

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.0322000

Opportunistic Mobile Networks Content Delivery for Important but Non-Urgent Traffic

Chun Pong Lau¹ (Member, IEEE), Guoqing Ma¹ (Student Member, IEEE),
Hengky Susanto^{2,5} (Member, IEEE), Shuping Dang^{1,3} (Member, IEEE),
Kam Shing Ng⁴ (Member, IEEE), and Basem Shihada¹ (Senior Member, IEEE)

¹Computer, Electrical and Mathematical Science Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

²Computer Science Department, University of Massachusetts Lowell, MA 01854, USA

³Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, UK

⁴Department of Science Department, City University of Hong Kong, Hong Kong, China

⁵Department of Science and Environmental Studies, Education University of Hong Kong, Hong Kong, China

Corresponding author: Basem Shihada (e-mail: basem.shihada@kaust.edu.sa).

This work was funded by King Abdullah University of Science and Technology.

ABSTRACT

As delay-tolerant and large-size content, for example, software updates, TV series, and virtual reality related content, become more prevalent in mobile networks, the need for efficient content delivery mechanisms becomes increasingly important, the traffic that carries these contents is not suitable to be evaluated using traditional network performance metrics, e.g., delay, throughput, and jitter. Based on this insight, we propose the solution of content dissemination from opportunistic mobile social communications (CODOMOC) which utilizes energy cost as an alternative performance metric and exploits daily human activity mobility pattern to determine how, when, and where the contents should be disseminated. Then, we introduce two options in CODOMOC to achieve different network operators' objectives. The two options are the Only Dense (OD) option which aims at minimizing energy consumption for network operators and the Broadcast Efficiency (BE) option to further reduce the total carbon footprint of network operators. CODOMOC is evaluated by comparing with a mobility-based broadcast method. The results show that CODOMOC reduces the average energy consumption by 51% and 60% in the OD and BE options respectively. The proposed solution equipped with the two modes is expected to provide a higher degree of flexibility and reduce energy consumption for mobile networks, while, admittedly, the application scope of the solution and the associated methodologies proposed in this paper is restricted to important but non-urgent traffic delivery.

INDEX TERMS Content delivery, important but non-urgent traffic, ad-hoc mobile communications, energy efficiency, human mobility.

I. INTRODUCTION

MOBILE networks have become the fastest growing communication network in recent years. This is due to the dramatic increase in both device connectivity and data usage [1], [2]. This in turn results in significant increase in network traffic, including multimedia, e.g., TV serials download, game packages, and virtual reality [3], [4], and more frequent software updates, e.g., operating system updates, software patches, bug fixes, application updates, security updates, deep learning model update for AI based application. These are types of traffic which cannot be directly evaluated using traditional network performance metrics, such as delay, throughput, and jitter. Traditional metrics are chosen to satisfy subscriber demands over specific applications that are sensi-

tive to network rate, bandwidth availability, and/or variation in the latency. However, software updates and multimedia do not necessarily mean that they demand urgency and immediacy as required by applications such as real-time streaming applications. Thus, traditional metrics might not be suitable for handling these types of traffic.

Studies from [5]–[8] support this as it identifies a network traffic perceived as non-urgent by many subscribers (users) because they are satisfied with what they have now, including traffic that carries software updates. In such situations, subscribers are less stringent can tolerate the low throughput when requesting content. Similar observation is also made in subscribers downloading multimedia content to be viewed later when they have free time, for instance after school or

work. These subscribers are also less concerned with network performance during the download as long as the content is available when they are ready to view it. Based on these observations, we refer this type of traffic as an important but non-urgent (I-NU) traffic in this paper.

In this study, we first investigate how to evaluate I-NU traffic in mobile networks by understanding the pain point of delivering the content when subscriber demand is not considered. Based on this understanding, we propose content dissemination scheme for I-NU traffic aiming to minimize the energy cost, which is the objective function adopted in the formulation. To solve the minimization problem, we exploit the insight that humans mobility is not purely random but often bound by geographical distance [9]–[12]. This means that mobile subscribers often follow regular patterns, such as commuting to work during weekdays. This correlation between human mobility and mobile traffic patterns is confirmed in [13].

With this insight on human mobility and subscribers being less immediately concerned with the traffic performance, we propose and refine our solution, termed the content dissemination from opportunistic mobile social communications (CODOMOC), to opportunistically determine *how*, *when*, and *where* content should be disseminated so that energy cost is minimized. CODOMOC factors in human mobility patterns, temporal restriction on when contents must be delivered, and cellular network infrastructure. CODOMOC's framework guides the design of content dissemination policies to achieve objectives in wireless network settings. To determine how, when, and where content should be disseminated, the policy consists of three stages that use the spatio-temporal insights from human activity and mobility patterns.

In the first stage, to determine how content should be disseminated, CODOMOC leverages the macroscopic understanding of human mobility in areas with high density of subscribers to broadcast the content. The objective is to opportunistically maximize the number of content recipients, while at the same time minimize the energy cost at the network infrastructure level. Here, we introduce two options: the Only Dense (OD) and Broadcast Efficiency (BE) options, which provide network operators with greater flexibility to balance infrastructure energy consumption and total energy consumption. Then, we use a network graph to determine where and when content should be broadcast based on the density of subscribers at different geographical locations and time. This simplifies the search algorithm and mechanics, particularly in large cellular networks. In the second stage, content recipients become carriers and forward the content through ad-hoc communications to remaining subscribers. This is achieved by incentivizing users with bonuses, credits, rewards, etc. [14]–[17]. Offloading content dissemination to users lowers infrastructure energy consumption. At the final stage, the scheme performs trickle broadcast to users that have not received the content in the previous two stages, while ensuring energy usage is minimized. This is accomplished by opportunistically utilizing the standard temporary IP address,

e.g., IP Care of Address used in managing mobile devices.

To evaluate the robustness of our proposed solution, we have implemented CODOMOC in a simulator that utilizes insights from real data trace collected for human mobility studies in [11], [18]. We demonstrate that CODOMOC reduces energy consumption by up to 63% at the network infrastructure level in comparison to the generic broadcast (GB) method. In our evaluation, we further demonstrate that an additional 9% decrease of energy cost is achieved when the BE option is applied. Additionally, our evaluation also indicate that CODOMOC also reduces the average energy consumption by 51% and 60% in the OD and BE options respectively.

Specifically, the contributions of this paper can be summarized as follows:

- We identify a class of traffic, i.e., the I-NU traffic that should be evaluated using energy cost, instead of traditional performance metrics.
- We propose a three-stage policy that determines how, when, and where contents should be disseminated opportunistically, exploiting human mobility pattern so that the dissemination cost is minimized.
- We design a network graph abstraction that captures the dynamic relationship between human mobility patterns, location, and time. This abstraction provides the building blocks for a polynomial-time bounded algorithm that decides when and where the content must be disseminated.
- Pseudo Broadcast Scheme to perform content forwarding in mobile ad-hoc network setting.
- We carry out extensive simulations using insights from real data trace of human mobility to demonstrate the effectiveness of CODOMOC in achieving energy-efficient and low-cost content delivery.

II. RELATED WORKS

Currently, there are very few studies on how I-NU content should be disseminated, particularly in mobile networks. The majority of existing studies focus on minimizing the volume of I-NU traffic with the intention to reduce its impact on the performance of higher priority traffic [19], [20]. In other words, most studies consider network traffic carrying I-NU contents as lower priority traffic because they are perceived as less important by subscribers.

From content dissemination perspective, there have been studies that exploit urban-scale mobility for content delivery. Lee *et al.* design a service to carry digital contents by mobile users to multimedia kiosks [21]. Furthermore, the authors [22] put forth a methodology that leverages a large pool of collected data to forecast and effectively manage the energy consumption of 5G wireless networks by identifying patterns in the data. In [23] reinforcement learning is used to ensure that communication and computing resources are allocated as efficiently as possible. In [24], Wu *et al.* investigate the challenges and characteristics of mobile social device caching that relies on user mobility. The study in [25] analyzes human activities for content broadcast in 5G networks. In [26], the

authors propose a long-term radio resource allocation scheme in multiple cells by predicting user movement. The authors of [27] further extend the study in [26] by applying machine learning to predict where to broadcast based on user behavior.

Additionally, as suggested by the authors [28], power and channel allocation scheme utilizes the Hungarian method to obtain an optimal solution that utilizes both uplink and downlink channels for channel allocation. [29] investigate the radio latency and reliability costs that are incurred as a result of a specific 5G new radio power-saving feature on specifically internet of things (IoT)s devices. Although these approaches reduce the energy cost in comparison to a generic broadcast scheme that uses all cells in networks, they still consume a considerable amount of energy. This is because the solutions may avoid broadcasting at the cells with no subscribers but the content still has to be broadcast at cells with a small number of subscribers (e.g., a single subscriber in a cell), resulting in energy inefficiency. Additionally, according to the studies in [30], distributed cooperative spectrum sensing based on energy correlation can effectively shorten sensing time, which conserve energy on user devices. Study in [31] discovers that altering user behavior on video streaming may reduce the carbon dioxide emissions.

Different from existing studies, our studies show that today's I-NU traffic also must meet stringent requirements imposed by the content providers for reliability and credibility purposes. In order to address the needs of content providers, we present a novel approach that takes into account human mobility patterns. The current state-of-the-art method for disseminating content based on human mobility, termed the human mobility-based content distribution (MOOD) strategy [32], which minimizes the number of transmissions by broadcasting content when the largest number of subscribers congregate. Our proposed methodology, however, distinguishes itself from MOOD by leveraging the dynamic nature of human mobility, which leads to the formation of clusters as people move about. Furthermore, an individual can belong to multiple clusters at various times, such as during working hours in an office setting or while dining out at restaurants. These clusters can give rise to pervasive networks comprised of mobile devices, which is a prevalent characteristic in modern mobile technology. By capitalizing on these insights, our approach is able to distribute content to a larger number of subscribers than MOOD can achieve.

III. MOTIVATION AND PROBLEM FORMULATION

A. IMPORTANT BUT NON-URGENT TRAFFIC

Studies from [5]–[8] show that despite the importance of software patches and updates in alleviating software problems and vulnerability, in practice, subscribers often tolerate delay in installing the patches or updates. Because of the compromised device performance and operating disruption, subscribers often prefer to schedule software updates when they are not using the corresponding devices. As such, content providers often turn to automatic updates for the cases of continuous update delay by subscribers [8]. As another exam-

ple, TV scheduling has long been arranged to suit subscriber schedules, as noted in [33]. For instance, children program is shown after school hours or programs targeted at older viewers will be shown after working hours. Today, technology allows subscribers to access content at their best convenience, such as Netflix and AppleTV. As a direct result, immediate availability is no longer necessary. The requirement is for the content to be available when being requested.

Generally, the aforementioned observations of digital content collectively indicate that with regards to content dissemination, while content providers may be concerned about reliability and *immediate availability*, subscribers are more concerned about their convenience and *timely availability* as well as reliability. This insight gives us a new perspective of how the I-NU content should be disseminated.

B. HUMAN MOBILITY

Empirical studies confirm that human mobility exhibits a high level of regularity and is bounded by geographical distance, as observed in [9]–[12], [18]. For example, Tokyo residents travel from residential to commercial districts for work, followed by a reverse pattern after working hours. Similar patterns were observed in Beijing for mobile subscribers [11], [18]. A study of Foursquare users also found similar patterns in Singapore, San Francisco, and Houston [12]. Additionally, human mobility patterns have a direct correlation with mobile traffic patterns, as confirmed in [13]. The study in [25] further concludes that human activities and their routine follow identifiable patterns by modeling the state transition of mobility patterns and life routine of a mobile subscribers as a chain of activity stages using the Markov Chain model. All these empirical evidences corroborate the spatio-temporal correlation between user mobility and data traffic¹.

Based on these findings, we propose that human mobility pattern can be leveraged to derive an effective strategy to disseminate content at lower energy cost. For instance, Fig. 1 illustrates that the optimal timing for content dissemination is at 3 p.m. because this is the time segment that most of the content subscribers congregate. During this time broadcasts in only four cells is sufficient, which is the minimum number of transmissions compared to other time segments. This observation leads to the next questions of *how*, *when*, and *where* content must be disseminated such that energy consumption is minimized.

C. PROBLEM FORMULATION

To understand I-NU content dissemination in mobile settings, we investigate the pain point from the mobile network operator's perspective, which is the cost of energy usage for base stations as confirmed in [34], [35]. For instance, China Mobile, one of the largest network operators in Mainland China, reduced yearly energy utilization by 36 million kWh after

¹This study acknowledges the potential privacy concerns that may arise from probing human mobility and data traffic for discovering the spatio-temporal correlation.

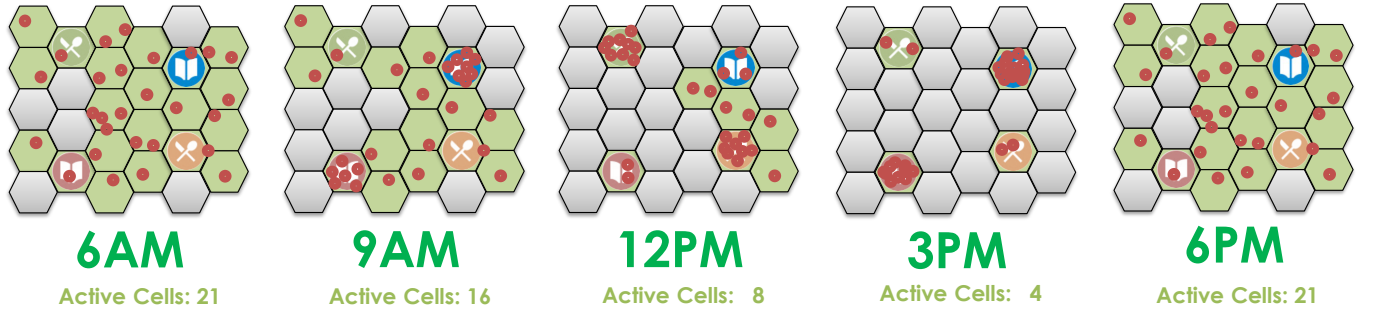


FIGURE 1. Example of human mobility and active cells varying by time. A green cell is denoted as an active cell where there is at least a content subscriber is currently located in that cell. The red circles represent mobile users. The symbols in the cells indicate the location of schools and restaurants. The most congregated time in this example is at 3 p.m

applying a simple on-off switching strategy at base stations [34]. Energy cost is a valid alternative metric for evaluating I-NU traffic when traditional performance metrics are not applicable. Thus, the content dissemination problem can be reduced to an energy consumption minimization problem in this context.

