Joint Sponsor Scheduling in Cellular and Edge Caching Networks for Mobile Video Delivery

Lifeng Sun, *Member, IEEE*, Haitian Pang, *Student Member, IEEE*, and Lin Gao, *Senior Member, IEEE*

***Abstract*—The explosive growth of mobile video traffic introduces new challenges for the network infrastructure. Edge caching, as one of the key technologies in 5G wireless networks, has shown great potential to improve the quality of mobile video services by reducing the transmission overhead over backhaul links. With edge caching, content providers (CPs) need to decide not only the traditional data sponsoring strategy on cellular networks (where CPs cover part or all of the mobile users' cellular data cost), but also a novel cache sponsoring strategy on the edge caching networks (where CPs place part of contents on edge networks in advance). In this paper, we study the joint optimization of both sponsors on cellular and edge caching networks for a single CP, aiming at maximizing the CP's revenue. Specifically, we formulate the joint optimization problem as a two-stage sequential decision problem. In stage I, the CP determines the edge caching policy (for a relatively long time period). In stage II, the CP decides the real-time data sponsoring strategy for each content request within the period. We analyze this two-stage decision problem systematically. First, we propose an online sponsoring strategy in stage II based on Lyapunov optimization framework. Then, we propose an edge caching strategy in stage I via predicting the number of aggregate user requests. Simulations on real data traces show that such a joint optimization policy can increase the CP's revenue by 124%–154%, comparing with the traditional data sponsoring policy (i.e., without edge caching). Moreover, the proposed online strategy can achieve 90% of the maximum revenue in the offline benchmark.**

***Index Terms*—Data sponsoring, edge caching, mobile video streaming.**

I. Introduction

A. Background and Motivation

NOWADAYS, we are witnessing the explosive growth of global mobile data traffic, which has reached 7.2 exabytes per month in 2016, where mobile video traffic accounts for 60% of the total mobile traffic. According to Cisco [1], the percentage of mobile video traffic is expected to increase to 78% by 2021. Such a rapid increase of mobile video traffic brings new challenges for video content providers (CPs). For example, it introduces a tremendous traffic burden on the content delivery networks (CDNs) and hence may cause severe network congestion [2], [3]. Thus, various approaches (e.g., cooperative streaming [4]–[6]) have been proposed to reduce the negative effects brought by the explosion of mobile video traffic.

On the other hand, the competition among CPs is becoming increasingly intensive. Thus, it is important for a CP to implement novel incentive paradigms to attract more video users and traffic, so as to generate more revenue.

Data sponsoring is a novel and effective method recently introduced by CPs to expend user demand for video content [7]–[11]. The key idea is to allow CPs to subsidize video users' cost for downloading video data, thereby attracting more mobile video users and traffic. With data sponsoring, mobile users benefit from the free wireless access for video content, and CPs benefit from the increased video users and traffic (through, for example, selling more built-in advertisements). Thus, it can lead to a win-win situation for CPs and users, hence has attracted the interests of both academia and industry. As a real-world example, AT&T announced its sponsored data program in January 2014, in which AT&T allows advertisers to sponsor mobile data to entice users to watch ads that they might have otherwise avoided [12].

Traditional data sponsoring can potentially help a CP to attract more video traffic and generate more revenue, but at the same time, it will also increase the burden of CDNs due to the increased video traffic, and hence increase the CP's cost of delivering content to video users. **Edge caching** is emerging as a promising paradigm to alleviate the burden of CDNs, reduce the content delivery cost of the CP, and also reduce the energy consumption of mobile users [13]–[16]. The key idea is to cache popular video content on edge networks in advance and deliver the cached content on edge networks to the local users through WiFi or femtocell links directly. Edge caching can alleviate the traffic burden of congested cellular networks and reduce the energy consumption of mobile users. Thus, it can deliver content with a lower cost. As a real-world example, Xunlei, one of the largest online content download service providers in China, has adopted a new service that utilizes users' bandwidth and storage capacities to implement edge caching [17]. Xunlei offers the edge network resources to CPs, allowing them to replicate

Manuscript received September 6, 2016; revised May 2, 2017, December 25, 2017, and March 15, 2018; accepted April 8, 2018. Date of publication May 9, 2018; date of current version November 15, 2018. The work of L. Sun and H. Pang was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61472204 and Grant 61521002, in part by Beijing Key Laboratory of Networked Multimedia under Grant Z161100005016051, and in part by the Alibaba Cooperation Funding. The work of L. Gao was supported by the NSFC under Grant 61771162. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria G. Martini. *(Lifeng Sun and Haitian Pang contributed equally to this work.) (Corresponding author: Haitian Pang.)*

L. Sun and H. Pang are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: pht14@mails.tsinghua.edu.cn; sunlf@tsinghua.edu.cn).

L. Gao is with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen 150001, China (e-mail: gaol@hit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2018.2834861

content for the accesses of neighbor users, and the CPs pay Xunlei a fee in return.

However, all of the above works, to the best of our knowledge, considered sponsored data and edge caching separately. In this work, we aim to solve the CP payoff maximization problem by jointly considering data sponsoring and edge caching under a fixed budget. Joint optimization of data sponsoring and edge caching has the potential to further improve the CP payoff, but it also introduces new challenges. First, the CP has to distribute his own budget between data sponsoring and edge caching. Second, the decisions of data sponsoring and edge caching are in different time scales and tightly coupled with each other.

B. Model and Problem Formulation

In this work, we consider a simple yet representative model where one CP offers video on demand (VoD) services to a set of mobile users, using both data sponsoring and edge caching (within a certain budget constraint). Specifically, users request video content randomly, and the requested video content can be delivered to the target user in two different ways: (i) cellular transmission and (ii) edge caching transmission if the user is within the coverage of an edge caching network and the content is cached on the edge device in advance.

Moreover, if a content is sponsored (either via data sponsoring or edge caching), it will be delivered to the user with a certain attachment (e.g., additional advertisements) called *value-added content*; thus, it can bring additional value for the CP. Note that such value-added content may reduce the users' payoff for the video content; hence, users can choose to refuse sponsoring, in which case they download data from the cellular network directly and pay the corresponding data downloading cost by themselves as usual. Different video contents include different value-added contents, and thus, they have different added values for the CP (when sponsoring them).

The CP can decide whether to sponsor the content request as well as through which network to sponsor the request. Then, users receiving sponsor offers will decide whether to accept the sponsoring based on their own valuations. If a user accepts the data sponsoring, then the CP pays the data downloading cost in the cellular network for the user; otherwise, the user pays the data downloading cost by himself as usual. Similarly, if a user accepts the edge caching, then the user does not need to pay the data downloading cost (as he does not need to download data from the cellular network but rather receives data from the local edge caching network directly).

Fig. 1 illustrates a system model with different content delivery and sponsor schemes, where red users and green users download video content through the cellular network without sponsoring and with data sponsoring, respectively, while blue users download video content through edge caching.

In such a model, there are two key problems for the CP:

- *Edge (WiFi) Caching Problem*: Whether to cache a video content on the edge caching network at a particular location for a certain time period.
- *Real-time Sponsoring Problem*: Whether and, if so, how (i.e., via data sponsoring or edge caching) to sponsor a video content request at an instantaneous time slot.

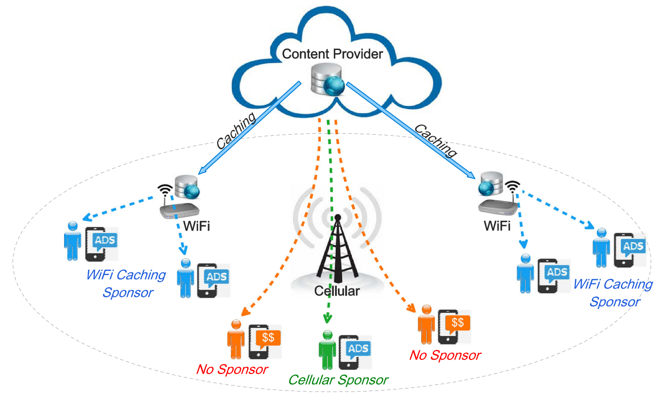

Fig. 1. System model.

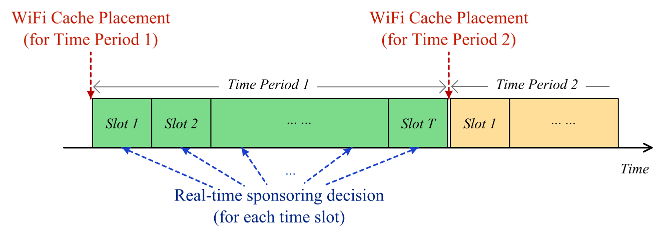

Fig. 2. Timing structure of our model.

Note that the time period of WiFi caching is on a relatively large time scale (e.g., one day), which consists of many time slots of relatively small time scales (e.g., one minute). Moreover, the WiFi caching solution has to be decided at the beginning of each time period and fixed in the entire time period due to the large overhead and delay of video cache placement on edge WiFi networks. Users request video content in each time slot in a real-time manner, and hence, the data sponsoring solution also has to be determined in real time. For clarity, we illustrate this timing structure in Fig. 2.

C. Solution and Contributions

We will study the joint optimization of data sponsoring and edge caching systematically, aiming at maximizing the CP's revenue. Specifically, we formulate the joint optimization problem as a two-stage decision problem for the CP. In stage I, the CP determines the edge caching policy in each time period at the beginning of that time period. In stage II, the CP determines the real-time data sponsoring solution for each content request in each time slot. We analyze this two-stage decision problem systematically. For the real-time sponsoring in stage II, we propose a Lyapunov-based online sponsoring strategy for the CP, which operates in an online and real-time manner, without the need of future network information. For the edge caching in stage I, we develop an edge caching strategy for the CP by predicting the aggregate user request for each video content in each location.

The key contributions of this work are as follows:

- *Novel Model*: As far as we know, this is the first work that systematically studies the joint optimization of data sponsoring and edge caching for mobile video delivery. It exploits the benefits of both edge caching and data sponsoring, and hence can help the CP to improve revenue and reduce content delivery cost.

