 2017.
[232] Y. Qian, L. Shi, L. Shi, K. Cai, J. Li, and F. Shu, "Cache-enabled power line communication networks: Caching node selection and backhaul energy optimization," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 606–615, Jun. 2020.
[233] M. Waqas, Y. Niu, M. Ahmed, Y. Li, D. Jin, and Z. Han, "Mobility-aware fog computing in dynamic environments: Understandings and implementation," *IEEE Access*, vol. 7, pp. 38867–38879, 2019.
[234] M. Waqas, Y. Niu, Y. Li, M. Ahmed, D. Jin, S. Chen, and Z. Han, "Mobility-aware device-to-device communications: Principles, practice and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1863–1886, 3rd Quart., 2020.
[235] Y. Kwon, H. Suh, J. Oh, and T. Hwang, "Energy efficient communication for secure D2D underlaid cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9110–9123, Oct. 2017.
[236] S. Khodaparas, A. Benslimane, and S. Yousefi, "A software-defined caching scheme for the Internet of Things," *Comput. Commun.*, vol. 158, pp. 178–188, May 2020.
[237] H. Jin, H. Lu, Y. Jin, and C. Zhao, "IVCN: Information-centric network slicing optimization based on NFV in fog-enabled RAN," *IEEE Access*, vol. 7, pp. 69667–69686, 2019.
[238] S. He, J. Wang, W. Huang, Y. Huang, M. Xiao, and Y. Zhang, "Energy-efficient transceiver design for cache-enabled millimeter-wave systems," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3876–3889, Jun. 2020.
[239] T.-Y. Wu and T. Chang, "Interference reduction by millimeter wave technology for 5G-based green communications," *IEEE Access*, vol. 4, pp. 10228–10234, 2016.

[240] S. Vuppala, T. X. Vu, S. Gautam, S. Chatzinotas, and B. Ottersten, "Cache-aided millimeter wave ad hoc networks with contention-based content delivery," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3540–3554, Aug. 2018.
[241] J. Qiao, Y. He, and X. S. Shen, "Proactive caching for mobile video streaming in millimeter wave 5G networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 7187–7198, Oct. 2016.

**M. ISHTIAQUE A. ZAHED** received the B.Sc. and M.Sc. degrees in electrical & electronic engineering from the Bangladesh University of Engineering and Technology (BUET), in 2011 and 2015, respectively. He is currently pursuing the Ph.D. degree with the School of Engineering, Edith Cowan University, Australia. He is also associated with the School of Science, Engineering and Technology, East Delta University, Chattogram, Bangladesh, as an Assistant Professor. His research interests include green communications, the IoT, wireless networks, and network security.

**IFTEKHAR AHMAD** (Member, IEEE) received the Ph.D. degree in communication networks from Monash University, Australia, in 2007. He is currently an Associate Professor with the School of Engineering, Edith Cowan University, Australia. His research interests include 5G technologies, green communications, QoS in communication networks, software-defined radio, wireless sensor networks, and computational intelligence.

**DARYOUSH HABIBI** (Senior Member, IEEE) received the B.E. degree (Hons.) in electrical engineering and the Ph.D. degree from the University of Tasmania, Australia, in 1989 and 1994, respectively. His employment history includes Telstra Research Laboratories, Flinders University, Intelligent Pixels, Inc., and Edith Cowan University, where he is currently a Professor, the Pro Vice-Chancellor, and the Executive Dean of engineering. His current research interests include engineering design for sustainable development, smart energy systems, environmental monitoring technologies, and reliability and quality of service in communication systems and networks. He is also a Fellow of Engineers Australia and the Institute of Marine Engineering, Science and Technology.

**QUOC VIET PHUNG** (Member, IEEE) received the Ph.D. degree in communication networks from Edith Cowan University, Australia, in 2010. He is currently a Postdoctoral Research Fellow with the School of Engineering, Edith Cowan University. His current research interests include 5G technologies, green communications, wireless sensor networks, and computational intelligence.

**MD. MUNJURE MOWLA** received the Ph.D. degree in communication engineering from Edith Cowan University, Australia, in 2018. He is currently working as an Associate Professor with the Department of Electronics and Telecommunication Engineering, Rajshahi University of Engineering & Technology (RUET), Bangladesh. His research interests include beyond 5G/6G wireless communications, radio resource management, millimeter wave, software defined radio, mobile edge computing, device to device communication, and ad hoc networks.

**MUHAMMAD WAQAS** (Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Electrical Engineering, University of Engineering and Technology Peshawar, Pakistan, in 2009 and 2014, respectively, and the Ph.D. degree from the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing, China, in June 2019. From 2012 to 2015, he has served for the Sarhad University of Science and Information Technology, Peshawar, Pakistan, as a Lecturer and a Program Coordinator. He is currently a Research Associate with the Beijing Key Laboratory of Trusted Computing, Faculty of Information Technology, Beijing. He is also associated with the Faculty of Computer Science and Engineering, GIK Institute of Engineering Sciences and Technology, Pakistan, as an Assistant Professor. His current research interests include 5G networks, D2D communication resource allocation and physical layer security and information security, mobility investigation, fog computing, and MEC.