Additionally, beside minimizing energy consumption, the network operator must consider content provider demands, such as timely availability and reachability for all subscribers. With this observation, we formulate the problem of minimizing total energy consumption $\mathcal{E}_{\text{TOT}}(c_m)$ for disseminating I-NU content c_m as follows:

$$\begin{aligned} & \underset{t \leq \mathbb{T}}{\text{minimize}} && \sum_{b_n \in \mathbb{B}} \mathcal{E}_{\text{TOT}}(c_m) \\ & \text{subject to} && \forall, u \in U, \\ & && t, b_n, \mathbb{T}, \mathbb{B} > 0. \end{aligned} \quad (1)$$

The problem formulation above can be interpreted as finding a set of base stations in set $\mathbb{B}$ to disseminate c_m before the deadline $\mathbb{T}$ determined by the content provider expires, such that $\mathcal{E}_{\text{TOT}}(c_m)$ is also minimized, while ensuring the content reach all subscriber $u \in U$.

We develop a framework to solve the above problem by opportunistically carrying out I-NU content dissemination tasks to minimize energy consumption. The framework analyzes complex issues of content dissemination and considers the following components:

- 1) Human mobility pattern: To provide insights to *how* the content should be disseminated.
- 2) Temporal restriction on content delivery: To provide time limit of *when* the content should be disseminated.
- 3) Cellular Infrastructure: To provides geographical context of *where* the contents should be disseminated.

D. NETWORK INFRASTRUCTURE

Next, we describe the cellular Infrastructure considered in our study, as illustrated in Fig. 2. The Infrastructure is composed of cloud radio access network (CRAN) and Evolved Multime-

dia Broadcast Multicast Service (eMBMS) technology [36], [37].

The motivation of CRAN is to enable critical coordination among cells and generalized virtual function processing in the cloud. This is realized by decomposing the traditional base station (BS) into a signal processing unit, named as baseband unit (BBU), and a radio unit, named as remote radio head (RRH). The BBU is responsible for receiving data, controlling signal from the core network through the backhaul link, and transforming the signal to baseband signal and send to the RRH through a line-card (LC) by the fronthaul optical fiber link. The BBUs of multiple cells are centralized into the BBU Pool, which enables the provisioning of network function virtualization for baseband processing. A RRH consists of an optical network unit (ONU) to convert the optical signal from the BBU to radio signal and transmit it through a power amplifier to the antenna. In the core network, eMBMS system architecture is considered in this paper. It consists of Broadcast/Multicast Service Center (BM-SC), MBMS Gateway (MBMS-GW), Mobility Management Entity (MME), and Multi-cell/multicast Coordinating Entity (MCE).

A content provided by the content provider first arrives at the BM-SC from the external network to the mobile network. The BM-SC is responsible for announcing the arrival of the content to the subscriber in the mobile network. The MBMS-GW is a logical node for forwarding and caching user plane data and sending control signals to MME. The MME is responsible for authentication, security, and mobility management procedures. The MCE performs admission control and radio resources allocation for the radio access network (RAN).

IV. CODOMOC FRAMEWORK

In this section, we present our solution designed for energy-efficient I-NU content dissemination, including a three-stage policy based on subscriber density at different times and locations. The flowchart of the algorithms used in the broadcasting and forwarding phases of CODOMOC is presented in Fig. 3 for clarity.

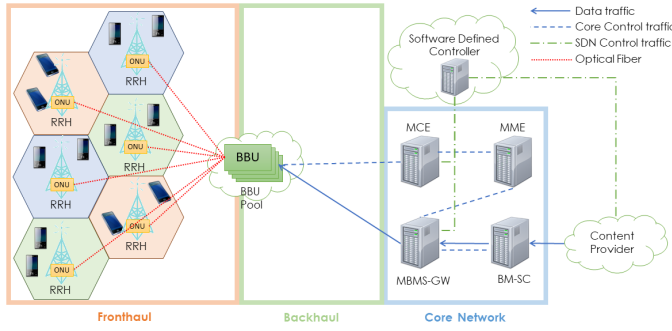


FIGURE 2. Architecture of reference system in the CRAN for content delivery.

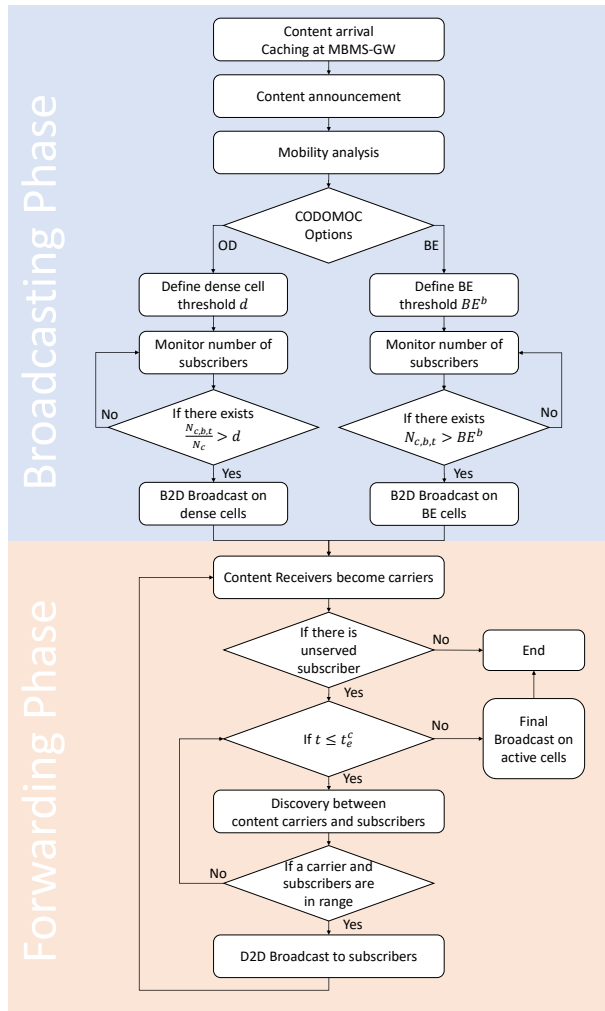


FIGURE 3. Flowchart of the algorithms used in the broadcasting and forwarding phases of CODOMOC.

A. STAGE 1: BROADCAST

In this stage, the content is to opportunistically broadcast aiming at maximizing the number of subscribers receiving the content through a selected number of base stations, while at the same time minimizing the number of base stations used for broadcasting.

To determine how the content c_m th should be broadcast, CODOMOC firstly counts number of mobile users within a cell by opportunistically exploiting the registration mechanism used by a mobile device to attach itself to a BS. This is possible because the registration scheme is utilized by the base station to collect mobile user's basic information of mobile users, as well as to establish connection between mobile device and base station [38]. The basic information includes the geographic information, channel state information (CSI), and service requirement, etc. Thus, the cell density is determined by the number of registered mobile devices at a BS. In a larger picture, this mechanism provides the building block to take advantage of human mobility patterns and network infrastructure information. The geographical density model that describes human mobility patterns within the network infrastructure context is provided in Appendix A.

Then, CODOMOC broadcasts content in dense cells and remains silent in less-dense cells. Content dissemination can be executed through the OD or BE option, providing flexibility to network operators to minimize infrastructure energy consumption or achieve a reduced carbon footprint by exploiting human mobility patterns. This helps network operators alleviate the environmental impacts of the telecommunication industry and ultimately achieve the net-zero goal in the telecommunication sector.

For those subscribers who are attached the silent base stations and do not receive the content, they receive notification, e.g., heartbeat packet, to initiate Ad-Hoc network transmission to obtain the content (Stage 2) or wait for the trickle dissemination (Stage 3). The notification is a small size packet containing information that c_m th has been disseminated and threshold time instant when stage 3 will be conducted if the subscribers aren't able to obtain the content in stage 2.

B. STAGE 2: LOCAL AD-HOC NETWORK TRANSMISSION

In stage 1, subscribers residing in silent cells may not receive the content through broadcast. To address this, CODOMOC opportunistically exploit human mobility pattern together with local ad-hoc wireless network technology, e.g., device-to-device (D2D), to facilitate content dissemination to these remaining subscribers. The content delivery is accomplished by subscribers performing content offloading, dissemination, and sharing, e.g., advertisement, sharing information between groups, and document distribution, as well as coverage extension [39]–[43]. By doing so, mobile subscribers can download content from other subscribers within proximity through short-range unlicensed (ISM) band wireless connection, such as Bluetooth or Wi-Fi connection.

The practice of forwarding scheme between two user ISM devices, e.g., mobile phone and table, for content delivery is possible through a variety of economic incentives. For example, content providers may provide subscription fee discount or bonus credits [17], or allowing the content carrier to resell the content to other subscribers in a secondary data market setting [14]–[16].

Without loss of generality, we assume that large-size content is divided into chunks for transmission to subscribers, and each chunk can be individually transmitted from a carrier to subscribers. If the encounter period is not enough for complete transmission, remaining chunks can be received from another encounter. We also assume that all subscribers are moving at a relatively slow speed compared to data transmission rate.

The objective of content forwarding is to maximize the number of recipients, but traditional broadcast schemes in ad-hoc networks may not be suitable due to the risk of broadcast storms, varying data transmission rates, and subscribers may be interested in the same content but require different parts. In this regard, we propose a pseudo broadcast scheme to enable carriers to forward content to multiple subscribers simultaneously and overcome these challenges. By doing so, CODOMOC opportunistically takes advantage of user clusters formed during daily activities, such as those sharing public transportation, dining together, and trapped in a traffic jam during peak hours. Carriers who receive content in the previous stage may forward it to nearby subscribers in need within radio ranges.

In this scheme, the carrier advertises at every channel for subscribers' device to discover and establish a unicast connection with the carrier. In other words, a single content carrier can be paired with multiple subscribers simultaneously where each pair is connected through an unicast connection. By doing so, the carrier may cater to the specific service of data forwarding according to the subscriber's device capacity, e.g. download rate, and specified part of the content. The advantage of Pseudo Broadcast Scheme is that the unicast connection provides reliability, which is not available in traditional broadcast scheme. The maximum number of pairs supported depends on the number of channels available and the processing capacity of the carrier's devices. For example, Bluetooth and Wi-fi technology that divide the band to 40 and 11 channels respectively [44].

In our design, we also considered the scenario where the transmission between two devices is interrupted due to a lack of power, and the content delivery cannot be completed promptly. To address this issue, CODOMOC allows subscribers to resume the transmission at a later time, once they regain and resume their connection with the provider or a new provider who has the desired content. This feature ensures that the user experience is not significantly impacted by power interruptions and improves the overall efficiency and reliability of the system.

C. STAGE 3: TRICKLE DISSEMINATION

To ensure the content reaches all subscribers and guarantee the fairness of the minorities that are not included in the first two stages, CODOMOC opportunistically disseminate the content to the last trickle group of subscribers by taking advantage of *Mobile IP Care of Address* (CoA) to broadcast the content at specific base stations. The IP CoA is a temporary IP address for mobile device that is used to identify the

geographical location of the device within the IP based network. Trickle Dissemination is accomplished by subscribers sending request for the desired content to the network operator. Then, the operator may utilize CoA to decide which base stations should to conduct broadcast. However, when the subscribers are in their home network, they may send their request with their IP home of address (HoA).

The trickle dissemination is triggered for content c_m when reaching a threshold time instant $t_e(c_m) - \tilde{T}_{FBC}$, where $\tilde{T}_{FBC}$ is a *controllable* broadcasting duration and $t_e(c_m)$ denotes when the content deadline c_m that must be delivered. In other words, the final stage is triggered when the deadline for second stage to be completed expires. The deadline $t_e(c_m)$ is generally determined by the content provider.

D. TOTAL DELIVERY COST

The total energy consumption over the entire allowed transmission frame for disseminating content c_m can be determined as followed.

$$\mathcal{E}_{TOT} = \mathcal{E}_{CBC}(c_m) + \mathcal{E}_{D2D}(c_m) + \mathcal{E}_{FBC}(c_m), \quad (2)$$

where $\mathcal{E}_{CBC}(c_m)$, $\mathcal{E}_{D2D}(c_m)$, and $\mathcal{E}_{FBC}(c_m)$ denote the amount of energy consume at each stage respectively. The details of the energy cost model for each stage is described in Appendix B.

V. TIME AND LOCATIONS OF DISSEMINATION

To determine when and where the content should be disseminated, we explore two different options to achieve different objectives of network operators. They are the OD and the BE option.

Before we discuss the two options in details, we first introduce an abstraction that captures the relationship between human mobility pattern, location, and time.

A. TIME BASED DENSITY GRAPH

To effectively explore the density at different times and locations, we introduce an abstraction as one of the building blocks in the design of our broadcast solution that minimizes the energy consumption. Given cellular infrastructure, we construct a directed acyclic time based density graph TD-G = (V, E) as illustrated in Fig. 4b. Let a vertex $v \in V$ represent an orthohexagonal cell covered by base station $b \in \mathbb{B}$ at time t . The index used for identifying b can be also utilized to index vertex v , as illustrated in Fig. 4a.

