Motivating Human-Enabled Mobile Participation for Data Offloading

Xiaonan Zhang, *Student Member, IEEE*, Linke Guo, *Member, IEEE*, Ming Li, *Member, IEEE*, and Yuguang Fang, *Fellow, IEEE*

Abstract—The exploding popularity of mobile devices enables people to enjoy benefits brought by various interesting mobile apps. However, the ever-increasing data traffic has exacerbated the congestion on current cellular networks, which results in users' dissatisfaction, especially in crowded areas. Hence, how to alleviate data traffic in cellular networks becomes a challenging problem. Traditional methods rely on mobile offloading techniques to deviate the data traffic originally targeted to cellular networks, such as the small cell, Wi-Fi, and opportunistic communication. Unfortunately, mobile users still experience severe congestion when a large number of users request for data. Facing these challenges, we introduce the concept of mobile participation to assist data offloading by leveraging the mobility of users and the social features among a group of users. A mobile caching user, who pre-caches a certain amount of contents, will roam around congested areas to participate in content dissemination in order to satisfy users' requests, which is expected to benefit both himself and users in the crowd simultaneously. To motivate such human-enabled mobile participation for data offloading, a Stackelberg game is deployed with joint considerations on social effect and delay effect. Based on detailed performance analysis, we demonstrate the feasibility and efficiency of the proposed approach.

Index Terms—Data offloading, mobile participation, homophily phenomenon, Stackelberg game

1 Introduction

The soaring popularity of mobile devices enables people to communicate with their social ties at any time and from anywhere. People use mobile apps to create and exchange a huge amount of data with their social interactions in cyberspace. Reports warn that monthly global mobile data traffic will surpass 24.3 exabytes and the mobile data traffic from smartphones will reach three-quarters by 2019 [1]. Although cellular network operators exploit their efforts to provide better services in terms of higher data rate and lower costs, users are still facing poor performance in their daily life, especially in some crowded areas, such as football stadiums, theme parks, and airports. However, the above crowded areas are the places that highly need reliable wireless communication, e.g., broadcasting evacuation information for safety purpose. As a promising solution, mobile data offloading takes advantages of small cell, Wi-Fi, and opportunistic communication to pro-actively reduce the data traffic targeted for cellular networks [2]. Unfortunately, although various types of mobile offloading schemes have been proposed in both academia and industry, we are still lacking effective methods. For example, utilizing small cells is not an effective method due to the scarcity of licensed spectrum bandwidths. Even worse, deploying more small cells will incur significant costs. Regarding Wi-Fi offloading, the service provider has access to much larger free spectrum to cater the Wi-Fi deployment. However, Wi-Fi offloading cannot provide guaranteed QoS, and Wi-Fi-enabled devices may experience increased battery drainage since it has to operate on two different radio interfaces [3]. To perform mobile offloading, opportunistic communication has been identified as another approach, which increases communication chances by utilizing the potential social connections among users and thus is beneficial to deliver contents. In particular, some works [4], [5] apply social-based approaches to help data dissemination among social ties or users with similar social profiles. Apparently, the opportunistic communication is not reliable for data delivery in an ad hoc mode because there is lack of incentives for source users to coordinate the data dissemination. Clearly, mobile offloading has not been well developed nor widely applied.

Facing these challenges and existing solutions, we take a step further to reconsider the human-enabled approach for mobile offloading, which takes human social behaviors and human activities into consideration. Intuitively, users with similar social interests often group together at certain location [6], which potentially results in similar content requests. For example, users gathered in specific attractions in the Disneyland may request the similar contents related to those attractions. When they request similar contents, network congestion would be caused due to limited bandwidth. Such congestion potentially prevents users from getting their requested contents. The above phenomenon leads us to consider how to avoid repeated requests/retrievals in order to reduce the number of accesses to the service provider (SP). A possible solution is to leverage users' similar social attributes to design a human-enabled data offloading scheme. In sociology [7], *homophily phenomenon* describes that people with more similar attributes contact more frequently than

- *X. Zhang and L. Guo are with the Department of Electrical and Computer Engineering, Binghamton University, State University of New York, Binghamton, NY 13902. E-mail: {xzhan167, lguo}@binghamton.edu.*
- *M. Li is with the Department of Computer Science and Engineering, University of Nevada, Reno, NV 89557. E-mail: mingli@unr.edu.*
- *Y. Fang is with the Department of Electrical and Computer Engineering and University of Florida, Gainesville, FL 32611. E-mail: fang@ece.ufl.edu.*

Manuscript received 31 Mar. 2017; revised 24 Oct. 2017; accepted 7 Nov. 2017. Date of publication 14 Nov. 2017; date of current version 1 June 2018. (Corresponding author: Linke Guo.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2017.2773087

complete strangers. The interactions between users with more contacts bring more *social effect*, which captures the advantages of word-of-mouth communication [8]. *Specifically, users typically form their opinions about the quality of the contents based on the information they obtain from other users. Thus, when a user demands more contents, his social friends would also request more contents due to the similarity of their interests.* Meanwhile, users with identical attributes could share their contents with each other using free device-to-device (D2D) communication. As for human activities, an observation is that users in crowded areas either walk around or go to their interested attractions. Hence, we can take advantage of the mobility of users to alleviate the congestion.

In this work, we propose a *human-enabled mobile participation* approach in data offloading by introducing a mobile caching user (MCU), who bridges the gap between the SP and users when the above congestion happens. Our approach is mainly divided into two steps. In the first step, we consider the data offloading between the MCU and the representing users (RUs) with similar content requests in crowded areas. Specifically, an MCU pre-caches a number of large volume contents in advance. After receiving congestion information (e.g., congestion area, requested contents, .etc) from the SP, the MCU chooses a specific crowded area where requested contents are similar with his own interests and is near to his current location, physically moves to the RUs in the chosen area and transfers the contents to them. In the second step, the RUs with obtained contents further disseminate content copies via D2D communication to other users opportunistically, who have the identical content requests with them. We mainly consider the first step, where delay-tolerant scenario and delay-sensitive scenario are discussed. In the delay-tolerant scenario, RUs would like to wait until they download the requested contents. Whereas in the delay-sensitive scenario, RUs are urgent to get the requested contents. They will be more dissatisfied with the increasing of the waiting time. Compared to traditional data offloading approaches, the proposed approach is significantly cheaper than the small cell build-out. Moreover, by physically moving to the crowd, the MCU makes data transmission more reliable and flexible than either Wi-Fi or pure D2D communication.

To motivate above human-enabled mobile participation, we design an incentive mechanism. While participating in human-enabled data offloading, the MCU spends a few time in moving and consumes his own resources such as battery and storage. Hence, he would not be interested in it unless he receives a satisfying revenue. As for RUs, they not only get the originally requested contents, but also harvest additional contents they may be interested in due to the similarity of their interests with other RUs, which largely improves their satisfactions. Since RUs request similar contents and pay for them individually, it is reasonable to assume that RUs are selfish and rational. Hence, each RU only wants to maximize his own satisfaction. To increase the MCU's total revenue and provide RUs' satisfaction, we will thoroughly investigate RUs' content requests, social effect, delay effect, and unit payment strategy for both the MCU and RUs in the proposed incentive mechanism.

Our Contributions: We highlight our major contributions as follows,

- We propose a new data offloading scheme that takes advantages of both *homophily phenomenon* and *mobile participation* to greatly reduce the congestion in crowded areas where users with similar interests are normally grouped together.
- Specifically, we consider two system models: the delay-tolerant model and the delay-sensitive model. In both models, by considering RUs' interactions, we formulate the communication between the MCU and RUs as a two-stage *Stackelberg game*. In Stage I, the MCU chooses a unit payment to maximize his total revenue. In Stage II, each RU chooses a requested content level given the unit payment to maximize his satisfaction on the received contents.
- For the delay-tolerant scenario, the interactions between RUs bring social effect. We first give an assumption under which we show the existence and uniqueness of the Nash equilibrium in Stage II. Then, we present an effective algorithm to compute the unique Stackelberg equilibrium in Stage I, at which the revenue of the MCU is maximized, and none of the RUs continue requesting contents by unilaterally deviating from his current strategy
- For the delay-sensitive scenario, the interactions between RUs not only bring social effect but also delay effect. We extend the Stakelberg game to the delay-sensitive model. To alleviate the serious delay effect, we propose two improved delay-sensitive models by further taking advantages of users' mobility, where the first one considers the queueing delay and the other introduces multiple MCUs.

The rest of this paper is organized as follows: In Section 2, we briefly review the existing data offloading approaches, economical incentives for performing data offloading and the social effect due to similar interests between RUs based on their social relationship. In Section 3, we explain our motivations of leveraging the homophily phenomenon and the mobile participation. Following with that, a detailed description of our proposed data offloading system models is given in Section 4, which are formulated as two-stage Stackelberg games respectively. In Section 5, we study the proposed Stackelberg game in the delay-tolerant scenario. To better adapt to the practical situation, we extend the Stackelberg game to the delay-sensitive scenario in Section 6. In Section 7, the performance of our data offloading approach is evaluated, followed by a conclusion in Section 8.

2 Related Work

2.1 Mobile Data Offloading

Mobile data offloading [3] is a promising way to alleviate traffic congestion and reduce the energy and bandwidth consumption. For example, Liang et al. in [9] offload their applications and data from mobile devices to the cloud to improve users' experience in terms of longer battery lifetime, larger data storage, faster processing speed and more powerful security services. Zhang et al. in [10] offload mobile users' applications to nearby mobile resource-rich devices (i.e., cloudlets) in an intermittently connected system to reduce energy consumption and improve performance. In this paper, we generally discuss the mobile offloading for cellular networks, which is classified into two categories [11]. Infrastructure-based mobile data offloading [12] refers to deploying small cell base stations and Wi-Fi hotspots for mobile users [7], [13]. The connection between mobile users and the base station is proposed to achieve

flow level load balancing under spatially heterogeneous traffic distributions in [14], [15]. However, the lack of cost-effective backhaul associations for base station often impairs their performance in terms of offloading mobile traffic. The second category is the ad-hoc-based mobile traffic offloading, which refers to applying short range communication as the underlay to offload mobile traffic [4], [5], [16], [17], [18].

2.2 Economic Incentives for Data Offloading

The above works mainly focus on the technical perspective adoption of data offloading without considering economic incentives. The incentive issue is significant for the case where Wi-Fi or small cell is privately owned by third-party entities, who are expected to be reluctant to admit non-registered users' traffic without proper incentives [19]. The incentive framework for the so-called user-initiated data offloading is considered in [20], [21], where users initiate the offloading process and offer necessary incentives in order to obtain their contents. Gao et al. in [19] consider the network-initiated data offloading, where cellular networks initiate the offloading process, and hence the network operators are responsible for incentivizing Wi-Fi.

2.3 Attribute-Based Social Effect

The above works do not consider homophily phenomenon [7]. Reingen et al. in [22] conduct a survey of the members of a sorority in which they measure brand preference congruity as a function of whether they live in the sorority house. They find that those who live together as a group have more congruent brand preferences than those who do not. Presumably, living together provides more opportunities for interaction and communication. Taking a further step, they note that information obtained from social tie connections will influence in decision making in [23].