- *Real-Time Data Sponsoring Strategy:* We propose a Lyapunov-based online sponsoring strategy for the CP in stage II, which operates in an online and real-time manner (without future information), and can achieve on average 90% of the maximum performance in the offline benchmark with the complete future information.
- *Optimal Edge Caching Strategy:* We show that the optimal edge caching strategy depends on the aggregate user request for each video content in each location, based on which we can formulate the edge caching problem in stage I as a mixed-integer linear programming problem. We further propose several methods to estimate such aggregate requests in different model scenarios.
- *Performance Evaluation:* We perform extensive simulations built on large-scale real data traces. The simulation results show that such a joint sponsoring policy can increase the CP's revenue by 124%$\sim$ 154%, comparing with the pure data sponsoring policy without edge caching.

The remainder of this paper is organized as follows. We review the literature in Section II. In Section III, we present the system model and problem formulation. In Section IV, we provide the offline solution as a benchmark. In Section V, we propose the joint online solution. In Section VI, we provide the simulation results, and we finally conclude in Section VII.

II. Literature Review

There have been extensive works studying data sponsoring [7]–[11]. Most of these works focused on studying data sponsorship in cellular networks. For example, in [7], Joe-Wong *et al.* studied the interplay among ISPs, CPs, and end users, and proposed a framework for CPs to decide which content to sponsor. In [8], Zhang *et al.* analyzed the impact of sponsored content on the existing wireless data market, and they found that the imbalance in CP revenue will be enlarged. In [9], [11], authors focused more on the interplay between CPs and ISPs. However, all of the above works studied data sponsorship in traditional cellular networks, without considering data sponsorship in an edge caching network.

Recent studies [13]–[16] have shown that edge caching is a promising approach to alleviate the burden of CDNs and to reduce the content delivery cost in the 5G wireless network. The edge cache node should be equipped with a large storage capacity, high-rate delivery capacity and low-rate backhaul link. In [14], Wang *et al.* utilized edge caching to alleviate the backhaul congestion and achieve better user QoE due to low latency. In [16], Bastug *et al.* analyzed the role of edge caching in 5G networks, and showed that edge caching can alleviate the backhaul congestion.

Some other works [18]–[23] studied different realizations of edge caching, e.g., on base station, femtocell, and WiFi access point (AP). In [18], Golrezaei *et al.* proposed a concept of femtocaching with a large storage capacity in the edge network, and they showed that with a proper caching strategy, the latency can be dramatically decreased. In [19], Purkayastha *et al.* designed a practical architecture for a small cell network operator to manage the content placement. In [20], Gharaibeh *et al.* proposed an online caching algorithm to minimize the total cost of CPs, assuming that the CPs pay ISPs for caching contents. In [21], Liu *et al.* proposed caching popular content at base stations, and they showed that the edge cache outperforms pico base stations with backhaul. In [22], Sung *et al.* proposed conducting caching in a novel two-tier wireless network. In [23], Li *et al.* proposed a novel caching strategy and proved its efficiency in a real data trace.

However, none of the works related to edge caching considered the problem of data sponsorship via an edge caching network. In our previous works [24] and [25], we considered the joint optimization of edge caching and data sponsoring, but we didn't study the online sponsoring problem in [24] and didn't study the joint optimization without complete information in [25].

III. System Model

A. Network Model

We consider a mobile video CDN, where one CP provides VoD services to a set $\mathcal{U} = \{1, 2, ..., U\}$ of mobile users. Each user moves randomly in a certain area and requests a video content randomly at a particular time. The video content can be delivered to the target user in two different ways:

- *Cellular Direct Delivery*: The video content (located on the remote server) will be delivered to the user directly through the cellular link;
- *Edge Caching Delivery*: The video content is cached in an (third-party) edge caching device in advance, and it will be delivered to the local user directly through the local link (e.g., Bluetooth and WiFi).

Without loss of generality, we assume that the cellular network covers the entire area, while each edge caching device covers a small local area. We further assume that the coverage areas of different edge caching devices are non-overlapping.[1] Let $\mathcal{L} = \{1, 2, ..., L\}$ denote the set of areas (locations) covered by all L edge devices. Let $\mathcal{S} = \{1, 2, ..., S\}$ denote the set of all video content.

Motivated by the cache refresh policy deployed by Hulu [26], we consider the cache operation on a *daily basis*, where the CP makes the cache decision at the beginning of each day, and each video content cached in a device will be available for one whole day. To facilitate the later analysis, we further divide the daily period into a set $\mathcal{T} = \{1, 2, ..., T\}$ of small time intervals (e.g., one minute), called *time slots*. Users request video content slot by slot, and the CP delivers the content to the users slot by slot.

B. Sponsoring Scheme

When a user requests a content in a time slot, the CP can decide whether to sponsor the request and, if so, how to sponsor the request (i.e., data sponsoring or edge caching):

- *Data Sponsoring*: The CP delivers the content (on the remote server) to the target user through the cellular link, and it pays the cellular data downloading cost for the user.

[1]When multiple devices are overlapping, the CP can simply choose one of them for video caching.

- *Edge Caching*: The CP delivers the content (cached in the local WiFi network) to the target user through the local WiFi link, and it pays the WiFi caching cost.

Each user can choose whether to accept the CP's sponsoring based on his individual payoff (to be defined in Section III-C). Moreover, if the CP decides *not* to sponsor a content request, or if the user refuses to accept the CP's sponsoring, then the content (on the remote server) will be delivered to the target user through the cellular link as usual, and the user himself pays the cellular data downloading cost.

The CP can achieve a certain basic value from each content that a user plays through, for example, built-in advertisements. If the CP decides to sponsor a content request and the user accepts the CP's sponsoring, then the CP, as compensation, will deliver the content to the user together with certain *value-added content* (e.g., additional built-in advertisements) and hence achieve additional value (called *sponsoring value*) from the content. Note that such value-added content may reduce users' valuations (to be defined in Section III-C), and hence, users may decide to refuse the CP's sponsoring.

C. User Model

Each user moves and requests a content randomly in each time slot $t \in \mathcal{T}$. Let $s_u[t] \in \mathcal{S} \bigcup \{0\}$ denote the content request of user u in time slot t, where $s_u[t] = 0$ denotes that the user does *not* request any video content. Let $l_u[t] \in \mathcal{L} \bigcup \{0\}$ denote the location of user u in time slot t, where $l_u[t] = 0$ denotes that the user is in an area without any device coverage. Then, we can write the service request vector and location vector of user $u \in \mathcal{U}$ in the whole period as $\mathcal{S}_u = \{s_u[1], \ldots, s_u[T]\}$ and $\mathcal{L}_u = \{l_u[1], \ldots, l_u[T]\}$, respectively.

When user u downloads the content $s_u[t]$ through the cellular network, there is a certain *cellular data downloading cost*, denoted by $C_{s_u[t]}$ or $C_u[t]$, due to the payment to the cellular network operator. This cellular data downloading cost mainly depends on the content size, denoted by $D_{s_u[t]}$ or $D_u[t]$, and it can be undertaken by the user himself without data sponsoring or by the CP with data sponsoring.[2]

When user u downloads the content $s_u[t]$ through the cellular network (either with no sponsoring or with data sponsoring), he incurs a certain *energy cost* $E_u[t]$, which primarily depends on the size of the content $D_u[t]$. When user u downloads the content $s_u[t]$ through the local WiFi network, he incurs a smaller energy cost $E_u^w[t] < E_u[t]$, which also depends on the content size $D_u[t]$. This is supported by [28], which indicated that WiFi links consume less power than cellular links when transmitting the same amount of data. For notational convenience, we denote $\eta \in (0, 1)$ as the ratio of WiFi energy cost and cellular energy cost (for delivering the same amount of data), i.e., $E_u^w[t] = E_u[t] \cdot \eta$.

Each user u is associated with an instantaneous *valuation*, denoted by $W_u[t]$, when viewing the video content $s_u[t]$ in time slot t. $W_u[t]$ is used to evaluate the overall benefit of watching a content by the user, depending on the user satisfaction, real-time demand, and video quality. The evaluation of $W_u[t]$ is a well-studied topic [29], and hence, it is beyond the scope of this paper.

We assume that users are self-motivated,[3] that is, each user will decide whether to accept the sponsorship based on his own benefit. When the user accepts the data sponsoring, such a valuation is discounted by a factor δ (where $0 < \delta < 1$) due to the discomfort caused by the added advertisements. When the user accepts the edge caching, such a valuation is further affected by a factor $\delta^w = \delta^{\text{neg}} \cdot \delta^{\text{pos}}$ (where $0 < \delta^{\text{neg}} < 1$ and $\delta^{\text{pos}} > 1$) due to the cost induced by handover denoted as δ^{neg}.[4] Specifically, the handover cost is a combination of the discomfort caused by the user rebuffering and long join time as well as the additional energy cost and the lower latency of the edge caching network denoted as δ^{pos}. Consequently, $\delta^w > 0$ is determined by the user experience of both lower delay and network mode change discomfort. When a user evaluates the lower delay of the edge caching network, $\delta^w > 1$, while when a user is sensitive to the network mode change, $0 < \delta^w < 1$. Hence, the total valuation discount under edge caching is $\delta \cdot \delta^w$.

The network information consists of the location, the content request, and the instantaneous valuation of each user. Formally, we define the *network information* in time slot t as follows:

$$\theta[t] = \{s_u[t], l_u[t], W_u[t], \quad \forall u \in \mathcal{U}\}. \tag{1}$$

Note that the energy costs $E_u[t]$ and $E_u^w[t]$ and the cellular data downloading cost $C_u[t]$ are not included in the network information vector because they are determined by the content size $D_u[t]$.