To construct graph TD-G, we first construct and map $v \in V$ to a tuple $(d(b, t), b, t)$, where $d(b, t)$ denotes the density of subscribers that have not received the content c_m in the region covered by base station b at time t . This mapping can be expressed as follows.

$$v(t, b) \leftarrow (d(b, t), b, t), \text{ for } \forall v \in V, \forall b \in \mathbb{B}. \quad (3)$$

The tuple (3) can be interpreted as v , it is a representation of an orthohexagonal cell covered by b at time t . Additionally,

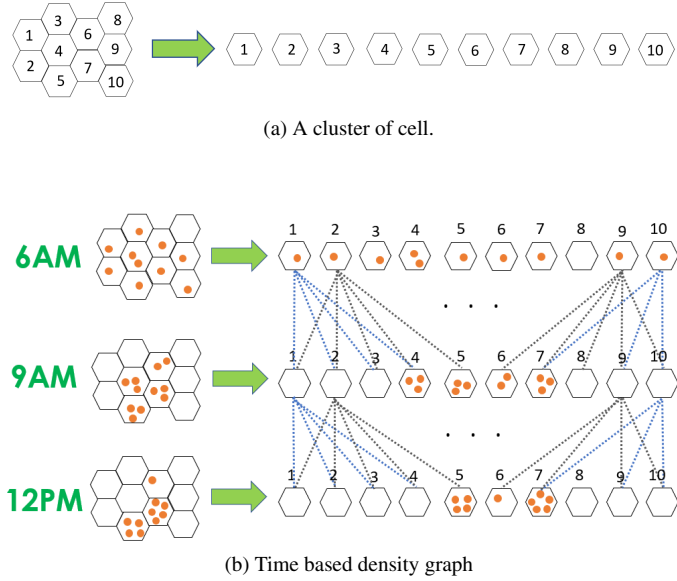


FIGURE 4. Time based density graph as an abstraction .

t describes the timing when the density of subscriber without the content is sampled. Since the operator may perform sampling K times between the content becomes available at time t_0 and its deadline at time $\mathbb{T}$, we construct $v(t, b) \in V$, such that $t = t_0 + k \times \phi$, for $t_0 \leq t \leq \mathbb{T}$ and $b \in \mathbb{B}$, where $k = 0, 1, 2, \dots, K$ and $\phi \geq 0$ denote the time interval of when the density is sampled. Thus, the number of vertices in V is $|V| = K \times |\mathbb{B}|$.

Next, we construct a directed edge $e \in E$ by connecting vertex $v(t, b_i)$ to $v_j(t + \phi, b_j)$ with edge e if b_i and b_j are geographically adjacent, where $v(t, b_i), v_j(t + \phi, b_j) \in V$, $i \neq j$, and $t_0 < t + \phi < \mathbb{T}$. The directed edge represents time chronology from time t to $t + \phi$ in TD-G. For this reason, the edge also connects $v(t, b_i)$ and $v(t + \phi, b_i)$, where both vertices are in V .

The weight of edges in E are set to 0 in this paper. This allows us to focus on how the density dynamics at different times and locations may impact energy consumption. However, our future study will leverage the weight of edges to explore how subscriber utility (satisfaction) may influence energy consumption. For instance, the weight can be interpreted as how much subscriber utility decreases as provider delays delivery time to reduce energy cost.

Cell density without the content ($d(b, t), b, t$) can be estimated as follows. First, CODOMOC selectively exploits the built-in mechanism available in MME to obtain information on the number of open sessions and which of these open sessions are associated to which BS. Then, it counts the number of users attached to each base station [45]. This information is accessible from software defined controller (SDC) that manages the core network.

Next, subscribers with recent received content send a heartbeat packet to inform the core network about their status as carriers. This is accomplished by CODOMOC opportunisti-

cally exploiting the ECN bits in the IP header that have not been used in Internet and mobile network since its deployment in year 2001 due to its incompatibility with the current Middleboxes [46]. CODOMOC sets these ECN bits in the heartbeat packet to "10" and this packet is sent when the carrier device attaches itself to a BS. Additionally, a heartbeat packet may use the broadcast IP address as its destination address and set the DHCP bits to "11111". Then, any packet that arrives from base station with these three characteristics will be treated as a heartbeat packet.

When the heartbeat packet arrives at MBMS-GW, it is forwarded to SDC after it inspects the packet. This can be done by SDC setting the forwarding policy at MBMS-GW. Next, SDC estimates number of carriers at each base station according to the number of heartbeat packets received together with the information from MME on open sessions associated to each BS. In other words, $(d(b, t), b, t)$ is the difference between total number of subscribers and number of carriers attached to BS. When the carrier leaves the BS, the carrier sends another heartbeat packet with ECN bit set to "01" when cellular handover process is initiated, and SDC adjust the estimation accordingly.

B. CODOMOC: ONLY DENSE OPTION

When a content arrives in the mobile network, it is cached in the MBMS-GW in the core network rather than being delivered to the BBU pool in the backhaul network. The mobile network then announces the arrival of the content to all of the mobile users in the region and collects subscription requests from mobile users. At the beginning of each selection time, base station b_n compares $f(n)$, the estimated total energy consumption, according to the estimated number of subscribers who have not received contents yet.

Based on the heuristic function selection scheme for the OD option, we can compare the broadcasting energy consumption $\mathcal{E}_{\text{CBC}}(c_m, b_n)$ and the estimated average D2D communication consumed energy $\sum_{u(u_k, c_m), u_k \in \mathbb{U}(b_n)} \mathcal{E}_{\text{D2D}}(u(u_k, c_m), u_k)$ in base station b_n to decide whether broadcast or not.

Since the average broadcasting and D2D communication energy consumption are both constant values, a simplified method to compare two kinds of energy consumption is to set the number of un-subscribers as a threshold. A *dense cell threshold* $d \in [0, 1]$ is set by the network operator to define how many percentages of subscribers are congregated in a cell to be labeled as a dense cell. A cell is called a dense cell, for content c_m , if the percentage of subscribers at time t is larger than the dense cell threshold d . Otherwise, it is a sparse cell. Fig. 5 illustrates an example of dense and sparse cells in a geographical scenario.

Search Methodology. Next, we design a search algorithm CODOMOC-Search (CSearch) that utilizes the TD-G graph to determine when and which base stations should broadcast (where) the content. CSearch leverages the classical Breath First Search (BFS) algorithm to explore TD-G graph starting from a designated source vertex then traverses to its

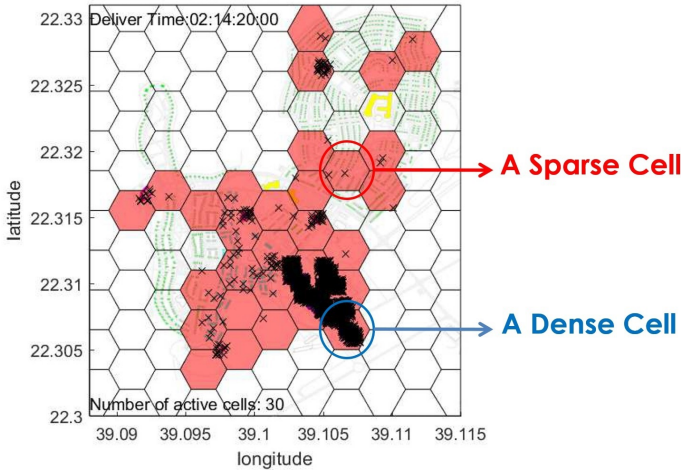


FIGURE 5. Sparse and dense cells illustration. A cell with few subscribers is considered to be a sparse cell.

neighboring vertices [47]. When CSearch visits vertex $v(t, b)$ during the exploration, CSearch compares density $d(v(t, b))$ recorded in $v(t, b)$ to density threshold d . If $d(v(t, b)) \geq d$, then CODOMOC broadcasts the content in region covered by b at time t . Moreover, CSearch allows multiple source vertices when they are not reachable from the designated starting vertex. CSearch stops exploring the graph when all the vertices in TD-G have been visited.

1) Discussion

In general, *dense cell threshold* is a function of the number of subscribers, the number of cells in the region, the mobile network topology, and the density of subscribers. Specifically, the network monitors the number of subscribers in each cell and calculate the percentage of all subscribers.

If a cell has a percentage of subscribers above the dense cell threshold, broadcasting is triggered at these cells. The network broadcasts for the first time in all dense cells, which have high subscriber density and energy consumption. These users become content carriers for the second stage, and the first stage ends with the first broadcasting stage completed.

There are several ways for the network operator to define *dense cell threshold*. For example, the threshold can be determined using economy approach, such the threshold is decided by minimum number of subscribers required to cover the operation cost per BS or to achieve the minimum profit. The threshold can also be determined according to the energy cost per subscriber that content provider is willing to pay. Another example, threshold could be learned by machine learning techniques to achieve specific objective. The specific study of how the threshold is not the focus in the paper.

In the forwarding stage, the mobile users who have received the content from the first broadcasting stage act as content carriers and forward the content to remaining subscribers through ad-hoc communication. Content carriers will discover subscribers by exchanging simple node discovery

process when their device within the radio range. Once there is a match of content carrier and subscriber, the content will then be transmitted through low-energy through ad-hoc communication technology, such as bluetooth or WiFi. The content will be broadcast by the content carrier device to subscriber devices who are within the transmission range. The subscribers who have completely received the content will become a new content carrier. This process continues until all subscribers are served before the time limit.

Finally, in order to ensure the completeness of content delivery, a final broadcast will be conducted if there are remaining subscribers. The final stage of content dissemination should be limited to a few BS-to-device (BS) transmissions since most of the subscribers are served in the two previous stages.

For example, in Fig. 1, let the network operator set the dense cell threshold as 0.2. At 9 a.m., two active cells are identified as dense cells, which have more than 20% of total subscribers in a cell in the region. Therefore, CODOMOC-OD broadcasts at 9 a.m. in only two cells in the first broadcasting stage. It reduces energy consumption in the RRH compared to the method that broadcasting to all of the 16 active cells. The other subscribers will be served by the content carriers through ad-hoc communications in the forwarding stage.

The characteristic of CODOMOC-OD is to offload part of the content delivery process from high-energy network infrastructures to low-energy mobile devices. CODOMOC-OD broadcasts the content in only dense cells, where subscribers are congregated and rely on content carriers to disseminate the content to the remaining subscribers through mobility. It releases the energy and radio resource demands for network infrastructures by not transmitting in sparse cells. However, as shown in the later evaluation section, when the number of subscribers increases, the energy consumption by ad-hoc communications increases as well as the overall energy consumption. The overall energy consumption for the delivery process could be reduced when network operators take the responsibility for broadcasting wisely to reduce the entire carbon footprint generated by content carriers. This idea motivates us to introduce the second approach of CODOMOC: The BE option.

C. CODOMOC: BROADCAST EFFICIENCY OPTION

The cost for the right side equals $\mathcal{E}_{CBC}(c_m, b_n)$, which is the base station broadcast energy consumption. Path cost for the left side equals $\sum_{u(u_k, c_m), u_k \in \mathbb{U}(b_n, t_i)} \mathcal{E}_{D2D}(u(u_k, c_m), u_k)$, which is the ad-hoc communication energy consumption in base station b_n during the time period t_i . Even though we derive a closed-form expression of energy consumption for both broadcasting and ad-hoc communications, it is still computationally inefficient for BSs to compare them at the beginning of each time period.

1) Broadcast Efficiency Threshold

Instead of using the dense cell threshold, as a criterion for determining when and which cells for broadcasting, the BE

option considers the broadcast efficiency when making the decision. Here, we define the broadcast efficiency as the number of served subscribers per unit energy to introduce the *broadcast efficiency threshold* for decision making in the first broadcasting stage. If the current number of subscribers requesting a content in a cell is larger than the BE threshold of that cell, it will be defined as a BE cell. CODOMOC-BE will broadcast the content in BE cells only. The forwarding stage remains the same as the OD option.

The BE threshold for each cell is calculated as follows. First, the network estimates the lowest SNR receiver in the cell according to a path loss model or historical record. The network bases on the lowest SNR to calculate the maximum throughput for broadcasting a content in the cell. Then, the average energy consumption per bit by B2D communications $\mathcal{E}_{\text{CBC}}(c_m)$ is calculated according to (16). Meanwhile, the network can also obtain the average energy consumption per bit by ad-hoc mobile communications according to (19), by the maximum distance between mobile nodes with minimum throughput or from averaging the historical record. Finally, the BE threshold BE^{b_n} of cell b_n is computed by dividing the energy consumption per bit by B2D communications by the energy consumption per bit in ad-hoc mobile communication as follows:

$$BE^{b_n} = \gamma \frac{\mathcal{E}_{\text{CBC}}(c_m, b_n)}{\sum_{u(u_k, c_m), u_k \in \mathbb{U}(b_n, t_i)} \mathcal{E}_{\text{D2D}}(u(u_k, c_m), u_k)} \quad (4)$$

where γ is an adjustable parameter according to the length of duration between two decisions. Since our proposed pseudo broadcast scheme is a set of unicast communication between the content carrier and subscribers. For example, a carrier doing a pseudo broadcast with N subscribers means there are N unicast communication. For this reason, $\mathcal{E}_{\text{D2D}}$ can be interpreted as the energy cost of a unicast pairing between a carrier with one of the subscribers through D2D communication in pseudo broadcast scheme. The definition and details of $\mathcal{E}_{\text{D2D}}$ is presented in Appendix B.

BE Option - Search Methodology. To incorporate the BE option in the search methodology, we extend CSearch as follows. When CSearch visits vertex v , if the density of subscribers $d(v(t, b)) \geq BE^{b_n}$, then insert v to set V_d . Otherwise, CSearch ignores v . Once CSearch has visited all vertices in TD-G, CSearch executes the following procedure. (i) Let $v_{\max}$ be a vertex with the highest density of subscribers that have not received the content in V_d . CSearch dequeues $v_{\max}$ from V_d . (ii) After that, information on b and t are retrieved from $v_{\max}$, the content is then broadcast in the region covered by b at time t . Next, (iii) CSearch adjusts the density of subscribers that have not received content in V_d , $\forall v \in V_d$; and then removes vertices in V_d if the updated $d'(v(t, b)) < BE^{b_n}$. Then, CSearch repeats the step (i) until V_d is empty. The total computation cost is the time required to visit every vertex in graph TD-G and the time required to select $v_{\max}$, which is $\mathcal{O}((|V| + |E|) + (|V_d|^2))$, $|V_d| \leq |V|$. From this calculation, it is clear that CSearch achieves polynomial-time complexity.