The above observations and inference are deployed in several works. In [24], [25], [26], different privacy-preserving authentication schemes for mobile health networks are designed from a social perspective view. Users in online social networks apply their attributes to find matched friends and establish social relationships with strangers in [27]. Gong et al. in [28] study users' behaviors by jointly considering congestion effect in the physical wireless domain and social effect based on users' social relationship. In [29], [30], a social group utility maximization framework, which captures the impact of mobile users' diverse social ties, is studied. Considering the social effect brought by social ties among users, different pricing strategies of a monopolist have been studied in [31]. In our previous work [32], the social effect brought by users' similar social attributes is deployed to assist data offloading. However, the introduction of the MCU brings severe delay effect, which negatively affects the data offloading performance. To alleviate delay effect, we take the queue and multi-leader Stackelberg game into consideration now, which differentiates our paper with [28]. We focus on incentive mechanisms to motivate human-enabled mobile participation for data offloading under both social effect and severe delay effect.

3 Motivations and Preliminaries

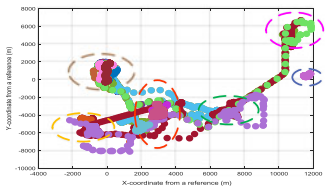
3.1 Social Enabled Data Offloading

Given a pair of strangers, one cannot push another to help recommend/forward his contents if they do not have any pre-established relationship. However, comparing with complete strangers, people may intend to help the one that shares some similarities in terms of attributes, e.g., language, nationality, affiliation, etc. As discussed in [33], it is a well-accepted nature of human interaction that people like to interact with those who are similar to themselves, which is often termed the "like me" principle. In [34], [35], the authors conduct an experiment based on the trace file collected during the INFOCOM 2006 [36], which analyzes the relationship between the contact rate and the number of identical attributes. The result shows that the contact rate in terms of the number of contacts between two users increases with the increment of identical attributes, which further validates the "like-me" principle. Therefore, a potential social tie can be set up based on the attribute similarity. Furthermore, Reingen et al. in [23] find that information obtained from strong tie connections are more influential in decision making than weak tie connections at a micro level (information flows within dyads or small groups). Motivated by it, content dissemination would be more efficient given the assumption that more attribute similarities exist between users. In addition, users who share similar interests intend to form a group and they can forward messages to others in the group more efficiently according to [37]. Hence, we infer that the social-enabled content dissemination would be much more efficient if users apply attribute similarity to form the attribute-similar group.

Motivated by the above discussions, we consider human's similar social attributes. In the scenarios where users group together based on their similar social attributes, such as interests, their requested contents have a higher probability to be similar even identical due to their influence on each other. Hence, we could select RUs to request contents and further disseminate them to other users via D2D communication. Thus, users can obtain more interested contents and their satisfactions are improved.

3.2 Mobile Participation

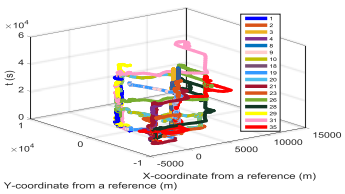
We conduct an experiment analyzing human mobility traces using the real data trace file [38] in order to show the feasibility of mobile participation. The human traces are obtained every 30 seconds from 40 volunteers who spent their Thanksgiving and Christmas holidays in Disney World, Florida, US. We describe all the locations the volunteers have gone to as shown in Fig. 1a, in which we circle the locations that are visited most. By comparing it with the real Disney World map [39] in Fig. 1b, we find that those circled locations are exactly the crowded attraction areas, where users with similar interests get together and request similar contents. For example, at the Rock 'n' Roller Coaster Starring Aerosmith attraction, many young visitors who enjoy the trilling feelings group together and they are more interested in exciting contents. In addition, we draw 17 volunteers' mobile traces as time changes in Fig. 2a, which verifies the mobility of volunteers. Meanwhile, we illustrate volunteers' locations in different time-slots in Fig. 2b, where we see that volunteers are distributed in all crowded attraction areas in each time-slot. Inferring from the observations in Figs. 1 and 2, we conclude that: 1). volunteers move as time changes; 2), there always exist volunteers in each attraction in each time-slot. Therefore, leveraging mobile participation is feasible to achieve content delivery and dissemination.



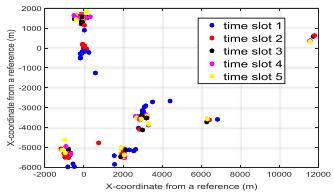
(a) Potential location in Real Trace

(b) Disney Map

Fig. 1. Potential location of the MCU.



(a) Potential Location vs. Time



(b) Locations in Different Time Slot

Fig. 2. Time changes versus potential location.

4 System Model and Problem Formulation

4.1 Overview

To assist the description, we continue the example in Disney World as shown in Fig. 3, where the yellow area is denoted as the Rock 'n' Roller Coaster Starring Aerosmith attraction. It is divided into two time-slots. In time-slot 1, no congestion exists in the yellow area. David downloads numbers of contents and continues to visit other attractions. In time-slot 2, an increasing number of users with similar interests group together and request for contents related to the attraction, which results in severe congestion. As a result, users cannot get the requested contents from the SP. The SP asks David for help via transmitting him the short message related to the congestion information. Since David is interested in the same attraction and can obtain extra revenue, he moves back to disseminate the contents after checking the distance availability between himself and the chosen attraction. He first announces the unit payment for the requested contents. Each RU chooses a requested content quantity to maximize his satisfaction based on the unit payment and other RU's choices, which is submitted to David. David maximizes the total revenue and computes the corresponding unit payment which is returned to RUs. Such communication between David and RUs is processed iteratively until David and RUs reach an agreement, in which David gets the maximized revenue and RUs satisfy the content obtaining experience. Finally, RUs disseminate their contents to other users in the crowd via D2D communication.

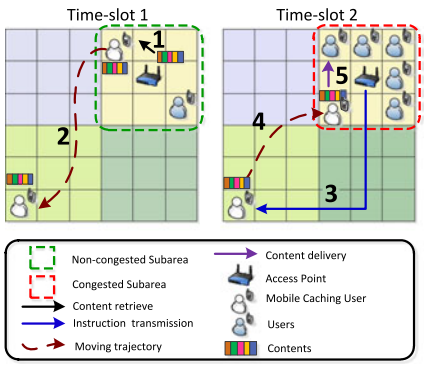


Fig. 3. System model of mobile participation.

4.2 System Model

Depending on RUs' sensitiveness to the waiting time for the requested contents, two models are considered: *delay-tolerant model* and *delay-sensitive model*.

4.2.1 Delay-Tolerant Model

In the delay-tolerant model, RUs do not care their waiting time. Assume a set of RUs $\mathcal{N} = \{1, 2, \ldots, i, \ldots N\}$ group together and cannot get their requested contents from the SP directly, where N denotes the total number of RUs. Their corresponding requested content level profile is represented as $\mathbf{x} = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}^T \in [0, \infty)^N$, which quantifies the contents they request from the MCU. Let $x_i \in [0, \infty)$ denote the requested content level of the RU i and $\mathbf{x}_{-i}$ denote the requested content levels of other RUs except for the RU i. According to [31], the RU i's satisfaction consists of the following two parts: 1), internal characteristics, represented by the maximum internal demand rate $a_i > 0$ and the internal demand elasticity factor $b_i > 0$. The internal demand rate represents the maximum satisfaction that each RU gets given unit content level whereas the elasticity factor measures the sensitivity of the RU's satisfaction to changes in content levels [40]. 2), external characteristics, represented by social effect that RU j brings to RU i, quantified by $g_{ij} > 0, \forall j \in \mathcal{N}$ and $j \neq i$. Since utility is a terminology in game theory and economics to represent the satisfaction experienced by the consumer of a good [41], the satisfaction of each RU is quantified by utility hereinafter. Given the unit payment p the MCU charges RUs, the utility of RU i is quantified as

$$u_i(x_i, \mathbf{x}_{-i}, p) = a_i x_i - \frac{1}{2} b_i x_i^2 + \sum_{j \neq i} g_{ij} x_i x_j - p x_i, \forall i. \quad (1)$$

The quadratic form in (1) not only allows for tractable analysis but also serves a good second-order approximation for a broad class of concave utility functions [31].

Given RUs' requested content levels, the total revenue of the MCU is

$$R(\mathbf{x}, p) = \sum_{i \in \mathcal{N}} (p - c) x_i, \quad (2)$$

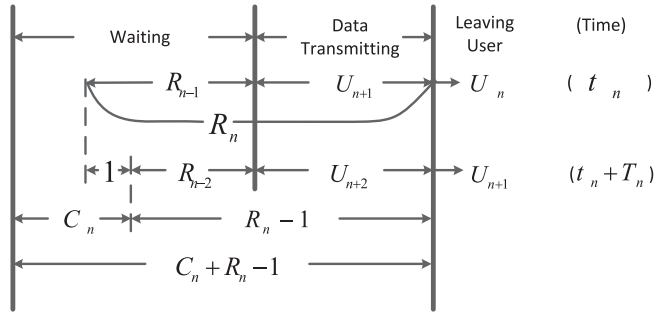


Fig. 4. $M/G/1$ queue in delay-sensitive model.

TABLE 1
Notations in $M/G/1$ Queue

Symbols	Meaning
R_n	the remaining requested content levels in the queue after the content delivery to user n
T_n	the content transmission period for user n
C_n	the content requests newly coming to the queue while user $n+1$ is receving the requested contents
t_n	the time at which the content transmission for user n is finished
$t_n + T_n$	the time at which the content transmission for user $n+1$ is finished

where c is the unit cost the MCU spends when transmitting contents to RUs, including energy and move consumption.

4.2.2 Intuitive Delay-Sensitive Model

Due to the difference of RUs' requested contents, the MCU moves to RUs and delivers contents to them one by one. As a result, each RU has to wait for the content transmission from the MCU when multiple RUs request contents. If they are urgent to obtain the requested contents, their utilities would be lowered due to long waiting time.

Assume RUs do not know the transmission order of the MCU in advance. Each RU would consider the worst case that he is the last one to receive the contents. To clearly show the time delay effect, we assume the transmission rates between the MCU and RUs are normalized and the same. The utility of the RU i in the delay-sensitive model is

$$\begin{aligned}\overline{u}_i(\overline{x}_i, \overline{\mathbf{x}}_{-i}, \overline{p}) = a_i\overline{x}_i - \frac{1}{2}b_i{\overline{x}_i}^2 + \sum_{j\in\mathcal{N}} g_{ij}\overline{x}_i\overline{x}_j \\ - \frac{1}{2}d\left(\sum_{j\in\mathcal{N}}\overline{x}_j\right)^2 - \overline{p}\overline{x}_i, \forall i,\end{aligned} \tag{3}$$

where d is the delay effect coefficient determined by the SP. Compared (3) with (1), the social relationship between RUs brings not only positive social effect but also severe delay effect in the intuitive delay-sensitive model.

The total revenue of the MCU keeps unchanged

$$\overline{R}(\overline{\mathbf{x}}, \overline{p}) = \sum_{i\in\mathcal{N}} (\overline{p} - c)\overline{x}_i. \tag{4}$$

4.2.3 Queue Delay-Sensitive Model

The potential assumption in the above intuitive delay-sensitive model is that the MCU begins transmission after the SP receives content requests from all RUs. If the SP can predict the potential congestion effect at some locations, it could arrange the MCU to move to these locations in advance instead of asking the MCU for help after congestion effect appears. Because the SP keeps the historical data monitoring records, the above assumption is easily satisfied. Thus, when an RU broadcasts a content request, the MCU could transmit the content to him on time. Simultaneously, the content requests from other RUs continuously arrive at the MCU. Content transmission from the MCU to RUs forms a First In First Out (FIFO) queue model in Fig. 4. The notations are listed in Table. 1.