Based on the above discussion, we can define the user payoff as a sum of valuation and costs with the aid of normalizing coefficients μ and ρ:

- Without sponsoring, the payoff of user u in time slot t is $\Pi_u[t] = W_u[t] - \mu C_u[t] - \rho E_u[t]$, where the user pays both the data downloading cost and the energy cost on the cellular link;
- With data sponsoring, the payoff of user u in time slot t is $\Pi_u^c[t] = W_u[t] \cdot \delta - \rho E_u[t]$, where the user achieves a discounted valuation and pays the energy cost on the cellular link;
- With edge caching, the payoff of user u in time slot t is $\Pi_u^w[t] = W_u[t] \cdot \delta \cdot \delta^w - \rho E_u[t] \cdot \eta$, where the user achieves a discounted valuation and pays the energy cost on the WiFi link.

Clearly, the user will accept data sponsoring only when $\Pi_u^c[t] > \Pi_u[t]$ and accept edge caching only when $\Pi_u^w[t] > \Pi_u[t]$. For notational convenience, we denote $\bar{w}_u[t] = C_u[t] + E_u[t]$ as the minimum valuation threshold of initiating the content request, that is, the user is willing to request the content $s_u[t]$ only if $W_u[t] \geq \bar{w}_u[t]$ (hence, $\Pi_u[t] \geq 0$). We denote $\hat{w}_u^c[t] = \frac{C_u[t]}{1-\delta}$ as the maximum valuation threshold of

[2]In reality, one specific video content may have many resolutions (as shown in [27]), but in this work, to facilitate the analysis, we model different resolutions of one content as different contents.

[3]Self-motivation and selfishness are common assumptions in game theoretic models (e.g., [30]–[35]).

[4]Note that δ^{neg} is positive, which has a "negative effect" when multiplied by the user valuation.

accepting data sponsoring, that is, the user is willing to accept data sponsoring only if $W_u[t] \le \hat{w}_u^c[t]$ (hence, $\Pi_u^c[t] \ge \Pi_u[t]$). Similarly, we analyze when the user is willing to accept edge caching. When $\delta \cdot \delta^w < 1$, we denote $\hat{w}_u^w[t] = \frac{C_u[t]+E_u[t](1-\eta)}{1-\delta\cdot\delta^w}$ as the maximum valuation threshold of accepting edge caching, that is, the user is willing to accept edge caching only if $W_u[t] \le \hat{w}_u^w[t]$ (hence, $\Pi_u^w[t] \ge \Pi_u[t]$). When $\delta \cdot \delta^w \ge 1$, we derive $\Pi_u^w[t] \ge \Pi_u[t]$ for any $W_u[t]$, and $\hat{w}_u^w[t] = \infty$. Let $f_u(w)$ denote the probability distribution function (pdf) of user u's valuation $W_u[t]$ in time slot t (for content $s_u[t]$). Then, we can compute the average acceptance probabilities of user u for data sponsoring and edge caching in time slot t:

$$P_u^w[t] = \int_{\bar{w}_u[t]}^{\hat{w}_u^w[t]} f_u(w)\mathrm{d}w, \tag{2}$$

$$P_u^c[t] = \int_{\bar{w}_u[t]}^{\hat{w}_u^c[t]} f_u(w)\mathrm{d}w. \tag{3}$$

To provide a concrete analysis, we adopt the following distribution function for $W_u[t]$ suggested in [8]:[5]

$$f_u(w) = \lambda_u \exp^{-\lambda_u(w-\bar{w}_u[t])},$$

where $\lambda_u > 0$ is called the *valuation density* of user u. A larger λ_u represents a higher valuation density for user u, that is, the valuation concentrates on a smaller range of values with a higher probability.

D. CP Model

As mentioned above, when sponsoring a content request, the CP needs to pay for the cellular data downloading cost or the edge caching cost; meanwhile, the CP can achieve a certain sponsoring value from the value-added content. Let V_s denote the sponsoring value for the CP when sponsoring a video content $s \in \mathcal{S}$, which depends on the video content only, not on the user who requests the content. For example, a popular video is generally associated with high-value advertisement and hence has a large sponsoring value. Let C_s denote the cellular data downloading cost for the CP when sponsoring a video content $s \in \mathcal{S}$ through the cellular network. Let C_s^w denote the edge caching cost for the CP when caching a video content $s \in \mathcal{S}$ in a caching device *for one whole time period*. Note that C_s^w is a one-shot caching cost, which depends on the length of the time period and the content size but not on the actual times of the edge caching sponsor in the time period. That is, after caching a content s in a caching device in a time period, the CP needs to pay the caching cost C_s^w, irrespective of how many requests (for content s) are delivered through this caching device. We further denote $\alpha = \frac{C_s^w}{C_s}$ as the caching-to-delivery factor, denoting the relative cost of edge caching to cellular downloading. In reality, $\alpha > 1$ is always satisfied as the caching cost consists of both downloading and storage.

For each content request, the CP decides whether to sponsor the request and, if so, how to sponsor the request. For this purpose, the CP needs to determine the following strategies:

[5]Note that our analysis holds for an arbitrary distribution of $W_u[t]$.

- *Edge Caching Sponsor Strategy*: Which video contents will be cached at which locations.
- *Cellular Sponsor Strategy*: Whether to sponsor a content request and with which sponsoring scheme.

Let $Z[l,s] \in \{0,1\}$ denote whether to cache a video content $s \in \mathcal{S}$ at location $l \in \mathcal{L}$. Then, given the caching strategy $\{Z[l,s], \forall l \in \mathcal{L}, \forall s \in \mathcal{S}\}$, the total *edge caching cost* for the CP is

$$\gamma = \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}} Z[l,s]\cdot C_s^w. \tag{4}$$

We define the capacity of the caching device as τ; hence, we have the following constraint for the device capacity:

$$\sum_{s\in\mathcal{S}} Z[l,s]\cdot C_s \le \tau \quad \forall l \in \mathcal{L}. \tag{5}$$

Let $y_u[t] \in \{0,1\}$ and $x_u[t] \in \{0,1\}$ denote whether the CP offers to sponsor the content request of user u in time slot t (i.e., $s_u[t]$) via edge caching and data sponsoring, respectively. Because the CP can offer only one sponsor scheme for each content request, we have the following constraint:

$$x_u[t] + y_u[t] \le 1, \quad \forall u \in \mathcal{U}, t \in \mathcal{T}. \tag{6}$$

Moreover, the edge caching scheme is available only when the content $s_u[t]$ is cached at the corresponding location $l_u[t]$ (i.e., the location of user u in time slot t). Thus, we have

$$y_u[t] \le Z\big[l_u[t], s_u[t]\big], \quad \forall u \in \mathcal{U}, t \in \mathcal{T}. \tag{7}$$

We also note that users may decide to refuse the sponsoring. Namely, a successful sponsoring occurs only when the target user is willing to accept the sponsoring. Hence, we have

$$x_u[t] = 0, \quad \forall u \notin \mathcal{U}^c[t], t \in \mathcal{T}. \tag{8}$$

$$y_u[t] = 0, \quad \forall u \notin \mathcal{U}^w[t], t \in \mathcal{T}, \tag{9}$$

where $\mathcal{U}^c[t] = \{u \in \mathcal{U} \mid \Pi_u^c[t] \ge \Pi_u[t]\}$ is the set of users who are willing to accept data sponsoring at time slot t and $\mathcal{U}^w[t] = \{u \in \mathcal{U} \mid \Pi_u^w[t] \ge \Pi_u[t]\}$ is the set of users who are willing to accept edge caching at time slot t.

Given the data sponsoring strategy $\{x_u[t], \forall u \in \mathcal{U}, t \in \mathcal{T}\}$, the total *cellular data downloading cost* that the CP pays for users in time slot t is

$$C[t] = \sum_{u\in\mathcal{U}} x_u[t]\cdot C_u[t], \quad \forall t \in \mathcal{T}. \tag{10}$$

where $C_u[t] = C_{s_u[t]}$ is the cellular data downloading cost for user u's content request $s_u[t]$ in time slot t.

Given the data sponsoring and edge caching strategies $\{x_u[t], y_u[t], \forall u \in \mathcal{U}, t \in \mathcal{T}\}$, the total *sponsoring value* (revenue) that the CP can achieve from all sponsored content requests in time slot t is

$$V[t] = \sum_{u\in\mathcal{U}}(x_u[t] + y_u[t])\cdot V_u[t], \quad \forall t \in \mathcal{T}, \tag{11}$$

where $V_u[t] = V_{s_u[t]}$ is the value from sponsoring user n's content $s_u[t]$ in time slot t.

Therefore, the CP's total (time-average) payoff is

$$R = \frac{1}{T} \cdot \left(\sum_{t=1}^{T} V[t] - \sum_{t=1}^{T} C[t] - \gamma \right), \tag{12}$$

where the first term is the total sponsoring value, the second term is the total cellular delivery cost, and the last term is the total edge caching cost. For convenience, we denote $R[t] = V[t] - C[t]$ as the instantaneous payoff for the CP in time slot t, which consists of the instantaneous sponsoring value $V[t]$ and the instantaneous cellular data downloading cost.

IV. Offline Joint Optimization Benchmark

In this section, we study the joint optimization of edge caching and data sponsoring with complete information or stochastic future information. In this case, we can formulate the joint optimization problem as an offline optimization problem, and the solution will serve as a benchmark for the online algorithm derived in Section V.

A. Complete Future Information

With complete information, the CP can derive the payoff R under any edge caching and data sponsoring solution. Hence, the CP's objective is to maximize its payoff R with a given budget B (for both caching and sponsoring video content). Namely, we have the following *budget constraint* for the CP:

$$\sum_{t=1}^{T} C[t] + \gamma \leq B. \tag{13}$$

Therefore, the joint caching and sponsoring problem for the CP can be formulated as follows:

$$\begin{aligned} \max \quad & R = \frac{1}{T} \cdot \left(\sum_{t=1}^{T} V[t] - \sum_{t=1}^{T} C[t] - \gamma \right) \\ s.t. \quad & (5), (6), (7), (8), (9), (13); \\ var. \quad & x_u[t] \in \{0,1\}, \ \forall u \in \mathcal{U}, t \in \mathcal{T}; \\ & y_u[t] \in \{0,1\}, \ \forall u \in \mathcal{U}, t \in \mathcal{T}; \\ & Z[l,s] \in \{0,1\}, \ \forall l \in \mathcal{L}, s \in \mathcal{S}. \end{aligned} \tag{14}$$

It is easy to observe that the above problem (14) is an integer linear programming (ILP) problem. Formulating and solving problem (14) requires the complete future information, which is infeasible in practice because we cannot derive the future information. Hence, we will study another benchmark solution based on stochastic future information.