The key insight is that CSearch is greedily maximizing the number of subscribers receiving the content at the first stage, but at the same time it is greedily minimizing the potential energy consumption in D2D communication scheme in the second stage, which eventually leads to further reduction D2D communication. As the result, energy cost is reduced.

2) Discussion

Here we discuss the key insights of how BE option addresses the problem encountered in OD option. Firstly, when the denominator increases in (4), broadcasting with BE option becomes more cost-effective, which also implies that the energy cost of D2D communication is too high in OD option. Secondly, the usage of per bit energy cost provides more granular calculation of the actual cost of disseminating content compared to the energy cost determined using subscriber head count. This is because the energy consumption for content dissemination is directly proportionate to the size of the content, whether it is through the broadcast or D2D communication. Thirdly, the broadcast decision through BE option reduces the energy consumption of both the broadcast operation and D2D communication. This is because the threshold in (4) provides the balance between reducing the energy cost through broadcast operation and energy cost consumed through D2D communication.

In summary, a cell will broadcast the content when it is efficient in terms of energy consumption. That is, the B2D energy consumption per user is lower than the energy consumption by D2D communications. When the number of subscribers in a cell is larger than the BE threshold, broadcasting in that cell will consume less energy compared to D2D communications.

VI. EVALUATION SETUP

To evaluate CODOMOC, our simulation experiments are conducted in weekday scenario and experiment setup leverages information extracted from real data on human mobility collected for the study in [11], [18]. Direct utilization of the data may not be possible due to the size of data set available (less than 200 users), while our experiments requires a larger data set (over 6000 users). Moreover, due to privacy protection purpose, the data does not reveal the exact location of their users, which may reduce the accuracy of our evaluation if the data is directly used. For these reasons, we derive insights from this data set and generated a more realistic trace data that is suitable for our experiments.

A. GEOGRAPHICAL LOCATION

A town located in Thuwal, Makkah Province, Saudi Arabia, is considered in the simulation. It is a moderate-density living compound that facilitates both working and living environment. In the simulation, there are about 2000 townhouses and 80 two-story apartment buildings. Each townhouse populates with a family of 3-8 people, and an apartment building populates with about 20-40 people. In the compound, the university campus is the major working area for the residents, and there are three schools for primary and secondary school

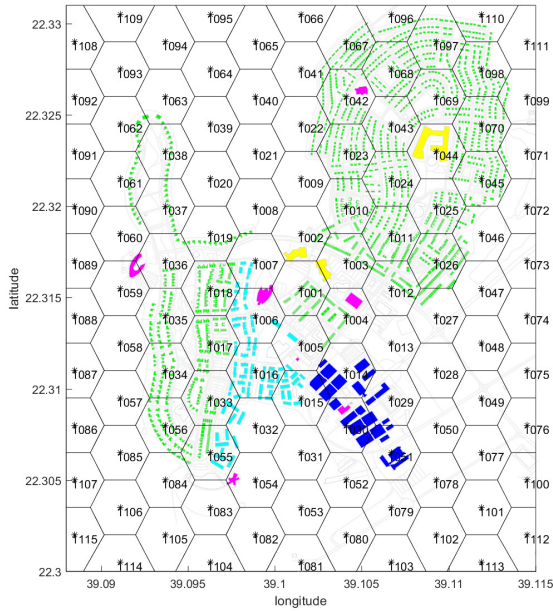


FIGURE 6. Simulation area with buildings by the following coloring: green for townhouses, cyan for apartment buildings, blue for university campus, magenta for recreational and dining areas, and yellow for primary and secondary schools. 115 cells with their names are shown in black asterisks (*) and black numbers. The black straight lines are the cell boundaries.

students. Furthermore, there are six buildings for recreation, dining, and shopping.

This simulation is conducted in a small-cell deployment, i.e., micro-cells, to show the performance of the proposed algorithm. The deployment is shown in Fig. 6. It is a typical hexagonal cell deployment with about 320 meters inter-BS distance and 115 cells in the entire region.

B. MOBILITY

In the simulation, we design four user groups, which are staff members, school students, university students, and dependants, with different daily mobility patterns for evaluation. The staying locations are randomly chosen in the simulation region, and the staying duration are randomly generated following the truncated normal distributions.

1) Staff members

The simulation includes 2,200 staff members, living in townhouses or apartment buildings and working in university campus buildings. Their daily mobility patterns involve sleeping at home before going to work in the morning, taking an average one-hour lunch break, and returning to work until evening before engaging in recreational and dining activities or returning home.

2) School students

There are 1,400 primary and secondary school students living in townhouses who attend school from 7 a.m. to 3 p.m. and

engage in active play in the region after school. They return home in the evening and have synchronized mobility patterns during school hours.

3) University students

There are 900 university students living in apartment buildings who attend classes and activities on campus during the day and engage in recreational and dining activities in the evening.

4) Dependants

The simulation includes 1,700 dependants with fixed home locations but unpredictable, random staying locations and durations. Mobility patterns start from midnight and continue until morning, with users staying at home, then traveling randomly in the simulation area for an average of three hours with one-hour variance.

5) Distinct patterns

The distinct mobility patterns of these four user groups imply various daily periodic patterns in terms of the number of users in each cell. For instance, the cells covering schools have a significant user drop after school hours, and the cells covering the university campus area have more users in working hours.

Fig. 7 shows the number of users of each user group in the five selected cells over a five-day simulation period. Each cell covers a specific type of buildings. Cell 1016 covers apartment buildings which are occupied by university students. Cell 1024 covers staff housings. Cell 1030 covers part of the university campus and the campus canteen. Cell 1044 covers a school and some staff housing. Cell 1054 covers a recreational and dining building.

In Fig. 7a, Cell 1030 shows a significant daily peak during the lunch hours when most of the staff members have their meals in the canteen. In Fig. 7b, Cell 1044, where the secondary school is located, shows a sharp increase of school students during the school hours. It shows the change from almost no users to close to 500 users during school hours. In Fig. 7c, Cell 1030 has a similar pattern observed in staff, having a sharp peak in lunch hours. Moreover, Cell 1016 which covers the student apartments, shows a wide daily peak at night when students are staying in the apartment buildings.

Finally, in Fig. 7d, the dependants have no static locations to travel or stay. Therefore, the number of users in each cell is chaotic. In the cells covering the dining area, Cell 1030 and Cell 1054, show distinct daily patterns in lunch and dinner hours. Fig. 8 shows the grouped numbers of users in these five cells. The cell patterns are highly correlated to the individual group patterns in Fig. 7. This means that a cell has an exclusive user group at a time segment.

C. CONTENTS

In the simulation, thirteen types of contents having different combinations of subscribers are evaluated. Table 1 shows the subscriber combinations for each content type. There are twelve contents in each type, and each content in a type has

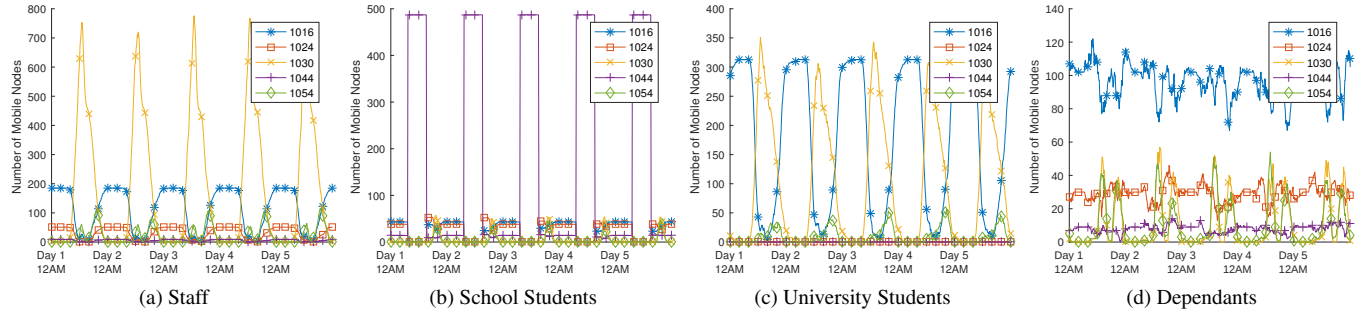


FIGURE 7. Number of users of each group in five selected cells: Cell 1016 covers university students housing; Cell 1024 covers staff housings; Cell 1030 covers part of the university campus and the campus canteen; Cell 1044 covers a school and some staff housing; Cell 1054 covers a recreational and dining building; (a) Cell 1030 has a peak for staff in every afternoon; (b) Cell 1044 has stable high numbers of students everyday in schools; (c) Cell 1016 and Cell 1030 has complimentary peaks every day for university students; (d) Cell 1054 has peaks at dining hours for dependents.

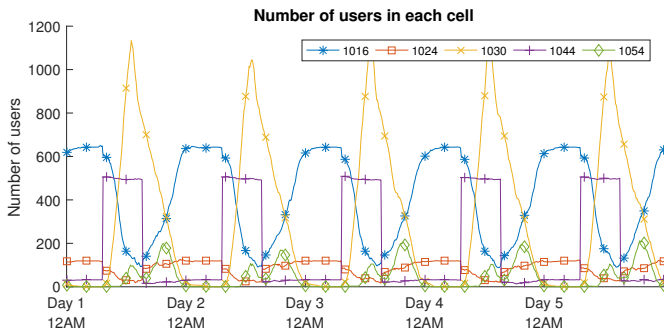


FIGURE 8. Number of users in five selected cells: Cell 1016: university students housing; Cell 1024: staff housing; Cell 1030: campus and canteen; Cell 1044: school; Cell 1054: recreational and dining.

TABLE 1. Number of Subscribers for Contents

# of users Type (ID)	Staff 2200	S. Students 1400	U. Students 900	Depend. 1700	Total subscribers
Type 1 (1-12)	✓				2200
Type 2 (13-24)		✓			1400
Type 3 (25-36)			✓		900
Type 4 (37-48)				✓	1700
Type 5 (49-60)	✓		✓		3100
Type 6 (61-72)		✓	✓		2300
Type 7 (72-84)				✓	3100
Type 8 (85-96)	✓			✓	3900
Type 9 (97-108)	✓		✓	✓	4800
Type 10 (119-120)		✓	✓	✓	4000
Type 11 (121-132)	✓	✓	✓		4500
Type 12 (133-144)	✓	✓		✓	5300
Type 13 (145-156)	✓	✓	✓	✓	6200

the same subscriber numbers. For the first content in each type, i.e., Content 1, Content 13, Content 25 etc., the arrival time is set as 12 a.m. Then, the arrival time is set as two hours later for each of the remaining contents recursively, until the last content in each type which the arrival time is set as 10 p.m., e.g., Content 1 & 13 at 12 a.m., Content 16 & 28 at 6 a.m., Content 32 & 44 at 2 p.m. The Deadline of delivering each content 72 hours upon the arrival, and each content has a size of 500 MB.

D. PARAMETERS AND ASSUMPTIONS

We set the parameters and make the assumptions for the simulation as follows: The dense cell threshold d is configured to be 0.1. There are totally 6200 mobile users in 115 cells, and 156 contents to be distributed. For the network configurations, we assume that the bandwidth B_w is 10 MHz with 2×2 MIMO for B2D communications and SISO for D2D communications. The cells are synchronized for multicasting contents as configured in MBMS single-frequency network (MBSFN). According to [48]–[51], we assume that a BBU in the BBU pool processes the baseband signal for one cell. Each BBU has one LC, and each LC has an optical channel connected to the ONU at each RRH. The BBU processing time for each radio frame is assumed to be 500 μ s. The optical line bit rate is 2457.6 Mbps, and the size of each optical frame is 38880 Byte, and the optical frame duration is 125 μ s. $P_{LC\gamma}$ is 20 W; $P_{BBU\phi}$ is 100 W; P_{CS} is 500 W; the number of transceiver chain N_{TRX} for each cell is 2, P_{ONU} is 20 W; P_0 is 56 W; P_{max} is 6.3 W; Δ_p is 2.6; P_{sleep} is 39 W, and P_{D2D} is 3 W.

VII. EVALUATION RESULTS

In this section, the performance of CODOMOC is compared with a GB method and a mobility-based (MB) method. The GB method delivers the content immediately in all active cells to subscribers once the content arrives, without attempting to minimize the number of active cells [37], [52]. The cells without subscriber for that content do not transmit. On the other hand, the MB method attempts to minimize the number of active cells with respect to different time segments within the content delivery period according to the mobility of subscribers [32].

A. CODOMOC: ONLY DENSE OPTION

1) Overall Energy Consumption

The comparison of total energy consumption for these methods is depicted in Fig. 9. Since the GB method delivered contents to all active cells immediately when the content arrives, it consumes the highest amount of energy. On average, the GB method consumes 595 kJ in the whole system

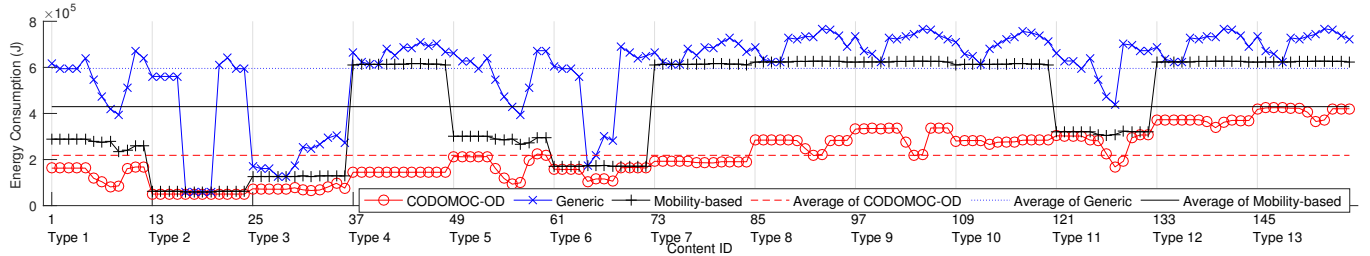


FIGURE 9. Comparison of energy consumption among GB, MB, and CODOMOC-OD: The x-axis is the Content ID, and the y-axis is the energy consumption in Joule; the red circle ($\circ$), blue cross ($\times$), and black plus (+) markers represent the total energy consumption used for disseminating a content by CODOMOC-OD, GB, and MB, respectively.