In the queue delay-sensitive model, we assume the levels of newly arrival requested contents C_n in a finite interval of length t follows the Poisson distribution with mean arrival rate λ: $P\{C_n = j | T_n = t\} = \frac{(\lambda t)^j}{j!}e^{-\lambda t}$. The Poisson process is a viable model when contents originate from a large population of independent RUs. Due to the similar interests of RUs at the same location, most of their requested content levels distribute in the same interval. Given unit content transmission speed, the content transmission time is modeled to follow the Gaussian distribution with mean $\mu \gg 0$ and variance σ^2. Assume the traffic intensity $\rho = \lambda/\mu < 1$ for stability. Based on Pollaczek-Khinchin (P-K) formula [42], the expected RU waiting time W_q for each RU is

$$W_q = \frac{\rho^2 + \lambda^2\sigma^2}{2\lambda(1-\rho)}. \tag{5}$$

Considering the waiting time, each RU's utility becomes

$$\begin{aligned}\hat{u}_i(\hat{x}_i, \hat{\mathbf{x}}_{-i}, \hat{p}) = a_i\hat{x}_i - \frac{1}{2}b_i{\hat{x}_i}^2 + \sum_{j\in\mathcal{N}} g_{ij}\hat{x}_i\hat{x}_j \\ - k\frac{\rho^2 + \lambda^2\sigma^2}{2\lambda(1-\rho)} - \hat{p}\hat{x}_i, \forall i,\end{aligned} \tag{6}$$

where k is the congestion coefficient. According to the historical records, the SP can predict the traffic mean arrival rate λ. One observation is that contents related to each attraction are time-invariant. Thus, the SP could also evaluate the current traffic intensity ρ. Since different RUs request contents when congestion effect happens, the variance σ^2 is unknown. Point estimation [43] is applied to estimate σ^2

$$\hat{\sigma^2} = \frac{1}{N-1}\sum_{j\in\mathcal{N}}\left(\hat{x}_j - \frac{1}{N}\sum_{m\in\mathcal{N}}\hat{x}_m\right)^2. \tag{7}$$

Substitute (7) into (6), the utility becomes

$$\begin{aligned}\hat{u}_i(\hat{x}_i, \hat{\mathbf{x}}_{-i}, \hat{p}) = a_i\hat{x}_i - \frac{1}{2}b_i{\hat{x}_i}^2 + \sum_{j\in\mathcal{N}} g_{ij}\hat{x}_i\hat{x}_j - k\frac{\rho^2}{2\lambda(1-\rho)} \\ - k\frac{\lambda}{2(1-\rho)}\frac{1}{N-1}\sum_{j\in\mathcal{N}}\left(\hat{x}_j - \frac{1}{N}\sum_{m\in\mathcal{N}}\hat{x}_m\right)^2 - \hat{p}\hat{x}_i, \forall i.\end{aligned} \tag{8}$$

The total revenue of the MCU is the same as that in the intuitive delay-sensitive model.

4.2.4 Multi-Leader Delay-Sensitive Model

Another observation in the intuitive delay-sensitive model is that only a single MCU satisfies RUs' content requests. If multiple MCUs cooperatively transmit contents to RUs

simultaneously, the waiting time for each RU is reduced. Therefore, we extend to the case where multiple MCUs assist content transmission.

Assume there are M MCUs denoted by $\mathcal{M} = \{m_1, m_2, \ldots, m_M\}$. Each RU is assigned to the nearest MCU. Denote $I_{i,m} = 0, 1, i \in \mathcal{N}, m \in \mathcal{M}$ as the connection indicator between RU i and MCU m. In particular, $I_{i,m} = 1$ implicits MCU m transmits contents to RU i. Otherwise, there is no connection between them. Meanwhile, each RU is restricted to connect one MCU whereas each MCU serves multiple RUs, $\sum_{m \in \mathcal{M}} I_{i,m} = 1$. All the $I_{i,m}$ compose a indicator matrix $\mathbf{I}$. Given the locations of both RUs and MCUs, the indicator matrix is known. Denote the number of RUs served by the MCU $m_i, i = 1, 2, \ldots, M$ as n_{m_i}. To ease the description, we put the RUs served by the same MCU together and reorder the RU set as $\mathcal{N} = \{x_1, \ldots, x_{n_{m_1}}, x_{n_{m_1}+1}, \ldots, x_{n_{m_1}+n_{m_2}}, \ldots, x_N\}$ with $\sum_{m_i \in \mathcal{M}} n_{m_i} = N$.

Because the introduction of multiple MCUs divides RUs into smaller piles whereas the $M/G/1$ queue model adapts to the case with a large number of RUs better. Taking the indicator matrix $\mathbf{I}$ into consideration, we model the utilities based on the intuitive delay-sensitive model instead of the queue model. The utility of each RU is

$$\begin{aligned}\tilde{u}_i(\tilde{x}_i, \tilde{\mathbf{x}}_{-i}, \tilde{\mathbf{p}}) &= a_i\tilde{x}_i - \frac{1}{2}b_i{\tilde{x}_i}^2 + \sum_{j \in \mathcal{N}} g_{ij}\tilde{x}_i\tilde{x}_j \\ &- \frac{1}{2}\tilde{d}\sum_{m=1}^{M} I_{i,m}\left(\sum_{j \in \mathcal{N}} I_{j,m}\tilde{x}_j\right)^2 - \sum_{m=1}^{M}\tilde{p}_m\tilde{x}_i, \forall i,\end{aligned} \tag{9}$$

where $\tilde{\mathbf{p}} = \{p_1\mathbf{1}^T_{n_{m_1}}, p_2\mathbf{1}^T_{n_{m_2}}, \ldots, p_M\mathbf{1}^T_{n_{m_M}}\}^T$ is the unit payment vector corresponding to each RU. Specifically, $\mathbf{1}_{n_{m_i}}$ represents $n_{m_i} \times 1$ vector with 1s, and $\tilde{p}_m$ is the unit payment at the MCU m. Since MCUs serve different RU piles, their unit payments are different.

Accordingly, the revenue of each MCU is

$$\tilde{R}_m(\tilde{\mathbf{x}}, \tilde{p}_m) = \sum_{i \in \mathcal{N}} (\tilde{p}_m - c) I_{i,m}\tilde{x}_i, \forall m \in \mathcal{M}. \tag{10}$$

Because all MCUs cooperate to offload data, they aim to achieve the maximum total revenue

$$\tilde{R}(\tilde{\mathbf{x}}, \tilde{\mathbf{p}}) = \sum_{m \in \mathcal{M}}\sum_{i \in \mathcal{N}} (\tilde{p}_m - c) I_{i,m}\tilde{x}_i. \tag{11}$$

5 Utility Maximization in Delay-Tolerant Model

5.1 Overview

In game theory, Stackelberg game [44] is a tool to model the scenario where a hierarchy of actions exists between two types of players: one is the leader, and the other is the follower. The leader makes its move first. After the leader chooses a strategy, the follower always chooses the best response strategy that maximizes its utility. Knowing this reaction from the follower, the leader strategically chooses a strategy to maximize its utility. This optimal strategy of the leader, together with the corresponding best response strategy of the follower, constitutes a Stackelberg equilibrium. At a Stackelberg equilibrium, no follower has an incentive to adjust its strategy unilaterally.

The communication between the MCU and RUs in the delay-tolerant scenario can be formulated as such a two-stage Stackelberg game, named as Utility Maximization game in delay-tolerant (UMDT).

Stage I (Unit Payment). The MCU chooses a unit payment p^* to maximize the total revenue R

$$p^* = \arg\max_{p \in [0,\infty)} \sum_{i \in \mathcal{N}} x_i(p - c).$$

Stage II (Requested Content Level). Each RU $i \in \mathcal{N}$ chooses a requested content level x_i to maximize the utility $u_i(x_i, \mathbf{x}_{-i}, p)$ given the unit payment p and the requested content levels of others $\mathbf{x}_{-i}$

$$x_i^* = \arg\max_{x_i \in [0,\infty)} u_i(x_i, \mathbf{x}_{-i}, p), \forall i.$$

In the UMDT game, the MCU is the leader with the unit payment p^* as the strategy and RUs are the followers. The strategy of RU i is the requested content level $x_i^*, \forall i$. Due to each RU is selfish, the game in Stage II is considered as a non-cooperative game, which we call Request Level Determination (RLD) game. Given the UMDT formulation, we are interested in the following questions:

- Q1: For a given unit payment p, is there a profile of stable strategies in the RLD game such that no RU can increase the utility by unilaterally changing his current strategy?
- Q2: If the answer to Q1 is affirmative, is the stable strategy profile unique? When it is unique, RUs will be guaranteed to select the strategies in the same stable strategy profile.
- Q3: How can the MCU select the value of p to maximize the total revenue?

The stable strategy profile in Q1 corresponds to the concept of Nash equilibrium [44].

Definition 1. *Nash equilibrium: A profile of strategies $\mathbf{x}^*$ is a Nash equilibrium of the RLD game if for any mobile RU i*

$$u_i(x_i^*, \mathbf{x}_{-i}^*, p) \geq u_i(x_i, \mathbf{x}_{-i}^*, p), \tag{12}$$

for any $x_i \geq 0$, where u_i is defined in (1).

The existence (Q1) and uniqueness (Q2) of a stable Nash equilibrium strategy profile not only ensure that no RU has an incentive to make a change unilaterally but also allow the MCU to predict the behaviors of RUs and thus to select the optimal unit payment. The answer to Q3 depends heavily on those to Q1 and Q2. Stackelberg equilibrium, which is the final solution to the UMDT game, consists of the optimal solution computed in Q3 and the corresponding strategies at the Nash equilibrium in the RLD game.

5.2 RU Utility Maximization

Backward reduction methods [44] are deployed to maximize the utilities of both RUs and MCUs. We answer above Q1 and Q2 first, followed by an algorithm to find the RUs' best response strategies in the RLD game.

Definition 2. *Best Response Strategy: Given p and $\mathbf{x}_{-i}$, a strategy is RU i's best response strategy, denoted by $\beta_i(\mathbf{x}_{-i})$, if it maximizes the utility function $u_i(x_i, \mathbf{x}_{-i}, p)$ in (1), over all $x_i \geq 0$.*

Based on the definition of Nash equilibrium, every RU plays his best response strategy at a Nash equilibrium. By setting the derivative $\frac{\partial u_i(x_i, \mathbf{x}_{-i}, p_i)}{\partial x_i} = 0$ as the first order condition in (1), we obtain the RU i's best response strategy

$$\beta_i(\mathbf{x}_{-i}) = \max\left\{0, \frac{a_i - p}{b_i} + \sum_{j \neq i} \frac{g_{ij}}{b_i} x_j\right\}, \forall i, \tag{13}$$

in which the max operation is to ensure RU i's strategy non-negative. Each RU's best response strategy consists of two parts: internal demand $(a_i - p)/b_i$ which is independent of other RUs, and external demand $\sum_{j \neq i} \frac{g_{ij}}{b_i} x_j$ indicating the social effect other RUs bring to the RU i. The coefficient g_{ij}/b_i represents the marginal increase of RU i's requested content level when RU j's requested content level increases. It implies that the increase of other RUs' strategies has a positive impact on the RU i's strategy.

5.2.1 Existence and Uniqueness of RUs' Best Response Strategies—the Answers to Q1 and Q2

Since each RU has a great incentive to unboundedly increase the requested content levels provided other RUs' request levels are sufficiently large, the Nash equilibrium cannot be ensured to exist. To circumvent such situation, we give a general assumption under which a Nash equilibrium exists.

Assumption 1. $\sum_{j \neq i} \frac{g_{ij}}{b_i} < 1, \forall i$.

The *Assumption* 1 is a sufficient condition for the existence of RUs' best response strategies. Assume that the maximum requested content level among all the other RUs is x'_j. Under the *Assumption* 1, the external demand is $\sum_{j \neq i} \frac{g_{ij}}{b_i} x_j \leq \sum_{j \neq i} \frac{g_{ij}}{b_i} x'_j < x'_j$. It implies that the social effect experienced by an RU from others is limited to the largest effect this RU can experience from an individual of the other RUs.