B. Stochastic Future Information

With stochastic future information only, the CP cannot determine the explicit sponsoring decision for each user request in each time slot in advance due to the lack of complete information (e.g., users' content requests in all time slots). In this case, we will formulate and solve the CP's *expected* payoff maximization problem built upon the stochastic information.

Let $\theta = \{l_u(\theta), s_u(\theta), W_u(\theta), \forall u \in \mathcal{U}\}$ denote a specific realization of network information. Let $x_u(\theta), y_u(\theta) \in \{0,1\}$ denote the corresponding sponsoring solutions under a particular realization θ. Then, the expected payoff maximization problem can be defined as follows:

$$\begin{aligned} \max \quad & R = \int_{\theta \in \Theta} (V(\theta) - C(\theta) - \gamma) g(\theta) d\theta \\ s.t. \quad & x_u(\theta) + y_u(\theta) \leq 1, \quad \forall u \in \mathcal{U}, t \in \mathcal{T}, \theta \in \Theta \\ & \sum_{s \in \mathcal{S}} Z[l,s] \cdot C_s \leq \tau \quad \forall l \in \mathcal{L} \\ & y_u(\theta) \leq Z\big[l_u(\theta), s_u(\theta)\big], \quad \forall u \in \mathcal{U}, t \in \mathcal{T}, \theta \in \Theta \\ & \int_{\theta \in \Theta} C(\theta) d\theta + \gamma \leq B. \\ var. \quad & x_u(\theta), y_u(\theta) \in \{0,1\} \quad \forall u \in \mathcal{U}, \theta \in \Theta \\ & Z[l,s] \in \{0,1\} \quad \forall l \in \mathcal{L}, s \in \mathcal{S}, \end{aligned} \tag{15}$$

where $f(\theta)$ is the probabilistic distribution function under information θ, $C(\theta) = \sum_{u \in \mathcal{U}} x_u(\theta) C_u(\theta)$ is the total cellular data downloading cost that the CP pays for users under information θ, and $V(\theta) = \sum_{u \in \mathcal{U}} (x_u(\theta) + y_u(\theta)) V_u(\theta)$ is the total revenue that the CP can achieve under information θ. Note that $g(\theta)$ is a general distribution representation, and it can be realized with specific distribution functions. Problem (15) is an ILP with an infinite number of decision variables (as θ is continuous), and it can be solved by many classic methods, e.g., KKT analysis.[6]

Next, we analyze the gap between the maximum payoff (denoted by R^0) derived from (14) with complete information and the maximum expected payoff (denoted by R^*) derived from (15) with stochastic information. Formally,

Lemma 1: If $T \to \infty$, then $R^* \to R^0$.

Proof: We note that problem (17) is time independent but information dependent. Hence, the sponsor decision in different time slots with the same network information must be identical, i.e., $x[t_1] = x[t_2]$, if $\theta[t_1] = \theta[t_2]$. Then, by the law of large numbers (LLN), we have the following:

if $T \to \infty$,

$$(1) \quad \frac{1}{T} \sum_{t=1}^{T} x_u[t] \to \int_{\theta \in \Theta} x_u(\theta) f(\theta) d\theta,$$

$$(2) \quad \frac{1}{T} \sum_{t=1}^{T} C[t] \to \int_{\theta \in \Theta} C(\theta) f(\theta) d\theta,$$

$$(3) \quad \frac{1}{T} \sum_{t=1}^{T} V[t] \to \int_{\theta \in \Theta} V(\theta) f(\theta) d\theta.$$

■

This implies that problems (16) and (17) are equivalent if $T \to \infty$; hence, their solutions are also equivalent.

We can derive from Lemma 1 that as long as the time period T is large enough, the payoff induced by the loss of complete

[6] We skip the details as the method is standard. Moreover, solving this stochastic optimization is not the main contribution of this work.

TABLE I
NOTATIONS OF IMPORTANT VARIABLES

Notation	Meaning
$R[t]$	The CP payoff in time slot t
$C[t]$	The CP cost in time slot t
$B[t-1]$	The total remaining budget in time slot t
$x_u[t]$	If user u's request is served by data sponsoring
$y_u[t]$	If user u's request is served by edge caching
$Z[l,s]$	If content s is cached in location l
$W_u[t]$	User u's valuation in time slot t
$P_u^c[t]$	User probability to accept data sponsoring
$P_u^w[t]$	User probability to accept edge caching
$\widetilde{C}[t]$	The expected data sponsoring cost in time slot t
$\widetilde{V}[t]$	The expected data sponsoring revenue in time slot t
$\widetilde{R}[t]$	The expected data sponsoring payoff in time slot t
$q[t]$	The queue backlog in time slot t
$J[t]$	The Lyapunov function in time slot t
$\delta[t]$	The Lyapunov drift in time slot t
ϕ	A non-negative control parameter
$\Pi[t]$	drift-plus-penalty function

information is negligible compared with that achieved with the stochastic information. Hence, R^0 and R^* can serve as the same benchmark for the online strategies in Section V.

Notably, problem (15) is also an *offline* optimization problem, and formulating (15) still requires certain (stochastic) future information. For example, the CP needs to know the stochastic network information in all time slots. In practice, however, the CP may not be able to obtain the stochastic network information when making caching and sponsoring decisions. To this end, in the next section, we will propose an *online* decision process for the CP to make the caching and sponsoring decision without future network information.

V. Online Optimization Framework

In this section, we study the CP's payoff maximization problem under incomplete network information (i.e., without complete and stochastic network information). In this case, the CP has to make the sponsoring decision in an online and real-time manner. We formulate the CP's joint caching and sponsoring problem as a two-stage decision process:

- Stage I: The CP determines the WiFi caching strategy at the beginning of each time period;
- Stage II: The CP determines the data sponsoring strategy for each content request in each time slot given the WiFi caching strategy in stage I.

Next, we will analyze this two-stage decision problem with backward induction. Because the edge caching decision in stage I is determined by the data sponsoring decision in stage II, we first analyze the data sponsoring decision in stage II and then analyze the edge caching decision based on it.

A. Stage II: Lyapunov-based Online Sponsoring Strategy

We first study the real-time sponsoring strategy for the CP in stage II. In this stage, the CP determines the best sponsoring strategy for each content request in each time slot in an online and real-time manner given the edge caching strategy $\{Z^*[l,s], \forall l \in \mathcal{L}, \forall s \in \mathcal{S}\}$ derived in stage I. The total budget (cost) for edge caching is

$$\gamma^* = \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} Z^*[l,s] \cdot C_s \cdot \alpha. \tag{16}$$

Thus, the remaining budget for real-time data sponsoring is

$$B_{\text{spon}} = B - \gamma^*. \tag{17}$$

Note that with the complete network information, we can formulate an offline sponsoring problem, which is the same as (14) with fixing $Z[l,s] = Z^*[l,s], \forall l \in \mathcal{L}, \forall s \in \mathcal{S}$. However, as mentioned previously, it is impossible for the CP to obtain the complete network information. Hence, it is necessary to propose the online sponsoring strategy without relying on the complete network information. In the following, we will propose a Lyapunov-based online sponsoring policy that optimizes a modified CP's payoff according to the remaining budget in each time slot. We further consider two different cases, depending on whether the CP can observe the instantaneous valuations of users realized in each time slot.

The Lyapunov optimization technique [36] is a promising technique for solving optimization problems with time average constraints and has been widely used in wireless network problems (e.g., [37]–[40]). In our model, the optimal sponsoring problem in all time slots can be formulated as such an optimization problem, *where the time average constraint is the budget constraint*. The key idea of Lyapunov optimization is to use the *stability* of the queue to ensure that the time average constraint is satisfied. Following this idea, we introduce a virtual queue for the CP, which is used to indicate the CP's *potential budget deficit*. Here, we use the prefix "potential" to denote that it is not the actual budget deficit of the CP; rather, it indicates that a potential deficit may occur with the current budget consumption rate.

Formally, we denote $q[t]$ as the queue backlog (potential budget deficit) at time slot t. Then, the queue backlog (potential budget deficit) at the next time slot $t+1$ is updated in the following way:

$$q[t+1] = \left[q[t] - \frac{B_{\text{spon}}}{T}\right]^+ + C[t]. \tag{18}$$

Here, $\frac{B_{\text{spon}}}{T}$ is the intended average budget for each time slot, $[x]^+ = \max(x,0)$, and $C[t]$ is the actually consumed budget in time slot t. Intuitively, if the actually consumed budget $C[t]$ is larger (or smaller) than the average budget, then the potential budget deficit will be increased (or decreased), indicating a higher (or lower) possibility that a budget deficit will occur at the end. In the above virtual queue system, we can view $C[t]$ as the arrival rate and $\frac{B_{\text{spon}}}{T}$ as the departure rate. By the queue stability theorem [36], a queue is stable if and only if the average arrival rate is no larger than the average departure rate, i.e.,

$$\frac{1}{T} \sum_{t\in\mathcal{T}} C[t] \le \frac{B_{\text{spon}}}{T}.$$

This implies that the stability of the virtual queue ensures the satisfaction of the CP budget constraint. Therefore, to guarantee the CP budget constraint, we only need to ensure that the virtual queue is stable under the proposed policy.