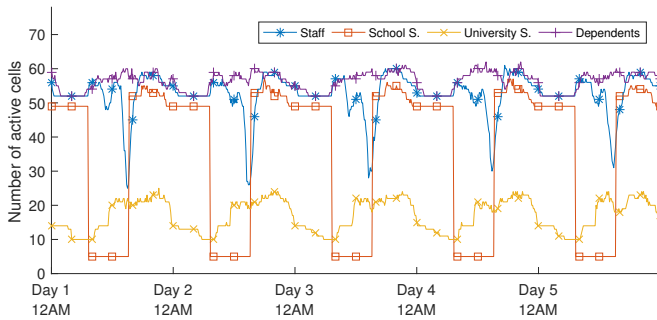


FIGURE 10. Number of active cells for each user group (equivalent to Type 1 to Type 4 contents). The statistics corresponding to staff, school students and university students have a significant local minimum on everyday. However, the local minimum with respect to the statistic of dependants is not obvious. Therefore, MB performs inappropriately for dependants in Type 4 contents.

to deliver a content to all subscribers. The content delivery time of the MB method is optimized to the time that has the minimum active cells for each content. Therefore, the total energy consumption in the system for broadcasting the contents through the network infrastructure was reduced compared to the GB method. It consumes 430 kJ on average for delivering each content. However, CODOMOC-OD further improves the performance by offloading part of the traffic from B2D communications to D2D communications. The average total energy consumption by CODOMOC-OD, including all of the B2D and D2D transmissions, is 218 kJ. It saves 63% and 51% of energy consumption from the GB and MB methods, respectively. Another observation is that the total energy consumption of CODOMOC-OD for Type 9 - 13 demonstrates an increasing trend but not by GB and MB. In CODOMOC-OD, the energy consumption increase is related to the increase of the number of subscribers. That is, there will be more frequently D2D communications established and energy consumed when the numbers of subscribers are high for Type 9 - 13.

By the MB method, the performance is highly affected by the human mobility pattern. This problem can be observed in Fig. 9 for delivering Type 1 to Type 4 contents. Type 1 to Type 3 contents are subscribed by staff, school students, and university students, respectively. These three user groups

have routine patterns in their daily lives and congregate at the campus buildings or schools every day in the office or school hours. This is interpreted as having a lower number of active cells every day as shown in Fig. 10. However, for Type 4 contents, which are subscribed by the dependants, the minimum number of active cells is much higher compared to the aforementioned content types because the dependants situate sparsely in the region. It requires more transmissions from BSs to deliver the content. Therefore, the energy consumption is relatively high for Type 4 contents when applying the MB method. This observation also suggests that the provider may achieve better energy efficiency when the subscribers are homogeneous, because of its predictability.

2) B2D Decision Comparison

The aforementioned problem is solved by CODOMOC-OD and the energy consumption has been significantly reduced for Type 4 contents. CODOMOC-OD does not broadcast in the cells which have few subscribers to reduce B2D transmissions. The improvement is shown by considering Content 44, as a case study. This content is subscribed by only dependants and arrives in the network at 2 p.m. on Day 2 in the simulation. The broadcasting decisions for this content of each method are depicted in Fig. 11.

When the content arrives at 2 p.m. on Day 2, the GB method broadcasts the content to every active cell which has subscribers. There are 57 cells transmitting the content and consuming 685 kJ in the whole system. By the MB method, the delivery time is delayed to Day 3 at 2:30 a.m. when the number of active cells is reduced to 52. It consumes 616 kJ by the MB method for these B2D transmissions. Both the GB and MB methods are not efficient since some cells consume a substantial amount of energy to serve only very few subscribers as shown in the figure.

However, CODOMOC-OD conducts B2D transmissions on the dense cells only to save energy. In this example, there is only one B2D transmission from the network infrastructure at 9 a.m. on Day 3 in Cell 1018, as marked as the red cell in the figure. There are 171 subscribers, which are 10% of all subscribers, in Cell 1018. Then, CODOMOC-OD decides to broadcast only in Cell 1018 and the subscribers become content carriers which are responsible for distributing Content 44 to other subscribers through D2D connections.

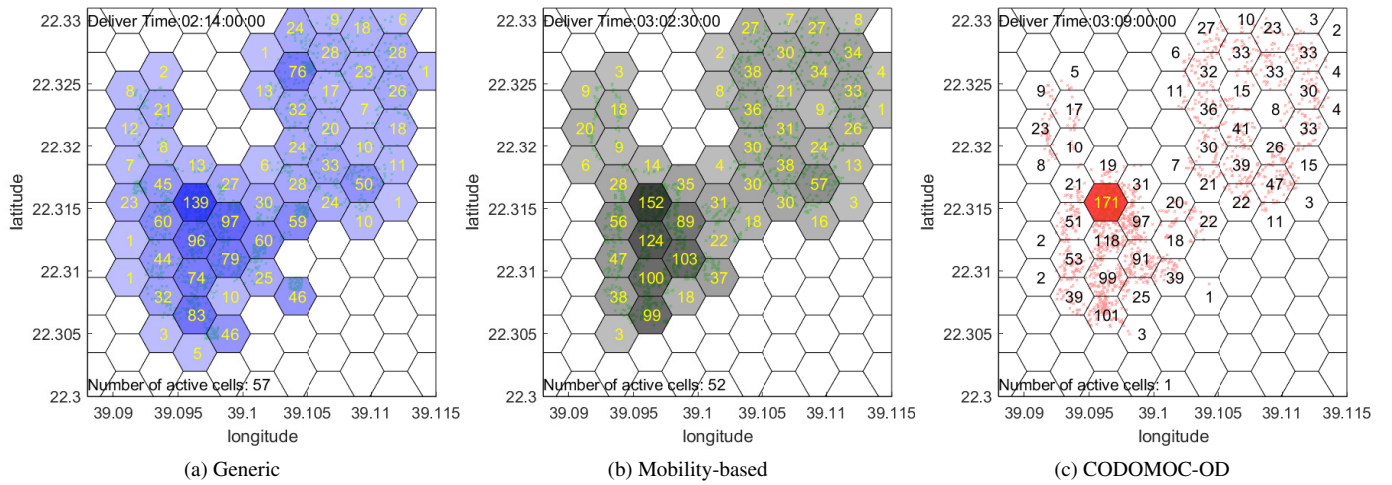


FIGURE 11. Comparison of broadcasting decision by three methods for Content 44. The number given in each cell is the number of subscribers. The colored cells are the cells conducted the B2D transmissions. The subscribers who receive the content are marked as green dots and red dots instead. CODMOC-OD significantly reduces the B2D transmissions to save energy.

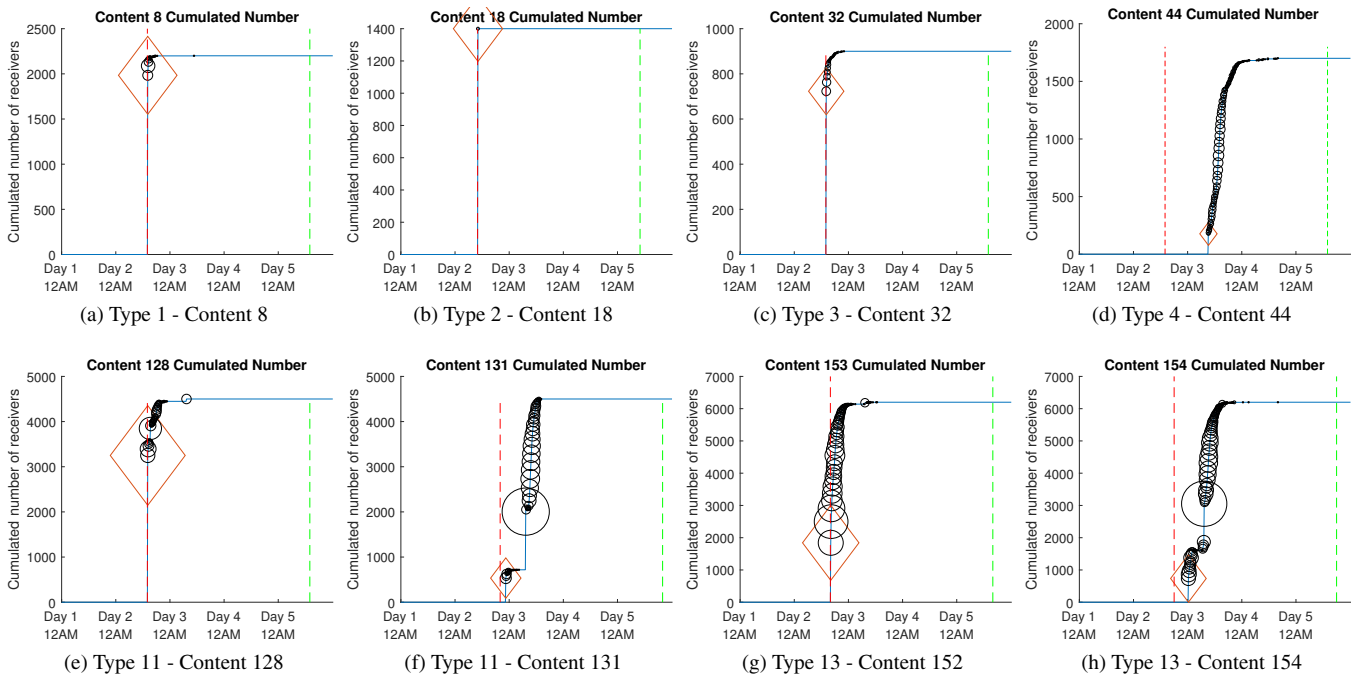


FIGURE 12. Accumulated number of receivers in CODMOC-OD. The x-axis is the time of the simulation period. The y-axis is the number of accumulated receivers. The red-dashed vertical line is the content arrival time. The green-dashed vertical line is the delivery deadline. The blue curve is the accumulated number of receivers. The diamonds indicate the first broadcasting occurs. The circles indicate the D2D transmissions. The size of the diamond and circles indicates the number of B2D and D2D transmissions in that time segment, respectively.

Therefore, CODMOC-OD consumes only 25.9 kJ on one B2D transmission in the network infrastructure, which significantly reduces the operating cost for network operators. The reduction in operating cost is due to CODMOC is able to minimize the broadcast energy cost and D2D energy usage at the same time. The average energy consumption for each D2D transmission is 77.8 J. The total energy consumption for the remaining 90% delivery by D2D transmissions is 119 kJ.

In summary, the overall energy consumption in CODMOC-OD for delivering Content 44 is 145 kJ. It reduces more than 75% on energy consumption compared to both the GB and MB methods.

Fig. 13 shows the total energy consumption of each content in B2D and D2D transmissions, respectively. It shows that the network infrastructure maintains a low energy consumption in B2D communications for all contents and offloads content

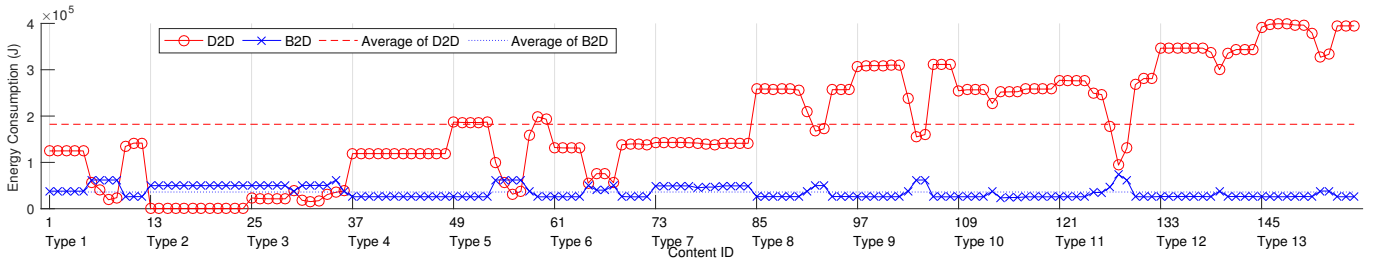


FIGURE 13. Total energy consumption by D2D and B2D communications by CODOMOC-OD. The B2D energy consumption is limited in low level. But the D2D energy consumption increases with the subscriber numbers

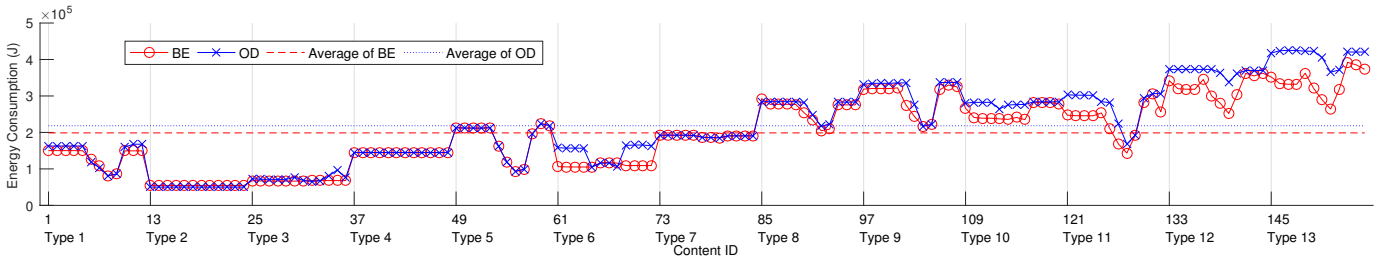


FIGURE 14. Comparison of energy consumption between CODOMOC-OD and CODOMOC-BE. CODOMOC-BE further reduces the total energy consumption.