Theorem 1. *Under* Assumption 1*, the RLD game in Stage II always admits a Nash equilibrium for RUs.*

We prove the Theorem 1 in Appendix. The main idea is to show our RLD game with unbounded content levels is equivalent to a game with bounded content levels that admits a Nash equilibrium.

Theorem 2. *Under* Assumption 1*, the RLD game in Stage II has a unique best response strategy.*

We prove the Theorem 2 in Appendix. According to [45], we try to demonstrate that the RLD game is a concave game.

5.2.2 Calculation of RUs' Best Response Strategies

We propose an algorithm to calculate RUs' best response strategies as shown in Algorithm 1.

Algorithm 1. Calculate the RUs' Best Response Strategies

Input: precision threshold ϵ
Output: $\mathbf{x}^*$

$x_i^{(0)} \leftarrow 0, \forall i \in \mathcal{N}$; $n \leftarrow 1$;
for $j = 1; j \leq N$ **do**
  $x_i^{(n)} = \max\left\{0, \frac{a_i - p}{b_i} + \sum_{j \neq i} \frac{g_{ij}}{b_i} x_j^{(n-1)}\right\}$;
end
if $||\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}|| < \epsilon$ **then**
  $\mathbf{x}^* = \mathbf{x}^{(n)}$;
  break;
else
  $n = n + 1$;
  go back to 2;
end
return $\mathbf{x}^*$;

Theorem 3. *Algorithm 1 calculates the Nash equilibrium in the RLD game.*

We prove the Theorem 3 in Appendix. The key is to prove that the best response strategy for each user is converged.

To ease the description, we express the best response strategies in a matrix format.

Lemma. *Denote $\mathcal{S}$ as the set of RUs with positive strategies and $\mathcal{N} - \mathcal{S}$ as the set of other RUs: $\mathcal{S} = \{i | x_i^* > 0\}$ and $\mathcal{N} - \mathcal{S} = \{i | x_i^* = 0\}$, the best response strategies are*

$$\mathbf{x}_{\mathcal{S}}^* = (\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}} - p\mathbf{1}_{\mathcal{S}}) \tag{14}$$

$$\mathbf{x}_{\mathcal{N}-\mathcal{S}}^* = \mathbf{0}_{\mathcal{N}-\mathcal{S}} \tag{15}$$

where $\mathbf{x}_{\mathcal{S}}^* = \{x_i^* | i \in \mathcal{S}\}$, $\mathbf{x}_{\mathcal{N}-\mathcal{S}}^* = \{x_i^* | i \in \mathcal{N} - \mathcal{S}\}$ and $\mathbf{a}_{\mathcal{S}} = \{a_i | i \in \mathcal{S}\}$. The matrices $\Lambda_{\mathcal{S}}, \mathbf{G}_{\mathcal{S}}$ are $|\mathcal{S}| \times |\mathcal{S}|$ matrices with elements in $\Lambda, \mathbf{G}$ with indices in $\mathcal{S} \times \mathcal{S}$, respectively. The vectors $\mathbf{1}_{\mathcal{S}}$ and $\mathbf{0}_{\mathcal{N}-\mathcal{S}}$ are $|\mathcal{S}| \times 1$ and $|\mathcal{N} - \mathcal{S}| \times 1$ vectors with 1s and 0s, respectively.

We prove the Lemma in Appendix. The important part is to show that $(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}$ is invertible.

5.2.3 Discussion on Social Effect

Proposition 1. *For the RLD game, when $a_i = a > p$ and the social effect is symmetric, $g_{ij} = g_{ji}, \forall i \neq j$, the social relationship between RUs brings a positive effect to Nash equilibrium.*

We prove the Proposition 1 in Appendix. The main idea is to show that the total requested content level at the Nash equilibrium increases when g_{ij} increases. In addition, the performance under asymmetric social effect is shown to be similar with that under symmetric social effect in Section 7.

5.3 The MCU Revenue Maximization

According to the above analysis, the MCU, as a leader, knows there exists the unique Nash equilibrium for the RUs given any unit payment. Hence, he can maximize the total revenue by choosing the optimal unit payment.

5.3.1 The Impact of Unit Payment

We first take the case with two RUs as an example. Without loss of generality, assume $a_1 > a_2$. Intuitively, in (13), both RU 1 and RU 2 have positive strategies when the unit payment p is in a low price regime. Their strategies are

$$\begin{cases} x_1 = \frac{a_1 - p}{b_1} + \frac{g_{12}}{b_1} x_2 & \text{(16a)} \\ x_2 = \frac{a_2 - p}{b_2} + \frac{g_{21}}{b_2} x_1, & \text{(16b)} \end{cases}$$

By solving above equations, we get the value of x_1 and x_2

$$x_1 = \frac{(a_1 - p)b_2 + (a_2 - p)g_{12}}{b_1 b_2 - g_{12} g_{21}} \tag{17}$$

$$x_2 = \frac{(a_2 - p)b_1 + (a_1 - p)g_{21}}{b_1 b_2 - g_{12} g_{21}}, \tag{18}$$

which show that the strategies of both RU 1 and RU 2 decrease as p increases. Based on the *Assumption* 1, $x_1 > x_2$. Thus, when increasing p, the strategy of RU 2, x_2, first decreases to 0. Denote the unit payment as pth at which RU 2's best response strategy is decreased to 0. According to (18), $p_{th} = \frac{a_2 b_1 + a_1 g_{21}}{b_1 + g_{21}}$. Continuing to increase p, the strategy of RU 1 then decreases to 0. Therefore, we have the Proposition 2.

Algorithm 2. Calculate the MCU's Optimal Revenue

Input: none
Output: $p^*, \mathbf{x}^*, r^*$
1 **calculate** the Nash equilibrium $\mathbf{x}^{*'}$ using Algorithm 1 when the unit payment is 0;
2 $\underline{p} \leftarrow 0; p^* \leftarrow 0; r^* \leftarrow 0; \mathcal{S} \leftarrow \emptyset$;
3 **for** $i = 1, i \leq N$ **do**
4 **if** $x_i^{*'} > 0$ **then**
5 $\mathcal{S} \leftarrow \mathcal{S} \bigcup \{i\}$;
6 **end**
7 **end**
8 **while** $\underline{p} \leq \max_{i \in \mathcal{N}} a_i$ and $\mathcal{S} \neq \emptyset$ **do**
9 $\mathcal{S}_1 \leftarrow \emptyset; \mathcal{S}_2 \leftarrow \emptyset$;
10 **foreach** $i \in \mathcal{S}$ **do**
11 **if** $\left[(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\right]_i \mathbf{1}_{\mathcal{S}} > 0$ **then**
12 $\mathcal{S}_1 \leftarrow \mathcal{S}_1 \bigcup \{i\}$;
13 $\hat{p}_i \leftarrow \frac{\left[(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\right]_i \mathbf{a}_{\mathcal{S}}}{\left[(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\right]_i \mathbf{1}_{\mathcal{S}}}$;
14 **end**
15 **end**
16 **foreach** $i \in \mathcal{N} - \mathcal{S}$ **do**
17 **if** $[\mathbf{G}]_{i,\mathcal{S}}(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}} < -1$ **then**
18 $\mathcal{S}_2 \leftarrow \mathcal{S}_2 \bigcup \{i\}$;
19 $\hat{p}_i \leftarrow \frac{[\mathbf{G}]_{i,\mathcal{S}}(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{a}_{\mathcal{S}} + a_i}{[\mathbf{G}]_{i,\mathcal{S}}(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}} + 1}$;
20 **end**
21 **end**
22 $\overline{p} = \min_{i \in \mathcal{S}_1 \cup \mathcal{S}_2} \hat{p}_i$;
23 $k = \arg_{i \in \mathcal{S}_1 \cup \mathcal{S}_2} \overline{p}$;
24 $p' = \frac{\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{a}_{\mathcal{S}} + c\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}}{2\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}}$;
25 **if** $p' \in \left[\underline{p}, \overline{p}\right]$ **then**
26 $\widetilde{p} = p'$;
27 **else if** $p' < \underline{p}$ **then**
28 $\widetilde{p} = \underline{p}$;
29 **else**
30 $\widetilde{p} = \overline{p}$;
31 **end**
32 $\widetilde{r} = (\widetilde{p} - c)\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}} - \widetilde{p}\mathbf{1}_{\mathcal{S}})$;
33 **if** $\widetilde{r} > r^*$ **then**
34 $p^* \leftarrow \widetilde{p}; r^* \leftarrow \widetilde{r}; \mathbf{x}_{\mathcal{S}}^* = (\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}} - p^*\mathbf{1}_{\mathcal{S}})$; $\mathbf{x}_{\mathcal{N}-\mathcal{S}}^* = \mathbf{0}_{\mathcal{N}-\mathcal{S}}, \mathbf{x}^* = \mathbf{x}_{\mathcal{S}}^* \bigcup \mathbf{x}_{\mathbf{N}-\mathbf{S}}^*$;
35 **end**
36 $\underline{p} \leftarrow \widetilde{p}$;
37 **if** $k \in \mathcal{S}$ **then**
38 $\mathcal{S} = \mathcal{S} \backslash \{k\}$;
39 **else**
40 $\mathcal{S} = \mathcal{S} \cup \{k\}$;
41 **end**
42 **end**
43 return $p^*, \mathbf{x}^*, r^*$

Proposition 2. *In RLD game, the impact that p brings to the two RUs' best response strategies $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ is as follows*

- *When we set p in a low regime: $0 \leq p < p_{th}$, the best response strategies of two RUs are listed in (17) and (18);*
- *When we set p in a medium regime: $p_{th} \leq p < a_1$, $x_1 = \frac{a_1 - p}{b_1}$ and $x_2 = 0$;*
- *When we set p in a high regime; $p \geq a_1$, RUs will not pick up their strategies: $x_1 = x_2 = 0$.*

Based on the *Assumption* 1, $p_{th} = \frac{a_2 b_1 + a_1 g_{21}}{b_1 + g_{21}} > a_2$. It implies that RU 2 would like to take part in the game ($x_2 \in 0$) although the unit payment he has to pay is larger than the internal effect. This gives the credits to the social effect that RU 1 brings to, which verifies that social effect brings benefits in our scheme.

Next, we extend our discussion on the impact of p to a general case where more RUs request contents.

Proposition 3. *In RLD game, the impact that p brings to the RUs' best response strategies $\mathbf{x}^*$ is as follows*

- *When we set p in a low regime $0 \leq p \leq \max_{i \in \mathcal{M}} a_i$: there is a set of prices $p_0 \triangleq 0 < p_1 < p_2 < \cdots < p_M < p_{M+1} \triangleq \max_{i \in \mathcal{N}} a_i$. For each $k \in \{0, 1, 2, \ldots, M\}$, there is a set $S_k \subseteq \mathcal{N}$ such that for any $p \in [p_k, p_{k+1}]$ such that $x_i^* = [(\Lambda_{S_k} - \mathbf{G}_{S_k})^{-1}(\mathbf{a}_{S_k} - p\mathbf{1}_{S_k})]_i, \forall i \in S_k$ and $x_i^* = 0, \forall i \notin S_k$*
- *When we set p in a high regime $p \geq \max_{i \in \mathcal{N}} a_i$, $x_i^* = 0, \forall i$*

We prove the Proposition 3 in Appendix. It shows that each RU' best response strategy is a piecewise linear function of the price, which motivates us to propose the Algorithm 2 to calculate the MCU's optimal revenue.