We study the queue stability using the Lyapunov drift theorem [36]. First, we define the following Lyapunov function in each time slot t:

$$J[t] = \frac{1}{2} \cdot q[t]^2. \tag{19}$$

The Lyapunov drift in each time slot t is the change of the Lyapunov function from one slot to the next, i.e.,

$$\triangle[t] = J[t+1] - J[t]. \tag{20}$$

By the Lyapunov drift theorem (Th. 4.1 in [36]), if a policy greedily minimizes the Lyapunov drift $\triangle[t]$ in each slot t, it potentially maintains the stability of the queue (i.e., guarantees the budget constraint).

Next, we analyze the joint queue stability and payoff maximization. By the Lyapunov optimization theorem (Th. 4.2 in [36]), to stabilize the queue while optimizing the payoff, we can use such a policy that greedily minimizes the following drift-plus-penalty function:

$$\begin{aligned} \min \quad & \Pi[t] = \triangle[t] - \phi \cdot R[t] \\ s.t. \quad & (6), (7), (8), (9), (13), (20) \\ & C[t] \le B[t-1] \\ var. \quad & x_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}; \\ & y_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}; \end{aligned} \tag{21}$$

where $B[t-1] = B_{\text{spon}} - \sum_{\tau \in [1:t-1]} C[\tau]$ is the total remaining budget at time slot t and $\phi \ge 0$ is a non-negative control parameter that achieves a tradeoff between optimality and queue backlog.

Note that $\triangle[t]$ is a quadratic function. We can find an upper bound of $\triangle[t]$ as follows:

$$\begin{aligned} \triangle[t] &\le \tfrac{1}{2}\left[\left(\tfrac{B_{\text{spon}}}{T}\right)^2 + C[t]^2 + 2q[t] \cdot \left(C[t] - \tfrac{B_{\text{spon}}}{T}\right)\right] \\ &\le F + q[t] \cdot \left(C[t] - \tfrac{B_{\text{spon}}}{T}\right), \end{aligned}$$

where $F = \frac{1}{2}(\frac{B_{\text{spon}}}{T})^2 + \frac{1}{2}(\sum_{u \in \mathcal{U}} C_{s_u[t]})^2$ is a constant. The first inequality follows because $([q-x]^+ + c)^2 \le q^2 + x^2 + c^2 + 2q(c-x)$, and the second inequality follows because $C[t] = \sum_{u \in \mathcal{U}} x_u[t] C_u[t] \le \sum_{u \in \mathcal{U}} C[t]$. Note that greedily minimizing $\Pi[t]$ in (21) is equivalent to maximizing the above upper bound of $\Pi[t]$. Hence, the policy defined in (21) is equivalent to the following policy:

$$\begin{aligned} \text{(Lyapunov)} \quad \min \quad & \Pi[t] = q[t] \cdot C[t] - \phi \cdot R[t] \\ s.t. \quad & (6), (7), (8), (9), (13), (20) \\ & C[t] \le B[t-1] \\ var. \quad & x_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}; \\ & y_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}. \end{aligned} \tag{22}$$

Here, we assume that the CP can observe the instantaneous valuations of users in each time slot and can hence compute the instantaneous payoff $R[t]$ precisely.

Similarly, when the CP cannot observe the instantaneous valuations of users realized in each time slot (but knows only the stochastic distribution information regarding user valuation), the CP can only compute the expected cellular data downloading cost $\widetilde{C}[t]$ by (32) and the expected payoff $\widetilde{R}[t]$ by (34). Hence, the Lyapunov-based sponsoring policy in (22) can be redefined in the following way:

$$\begin{aligned} \text{(Lyapunov*)} \quad \min \quad & \widetilde{\Pi}[t] = q[t] \cdot \widetilde{C}[t] - \phi \cdot \widetilde{R}[t] \\ s.t. \quad & (6), (7), (13), (20) \\ & \widetilde{C}[t] \le \widetilde{B}[t-1] \\ var. \quad & x_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}; \\ & y_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}, \end{aligned} \tag{23}$$

where $\widetilde{B}[t-1] = B_{\text{spon}} - \sum_{\tau \in [1:t-1]} \widetilde{C}[\tau]$ is the total remaining budget at time slot t.

1) Performance Analysis: We now show the performance of the above Lyapunov-based policies. Let $R^\dagger[t]$ and $R^\ddagger[t]$ denote the payoff achieved in each slot t by the Lyapunov and Lyapunov* policies, respectively.

Theorem 1: Let R^* denote the offline maximum payoff derived from (14) or (15) in Section IV. Then,

$$\begin{aligned} \lim_{T \to \infty} \frac{1}{T} \sum_{t \in \mathcal{T}} E(R^\dagger[t]) &\ge R^* - \frac{F}{\phi}, \\ \lim_{T \to \infty} \frac{1}{T} \sum_{t \in \mathcal{T}} E(R^\ddagger[t]) &\ge R^* - \frac{F}{\phi}. \end{aligned} \tag{24}$$

The proof follows the standard Lyapunov optimization theory [36]. From Theorem 1, we can find that the proposed Lyapunov policies converge to the maximum payoff benchmark asymptotically, with a controllable approximation error bound $\mathcal{O}(1/\phi)$.

B. Stage I: Best Caching Strategy

In contrast to the sponsoring strategy in stage II, which is performed per time slot, the caching strategy in stage I is performed per time period. Meanwhile, the caching strategy in stage I is dependent on the sponsoring strategy in Stage II, but we cannot obtain the closed form of the optimal online sponsoring solution in stage II. Now, we examine the offline optimization problem (14) to obtain certain meaningful insights into the design of our desired WiFi caching strategy.

Let $N^w[l,s]$ denote the total requests for content s at location l from those users *who are willing to accept WiFi sponsoring*, i.e.,

$$N^w[l,s] = \sum_{t \in \mathcal{T}} \sum_{u \in \mathcal{U}^w[t]} \mathbf{1}(s_u[t] = s \,\&\, l_u[t] = l), \tag{25}$$

where $\mathbf{1}(x) = 1$ if x is true and 0 otherwise. Then, we can derive that if $Z[l,s] = 1$, then $x_u[t] = 0$ and $y_u[t] = 1$, for all $u \in \mathcal{U}^w[t], t \in \mathcal{T}$ with $s_u[t] = s, l_u[t] = l$. In other words, if a content s is cached in location l, then all of the user requests (for content s at location l) from those users who are willing to accept WiFi sponsoring will be sponsored by WiFi. Hence, the CP's payoff achieved from caching a content s in location l can be computed as follows:

$$R^w[l,s] = Z[l,s] \cdot (N^w[l,s] \cdot V_s - \alpha \cdot C_s); \tag{26}$$

Similarly, if $Z[l,s] = 0$, then $y_u[t] = 0$, for all $u \in \mathcal{U}, t \in \mathcal{T}$ with $s_u[t] = s, l_u[t] = l$. In other words, if content s is *not* cached in location l, then none of the user requests for content s at location l can be sponsored via WiFi. Then, we denote $N^c[l,s]$ as the total requests for content s at location l from those users *who are willing to accept cellular sponsoring*, i.e.,

$$N^c[l,s] = \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}^c[t]} \mathbf{1}(s_u[t] = s \;\&\; l_u[t] = l).$$

Hence, the CP's payoff achieved from sponsoring content s in location l via cellular is

$$\begin{aligned} R^c[l,s] &= \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}^c[t]} \mathbf{1}(s_u[t]=s \;\&\; l_u[t]=l) \cdot x_u[t] \cdot (V_s - C_s) \\ &= X[l,s] \cdot N^c[l,s] \cdot (V_s - C_s), \end{aligned}$$

where $X[l,s] = \frac{\sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}_c[t]} \mathbf{1}(s_u[t]=s \;\&\; l_u[t]=l)\cdot x_u[t]}{N^c[l,s]} \in [0,1]$ denotes the *cellular sponsoring ratio* (of content s at location l), i.e., the percentage of the cellular sponsored requests (for content s at location l) among the total requests (for content s at location l) that can be sponsored via cellular networks.

Based on the above analysis, we can transform the original problem (14) into an equivalent problem with respect to the WiFi caching decision and the cellular sponsoring ratio regarding each content at each location, i.e., $Z[l,s] \in \{0,1\}$ and $X[l,s] \in [0,1]$. We define the constraints as follows:

$$X[l,s] + Z[l,s] \le 1, \;\forall l, s; \tag{27}$$

$$\sum_{s\in\mathcal{S}} Z[l,s] \cdot C_s \le \tau \quad \forall l; \tag{28}$$

$$\gamma + \beta \le B; \tag{29}$$

where $\gamma = \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} Z[l,s] \cdot \alpha \cdot C_s$ is the total WiFi caching cost and $\beta = \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} X[l,s] \cdot N^c[l,s] \cdot C_s$ is the total cellular sponsoring cost.

$$\begin{aligned} \max \quad & R = \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} (R^w[l,s] + R^c[l,s]) \\ s.t. \quad & (27), (28), (29), \\ var. \quad & X[l,s] \in [0,1], \;\forall l, s; \\ & Z[l,s] \in \{0,1\}, \;\forall l, s; \end{aligned} \tag{30}$$

Problem (30) is a mixed-integer linear programming problem, and it can be solved by using classic methods.

Note that problem (30) depends only on $N^c[l,s]$ and $N^w[l,s]$ (i.e., the total requests for content s at location l from those users who are willing to accept cellular sponsoring and WiFi sponsoring, respectively), not on the detailed request and location of each user in each time slot. However, it still requires the complete information to compute the exact $N^c[l,s]$ and $N^w[l,s]$. Nevertheless, **it inspires us to find some estimations for $N^c[l,s]$ and $N^w[l,s]$ and design the caching policy based on the estimated $N^c[l,s]$ and $N^w[l,s]$**. Next, we provide two estimation methods for $N^c[l,s]$ and $N^w[l,s]$ under two different model scenarios. One is an independent and identically distributed (i.i.d.) model, applicable for content such as news, which appears and vanishes quickly. The other one is a discrete-time Markov model, applicable for content such as movies, which contain a relatively long time sequence.