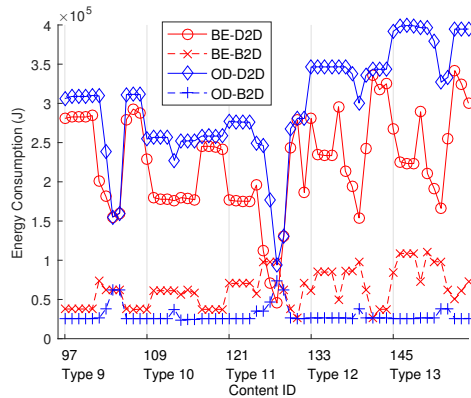


FIGURE 15. Comparison of energy consumption by D2D and B2D communications between CODOMOC-OD and CODOMOC-BE. CODOMOC-OD (blue) has a big consumption different between B2D and D2D. CODOMOC-BE (red) reduces the gap.

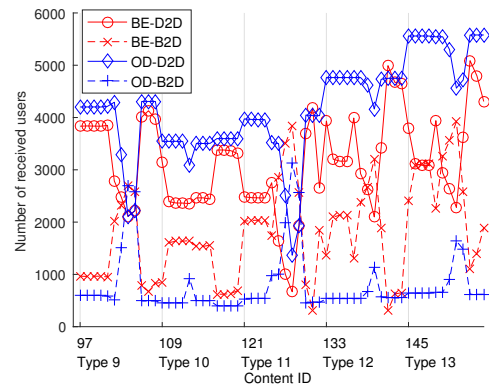


FIGURE 16. Comparison of the number of receivers by D2D and B2D communications between CODOMOC-OD and CODOMOC-BE. CODOMOC-OD (blue) relies too much on D2D. CODOMOC-BE (red) has more even distribution.

delivery tasks to content carriers.

3) Delivery Latency Analysis

Next, we evaluate the delivery latency measurement in details. Fig. 12 illustrates the delivery pattern of contents. The plots present various broadcasting and D2D distribution results for various content types and arrival time.

In Fig. 12a, Type 1 - Content 8 is subscribed only by university staff. Staff members are most congregated in the afternoon at 2-4 pm. Therefore, when Content 8 arrived at 2 pm on Day 2, 1927 out of 2200 staff members received the content in the first broadcasting stage. The figure shows a big diamond at the content arrival time. The remaining subscribers obtained the content by D2D as shown as circles in the figure. The average delivery latency for Content 8 is 3

minutes.

In Fig. 12b, Type 2 - Content 18 is subscribed only by school students. The school students arrived school at 9 am sharply. When the content arrived at 10 am on Day 2, almost all students received the content by the first broadcast. The average delivery latency for Content 18 is almost zero.

In Fig. 12d, Type 4 - Content 44 is subscribed only by dependants, which have the most random mobility without congregation. Therefore, the content does not broadcast at the content arrival time at 2 pm on Day 2. Instead, it waits until Day 3 at 9 am to perform the first broadcast to 171 subscribers. Then, rely on D2D transmissions to complete the dissemination. The average delivery latency for Content 44 is 23 hours 55 minutes.

Fig. 12e and Fig. 12f show two extreme cases for the same

type of content but with different arrival time. Type 11 is general academic content that subscribed by both university staff, school students, and university students. Content 128 arrived at 2 pm on Day 2 when most subscribers (3133 out of 4500) are staying in the university or schools. It was the best timing for the first broadcast. However, Content 131 arrived at 8 pm on Day 2 when most subscribers had been returned home. The first broadcast occurred at 10:20 pm on Day 2 to only 464 subscribers. The remaining 4036 subscribers received the content by D2D, especially at 9 am on Day 3 when school students went back to schools. The average delivery latency for Content 128 and Content 131 are 45 minutes and 11 hours 18 minutes, respectively.

Fig. 12g and Fig. 12h show the similar situation as previous. Type 13 is operating system update of mobile device that subscribed by all users in the region. Content 153 arrived at the best timing (4 pm on Day 2) that minimize the average delivery latency (1 hour 21 minutes). But Content 154 arrived at the worst timing (6 pm on Day 2) that having the largest delivery latency (12 hours 50 minutes).

In brief, the average delivery latency of a content is correlated to the arrival time. If the arrival time is at the time when subscribers are highly congregated, the average delivery latency will become low. Otherwise, the delivery latency is significant because it needs to wait until the next congregation happens according to the daily mobility pattern.

In summary, it indicates that broadcasting in dense cells minimizes the energy consumption for the network operators on their maximum benefit in terms of the cost on electricity and transfers the cost to the mobile users who act as content carriers. However, as the number of subscribers increases, the energy consumption in D2D communications increases as well as the overall energy consumption. In the next subsection, we evaluate the BE option to further reduce the total carbon footprint.

B. CODOMOC: BROADCAST EFFICIENCY OPTION

We compare the energy consumption and saving in percentage for all methods and content type in Table 2. In average, CODOMOC-OD saves 63% and CODOMOC-BE saves 65% from the GB method, saves 43% and 48% from the MB method.

The aim of the BE option is to reduce the total energy consumption by transmitting the content to the broadcast efficient cells, and thus relies more on B2D communications than D2D communications. Fig. 14 shows the total energy consumption of both options. The energy consumption can be reduced by the BE option with a larger number of subscribers in Type 12 and Type 13 contents. On average, the BE option reduces 9% of total energy consumption compared to the OD option.

In Fig. 15, the energy consumption from Type 9 to Type 13 contents of both options between B2D and D2D communications is depicted. CODOMOC-OD broadcasts in one to two cells normally in the first broadcasting stage, and let the contents be distributed by content carriers through D2D

TABLE 2. Average energy consumption and saving in percentage compared to the GB method.

kJ (saving in %)	GB	MB	CODOMOC	
			OD	BE
Type 1	558	272 (51%)	142 (75%)	133 (76%)
Type 2	410	62 (85%)	50 (88%)	55 (87%)
Type 3	213	128 (40%)	74 (65%)	67 (68%)
Type 4	666	614 (8%)	145 (78%)	145 (78%)
Type 5	571	292 (49%)	181 (68%)	181 (68%)
Type 6	498	172 (65%)	144 (71%)	110 (78%)
Type 7	669	614 (8%)	190 (72%)	190 (72%)
Type 8	704	625 (11%)	270 (62%)	261 (63%)
Type 9	717	625 (13%)	311 (57%)	294 (59%)
Type 10	702	614 (13%)	280 (60%)	255 (64%)
Type 11	612	318 (48%)	272 (56%)	233 (62%)
Type 12	704	625 (11%)	367 (48%)	322 (54%)
Type 13	717	625 (13%)	412 (43%)	338 (53%)

communications. CODOMOC-BE broadcasts in more cells which have sufficient subscribers to compensate the energy consumption for D2D communications. Broadcasting in these cells is more energy efficient than D2D communications, especially for Type 12 and 13 contents with a larger number of receivers. By taking more responsibility from the network operators, it significantly reduced the energy consumption by D2D communications.

Fig. 16 shows the number of receivers by B2D and D2D communications for both options. In the OD option, the contents mainly relies on D2D communications to disseminate. However, in the BE option, the number of B2D receivers considerably increases. For Type 13 content, in most of the cases, half of the subscribers are served by the broadcast.

The broadcast decision comparison of Content 152 is depicted in Fig. 17 for demonstrating the differences between the MB method, CODOMOC-OD, and BE options. The MB method broadcasts in 53 cells and consumes 627 kJ in the network infrastructure. The OD option broadcasts in only two dense cells and consumes only 38.2 kJ in the infrastructure. However, the energy consumption for D2D communication is 327 kJ, and the total energy consumption is 365 kJ. In the BE option, broadcast efficiency thresholds are calculated for each cell. In the simulation, most of the cells have a similar shape and size. Therefore, the broadcast efficiency threshold for all cell is similar, which was about 300 subscribers. It means that the energy consumption by D2D communications for more than 300 transmissions is higher than the energy consumption for a single broadcast in a cell. In the first broadcasting stage, the network broadcasts in 7 cells which has more than 300 subscribers. It consumes 98 kJ in the infrastructure, but only 166 kJ for the remaining D2D communications.

The BE option further outperforms the MB method and OD option by saving 57.9% and 27.6% energy, respectively. The accumulated number of receivers for both OD and BE options are shown in Fig. 18. The BE option shows a larger number of broadcast receivers and a smaller number of D2D receivers. It reduces the waiting time for most of the subscribers compared to the OD option. In other words, by considering the impact of the broadcast operation to D2D communication,

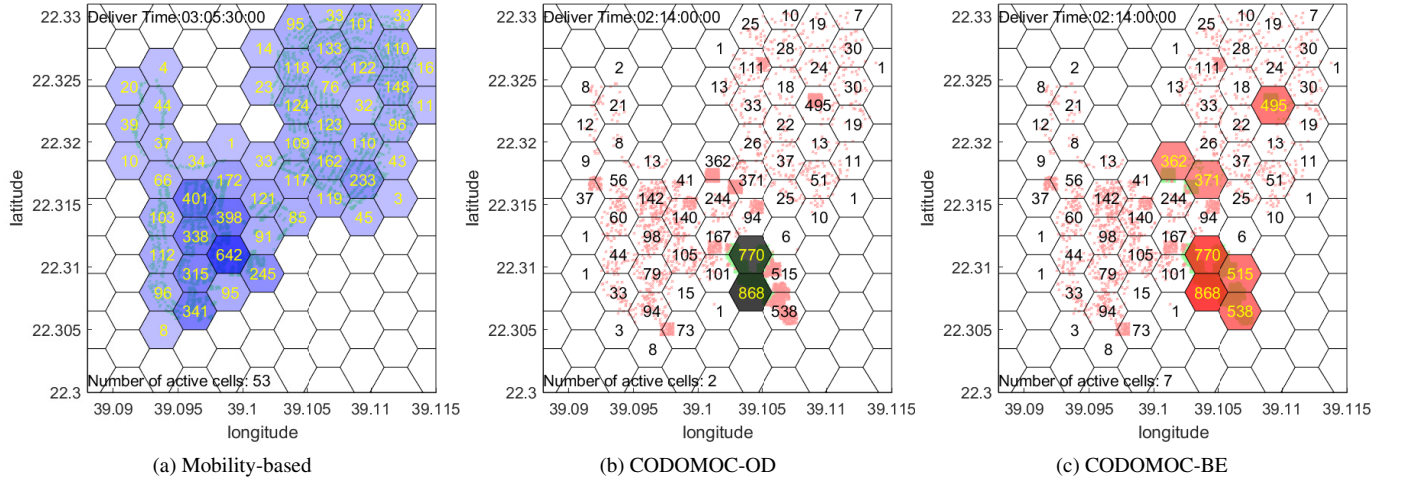


FIGURE 17. Comparison of broadcasting decision for Mobility-based, CODOMOC-OD, and CODOMOC-BE with Content 152. The number given in each cell is the number of subscribers. The colored cells are the cells conducted the B2D transmissions. The subscribers who receive the content are marked as green dots and red dots instead. CODOMOC-BE increases the B2D transmissions to take more responsibility.

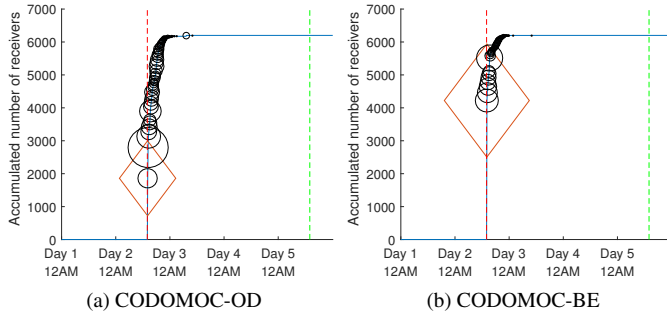


FIGURE 18. Accumulated number of receivers of Content 152 by the OD and BE options. CODOMOC-BE has a larger diamond indicates more receivers in the first broadcast.

CODOMOC-BE further reduces the overall energy consumption by increasing number of subscribers receiving the content in the first stage and at the same time reducing the reliance on D2D communication.

Overall, the OD option minimizes network infrastructure energy consumption, while the BE option reduces overall carbon footprint by increasing energy consumption slightly. CODOMOC is flexible for network operators to choose a delivery strategy that meets business and operational requirements.

VIII. CONCLUSION

In this paper, we identify a class of traffic which should not be evaluated using traditional network performance metrics. We then propose CODOMOC, a content dissemination solution that leverages human mobility patterns to minimize energy consumption. This solution combines broadcasting and ad-hoc communication to reduce transmissions, engage mobile users as content carriers, and achieve energy cost savings.

The OD and BE options, minimizing energy consumption for network operators and reducing carbon footprint in content delivery, respectively, were proposed and evaluated. Simulations with four mobility patterns showed that CODOMOC outperforms existing methods by over 50% in energy efficiency. These results validated the effectiveness and efficiency of the proposed solution and the relevant methods for I-NU content delivery in opportunistic mobile networks.

APPENDIX A GEOGRAPHICAL DENSITY

Here, we present model that describes human mobility pattern together with information on network infrastructure and time deadline to derive density of subscribers at different time and locations that is covered by various base stations. We assume that cellular network fully covered by N orthohexagonal cells, which are coordinated according to the MBSFN model in the 3GPP standard. An omnidirectional and isotropic BS is located at the center of each cell. Therefore, there are N BSs in total deployed in the ideal cellular network. Because of the bijective relation between cells and base stations, we consistently denote the sets of these N cells/BSs as $\mathbb{B} = \{b_1, b_2, \dots, b_N\}$, which are further assumed to be homogeneous.

At an arbitrary time instant t , there are M possible contents regulated by the set $\mathbb{C} = \{c_1, c_2, \dots, c_M\}$ required to be delivered to corresponding subscribers over multiple cells. Because of timeliness, an arbitrary content numbered by c_m needs to be delivered to all its subscribers between the arrival time $t_a(c_m)$ and deadline $t_e(c_m)$. Supposing there are K mobile users randomly distributed over N cells by certain mobility patterns, we let $\mathbb{U} = \{u_1, u_2, \dots, u_K\}$ be the set of mobile users. It is allowed that a mobile user does not demand any content or demand one or more contents at time instant t . The content demanded by a mobile user can be regarded as an

independent random event.