5.3.2 Calculation of the MCU's Optimal Revenue—The Answer to Q3

Based on the Lemma, the piecewise unit payment p is linear with the total best response strategies $\mathbf{1}^T\mathbf{x}^*$ at the Nash equilibrium. Hence, the total revenue of the MCU $(p - c)\mathbf{1}^T\mathbf{x}^*$ is a quadratic function with the unit payment p according to (2). Given above characteristics, we propose the *Algorithm* 2. Inspired by Proposition 3, we first determine the unit payment interval in which the set of RUs with positive strategies does not change when the unit payment increases or decreases. Within each determined unit payment interval, we calculate the optimal unit payment to maximize the total revenue of the MCU. Finally, by comparing total revenues in each interval, we obtain the final unit payment, which makes largest total revenue for the MCU. The final unit payment, together with the corresponding RUs' requested content levels, composes the Stackelberg equilibrium.

Specifically, the *Algorithm* 2 is initialized by calculating the RUs' best response strategies when the unit payment $p = 0$, as shown in Step 1. From Step 3 to Step 7, it finds the set $\mathcal{S}$ composed of RUs with positive strategies, which serves the initial conditions in the following steps. As the unit payment p increases from 0 to $\max_{i \in \mathcal{N}} a_i$, it iteratively finds the critical unit payment at which the set $\mathcal{S}$ changes as illustrated from Step 10 to Step 22. Because the change of the set means either adding or dropping an eligible RU, the process of finding the critical unit payment can be divided into the following three parts:

- Step 10 to Step 15 investigates the critical unit payment in the set $\mathcal{S}$, which makes at least one RU's positive strategy decreases to 0. Since RU i is in the set $\mathcal{S}$, according to (14), his positive strategy x_i is

$$x_i = \left[(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\right]_{i,\mathcal{S}}(\mathbf{a}_{\mathcal{S}} - p\mathbf{1}_{\mathcal{S}}) > 0, \tag{19}$$

where $[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}]_{i,\mathcal{S}}$ denotes a $1\times|\mathcal{S}|$ vector with elements in the ith row of the matrix $(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}$ and the columns with indices in $\mathcal{S}$. If $[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}]_{i,\mathcal{S}}\mathbf{1}_{\mathcal{S}}>0$, the RU i's positive strategy decreases as p increases. Assuming when the unit payment increases to $\hat{p}_i$, the RU i's positive strategy x_i decreases to 0. We have

$$\Big[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\Big]_{i,\mathcal{S}}\mathbf{a}_{\mathcal{S}}=\hat{p}_i\Big[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\Big]_{i,\mathcal{S}}\mathbf{1}_{\mathcal{S}}$$

$$\hat{p}_i=\frac{\Big[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\Big]_{i,\mathcal{S}}\mathbf{a}_{\mathcal{S}}}{\Big[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\Big]_{i,\mathcal{S}}\mathbf{1}_{\mathcal{S}}}. \tag{20}$$

- Step 16 to 21 investigates the critical unit payment in the set $\mathcal{N}-\mathcal{S}$, which makes at least one RU's strategy become positive When RU i is in the set $\mathcal{N}-\mathcal{S}$, $x_i=0>\frac{a_i-p}{b_i}+\sum_{j\neq i}\frac{g_{ij}}{b_i}x_j$. If $x_j>0$, $x_j=[(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}]_{j,\mathcal{S}}(\mathbf{a}_{\mathcal{S}}-p\mathbf{1}_{\mathcal{S}})$. Denote $\mathbf{G}_{i,\mathcal{S}}$ as a $1\times|\mathcal{S}|$ vector composed of the element of the ith row of the matrix $\mathbf{G}$ with column indices in $\mathcal{S}$

$$\begin{aligned}x_i=0&>\frac{a_i-p}{b_i}+\frac{1}{b_i}[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}}-p\mathbf{1}_{\mathcal{S}})\\&=\frac{1}{b_i}[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{a}_{\mathcal{S}}+\frac{a_i}{b_i}\\&\quad-\frac{p}{b_i}\Big([\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}+1\Big),\end{aligned}$$

If $[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}<-1$, $\frac{a_i-p}{b_i}+\frac{1}{b_i}[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}}-p\mathbf{1}_{\mathcal{S}})$ increases as p decreases. It becomes positive when the unit payment decreases to

$$\hat{p}_i=\frac{[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{a}_{\mathcal{S}}+a_i}{[\mathbf{G}]_{i,S}(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}+1}. \tag{21}$$

- By comparing both the critical unit payments in set $\mathcal{S}$ and $\mathcal{N}-\mathcal{S}$, we choose the minimized one as the final critical unit payment as illustrated in Step 22.

From Step 24 to Step 31, we calculate the unit payment $\widetilde{p}\in[\underline{p},\overline{p}]$ such that the MCU's revenue $R(\mathbf{x},p)$ is maximized, in which $R(\mathbf{x},p)=R(\mathbf{x}_{\mathcal{S}},p)=\sum_{i\in\mathcal{S}}x_i(p-c)=(p-c)\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}}-p\mathbf{1}_{\mathcal{S}}),p\in[\underline{p},\overline{p}]$. By setting the first order derivative of $R(\mathbf{x},p)$ to 0, we find the potential optimal unit payment p' in the interval $[\underline{p},\overline{p}]$

$$p'=\frac{\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{a}_{\mathcal{S}}+c\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}}{2\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}}, \tag{22}$$

if $p'\in[\underline{p},\overline{p}]$, the optimal unit revenue $\widetilde{p}=p'$. Otherwise, the optimal unit payment is $\widetilde{p}=\underline{p}$ if $p'<=\underline{p}$, or $\widetilde{p}=\overline{p}$ if $p'<=\overline{p}$. The local optimal revenue r' is

$$r'=(\widetilde{p}-c)\mathbf{1}_{\mathcal{S}}^T(\Lambda_{\mathcal{S}}-\mathbf{G}_{\mathcal{S}})^{-1}(\mathbf{a}_{\mathcal{S}}-\widetilde{p}\mathbf{1}_{\mathcal{S}}),\widetilde{p}\in\Big[\underline{p},\overline{p}\Big]. \tag{23}$$

Meanwhile, the set $\mathcal{S}$ is updated as shown from Step 37 to Step 41 by adding or deleting the RU k found in Step 23. The renewed set $\mathcal{S}$ is deployed to continue finding another local optimal unit payment.

Finally, by comparing the local optimal revenues in each unit payment interval, we find the global optimal revenue r^* and its corresponding unit payment p^* as illustrated in Step 32 to Step 35. The related RUs' best response strategies $\mathbf{x}^*$ are calculated.

6 Utility Maximization in Delay-Sensitive Model

In this section, we model the delay-sensitive cases as three two-stage Stackelberg games to maximize the utilities of RUs and MCUs, respectively. Specifically, the delay effect considered in the intuitive delay-sensitive model is essentially a specific form of the congestion effect studied in [28]. Therefore, we mainly discuss the other two delay-sensitive models.

6.1 Intuitive Delay-Sensitive Model

Refering to [29], the RU i's best response strategy is

$$\beta_i(\overline{\mathbf{x}}_{-i})=\max\left\{0,\frac{a_i-\overline{p}}{b_i+d}+\sum_{j\neq i}\frac{g_{ij}-d}{b_i+d}x_j\right\},\forall i. \tag{24}$$

By comparing (13) and (24), each RU suffers both positive social effect and negative delay effect brought by other RUs. When $g_{ij}<d$, the RU j even brings negative external effect to the RU i. Otherwise, the RU j puts positive external effect. Under the assumption $\sum_{j\neq i}\frac{|g_{ij}-d|}{(b_i+d)}<1,\forall i$, the utility maximization is obtained according to Algorithm 3 in [29].

6.2 Queueing Delay-Sensitive Model

By setting the derivative $\frac{\partial\hat{u}_i(\hat{x}_i,\hat{\mathbf{x}}_{-i},\hat{p})}{\partial\hat{x}_i}=0$ in (8), the RU i's best response strategy is obtained as

$$\beta_i(\hat{\mathbf{x}}_{-i})=\max\left\{0,\frac{a_i-\hat{p}}{b_i+\hat{d}}+\sum_{j\neq i,j\in\mathcal{N}}\frac{g_{ij}-\frac{\hat{d}}{N-1}}{b_i+\hat{d}}x_j\right\}, \tag{25}$$

where $\hat{d}=\frac{k\lambda}{N(1-\rho)}$ is assumed as a system parameter estimated by the SP. Comparing (24) and (25), given $\hat{d}=d$, the delay effect in the queueing delay-sensitive model is relieved from d to $\frac{d}{N-1}$, which theoretically proves that our queue model lowers the delay effect. Meanwhile, the content mean arrival rate λ brings a negative effect to RUs' utilities. It is because larger λ increases the queue length given the fixed average content transmission time and thus puts RUs to the longer waiting time. Similarly, the traffic intensity ρ puts a negative delay effect to RUs' utilities.

Since each RU's utility in (25) is similar to that in (13) and the MCU's utility keeps unchanged, we could simply apply the Algorithm 2 to obtaining the best strategies for both RUs and MCU under the following assumption:

Assumption 2. $\sum_{j\neq i}\frac{|g_{ij}-\frac{\hat{d}}{N-1}|}{(b_i+\hat{d})}<1,\forall i.$

6.3 Multi-Leader Delay-Sensitive Model

Due to the participation of multiple MCUs, the previous single-leader Stackelberg game is extended to a multi-leader two-stage Stackelberg game as follows:

Stage I (Unit Payment). Each MCU announces its unit payment $\tilde{p}_m$ to maximize their total revenues

$$\tilde{\mathbf{p}}*=\arg\max_{\tilde{p}\in[0,\infty)^M}\tilde{R}(\tilde{\mathbf{x}},\tilde{\mathbf{p}}).$$

Stage II (Requested Content Level). Each RU $i\in\mathcal{N}$ strategies the required content level $\tilde{x}_i$ to maximize his own utility given the price $\tilde{\mathbf{p}}$ and the requested content levels of others $\tilde{\mathbf{x}}_{-i}$

$$\tilde{x}_i^*=\arg\max_{\tilde{x}_i\in[0,\infty)}\tilde{u}_i(\tilde{x}_i,\tilde{\mathbf{x}}_{-i},\tilde{\mathbf{p}}),\forall i.$$

6.3.1 Utility Maximization for RUs

Similar with (13), the best response strategy for RU i is

$$\beta_i(\tilde{\mathbf{x}}_{-i},\tilde{\mathbf{p}}) = \max\left\{0, \frac{a_i-\sum_{m=1}^{M} I_{i,m}\tilde{p}_m}{b_i+\tilde{d}} + \sum_{j\neq i, j\in\mathcal{N}} \frac{g_{ij}-\tilde{d}\sum_{m=1}^{M} I_{i,m}I_{j,m}}{b_i+\tilde{d}} x_j\right\}. \quad (26)$$

Formula (26) shows that the introduction of multiple MCUs reduces each RU's delay effect by serving them locally whereas does not affect their global positive social effect. With known indicator matrix, (26) is similar with (3). Therefore, if we have the following assumption, the existence and uniqueness can be proved referring to the previous proof.