1) Independent and Identically Distributed Model: In this scenario, each user's locations and content requests are independent and identically distributed across different time slots. Let $\mu_{u,s}$ denote the probability of user u requesting content s at any time slot and $\eta_{u,l}$ denote the probability of user u moving to location l at any time slot. Then, we can compute the *expected* requests for content s at location l from those users who are willing to accept WiFi sponsoring as[7]

$$\widetilde{N}^w[l,s] = \sum_{u\in\mathcal{U}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \mu_{u,s} \cdot \eta_{u,l} \cdot P_u^w.$$

Similarly, we can compute the *expected* requests for content request s at location l from those users who are willing to accept cellular sponsoring as

$$\widetilde{N}^c[l,s] = \sum_{u\in\mathcal{U}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \mu_{u,s} \cdot \eta_{u,l} \cdot P_u^c.$$

Substituting the above expected numbers $\widetilde{N}^c[l,s]$ and $\widetilde{N}^w[l,s]$ into (30), we can derive the estimated caching strategy in this i.i.d. model scenario.

2) Discrete-Time Markov System: In this scenario, each user behaviors rather irregularly within one time period, but the crowd behavior shows a strong positive correlation among two successive time periods. This is because a video's popularity is strongly correlated among two successive days. For example, a popular video on one day is very likely to be popular on the next day.

Let $\widehat{N}^c[l,s]$ and $\widehat{N}^w[l,s]$ denote the corresponding requests for content s at location l from users who are willing to accept cellular sponsoring and WiFi sponsoring, respectively, in the *previous* time period. Let $Pr(M|N)$ denote the one-step transition probability of such requests. Then, we can compute the *expected* number of requests (for each content s on each location l from those users who are willing to accept WiFi sponsoring and cellular sponsoring) as follows:

$$\widetilde{N}^w[l,s] = \sum_{M=0}^{\infty} Pr(M|\widehat{N}^w[l,s]) \cdot M.$$

$$\widetilde{N}^c[l,s] = \sum_{M=0}^{\infty} Pr(M|\widehat{N}^c[l,s]) \cdot M.$$

Substituting the above expected numbers $\widetilde{N}^c[l,s]$ and $\widetilde{N}^w[l,s]$ into (30), we can derive the estimated caching strategy in this Markov model scenario.

[7]For convenience, we ignore the time index t in the acceptance probability $P_u^w[t]$ because the system is i.i.d.

Remark 1: We analyze a large-scale dataset with 3 million viewing sessions provided by iQiYi,[8] and we combine it with the WiFi stations' location information collected by Tencent[9] (see Section VI for more details). Through these real data traces, we observe that the popularity of each content at each location shows a strong correlation in time periods, which enhances the feasibility of this Markov model. The reason is that most popular video contents in this dataset are TV shows and TV series which burst and fade away soon.

VI. Experiments and Simulations

To evaluate the performance (in terms of the CP's payoff, the user revenue, and the backhaul traffic) of our proposed joint optimization framework, we conduct extensive experiments **on large-scale real data traces**, including both mobile video traces and WiFi AP traces.

A. Dataset and Important Parameters

1) Mobile Video Streaming Trace: We use the mobile video streaming trace collected by iQiYi [?], one of the most popular online video providers in China with approximately 200,000,000 active mobile users. The dataset contains approximately 190,000 mobile video users with a total of 3,000,000 viewing sessions in Beijing within 2 weeks in May 2015. Each trace records the user ID, the requested video content ID, the timestamp of the request, and the location of the user (provided with longitude and latitude).

2) WiFi Station Data Trace: We also collected the locations of free WiFi APs in Beijing via Tencent Mobile Manager [?], a mobile phone app with over 400,000,000 users in China, which helps users discover and connect to free WiFi. The dataset contains the information of over 166,000 WiFi APs in Beijing. Each trace records the basic service set identifier (BSSID) of a WiFi AP and the location of the WiFi AP (also provided with longitude and latitude).

3) Mapping of the Two Datasets: In our experiments, we combine these two datasets to explore the WiFi APs that are available to users when they request video content. We assume that the signal range of each WiFi AP is 100 m, within which mobile users can connect to the AP. We map the location of each user at each content request to the nearest AP if the distance between the user and the AP is less than 100 m; otherwise, we map the user to a blank location (i.e., $l = 0$) without any WiFi coverage. Such a user-AP mapping is illustrated in Fig. 3.

4) Important Parameters: We choose the value and cost of each video content based on the video length and the embedded value-added content. Moreover, according to the measurement in [28], the average energy consumption is 32 J/MB for 3G cellular network transmission and 1 J/MB for WiFi transmission. Hence, we set the energy consumption ratio of WiFi transmission and cellular transmission to $\eta = 3.1\%$. We set the capacity of the edge host as $32G$.

[8]http://www.iQiyi.com
[9]http://m.qq.com/

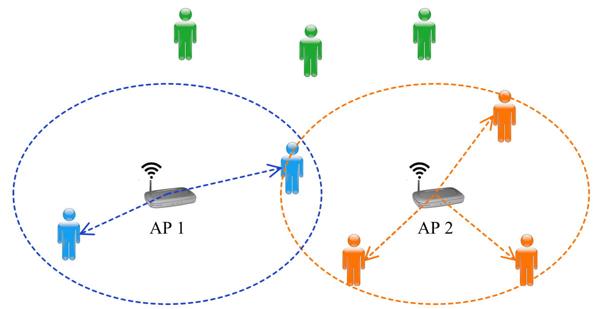

Fig. 3. Mapping of mobile users and wifi APs. Green users: No wifi coverage; blue users: AP 1; red users: AP 2.

B. Performance Analysis

We first conduct experiments for the scenario where the CP knows the instantaneous valuation of users in each time slot t, i.e., $W_u[t]$, $u \in \mathcal{U}$.

Fig. 4(a) presents the CP's payoff under different online sponsoring policies (in stage II). For a clear comparison, we do not consider edge caching in this experiment. To evaluate the performance of the Lyapunov online sponsoring policy, we choose the Greedy policy and the SemiGreedy policy as baselines, the details of which are provided in Appendix A. The Greedy policy aims to maximize the performance in each time slot with all remaining budget, whereas the SemiGreedy policy aims to maximize the performance in each time slot with part of the remaining budget. To draw a more convincing conclusion, we compare our proposed algorithm with the offline sponsoring strategy under complete information, which serves as an upper bound. As shown in Fig. 4(a), Lyapunov outperforms the two baselines under different sponsoring budgets (with an average gain of $15\% \sim 75\%$ in terms of the CP's payoff), and it can reach an average 90% of the offline performance under complete information. Moreover, the SemiGreedy algorithm outperforms the Greedy algorithm, as it can achieve a better budget allocation among time slots.

Next, we provide the analysis of joint optimization, where we set the Lyapunov algorithm as the cellular sponsor policy. We first define the following scenarios for comparison:

- Joint Best Cache (JBC): the CP jointly uses edge caching and data sponsoring, with the knowledge of real content popularity.
- Joint Opportunistic Cache (JOC): the CP jointly uses the opportunistic edge caching strategy and data sponsoring. The key idea of the opportunistic caching strategy is caching a content when it is accessed for the first time by some users.
- Separate Best Cache (SBC): the CP uses the best caching strategy and data sponsoring separately, i.e., the CP caches content without considering the data sponsoring and uses the remaining budget for data sponsoring.
- Separate Opportunistic Cache (SOC): the CP uses the opportunistic caching strategy and data sponsoring separately, i.e., the CP caches content without considering the data sponsoring, and it uses the remaining budget for data sponsoring.

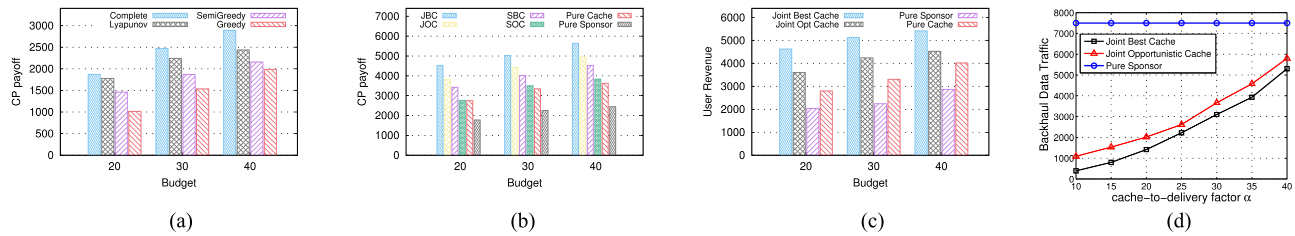

Fig. 4. (a) The CP's payoff under different sponsoring policies without edge caching. (b) The CP's payoff under Lyapunov sponsoring policy with different edge caching policies. (c) The user revenue under both joint and pure policies. (d) The backhaul data traffic under different edge caching policies in the case that **the CP can observe the users' valuations**.

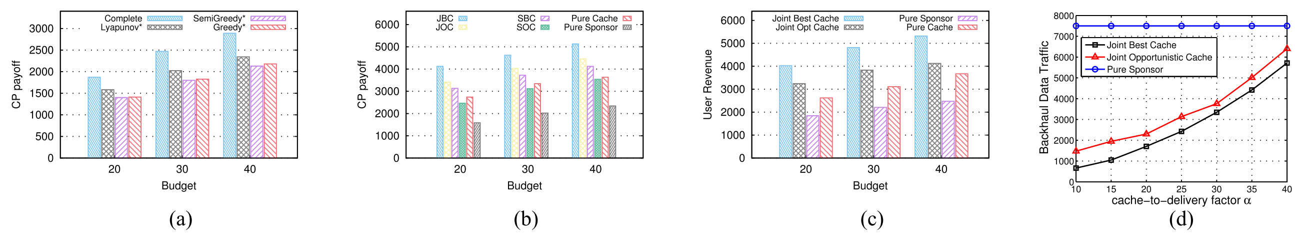

Fig. 5. (a) The CP's payoff under different sponsoring policies without edge caching. (b) The CP's payoff under Lyapunov sponsoring policy with different edge caching policies. (c) The user revenue under both joint and pure policies. (d) The backhaul data traffic under different edge caching policies in the case that **the CP cannot observe the users' valuations**.