For clarity, we define a mobile user, who demands content c_m and content c_m has not been delivered yet, as the *subscriber* of content c_m . To explicitly characterize the demand, we construct a binary demand indicator $q(u_k, c_m)$, which follows the random distribution infra:

$$q(u_k, c_m) = \begin{cases} 1, & \text{with probability } \psi_c(u_k, c_m) \\ 0, & \text{with probability } 1 - \psi_c(u_k, c_m) \end{cases}, \quad (5)$$

where $\psi_c(u_k, c_m)$ is termed the content demanding probability, which belongs to the user profile information and can be extracted by analyzing the historical data of user u_k . $q(u_k, c_m) = 1$ indicates that the u_k th user demands the c_m th content, otherwise $q(u_k, c_m) = 0$. Further considering the temporal and spatial information, we equip the binary demand indicator with time instant t and location represented by the index of base station b_n and finally have $q(u_k, c_m, b_n, t)$ indicating that the u_k th subscriber located within the cell of the b_n th base station demands the c_m th content at time instant t . As a result, the total number of subscribers demanding content c_m in the cell administrated by the b_n th base station at time instant t can be formulated as follows.

$$U(c_m, b_n, t) = \sum_{u_k \in \mathbb{U}} q(u_k, c_m, b_n, t). \quad (6)$$

If all cells are taken into consideration, we can easily express the number of subscribers of content c_m at time instant t as

$$U(c_m, t) = \sum_{b_n \in \mathbb{B}} U(c_m, b_n, t) = \sum_{b_n \in \mathbb{B}} \sum_{u_k \in \mathbb{U}} q(u_k, c_m, b_n, t). \quad (7)$$

For content c_m , the b_n th cell is denoted as an active cell at time instant $t \leq T$ if there is at least one content subscriber located in that cell. The number of active cells in terms of content c_m at time instant t is denoted as $B(c_m, t)$. From the above definitions and rationale, it is clear that $B(c_m, t)$ is a joint random variable directly associated with two stochastic mechanisms: the demands of content c_m by mobile users and the locations of these mobile users.

The derivation of subscribers density at base station can be performed at BM-SC in the core network

$d(b, t)$ denotes the density of subscribers

APPENDIX B ENERGY CONSUMPTION MODEL

Here, we design energy consumption models for each of three dissemination stages described in the policy.

A. FIRST STAGE

For the energy consumption at base stations, we first introduce three important concepts: the power of BBUs P_{BBU} , power of the fronthaul network (FHN) P_{FHN} , and power of active radio units (ARUs) at the b_n th base station $P_{\text{ARU}}(c_m)$. The power of BBUs P_{BBU} can be formulated as [53]

$$P_{\text{BBU}} = \sum_{\phi} P_{\text{BBU}}(\phi) + P_{\text{CS}}, \quad (8)$$

where $P_{\text{BBU}}(\phi)$ is the power of a single active BBU ϕ and P_{CS} is the power required to maintain the communication infrastructure at the central site, e.g., cooling and monitoring. The power of the FHN can be modeled as

$$P_{\text{FHN}} = \sum_{\gamma} P_{\text{LC}}(\gamma) + \sum_{\omega} P_{\text{ONU}}(\omega), \quad (9)$$

where $P_{\text{LC}}(\gamma)$ is the power of a single active wavelength γ in a LC and $P_{\text{ONU}}(\omega)$ denotes the power of a single active ONU. For the power of ARUs at the b_n th BS, we adopt the linear approximation model proposed in [49] that diversifies the power requirements for operating and sleeping modes:

$$P_{\text{ARU}}(c_m) = \begin{cases} N_{\text{TRX}} P_{\text{MPO}} \\ \quad + \Delta_{\text{LDP}} P_{\text{OUT}}(c_m), & \text{Operating mode} \\ N_{\text{TRX}} P_{\text{SLP}}, & \text{Sleeping mode} \end{cases} \quad (10)$$

where N_{TRX} is the number of transceiver chains; P_{MPO} and P_{SLP} are the minimum output power in the operating and sleeping mode, respectively; Δ_{LDP} is a slope of the load dependent power consumption; $P_{\text{OUT}}(c_m)$ is the uniform radio frequency output power at all N base stations and should be upper bounded by the maximum radio frequency output power at maximum load P_{BMAX} .

In this paper, we adopt the Shannon capacity to estimate the transmission rate and duration, which is calculated by

$$R(u_k, b_n, t) = W \log_2(1 + \text{SNR}(u_k, b_n, t)) \quad (11)$$

for the broadcasting transmission from the b_n th base station to the u_k th subscriber, where W is the bandwidth in Hertz; $\text{SNR}(u_k, b_n, t) = P_{\text{OUT}}(c_m)G(u_k, b_n, t)/N_0$ is the instantaneous received signal-to-noise ratio (SNR) and N_0 is the average noise power. To successfully convey the signal from a base station to a content subscriber, we need to ensure

$$R(u_k, b_n, t) \geq \tilde{R}_{\text{th}}, \quad (12)$$

where $\tilde{R}_{\text{th}}$ is a *controllable* outage threshold depending on the requirement of quality of service (QoS). It can thereby be derived that

$$\text{SNR}(u_k, b_n, t) \geq 2^{\tilde{R}_{\text{th}}/W} - 1 \quad (13)$$

needs to be satisfied in order not to be in outage. Considering a MBSFN in the 3GPP standard, we suppose that the same transmit power is utilized by all base stations determined by

$$P_{\text{OUT}}(c_m) = \min \left\{ \frac{N_0 (2^{\tilde{R}_{\text{th}}/W} - 1)}{\min_{u_k \in \mathbb{U}(c_m)} \{G(u_k, b_n, t)\}}, P_{\text{BMAX}} \right\}, \quad (14)$$

where $\mathbb{U}(c_m)$ is the set of mobile users in the whole area covered by all N cells and $|\mathbb{U}(c_m)| = U(c_m, t)$.

Now, let us introduce three important time intervals. $T_{\text{TX}}(c_m)$ is the uniform radio transmission duration in all K cells for the content c_m , which is decided by the content size $\Xi(c_m)$, transmission rate and the network protocol; Again,

considering the MBSFN in the 3GPP standard, we set the uniform transmission duration to be

$$T_{TX}(c_m) = \min \left\{ \frac{\Xi(c_m)}{\min_{u_k \in \mathbb{U}(c_m)} \{R(u_k, b_n, t)\}}, t_e(c_m) - t \right\}; \quad (15)$$

T_{BBU} is the BBU processing duration; T_{OPT} is the optimal transmission duration, which can be determined by the content size, link rate, and equipment processing capability [50].

Based on the modeling of power consumption and the definitions of the three important time intervals, we can now determine the energy consumption for the first cellular broadcasting for delivering content c_m in the radio access, fronthaul, and backhaul networks as

$$\begin{aligned} \mathcal{E}_{CBC}(c_m) &= P_{BBU}T_{BBU} + P_{FHN}T_{OPT} \\ &+ B(c_m, t)(N_{TRX}P_{MPO} + \Delta_{LDP}P_{OUT}(c_m))T_{TX}(c_m) \\ &+ (N - B(c_m, t))N_{TRX}P_{SLP}T_{TX}(c_m). \end{aligned} \quad (16)$$

B. SECOND STAGE

In a similar manner, we can also have the Shannon capacity for the D2D transmission denoted as $R(u(u_k, c_m), u_k, t + \lambda\delta)$ as well as the condition of not being in outage in terms of $\text{SNR}(u(u_k, c_m), u_k, t + \lambda\delta)$. According to [54], the power consumption of the D2D transmission between the $u(u_k, c_m)$ th mobile device to the u_k th mobile device $P_{D2D}(u(u_k, c_m), u_k, t + \lambda\delta)$ can be modeled as the base power of the transmission chain in the radio frequency components and baseband components. For simplicity, we can let [55]

$$\begin{aligned} P_{D2D}(u(u_k, c_m), u_k, t + \lambda\delta) \\ = \min \left\{ \frac{N_0 (2^{\tilde{R}_{th}/W} - 1)}{\Gamma(u(u_k, c_m), u_k, t + \lambda\delta)}, P_{DMAX} \right\}, \end{aligned} \quad (17)$$

where

$$\begin{aligned} \Gamma(u(u_k, c_m), u_k, t + \lambda\delta) &= \frac{\mathbb{E}\{g(u(u_k, c_m), u_k, t + \lambda\delta)\}}{(\Delta(u(u_k, c_m), u_k, t + \lambda\delta))^\alpha} \\ &= 1/(\Delta(u(u_k, c_m), u_k, t + \lambda\delta))^\alpha \end{aligned} \quad (18)$$

Here, we have P_{DMAX} is the maximum allowed transmit power by D2D mobile user equipment. Note that, D2D communications do not traverse the base station or core network and is thereby non-transparent to the cellular protocols [56]. As a result, the constraints on MBSFNs in the 3GPP standard do not apply to the D2D forwarding stage, and the transmit power of different D2D pair is allowed to be different.

By defining $T_{D2D}(u(u_k, c_m), u_k)$ as the transmission duration between the $u(u_k, c_m)$ th mobile device to the u_k th mobile device, the energy consumption for the single D2D pair can be expressed as

$$\begin{aligned} \mathcal{E}_{D2D}(u(u_k, c_m), u_k) \\ = P_{D2D}(u(u_k, c_m), u_k)T_{D2D}(u(u_k, c_m), u_k), \end{aligned} \quad (19)$$

where²

$$\begin{aligned} T_{D2D}(u(u_k, c_m), u_k) \\ = \min \left\{ \frac{\Xi(c_m)}{R(u(u_k, c_m), u_k, t + \lambda\delta)}, t_e(c_m) - (t + \lambda\delta) \right\}. \end{aligned} \quad (20)$$

C. FINAL STAGE

In the final broadcasting stage, we need to account for the energy consumption by broadcasting from base stations to content subscribers with P_{BMAX} as

$$\begin{aligned} \mathcal{E}_{FBC}(c_m) &= \beta(c_m, t_e(c_m) - \tilde{T}_{FBC}) \\ &\times (N_{TRX}P_{MPO} + \Delta_{LDP}P_{BMAX})\tilde{T}_{FBC}, \end{aligned} \quad (21)$$

where $\beta(c_m, t_e(c_m) - \tilde{T}_{FBC})$ is a dependent random variable representing the number of cells, in which there are still unserved subscribers of content c_m at time $t_e(c_m) - \tilde{T}_{FBC}$.

Therefore, the total energy consumption over the entire allowed transmission frame for disseminating content c_m can be determined by

$$\begin{aligned} \mathcal{E}_{TOT}(c_m) &= \mathcal{E}_{CBC}(c_m) + \mathcal{E}_{FBC}(c_m) \\ &+ \sum_{b_n \in \mathbb{U}} \left(\sum_{u_k \in \mathbb{D}(c_m, b_n)} \mathcal{E}_{D2D}(u(u_k, c_m), u_k) \right), \end{aligned} \quad (22)$$

where $\mathbb{D}(c_m, b_n)$ denotes the sets of subscribers of content c_m in the cell coordinated by the b_n th BS. By adjusting $\tilde{\Delta}_{th}$, $\tilde{R}_{th}$, and $\tilde{T}_{FBC}$, as well as the criterion to determine active cells, we are able to control $\mathcal{E}_{TOT}(c_m)$ to some extent. However, it is obvious that $\mathcal{E}_{TOT}(c_m)$ contains stochastic terms depending on future uncertain states of mobile users, i.e., how, where, and when mobile users are moving. Also, these states are hardly predicted in an accurate way. As a consequence of the randomness, it is impossible to guarantee a minimum energy consumption without precisely knowing the future states. In this regard, we propose two heuristic optimization schemes in the next section to reduce $\mathcal{E}_{TOT}(c_m)$ by data analytics based on current states of the cellular network and the statistical information, e.g., user mobility patterns, which have already been known.

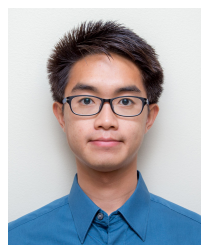
REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017 - 2022 White Paper," Cisco, Tech. Rep., Feb. 2019.
- [2] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.
- [3] C. P. Lau, A. Alabbasi, and B. Shihada, "An Efficient Live TV Scheduling System for 4G LTE Broadcast," *IEEE Systems J.*, vol. 11, no. 4, pp. 2737–2748, Dec. 2017.
- [4] L. Shi, K. W. Sung, and J. Zander, "Future TV Content Delivery Over Cellular Networks From Urban to Rural Environments," *IEEE Trans. Wireless Comm.*, vol. 14, no. 11, pp. 6177–6187, Nov. 2015.

²It is stipulated that once the entire content c_m has been received by the subscriber, an acknowledgement (ACK(c_m)) will be sent back to the D2D content provider to terminate the transmission, so that the D2D transmission time is adaptive. Again, due to the non-transparency of D2D communications to the cellular protocols, the transmission duration corresponding to each D2D pair is allowed to be different.