Assumption 3. $\sum_{j\neq i} \frac{|g_{ij}-\tilde{d}\sum_{m=1}^{M} I_{i,m}I_{j,m}|}{(b_i+\tilde{d})} < 1, \forall i$

Meanwhile, under the Assumption 3, the best response strategies for all RUs given the unit payment vector are

$$\tilde{\mathbf{x}}^*_{\mathcal{S}} = \left(\tilde{\Lambda}_{\mathcal{S}} - \tilde{\mathbf{G}}_{\mathcal{S}}\right)^{-1}\left(\mathbf{a}_{\mathcal{S}} - \tilde{\mathbf{p}}_{\mathcal{S}}\right) \quad (27)$$

$$\tilde{\mathbf{x}}^*_{\mathcal{N}-\mathcal{S}} = \mathbf{0}_{\mathcal{N}-\mathcal{S}}. \quad (28)$$

The corresponding matrices $\tilde{\Lambda} = \mathrm{diag}(b_1+\tilde{d}, b_2+\tilde{d}, \ldots, b_N+\tilde{d})$ and $\tilde{\mathbf{G}} = \mathbf{G} - \mathbf{D}$, where

$$D = \tilde{d}\begin{bmatrix} 0 & \sum_{m\in\mathcal{M}} I_{1,m}I_{2,m} & \cdots & \sum_{m\in\mathcal{M}} I_{1,m}I_{N,m} \\ \sum_{m\in\mathcal{M}} I_{2,m}I_{1,m} & 0 & \cdots & \sum_{m\in\mathcal{M}} I_{2,m}I_{N,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{m\in\mathcal{M}} I_{N,m}I_{1,m} & \sum_{m\in\mathcal{M}} I_{N,m}I_{2,m} & \cdots & 0 \end{bmatrix}$$

The implication for $\mathcal{S}$ has been explained previously.

6.3.2 Utility Maximization for MCUs

Due to the globally positive social effect and locally negative delay effect, we cannot simply deploy the Algorithm 2 to solve the Stackelberg game for each pile of RUs. However, owing to the existence and uniqueness of all RUs' best response strategies $\tilde{\mathbf{x}^*}$, MCUs can correctly predict the behaviors of all RUs given the unit price $\tilde{\mathbf{p}}$, which gives them opportunities to maximize their total revenues.

To ease the description, we consider the case where all RUs receive their requested data $\tilde{\mathbf{x}}^*_{\mathcal{S}} = \tilde{\mathbf{x}}^*$. The case in which some RUs receive no contents can be easily extended. With the known indicator matrix, (11) is rewritten as

$$\tilde{R}(\tilde{\mathbf{x}},\tilde{\mathbf{p}}) = (\tilde{\mathbf{p}} - c\mathbf{1}_N)^T\tilde{\mathbf{x}}^*. \quad (29)$$

Substitute (27) into (29), we have

$$\begin{aligned}\tilde{R}(\tilde{\mathbf{x}},\tilde{\mathbf{p}}) &= (\tilde{\mathbf{p}} - c\mathbf{1}_N)^T\left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}(\mathbf{a} - \tilde{\mathbf{p}}) \\ &= -\tilde{\mathbf{p}}^T\left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}\tilde{\mathbf{p}} + \tilde{\mathbf{p}}^T\left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}\mathbf{a} \\ &\quad + c\mathbf{1}_N^T\left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}\tilde{\mathbf{p}} - c\mathbf{1}_N^T\left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}\mathbf{a}.\end{aligned} \quad (30)$$

We ignore the last term in (30) since it has nothing to do with $\tilde{\mathbf{p}}$ in the following. To obtain the strategies for each MCU, we have the total utilities maximization problem as

$$\begin{aligned}\max_{\tilde{p}_1,\ldots,\tilde{p}_M} \quad & \tilde{R}(\tilde{\mathbf{x}},\tilde{\mathbf{p}})' = -\tilde{\mathbf{p}}^T\mathbf{A}\tilde{\mathbf{p}} + \tilde{\mathbf{p}}^T\mathbf{A}\mathbf{a} + c\mathbf{1}_N^T\mathbf{A}\tilde{\mathbf{p}} \\ \text{s.t.} \quad & 0 \le \tilde{p}_m \le \max_{i\in\mathcal{N}} a_i, \forall m,\end{aligned} \quad (31)$$

where $\mathbf{A} = \left(\tilde{\Lambda} - \tilde{\mathbf{G}}\right)^{-1}$. The constraints in (31) is to restrict each MCU's unit payment. Otherwise, RUs would not receive any contents from MCUs as shown in (27) and (28). Since $\tilde{\mathbf{p}}$ is piecewise, we divide the matrix $\mathbf{A}$ into blocks

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1M} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{M1} & \mathbf{A}_{M2} & \cdots & \mathbf{A}_{MM} \end{bmatrix}$$

where

$$\mathbf{A}_{uv} = \begin{bmatrix} a_{\sum_{u=1}^{i-1} n_{m_u}+1, \sum_{v=1}^{j-1} n_{m_v}+1} & \cdots & a_{\sum_{u=1}^{i-1} n_{m_u}+1, \sum_{v=1}^{j} n_{m_v}} \\ \vdots & \ddots & \vdots \\ a_{\sum_{u=1}^{i} n_{m_u}, \sum_{v=1}^{j-1} n_{m_v}+1} & \cdots & a_{\sum_{u=1}^{i} n_{m_u}, \sum_{v=1}^{j} n_{m_v}} \end{bmatrix}$$

$\mathbf{a} = \{a_1, \ldots, a_{n_{m_1}}, a_{n_{m_1}+1}, \ldots, a_{n_{m_1}+n_{m_2}}, \ldots, a_N\}^T = \{\mathbf{a}_1'^T, \mathbf{a}_2'^T, \ldots, \mathbf{a}_M'^T\}^T$ is rewritten, where $\mathbf{a}_i' = \{a_{\sum n_{m_{i-1}}+1}, \ldots, a_{\sum n_{m_i}}\}^T$. Substituting (32) into (31),

$$\begin{aligned}\tilde{R}(\tilde{\mathbf{x}},\tilde{\mathbf{p}})' &= \sum_{i=1}^{M}\sum_{j=1}^{M} \tilde{p}_i\tilde{p}_j\mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{1}_{n_{m_j}} \\ &\quad + \sum_{i=1}^{M}\tilde{p}_i\sum_{j=1}^{M}\left((\mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{1}_{n_{m_j}})^T + \mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{a}_j'\right) \\ &= \tilde{\mathbf{p}}'^T\mathbf{A}'\tilde{\mathbf{p}}' + \sum_{i=1}^{M}\tilde{p}_i\sum_{j=1}^{M}\left((\mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{1}_{n_{m_j}})^T + \mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{a}_j'\right)\end{aligned}$$

where $\tilde{\mathbf{p}}' = [\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_M]$ and $\mathbf{A}'$ is a new matrix with the ijth element $\mathbf{1}_{n_{m_i}}^T\mathbf{A}_{ij}\mathbf{1}_{n_{m_j}}$. According to [46] and [47], the total utilities maximization is a convex optimization problem as long as $\mathbf{A}' + \mathbf{A}'^T$ is positive semidefinite. Therefore, we can use convex toolbox cvx [48] to obtain the strategies of MCUs under the positive semidefinite assumption.

7 Performance Evaluation

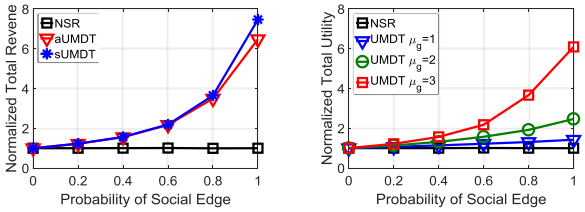
In this section, we evaluate the performance of the data offloading approaches in both the delay-tolerant scenario and the delay-sensitive scenario.

7.1 Simulation Settings

We consider a scenario with $N = 10$ RUs served by MCUs. Their internal characteristics follow a Gaussian distribution, where $a_i \sim \mathcal{N}(\mu_a, 2)$ and $b_i \sim \mathcal{N}(\mu_b, 2), \forall i$. To show the social effect brought by RUs' social relationship, we deploy the Erdős-Reńyi (ER) graph [49] model, in which a social edge between RUs exists with probability P_S in a group. If a social edge indeed exists, it is assumed to follow a normal distribution $\mathcal{N}(\mu_g, 2)$. To ensure the assumptions proposed in the paper, we set $\mu_a = \mu_b = 30$. In addition, the MCU's unit cost when delivering contents to RUs is constant, $c = 5$.

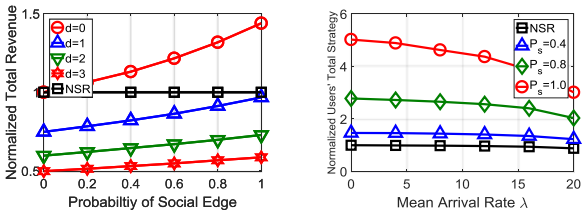
7.2 Simulation Results

In our simulations, we mainly compare the performance of the following cases: (1) No relationship case (NSR), in which



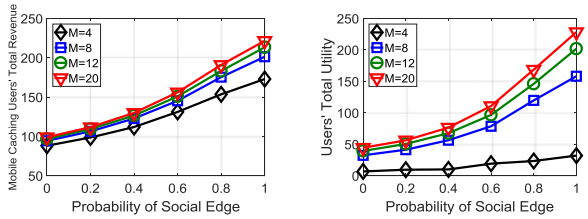
(a) Asymmetry Effect (b) Social Effect

Fig. 5. UMDT case.



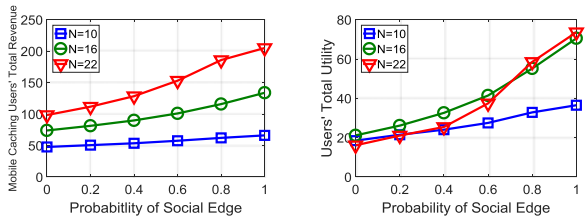
(a) Primitive Delay Effect (b) Queue Effect

Fig. 6. Delay effect.



(a) Total Revenue (b) Total Utility

Fig. 7. Effect from MCUs.



(a) Total Revenue (b) Total Utility

Fig. 8. Effect from RUs.

there are no interactions between RUs, $g_{ij} = 0, i, j \in \mathcal{N}$, $d = \hat{d} = \tilde{d} = 0$. (2) Delay-tolerant case (UMDT), in which the social effect exists among RUs due to their similar social attributes $g_{ij} \neq 0, \exists i, j \in \mathcal{N}$, $d = \hat{d} = \tilde{d} = 0$. (3) Intuitive Delay-sensitive case (iUMDS). (4) Queue Delay-sensitive case (qUMDS), and (5) Multi-leader Delay-sensitive case (mUMDS). Note that we normalize most simulation performance based on the NSR case, which means the performance value is divided by the corresponding value in the NSR case. In what follows, we show the impacts to which the social effect and delay effect will bring respectively.

7.2.1 The Impact of the Probability of Social Edge

To investigate the impact of social effect, we first consider the UMDT case in Fig. 5. Since two RUs in a social relationship could have different interests, we want to find whether such an asymmetry impacts RUs' utilities. Fig. 5a shows that it does not play an important role on RUs. Therefore, we choose the asymmetric social relationship in the followings as $g_{ij} \neq g_{ji}$ to be close to reality. Fig. 5a also tells us that the probability of the social relationship between RUs has a large impact. This is because the probability implies the contact opportunities between RUs, which would bring more social effects. Fig. 5b further demonstrates the above observation, which shows that the total utility of RUs increases as the increasing of the probability of social relationship. Hence, our motivation is verified that the homophily phenomenon truly brings positive social effects to data offloading scheme.

7.2.2 The Impact of Delay Effect

In iUMDS case, we consider the intuitive delay effect. From Fig. 6a, we find that such delay effect puts a serious negative impact on the MCU's total revenue. Specifically, when the delay effect is large, it could even cancel out the benefits brought by the social effect. When RUs are eager to obtain their requested contents, they have to wait for a long time. Thus, they would not request more contents even if the unit payment is low. The low unit payment and few contents decrease the total revenue of the MCU.