- Pure Cache: the CP uses only the best caching strategy.
- Pure Sponsor: the CP uses only the data sponsoring strategy (Lyapunov).

Fig. 4(b) presents the CP's payoff under different scenarios. As shown, the joint utilization of edge caching and data sponsoring can improve the CP's payoff by $124\% \sim 154\%$ compared with pure data sponsoring ($50\% \sim 65\%$ compared with pure edge caching). We also observe that using edge caching and data sponsoring jointly can improve the CP's payoff by $25\% \sim 32\%$ compared with using both separately. The reason for this result is that using both separately may induce an inefficient budget distribution between edge caching and data sponsoring.

Fig. 4(c) shows the total user revenue under joint and pure scenarios. The user revenue under joint utilization can outperform that under the pure strategy by $33\% \sim 130\%$. An observation is that the user revenue increases with the CP's payoff, indicating that a policy with a higher CP's payoff will also induce higher user revenue. Although we aim to optimize the CP's payoff, this observation implies a win-win situation between the CP and users.

Fig. 4(d) presents the backhaul data traffic under different scenarios. We observe that the backhaul data traffic can be reduced significantly with the edge caching. Moreover, the backhaul data traffic reduction decreases with the *cache-to-delivery* factor α, which implies that the joint framework can reduce more backhaul data traffic with a lower caching cost (where more requests will be served via the WiFi network). The joint sponsoring framework can save 87% backhaul data traffic for users in the best case.

We now conduct experiments for the scenario where the CP cannot observe the instantaneous valuation of users in each time slot t, i.e., $W_u[t]$, $u \in \mathcal{U}$. As shown in Fig. 5(a), the Lyapunov-based sponsoring algorithm outperforms the baselines in different CP budgets with a $13\% \sim 20\%$ gain in the CP's payoff and can reach approximately 88% of payoff under complete information. The SemiGreedy algorithm performs worse than the Greedy algorithm, possibly because without users' accurate decisions on accepting sponsoring, the Greedy algorithm consumes the CP's budget less quickly and becomes closer to the optimal value.

Fig. 5(b) shows that in the case where the CP does not know the instant user valuation, the joint optimization improves $119\% \sim 160\%$ in the CP's payoff compared to the pure sponsoring scenario ($38\% \sim 50\%$ compared to the pure caching scenario). In addition, the joint utilization outperforms the separate utilization by $24\% \sim 31\%$. Fig. 5(c) shows that the user revenue under joint utilization can outperform that under the pure strategy by $45\% \sim 118\%$.

We also compare the backhaul data traffic of users in this case. Fig. 5(d) illustrates that backhaul data traffic is reduced in the joint framework. The CP can achieve 80% backhaul data traffic savings when $\alpha = 10$ with joint optimization.

From the above two practical scenarios, we find that the Lyapunov online algorithm always outperforms other online algorithms, both in the pure sponsor case and the joint framework. When we deploy edge caching, end users significantly benefit from the energy savings for connecting to WiFi networks rather than cellular networks, which is particularly important for mobile devices with low battery capacity.

Fig. 6 illustrates the payoff versus user density under caching strategies with different accuracies. As shown, the payoff of edge caching schemes always increases with user density, and the rate of increase is larger when the user density is low. In the no caching case, the CP's payoff remains almost the same under different user densities. We further observe that given the user density, when the prediction of request number is accurate, edge

TABLE II
TIME CONSUMPTION OF ALGORITHMS IN PEAK HOUR

Algorithm	Greedy	Greedy*	SemiGreedy	SemiGreedy*	Lyapunov	Lyapunov*
Time (sec)	0.086	0.203	0.085	0.198	0.087	0.101

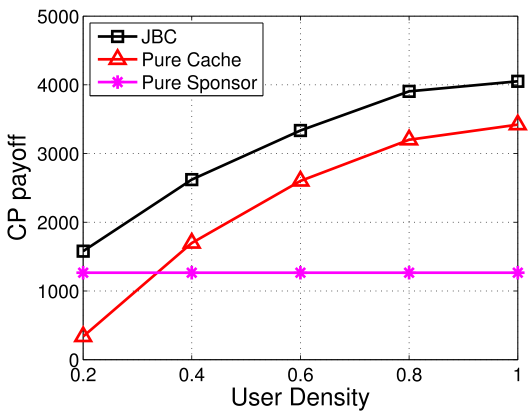

Fig. 6. The CP's payoff under different prediction error levels and user densities.

caching enhances the CP's payoff dramatically. Conversely, if the prediction error of request number is high, edge caching may even decrease the CP's payoff because the deployment of caching contents is costly if they cannot obtain enough access. This result implies that user density significantly affects the performance of edge caching. The CP's payoff can outperform pure data sponsoring by $30\% \sim 100\%$ under different user densities.

In terms of time consumption, we list the average scheduling computation times in peak hour in Table II, from which we see that the algorithms without user valuations require more computation time than the corresponding algorithms with user valuations. This is caused by the estimation process in the algorithms without user valuations. The overall processing time is relatively short, thereby satisfying real-time processing. The running time measurement was conducted using a desktop with an Intel Core i7-3770 CPU @ 2.7 GHz $\times$ 4.

VII. CONCLUSION

In this paper, we studied the joint optimization of data sponsoring and edge caching for mobile video delivery, aiming at maximizing the CP's revenue. We formulated the joint optimization problem as a two-stage decision problem, and we solved the problem using the Lyapunov optimization technique and the predicted edge caching principle. The simulation results on large-scale real data traces indicate that our design significantly improves the CP's payoff. The backhaul data traffic is dramatically reduced due to the edge caching.

APPENDIX

A. The Greedy/SemiGreedy Policy in Stage II

With the greedy policy, the CP makes the sponsoring decision in each time slot that maximizes the CP's instantaneous payoff in that time slot within the remaining budget.

We first introduce the greedy policy for the case in which the CP can observe the instantaneous valuations of users realized in each time slot t, i.e., $W_u[t], \forall u \in \mathcal{U}$.[10] In this case, the CP can compute the precise instantaneous payoff $R[t]$ in each time slot t under any sponsoring decision, and hence, it can greedily solve for the best sponsoring strategy that maximizes the precise payoff $R[t]$ in time slot t, i.e.,

$$\begin{aligned} \text{(Greedy)} \quad & \max \quad R[t] \\ & s.t. \quad (6), (7), (8), (9) \\ & \qquad C[t] \le B[t-1] \\ & var. \quad x_u[t], y_u[t] \in \{0,1\}, \ \forall u \in \mathcal{U}, \end{aligned} \tag{31}$$

where $B[t-1] = B_{\text{spon}} - \sum_{\tau \in [1:t-1]} C[\tau]$ is the total remaining budget at time slot t. In practice, $W_u[t]$ is generally user u's private valuation and cannot be directly observed by the CP. Hence, we will re-design the greedy policy based on the stochastic information regarding $W_u[t]$.

With the above greedy policy, the CP will make the sponsoring decision in each time slot using as much budget as possible. This may cause an imbalanced allocation of budget among time slots. To this end, we further introduce a modified greedy policy (called *SemiGreedy policy*), which restricts the budget consumption in each time slot and can hence achieve more even budget allocation. Formally, the CP greedily solves for the best sponsoring strategy that maximizes the precise payoff $R[t]$ in time slot t *using the reduced remaining budget*, i.e., we replace the constraint $C[t] \le B[t-1]$ with $C[t] \le \frac{B[t-1]}{T-t+1}$. Here, only a portion of the remaining budget can be used in time slot t.

We now introduce the *Greedy** policy in the case that the CP cannot observe the instantaneous valuations of users realized in each time slot but knows only the stochastic distribution information regarding user valuation. Specifically, with the stochastic information regarding $W_u[t]$, the CP can only compute the average probability of each user u accepting the cellular/WiFi sponsoring, i.e., $P_u^w[t]$ and $P_u^c[t]$. Recall that the estimations of $P_u^w[t]$ and $P_u^c[t]$ are provided in (2) and (3), respectively. Then, we can compute the expected cellular data downloading cost induced by cellular sponsoring as follows:

$$\widetilde{C}[t] \triangleq \sum_{u \in \mathcal{U}} x_u[t] \cdot C_u[t] \cdot P_u^c[t], \tag{32}$$

and the expected sponsoring value (revenue) that can be achieved from sponsoring as follows:

$$\widetilde{V}[t] \triangleq \sum_{u \in \mathcal{U}} (x_u[t] \cdot P_u^c[t] + y_u[t] \cdot P_u^w[t]) \cdot V_u[t]. \tag{33}$$

[10]Note that the content requests $s_u[t]$ and locations $l_u[t]$ of all users in time slot t will be naturally revealed to the CP.

Thus, the expected payoff that the CP can achieve in each time slot t can be computed as follows:

$$\widetilde{R}[t] = \widetilde{V}[t] - \widetilde{C}[t]. \tag{34}$$

Note that such an expected payoff based on the stochastic valuation information may be different from the precise instantaneous payoff based on the complete valuation information. Nevertheless, when the number of users is large enough, this expected payoff converges to the precise payoff asymptotically:

$$\lim_{|\mathcal{U}|\to\infty} \widetilde{C}[t] = C[t], \quad \lim_{|\mathcal{U}|\to\infty} \widetilde{V}[t] = V[t], \quad \forall t \in \mathcal{T}.$$

Proof: Note that problem $C[t]$ and $V[t]$ is time independent but user dependent. Hence, the sponsor decisions in different time slots with the same user valuation must be identical, i.e., $C_u[t_1] = C_u[t_2]$, if $\theta[t_1] = \theta[t_2]$.

$$(1) \quad C[t] = \sum_{u=1}^{|\mathcal{U}|} C_u[t] \cdot x_u[t], \quad s.t.\ x_u[t] = 0,\ \forall u \notin \mathcal{U}^c[t].$$

□

The probability for $u \in \mathcal{U}^c[t]$ is $P_u^c[t]$. Then, by the law of large numbers (LLN), we have the following: When $|\mathcal{U}| \to \infty$,

$$(2) \quad C[t] \to \widetilde{C}[t] = \sum_{u\in\mathcal{U}} x_u[t] \cdot C_u[t] \cdot P_u^c[t], \quad \forall u \in \mathcal{U}[t].$$

This implies that $C[t]$ and $\widetilde{C}[t]$ are equivalent if $|\mathcal{U}| \to \infty$.