- [5] K. Vaniea and Y. Rashidi, "Tales of software updates: The process of updating software," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, May 7-12, 2016*, J. Kaye, A. Druin, C. Lampe, D. Morris, and J. P. Hourcade, Eds. ACM, 2016, pp. 3215–3226.
- [6] M. Fagan, M. M. H. Khan, and R. Buck, "A study of users' experiences and beliefs about software update messages," *Comput. Hum. Behav.*, vol. 51, pp. 504–519, 2015.
- [7] K. Vaniea, E. J. Rader, and R. Wash, "Betrayed by updates: how negative experiences affect future security," in *CHI Conference on Human Factors in Computing Systems, CHI '14, Toronto, ON, Canada - April 26 - May 01, 2014*, M. Jones, P. A. Palanque, A. Schmidt, and T. Grossman, Eds. ACM, 2014, pp. 2671–2674.
- [8] A. Mathur, N. Malkin, M. Harbach, E. Péér, and S. Egelman, "Quantifying users' beliefs about software updates," *CoRR*, vol. abs/1805.04594, 2018. [Online]. Available: <http://arxiv.org/abs/1805.04594>
- [9] Z. Fan et al., "CityMomentum: An Online Approach for Crowd Behavior Prediction at a Citywide Level," in *Proc. 2015 ACM Int'l Joint Conf. Pervasive and Ubiquitous Computing (UbiComp 15)*, Sep. 2015, pp. 559–569.
- [10] X. Song, R. Shibasaki, N. J. Yuan, X. Xie, T. Li, and R. Adachi, "Deepmob: Learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data," *ACM Trans. Inf. Syst.*, vol. 35, no. 4, pp. 41:1–41:19, 2017.
- [11] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 791–800.
- [12] A. Noulas et al., "A Tale of Many Cities: Universal Patterns in Human Urban Mobility," *PLOS ONE*, vol. 7, May 2012.
- [13] F. Xu, Y. Li, M. Chen, and S. Chen, "Mobile cellular big data: linking cyberspace and the physical world with social ecology," *IEEE Network*, vol. 30, no. 3, pp. 6–12, 2016.
- [14] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A survey on big data market: Pricing, trading and protection," *IEEE Access*, vol. 6, pp. 15 132–15 154, 2018.
- [15] H. Susanto et al., "Effective Mobile Data Trading in Secondary Ad-hoc Market with Heterogeneous and Dynamic Environment," in *Proc. 2017 IEEE 37th Int'l Conf. on Distributed Computing Systems (ICDCS 17)*, June 2017, pp. 645–655.
- [16] H. Zhang, B. Liu, H. Susanto, G. Xue, and T. Sun, "Incentive mechanism for proximity-based mobile crowd service systems," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9.
- [17] D. Ying and F. Ye, "Incentive d2d off-loading scheme for energy efficiency in the next-generation mobile networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–5.
- [18] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th International Conference on Ubiquitous Computing*. New York, NY, USA: Association for Computing Machinery, 2008, p. 312–321.
- [19] S. Brown and C. J. Sreenan, "Software updating in wireless sensor networks: A survey and lacunae," *J. Sens. Actuator Networks*, vol. 2, no. 4, pp. 717–760, 2013.
- [20] L. Bettini, R. D. Nicola, and M. Loreti, "Software update via mobile agent based programming," in *Proceedings of the 2002 ACM Symposium on Applied Computing (SAC), March 10-14, 2002, Madrid, Spain*, G. B. Lamont, H. Haddad, G. A. Papadopoulos, and B. Panda, Eds. ACM, 2002, pp. 32–36.
- [21] G.M. Lee et al., "Mobile Video Delivery via Human Movement," in *Proc. 2013 IEEE 10th Int'l Conf. Sensing, Comm. and Networking (SECON 13)*, June 2013, pp. 406–414.
- [22] T. Rumeng, W. Tong, S. Ying, and H. Yanpu, "Intelligent energy saving solution of 5g base station based on artificial intelligence technologies," in *2021 IEEE International Joint EMC/SIPI and EMC Europe Symposium*, 2021, pp. 739–742.
- [23] X. Li, "A computing offloading resource allocation scheme using deep reinforcement learning in mobile edge computing systems," *Journal of Grid Computing*, vol. 19, 09 2021.
- [24] Y. Wu et al., "Challenges of Mobile Social Device Caching," *IEEE Access*, vol. 4, pp. 8938–8947, Dec. 2016.
- [25] C.P. Lau, A. Alabbasi, and B. Shihada, "On the Analysis of Human Mobility Model for Content Broadcasting in 5G Networks," in *Proc. 2017 IEEE 28th Ann. Int'l Symp. Personal, Indoor, and Mobile Radio Comm. (PIMRC 17)*, Oct. 2017.
- [26] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Optimal Predictive Resource Allocation: Exploiting Mobility Patterns and Radio Maps," in *Proc. 2013 IEEE Global Comm. Conf. (GLOBECOM 13)*, Dec. 2013, pp. 4877–4882.
- [27] Y. Qiao et al., "Mobile big-data-driven rating framework: Measuring the relationship between human mobility and app usage behavior," *IEEE Network*, vol. 30, no. 3, pp. 14–21, May/June 2016.
- [28] R. Dubey, P. K. Mishra, and S. Pandey, "An energy efficient scheme by exploiting multi-hop d2d communications for 5g networks," *Physical Communication*, vol. 51, p. 101576, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1874490721002780>
- [29] A. A. Esswie, "Power saving techniques in 3gpp 5g new radio: A comprehensive latency and reliability analysis," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE Press, 2022, p. 66–71. [Online]. Available: <https://doi.org/10.1109/WCNC51071.2022.9771950>
- [30] Y. Liu, X. Qin, Y. Huang, L. Tang, and J. Fu, "Maximizing energy efficiency in hybrid overlay-underlay cognitive radio networks based on energy harvesting-cooperative spectrum sensing," *Energies*, vol. 15, no. 8, 2022. [Online]. Available: <https://www.mdpi.com/1996-1073/15/8/2803>
- [31] P. Suski, J. Pohl, and V. Frick, "All you can stream: Investigating the role of user behavior for greenhouse gas intensity of video streaming," ser. ICT4S2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 128–138. [Online]. Available: <https://doi.org/10.1145/3401335.3401709>
- [32] C.P. Lau, A. Alabbasi, and B. Shihada, "An Efficient Content Delivery System for 5G CRAN Employing Realistic Human Mobility," *IEEE Tran. Mobile Computing*, vol. 18, no. 4, pp. 742–756, Apr. 2019.
- [33] I. Gross, B. Gross, and P. Perebinosoff, *Programming for TV, Radio and The Internet Strategy, Development and Evaluation*. London: Routledge, 2005.
- [34] Z. Niu, "TANGO: traffic-aware network planning and green operation," *IEEE Wirel. Commun.*, vol. 18, no. 5, pp. 25–29, 2011.
- [35] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surv. Tutorials*, vol. 13, no. 4, pp. 524–540, 2011.
- [36] A. Checko et al., "Cloud RAN for Mobile Networks - A Technology Overview," *IEEE Comm. Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.
- [37] D. Lecompte and F. Gabin, "Evolved Multimedia Broadcast/Multicast Service (eMBMS) in LTE-Advanced: Overview and Rel-11 Enhancements," *IEEE Comm. Magazine*, vol. 50, no. 11, pp. 68–74, Nov. 2012.
- [38] G. Chen et al., "Dual Antenna Selection in Secure Cognitive Radio Networks," *IEEE Trans. on Vehicular Technology*, vol. 65, no. 10, pp. 7993–8002, Oct. 2016.
- [39] M. Ji, G. Caire, and A. F. Molisch, "Wireless Device-to-Device Caching Networks: Basic Principles and System Performance," *IEEE Journal on Selected Areas in Comm.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.
- [40] A. Asadi and V. Mancuso, "On the Compound Impact of Opportunistic Scheduling and D2D Communications in Cellular Networks," in *Proc. 2013 ACM 16th Int'l Conf. on Modeling, Analysis; Simulation of Wireless and Mobile Systems (MSWiM 13)*, Nov. 2013, pp. 279–288.
- [41] A. Asadi and V. Mancuso, "Energy Efficient Opportunistic Uplink Packet Forwarding in Hybrid Wireless Networks," in *Proc. 2013 ACM 4th Int'l Conf. Future Energy Systems (e-Energy 13)*, May 2013, pp. 261–262.
- [42] Q. Wang and B. Rengarajan, "Recouping opportunistic gain in dense base station layouts through energy-aware user cooperation," in *Proc. 2013 IEEE 14th Int'l Symp. "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM 2013)*, Aug. 2013.
- [43] (M. Botsov et al., "Location Dependent Resource Allocation for Mobile Device-to-Device Communications," in *Proc. 2014 IEEE Wireless Comm. and Networking Conf. (WCNC 14)*, Apr. 2014, pp. 1679–1684.
- [44] B. Han, J. Li, and A. Srinivasan, "On the energy efficiency of device discovery in mobile opportunistic networks: A systematic approach," *IEEE Trans. Mob. Comput.*, vol. 14, no. 4, pp. 786–799, 2015.
- [45] Cisco, *MME Administration Guide, StarOS Release 21.9*. San Jose: Cisco, 2018.
- [46] A. M. Mandalari, A. Lutu, B. Briscoe, M. Bagnulo, and O. Alay, "Measuring ecn++: Good news for ++, bad news for ecn over mobile," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 180–186, 2018.
- [47] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, 3rd Edition*. MIT Press, 2009.

- [48] X. Wang, A. Alabbasi, and C. Cavdar, "Interplay of Energy and Bandwidth Consumption in CRAN with Optimal Function Split," in *Proc. 2017 IEEE Int'l Conf. Comm. (ICC 17)*, May 2017.
- [49] G. Auer et al., "How Much Energy Is Needed to Run a Wireless Network?" *IEEE Wireless Comm.*, vol. 18, no. 5, Oct. 2011.
- [50] A. Alabbasi and C. Cavdar, "Delay-Aware Green Hybrid CRAN," in *Proc. 2017 IEEE 15th Int'l Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 17)*, May 2017.
- [51] N. Nikaein, "Processing Radio Access Network Functions in the Cloud: Critical Issues and Modeling," in *Proc. 2015 ACM 6th Int'l Workshop Mobile Cloud Computing and Services (MCS 15)*. ACM, Sep. 2015, pp. 36–43.
- [52] T. Tran, D. Navrátil, P. Sanders, J. Hart, R. Odarchenko, C. Barjau, B. Altman, C. Burdinat, and D. Gomez-Barquero, "Enabling multicast and broadcast in the 5g core for converged fixed and mobile networks," *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 428–439, 2020.
- [53] A. Alabbasi, X. Wang, and C. Cavdar, "Optimal Processing Allocation to Minimize Energy and Bandwidth Consumption in Hybrid CRAN," *IEEE Trans. on Green Comm. and Networking*, vol. 2, no. 2, pp. 545–555, June 2018.
- [54] A.R. Jensen et al., "LTE UE Power Consumption Model: For System Level Energy and Performance Optimization," in *Proc. 2012 IEEE Vehicular Technology Conf. (VTC Fall 12)*, Sep. 2012.
- [55] S. Dang, G. Chen, and J. P. Coon, "Outage Performance Analysis of Full-Duplex Relay-Assisted Device-to-Device Systems in Uplink Cellular Networks," *IEEE Trans. on Vehicular Technology*, vol. 66, no. 5, pp. 4506–4510, May 2017.
- [56] A. Asadi, Q. Wang, and V. Mancuso, "A Survey on Device-to-Device Communication in Cellular Networks," *IEEE Comm. Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, Fourthquarter 2014.



CHUN PONG LAU (Member, IEEE) received the Bachelor's degree in Computer Engineering from Hong Kong University of Science and Technology (HKUST), Hong Kong in 2008. He also received Master's and Ph.D. degree in Computer Science from King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2010 and 2018, respectively. His research interests include content delivery, video broadcasting, and energy efficiency in mobile networks.



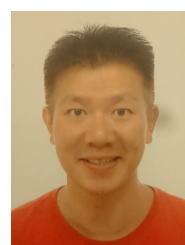
HENGKY SUSANTO (Member, IEEE) received B.Sci. degree in Computer Science from the University of Massachusetts Amherst, M.S. and PhD degree in computer science from the University of Massachusetts Lowell. He was a postdoctoral researcher/fellow at the University of Massachusetts Lowell and Hong Kong University of Science and Technology. After that role, he became a senior researcher at Future Network Theory Lab in Hong Kong. He is currently a principal research scientist in a startup mode research laboratory. He received several awards including the best USENIX LISA best paper award, EMC Achievement award, and StorageNetwork Shine Award. His current interests include datacenter network, IoT, mobile computing, and network economics



SHUPING DANG (Member, IEEE) received B.Eng (Hons) in Electrical and Electronic Engineering from the University of Manchester (with first class honors) and B.Eng in Electrical Engineering and Automation from Beijing Jiaotong University in 2014 via a joint '2+2' dual-degree program. He also received D.Phil in Engineering Science from University of Oxford in 2018. Dr. Dang joined in the R&D Center, Huanan Communication Co., Ltd. after graduating from University of Oxford and worked as a Postdoctoral Fellow with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST). He is currently a Lecturer with Department of Electrical and Electronic Engineering, University of Bristol. The research interests of Dr. Dang include wireless communications, wireless security, and machine learning for communications.



GUOQING MA (Student Member, IEEE) received his B.Sc degree in South University of Science and Technology of China, Shenzhen, China, in 2017. Now he is a Ph.D. candidate in computer science at King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. His main research interests include big data analysis and large-scale system design.



KAM SHING NG (Member, IEEE) received B.Sci. degree with first class honors in Information Technology from the City University of Hong Kong and a PhD degree in Electronics Engineering from the same university, supported by a Croucher Foundation Scholarship. His research interests include AI and data coding.



BASEM SHIHADA (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Waterloo. He is currently a professor of the Computer Science and Electrical and Computer Engineering Programs, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST). In 2009, he was appointed as a Visiting Faculty Member of the Department of Computer Science, Stanford University. His work received several awards and honors, including two best paper awards in AFIN and IEEE WiMob. He has coauthored more than 170 papers in the field of wireless networks. His research interests include the topics in energy and resource allocation in wired and wireless networks, software defined networking, cloud/edge computing, the Internet of Things, data networks, and underwater networks. He serves as a Senior Editor for IEEE Communications Letters, an Associate Editor for *Frontiers in Communication and Networks*, and an Editor for *Photonic Network journal*. He served as the Track Chair, the Organizing Chair, the Publication Chair, and the Technical Program Committee for several conferences, such as VTC, InfoCom, Globecom, ICC, Mobiquitous, WCNC, and others. He is a professional member of ACM. His work on underwater networks, specifically Aqua-Fi, got the world attention and resulted into wide international media coverage.

...