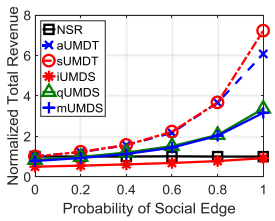


Fig. 9. Total levels versus number of RUs.

7.2.3 The Benefits Brought by Improved Models

In order to show the benefits in the qUMDS and mUMDS cases, we compare the MCU's total revenue as shown in Fig. 9. The worst situation is considered that the intuitive delay effect cancels the benefits brought by social effect completely, where $\mu_g = d = 3$. Fig. 9 demonstrates that the introduction of the queue and multiple MCUs indeed helps increase the total revenue.

qUMDS Case. We discuss the impact of the mean arrival rate shown in Fig. 6b. It impacts RUs' content levels negatively. Higher mean arrival rate indicates that more content requests come to the MCU while it is delivering contents, which would increase the content queue length. RUs have to wait for a longer time to obtain their contents and thus dissatisfy with the content transmission. Therefore, their requested content levels would decrease.

mUMDS Case. In Fig. 7, we draw the impacts to both MCUs and RUs' utilities brought by the number of MCUs. Assume there are $N = 25$ RUs requesting contents. As can be seen from Figs. 7a and 7b, more MCUs not only increase the utilities of RUs but also improve the total revenue of themselves. Fig. 8 shows an interesting phenomenon. Given the number of MCUs, each RU's waiting time will increase as the number of RUs becomes large, and thus their own utilities reduce. In the worst case, the total utilities of a larger number of RUs are lower than those of a smaller number of RUs as shown in Fig. 8b. However, since the number of RUs is large, the total avenue obtained from them can still be as high as shown in Fig. 8a. Both Figs. 7 and 8 demonstrate the effectiveness of our proposed multiple MCU delay sensitive model.

8 Conclusion

In this paper, we propose a data offloading approach by leveraging human's social behavior and human activities. To motivate the participation of MCUs, a two-stage Stackelberg game is deployed considering the interactions between RUs. In the delay-tolerant scenario, the interactions bring social effect owing to RUs' similar social attributes. We prove that the Stackelberg game has a unique Nash equilibrium and design an effective algorithm to compute the RUs' best response strategies. This enables the MCU to maximize the revenue. In the delay-sensitive scenario, by further taking advantages of RUs' mobility, we propose two improved approaches to lower RUs' delay effect due to their long waiting time, which introduces queue and extends the single-leader Stackelberg game to the multi-leader scheme, respectively. Based on the simulation results, we have shown the feasibility and effectiveness of our proposed approaches.

Appendix

Proof of Theorem 1 In the RLD game $\mathcal{G} = \{\mathcal{N}, \{u_i\}_{i\in\mathcal{N}}, [0,\infty]^N\}$, we denote $\mathbf{x}^*$ as a strategy profile and x_i^* as the largest requested content level in it, i.e., $x_i^* > x_j^*, \forall j \neq i$. Based on (13), when $x_i^* > 0$

$$x_i^* = \frac{a_i - p}{b_i} + \sum_{j\neq i}\frac{g_{ij}}{b_i}x_j^* \leq \frac{|a_i - p|}{b_i} + \sum_{j\neq i}\frac{g_{ij}}{b_i}x_i^*,$$

from which we get $x_i^* \leq |a_i - p|/(b_i - \sum_{j\neq i} g_{ij}) \leq \tilde{x}$. $\tilde{x}$ is any number that satisfies $\tilde{x} \geq \max_{i\in\mathcal{N}}|a_i - p|/(b_i - \sum_{j\neq i} g_{ij})$. Since x_i^* is the largest content level, all the content levels in game $\mathcal{G}$ are bounded, i.e., $x_j^* \in [0, \tilde{x}]$, $j \in \mathcal{N}$. Therefore, our game $\mathcal{G}$ is equivalent to a new game $\tilde{\mathcal{G}} = \{\mathcal{N}, \{u_i\}_{i\in\mathcal{N}}, [0,\tilde{x}]^N\}$ that has the same Nash equilbium stategy profile.

Taking the game $\tilde{\mathcal{G}}$ into consideration, the strategy space $[0,\tilde{x}]^N$ is compact and convex. The utility function $u_i(x_i, \mathbf{x}_{-\mathbf{i}}, p)$ is continuous in x_i and $\mathbf{x}_{-i}$. The second-order derivative of RU i's utility function $\frac{\partial^2 u_i(x_i,\mathbf{x}_{-\mathbf{i}},p)}{\partial^2 x_i} = -b_i$ is negative. Therefore, it is a concave game and admits a Nash equilibrium [45], [50]. Hence, the Nash equilibrium for our RLD game $\mathcal{G}$ exists. □

Proof of Theorem 2. The Jacobian matrix $\nabla\mathbf{u}(\mathbf{x})$ of RUs' utility profile $\mathbf{u}(\mathbf{x}) \triangleq \{u_1(\mathbf{x}), u_2(\mathbf{x}), \ldots, u_N(\mathbf{x})\}$ is given by $\nabla\mathbf{u}(\mathbf{x}) = -(\Lambda - \mathbf{G})$, where $\Lambda = \mathrm{diag}(b_1, b_2, \ldots, b_N)$ and

$$\mathbf{G} = \begin{bmatrix} 0 & g_{12} & \cdots & g_{1N} \\ g_{21} & 0 & \cdots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N1} & g_{N2} & \cdots & 0 \end{bmatrix}. \tag{32}$$

Based on *Assumption* 1, we have

$$[\Lambda - \mathbf{G}]_{ii} > \sum_{j\neq i}\Big|[\Lambda - \mathbf{G}]_{ij}\Big|, \forall i,$$

where $[\Lambda - \mathbf{G}]_{ij}$ denotes the element in the ith row and jth column in the matrix $[\Lambda - \mathbf{G}]$. Hence, $[\Lambda - \mathbf{G}]$ is strictly diagonal dominant. Assume social effect between RUs is symmetric, $g_{ij} = g_{ji}, \forall i, j \in \mathcal{N}$, $[\Lambda - \mathbf{G}]^T$ is also strictly diagonal dominant. Therefore, $\nabla\mathbf{u}(\mathbf{x}) + \nabla\mathbf{u}^T(\mathbf{x}) = -[\Lambda - \mathbf{G}] - [\Lambda - \mathbf{G}]^T$ is strictly diagonal dominant and symmetric. According to [46], a symmetric matrix that is strictly diagonally dominant with real nonnegative diagonal elements is positive definite. Thus, $-[\Lambda - \mathbf{G}] - [\Lambda - \mathbf{G}]^T$ is negative definite since the elements in it are negative. $\nabla\mathbf{u}(\mathbf{x})$ is diagonally strictly concave [45]. The RLD game $\mathcal{G}$ has a unique Nash equilibrium. □

Proof of Theorem 3. Let $\Delta x_i^{(n)} \triangleq x_i^{(n)} - x_i^*, \forall i$. According to step 3 in *Algorithm 1*

$$|\Delta x_i^{(n)}| \leq \left|\sum_{j\neq i}\frac{g_{ij}}{b_i}\Delta x_j^{(n-1)}\right| \leq \sum_{j\neq i}\frac{g_{ij}}{b_i}\Big|\Delta x_j^{(n-1)}\Big|, \forall i, \tag{33}$$

Denote $||\Delta x_i^{(n)}||_\infty$ as the l_∞-norm of vector $(\Delta x_1^{(n)}, \Delta x_2^{(n)}, \ldots, \Delta x_N^{(n)})$, $||\Delta x_i^{(n)}||_\infty = \max_{i\in\mathcal{N}}(\Delta x_1^{(n)}, \Delta x_2^{(n)}, \ldots, \Delta x_N^{(n)})$. According to (33), $||\Delta x_i^{(n)}||_\infty \leq \max_{i\in\mathcal{N}}\sum_{j\neq i}\frac{g_{ij}}{b_i}|\Delta x_j^{(n-1)}| \leq (\max_{i\in\mathcal{N}}\sum_{j\neq i}\frac{g_{ij}}{b_i})||\Delta x_i^{(n-1)}||_\infty$. Since $\max_{i\in\mathcal{N}}\sum_{j\neq i}\frac{g_{ij}}{b_i} < 1$, $||\Delta x_i^{(n)}||_\infty \leq ||\Delta x_i^{(n-1)}||_\infty$. It implies that *Algorithm 1* results in a contraction mapping of $||\Delta x_i^{(n-1)}||_\infty$ and thus converges to the Nash equilibrium. □

Proof of Lemma. According to (13) and *Algorithm 1*

$$x_i^* = \frac{a_i - p}{b_i} + \sum_{j\neq i}\frac{g_{ij}}{b_i}x_j^*, i, j \in \mathcal{S}. \tag{34}$$

The matrix format of (34) is

$$(\Lambda_\mathcal{S} - \mathbf{G}_\mathcal{S})\mathbf{x}_\mathcal{S}^* = (\mathbf{a}_\mathcal{S} - p\mathbf{1}_\mathcal{S}). \tag{35}$$

Because $\Lambda_\mathcal{S}$ is a positive diagonal matrix, it is invertible. Denote any eigenvalue and the corresponding eigenvector of $\Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}$ as λ and μ, respectively. Mathematically, $(\Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S})\mu = \lambda\mu$. Assume μ_i is the largest element in absolute value, $|\mu_i| \geq |\mu_j|, \forall j \neq i$

$$\begin{aligned} |\lambda\mu_i| &= \left|\sum_{j\in\mathcal{N}}\left[\Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}\right]_{ij}\mu_j\right| \\ &\leq \sum_{j\in\mathcal{N}}\left|\left[\Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}\right]_{ij}\right||\mu_j| \leq |\mu_i|\sum_{j\in\mathcal{N}}\frac{|g_{ij}|}{b_i} < |\mu_i|. \end{aligned} \tag{36}$$

From (36), the absolute values of all eigenvalues of $\Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}$ are less than 1. Since the eigenvalue values of the matrix $\mathbf{I} - \Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}$ are equaled to $1 - \lambda$, the matrix $\mathbf{I} - \Lambda_\mathcal{S}^{-1}\mathbf{G}_\mathcal{S}$ does not have 0 eigenvalues. Thus, $\Lambda_\mathcal{S} - \mathbf{G}_\mathcal{S} = \Lambda_\mathcal{S}(\mathbf{I} - \Lambda_\mathcal{S}^{-1}\mathbf{G})$ is invertible and $\mathbf{x}_\mathcal{S}{}^* = (\Lambda_\mathcal{S} - \mathbf{G}_\mathcal{S})^{-1}(\mathbf{a}_\mathcal{S} - p\mathbf{1}_\mathcal{S})$. □

Proof of Proposition 1. Based on *Lemma*, RUs' strategies at the Nash equilibrium is a continuous function of the matrix G_S. Thus, we can find a matrix G'_S, in which $g'_{ij} \geq g_{ij}, g'_{ij} \in G'_S, g_{ij} \in G_S$ and at least one strictly inequality exists, such that RUs with positive strategies $\mathbf{x}_\mathcal{S}^{*'}$ at the Nash equilibrium under G'_S are also in the set $\mathcal{S}$. According to (35)

$$(\Lambda_\mathcal{S} - \mathbf{G}_\mathcal{S})\mathbf{x}_\mathcal{S}^* = (\mathbf{a}_\mathcal{S} - p\mathbf{1}_\mathcal{S}) \tag{37}$$

$$\left(\Lambda_{\mathcal{S}} - \mathbf{G}'_{\mathcal{S}}\right)\mathbf{x}^{*'}_{\mathcal{S}} = (\mathbf{a}_{\mathcal{S}} - p\mathbf{1}_{\mathcal{S}}). \tag{38}$$