Based on the expected payoff $\widetilde{R}[t]$ in each time slot t under any sponsoring decision, the CP can greedily solve for the best sponsoring strategy that maximizes the expected payoff $\widetilde{R}[t]$ in time slot t, i.e.,

$$\begin{aligned} \text{(Greedy*)} \quad \max \quad & \widetilde{R}[t] \\ s.t. \quad & (6), (7), \\ & \widetilde{C}[t] \le \widetilde{B}[t-1] \\ var. \quad & x_u[t], y_u[t] \in \{0,1\},\ \forall u \in \mathcal{U}, \end{aligned} \tag{35}$$

where $\widetilde{B}[t-1] = B_{\text{spon}} - \sum_{\tau\in[1:t-1]} \widetilde{C}[\tau]$ is the total remaining budget at time slot t.

Similarly, we can also design a *SemiGreedy** algorithm in the case that the CP cannot observe the instantaneous valuations of users. The CP can greedily solve for the best sponsoring strategy that maximizes the expected payoff $\widetilde{R}[t]$ in time slot t *using the reduced remaining budget*, i.e., we replace the constraint $\widetilde{C}[t] \le \widetilde{B}[t-1]$ with $\widetilde{C}[t] \le \frac{\widetilde{B}[t-1]}{T-t+1}$. In this way, only a portion of the remaining budget can be used in time slot t.

REFERENCES

[1] "Cisco visual networking index: Global mobile data traffic forecast update 2016–2021," Cisco, San Jose, CA, USA, White Paper, 2017.

[2] Y. Wen, X. Zhu, J. J. P. C. Rodrigues, and C. W. Chen, "Cloud mobile media: Reflections and outlook," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 885–902, Jun. 2014.

[3] H. Hu *et al.*, "Joint content replication and request routing for social video distribution over cloud CDN: A community clustering method," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1320–1333, Jul. 2016.

[4] M. Tang, L. Gao, H. Pang, J. Huang, and L. Sun, "Optimizations and economics of crowdsourced mobile streaming," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 21–27, Apr. 2017.

[5] Lin Gao, M. Tang, H. Pang, J. Huang, and L. Sun, "Performance bound analysis for crowdsourced mobile video streaming," in *Proc. IEEE Conf. Inf. Sci. Syst.*, 2016, pp. 366–371.

[6] M. Tang, S. Wang, Lin Gao, J. Huang, and L. Sun, "MOMD: A multi-object multi-dimensional auction for crowdsourced mobile video streaming," in *Proc. IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.

[7] C. Joe-Wong, S. Ha, and M. Chiang, "Sponsoring mobile data: An economic analysis of the impact on users and content providers." in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1499–1507.

[8] L. Zhang, W. Wu, and D. Wang, "Sponsored data plan: A two-class service model in wireless data networks," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Model. Comput. Syst.*, 2015, pp. 85–96.

[9] M. Andrews, U. Ozen, M. Reiman, and Q. Wang, "Economic models of sponsored content in wireless networks with uncertain demand," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2013, pp. 345–350.

[10] R. ElDelgawy and R. J. La, "Interaction between a content provider and a service provider and its efficiency," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 5890–5895.

[11] Y. Wu, H. Kim, P. H. Hande, M. Chiang, and D. H. K. Tsang , "Revenue sharing among ISPs in two-sided markets," in *Proc. IEEE INFOCOM*, 2011, pp. 596–600.

[12] 2014. [Online]. Available: http://www.businessinsider.com/att-sponsored-data-2014-7

[13] G. Ma *et al.*, "Understanding performance of edge content caching for mobile video streaming," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1076–1089, May 2017.

[14] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the Air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[15] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled tactile internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

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

[17] L. Chen, Y. Zhou, M. Jing, and R. T. B. Ma, "Thunder crystal: A novel crowdsourcing-based content distribution platform," in *Proc. 25th ACM Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2015, pp. 43–48.

[18] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, 2012, pp. 1107–1115.

[19] D. Purkayastha, J. Li, B. Balazinski, J. Cartmell, and A. Reznik, "Edge caching in a small cell network," *Int. J. Comput. Netw. Technol.*, vol. 2, no. 1, pp. 41–46, 2014.

[20] A. Gharaibeh, A. Khreishah, and I. Khalil, "An O(1)-Competitive online caching algorithm for content centric networking," in *Proc. IEEE 35th Annu. Int. Conf. Comput. Commun.*, 2016, pp. 1–9.

[21] D. Liu and C. Yang, "Cache-enabled heterogeneous cellular networks: Comparison and tradeoffs," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–6.

[22] J. Sung, M. Kim, K. Lim, and J.-K. K. Rhee, "Efficient cache placement strategy in two-tier wireless content delivery network," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1163–1174, Jun. 2016.

[23] S. Li, J. Xu, M. van der Schaar, and W. Li, "Trend-Aware video caching through online learning," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2503–2516, Dec. 2016.

[24] H. Pang, Lin Gao, Q. Ding, and L. Sun, "When data sponsoring meets edge caching: A game-theoretic analysis," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[25] H. Pang, L. Gao, and L. Sun, "Joint optimization of data sponsoring and edge caching for mobile video delivery," in *Proc. IEEE Global Commun. Conf.*, 2016, pp. 1–7.

[26] D. K. Krishnappa, S. Khemmarat, L. Gao, and M. Zink, "On the feasibility of prefetching and caching for online TV services: A measurement study on hulu," in *Proc. 12th Int. Conf. Passive Active Netw. Meas.*, 2011, pp. 72–80.

[27] M. Tang, L. Gao, H. Pang, J. Huang, and L. Sun, "Multi-dimensional auction mechanism for mobile crowdsourced video streaming," in *Proc. IEEE 14th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw.*, 2016, pp. 1–8.

[28] N. Ristanovic, J.-Y. L. Boudec, A. Chaintreau, and V. Erramilli, "Energy efficient offloading of 3G networks," in *Proc. IEEE 8th Int. Conf. Mobile Ad-Hoc Sensor Syst.*, 2011, pp. 202–211.
[29] J. Xue, D. Q. Zhang, H. Yu, and C. W. Chen *et al.*, "Assessing quality of experience for adaptive HTTP video streaming," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, 2014, pp. 1–6.
[30] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 843–855, Apr. 2011.
[31] Q. Ma, L. Gao, Y. F. Liu, and J. Huang, "Incentivizing Wi-Fi network crowdsourcing: A contract theoretic approach," *IEEE/ACM Trans. Netw.*, pp. 1–14, 2018.
[32] C. Jiang, L. Gao, L. Duan, and J. Huang, "Scalable mobile crowdsensing via peer-to-peer data sharing," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 898–912, Apr. 2017.
[33] L. Gao, L. Duan, and J. Huang, "Two-Sided matching based cooperative spectrum sharing," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 538–551, Feb. 2017.
[34] Y. Luo, L. Gao, and J. Huang, "An integrated spectrum and information market for green cognitive communications," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3326–3338, Dec. 2016.
[35] L. Gao, Y. Xu, and X. Wang, "MAP: Multi-auctioneer progressive auction for dynamic spectrum access," *IEEE Trans. Mobile Comput.*, vol. 10, no. 8, pp. 1144–1161, Aug. 2011.
[36] M. J. Neely, *Stochastic Network Optimization With Application to Communication and Queueing Systems*. San Rafael, CA, USA: Morgan & Claypool, 2010.
[37] L. Gao, F. Hou, and J. Huang, "Providing long-term participation incentive in participatory sensing," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2803–2811.
[38] Y.-P. Hsu and L. Duan, "To motivate social grouping in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4880–4893, Aug. 2017.
[39] L. Gao, M. Tang, H. Pang, J. Huang, and L. Sun, "Multi-user cooperative mobile video streaming: Performance analysis and online mechanism design," *IEEE Trans. Mobile Comput.*, p. 1, 2018.
[40] Y.-P. Hsu, E. Modiano, and L. Duan, "Age of informaiton: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 561–565.

Lifeng Sun (M'05) received the B.S. and Ph.D. degrees in system engineering from the National University of Defense Technology, Changsha, China, in 1995 and 2000, respectively. He joined the Department of Computer Science and Technology (CST), Tsinghua University (THU), Beijing, China, in 2001. He is currently a Full Professor with CST, THU. His research interests include networked multimedia, video streaming, 3-D/multiview video coding, multimedia content analysis, multimedia cloud computing, and social media. He has authored or coauthored more than 100 high quality papers and one chapter of a book. He was the recipient of the Annual Best Paper Award of the IEEE Transactions on Circuits and Systems for Video Technology (2010), the Best Paper Award of the ACM Multimedia (2012), and the Best Student Paper Award of Multimedia Modeling (2015) and the IEEE Multimedia Big Data (2017).

Haitian Pang (S'16) received the B.E. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2014. He is currently working toward the Ph.D. degree in computer science with Tsinghua University. His research interests include deep learning for network, cellular-WiFi networking, video streaming system design, and mobile networking optimizations.

Lin Gao (S'08–M'10–SM'16) received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China in 2010. He is currently an Associate Professor with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China. From 2010 to 2015, he was a Postdoctoral Fellow with The Chinese University of Hong Kong. His main research interests include the area of network economics and games, with applications in wireless communications and networking. He was the co-receipt of three Best Paper Awards from WiOpt 2013, 2014, 2015, and one Best Paper Award Finalist from the IEEE INFOCOM 2016. He was the recepient of the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2016.