Subtract (37) from (38),

$$\mathbf{x}^{*'}_{\mathcal{S}} - \mathbf{x}^{*}_{\mathcal{S}} = (\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\Delta\mathbf{G}_{\mathcal{S}}\mathbf{x}^{*'}_{\mathcal{S}}, \tag{39}$$

where $\Delta\mathbf{G}_{\mathcal{S}} = \mathbf{G}'_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}}$. Thus, the total difference between $\mathbf{x}^{*'}_{\mathcal{S}}$ and $\mathbf{x}^{*}_{\mathcal{S}}$ is

$$\mathbf{1}^T_{\mathcal{S}}\left(\mathbf{x}^{*'}_{\mathcal{S}} - \mathbf{x}^{*}_{\mathcal{S}}\right) = \mathbf{1}^T_{\mathcal{S}}(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\Delta\mathbf{G}_{\mathcal{S}}\mathbf{x}^{*'}_{\mathcal{S}}. \tag{40}$$

Since $\mathbf{x}^{*}_{\mathcal{S}} = (\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}(a-p)\mathbf{1}_{\mathcal{S}}$ in (14), it follows that

$$\mathbf{1}^T_{\mathcal{S}}(\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1} = \left((\Lambda_{\mathcal{S}} - \mathbf{G}_{\mathcal{S}})^{-1}\mathbf{1}_{\mathcal{S}}\right)^T = \frac{1}{a-p}\mathbf{x}^{*T}_{\mathcal{S}}. \tag{41}$$

Substitute (41) into (40), we get the total difference as

$$\mathbf{1}^T_{\mathcal{S}}\left(\mathbf{x}^{*'}_{\mathcal{S}} - \mathbf{x}^{*}_{\mathcal{S}}\right) = \frac{1}{a-p}\mathbf{x}^{*T}_{\mathcal{S}}\Delta\mathbf{G}_{\mathcal{S}}\mathbf{x}^{*'}_{\mathcal{S}}. \tag{42}$$

Because $a > p$, $\mathbf{x}^{*'}_{\mathcal{S}}, \mathbf{x}^{*}_{\mathcal{S}} \succ 0$ and $\Delta\mathbf{G}_{\mathcal{S}} \succeq 0$, the total difference between $\mathbf{x}^{*'}_{\mathcal{S}}$ and $\mathbf{x}^{*}_{\mathcal{S}}$, $\mathbf{1}^T_{\mathcal{S}}(\mathbf{x}^{*'}_{\mathcal{S}} - \mathbf{x}^{*}_{\mathcal{S}}) > 0$, which implies that the total requested content levels at the Nash equilibrium increase when g_{ij} increases. The Proposition 1 verifies that the social effect between RUs with similar social attributes makes RUs get more interested contents. □

Proof of Proposition 3. For any unit payment $p \in [0, \max_{i\in\mathcal{N}} a_i]$, the requested content levels of the set of RUs $\mathcal{S}$ with positive strategies are given in (14). Meanwhile, according to (13), RU i's the requested content level $x^*_i = \frac{a_i - p}{b_i} + \sum_{j\neq i}\frac{g_{ij}}{b_i}x^*_j$ is continuous in p and RU j's requested content level x^*_j, $j \neq i$. When the unit payment p increases a small amount to p', the set of RUs with positive strategies at the Nash equilibrium does not change and their strategies are still given by (14) except that p is replaced by p'. Hence, the set of RUs with positive strategies is the same at any unit payment in a continuous unit payment interval. However, when the unit payment p increases a large amount to p'', some RUs' strategies decrease to 0 and thus they would not request any contents as shown in the two-RU example. Therefore, the interval of the unit payment is piecewise.

Assuming RU i has a maximized strategy $x^*_i > 0$ when $p \geq \max_{i\in\mathcal{N}}$. According to (13), $x^*_i = \frac{a_i - p}{b_i} + \sum_{j\neq i}\frac{g_{ij}}{b_i}x^*_j \leq \sum_{j\neq i}\frac{g_{ij}}{b_i}x^*_j \leq \sum_{j\neq i}\frac{g_{ij}}{b_i}x^*_i < x^*_i$, which is a contradiction. Therefore, $x^*_i = 0, \forall i$ when $p \geq \max_{i\in\mathcal{N}} a_i$. □

ACKNOWLEDGMENTS

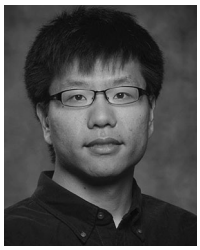
The work of L. Guo was supported by the U.S. National Science Foundation under grants ECCS-1710996, CNS-1744261, and IIS-1722731. The work of M. Li was supported by the U.S. National Science Foundation under grants CNS-1566634 and ECCS-1711991. The work of Y. Fang was supported by the U.S. National Science Foundation under grants IIS-1722791 and CNS-1423165. The preliminary version of this paper was published in IEEE Globecom 2016, Washington, DC, USA.

REFERENCES

[1] "Cisco visual networking index: Forecast and methodology, 2014–2019 white paper," Cisco, San Jose, CA, USA, 2015.
[2] iData Research, "Small cells market and wifi offloading opportunities for MNOs discussed in new 2015 research report." (2015). [Online]. Available: http://www.rnrmarketresearch.com/small-cells-and-wifi-offloading-now-mainstream-for-mnos-market-report.html
[3] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: Technical and business perspectives," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 104–112, Apr. 2013.
[4] W. Gao, Q. Li, and G. Cao, "Forwarding redundancy in opportunistic mobile networks: Investigation and elimination," in *Proc. IEEE INFOCOM*, 2014, pp. 2301–2309.
[5] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, Nov. 2011.
[6] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Sci.*, vol. 327, no. 5968, pp. 1018–1021, 2010.
[7] S. Mitra, S. Chattopadhyay, and S. S. Das, "Deployment considerations for mobile data offloading in lte-femtocell networks," in *Proc. Int. Conf. Signal Process. Commun.*, 2014, pp. 1–6.
[8] J. Brown, A. J. Broderick, and N. Lee, "Word of mouth communication within online communities: Conceptualizing the online social network," *J. Interactive Marketing*, vol. 21, no. 3, pp. 2–20, 2007.
[9] L. Xiao, C. Xie, T. Chen, H. Dai, and H. V. Poor, "A mobile offloading game against smart attacks," *IEEE Access*, vol. 4, pp. 2281–2291, 2016.
[10] Y. Zhang, D. Niyato, and P. Wang, "Offloading in mobile cloudlet systems with intermittent connectivity," *IEEE Trans. Mobile Comput.*, vol. 14, no. 12, pp. 2516–2529, Dec. 2015.
[11] T. Han and N. Ansari, "Offloading mobile traffic via green content broker," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 161–170, Apr. 2014.
[12] I. Ashraf, L. T. Ho, and H. Claussen, "Improving energy efficiency of femtocell base stations via user activity detection," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2010, pp. 1–5.
[13] T. Han, N. Ansari, M. Wu, and H. Yu, "On accelerating content delivery in mobile networks," *IEEE Commun. Surveys Tutorials*, vol. 15, no. 3, pp. 1314–1333, Jul.-Sep. 2013.
[14] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" in *Proc. 6th Int. Conf.*, 2010, Art. no. 26.
[15] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. 8th Int. Conf. Mobile Syst. Appl. Services*, 2010, pp. 209–222.
[16] S. Tombaz, Z. Zheng, and J. Zander, "Energy efficiency assessment of wireless access networks utilizing indoor base stations," in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2013, pp. 3105–3110.
[17] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *IEEE Commun. Surveys Tutorials*, vol. 15, no. 1, pp. 387–401, Jan.-Mar. 2013.
[18] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in location based services," in *Proc. IEEE INFOCOM.*, 2013, pp. 2985–2993.
[19] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics of mobile data offloading," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2013, pp. 351–356.
[20] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum leasing as an incentive towards uplink macrocell and femtocell cooperation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 617–630, Apr. 2012.
[21] S. Hua, X. Zhuo, and S. S. Panwar, "A truthful auction based incentive framework for femtocell access," in *Proc. Wireless Commun. Netw. Conf.*, 2013, pp. 2271–2276.
[22] P. H. Reingen, B. L. Foster, J. J. Brown, and S. B. Seidman, "Brand congruence in interpersonal relations: A social network analysis," *J. Consum. Res.*, vol. 11, no. 3, pp. 771–783, 1984.
[23] J. J. Brown and P. H. Reingen, "Social ties and word-of-mouth referral behavior," *J. Consum. Res.*, vol. 14, no. 3, pp. 350–362, 1987.
[24] L. Guo, X. Liu, Y. Fang, and X. Li, "User-centric private matching for ehealth networks-a social perspective," in *Proc. Global Commun. Conf.*, 2012, pp. 732–737.
[25] L. Guo, C. Zhang, J. Sun, and Y. Fang, "PAAS: A privacy-preserving attribute-based authentication system for ehealth networks," in *Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst.*, 2012, pp. 224–233.
[26] L. Guo, C. Zhang, J. Sun, and Y. Fang, "A privacy-preserving attribute-based authentication system for mobile health networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 9, pp. 1927–1941, Sep. 2014.

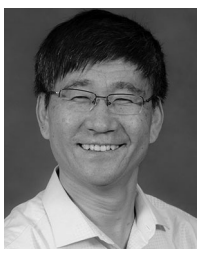
[27] L. Guo, C. Zhang, and Y. Fang, "A trust-based privacy-preserving friend recommendation scheme for online social networks," *IEEE Trans. Depend. Secure Comput.*, vol. 12, no. 4, pp. 413–427, Jul./Aug. 2015.
[28] X. Gong, L. Duan, and X. Chen, "When network effect meets congestion effect: Leveraging social services for wireless services," *Netw.*, vol. 1, no. 2, 2015, Art. no. 3.
[29] X. Gong, X. Chen, and J. Zhang, "Social group utility maximization game with applications in mobile social networks," in *Proc. 51st Annu. Allerton Conf. Commun. Control Comput.*, 2013, pp. 1496–1500.
[30] X. Chen, X. Gong, L. Yang, and J. Zhang, "A social group utility maximization framework with applications in database assisted spectrum access," in *Proc. IEEE INFOCOM*, 2014, pp. 1959–1967.
[31] O. Candogan, K. Bimpikis, and A. Ozdaglar, "Optimal pricing in networks with externalities," *Operations Res.*, vol. 60, no. 4, pp. 883–905, 2012.
[32] X. Zhang, L. Guo, M. Li, and Y. Fang, "Social-enabled data offloading via mobile participation-a game-theoretical approach," in *Proc. IEEE Global Commun. Conf.*, 2016, pp. 1–6.
[33] E. O. Laumann, *Prestige and Association in an Urban Community: An Analysis of an Urban Stratification System*. Indianapolis, IN, USA: Bobbs-Merrill Company, 1966.
[34] L. Guo, C. Zhang, H. Yue, and Y. Fang, "A privacy-preserving social-assisted mobile content dissemination scheme in dtns," in *Proc. IEEE INFOCOM.*, 2013, pp. 2301–2309.
[35] L. Guo, C. Zhang, H. Yue, and Y. Fang, "PSaD: A privacy-preserving social-assisted content dissemination scheme in DTNs," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2903–2918, Dec. 2014.
[36] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD dataset cambridge/haggle (v. 2009-05-29)," May 2009. [Online]. Available: http://crawdad.org/cambridge/haggle/20090529
[37] W.-J. Hsu, D. Dutta, and A. Helmy, "Profile-cast: Behavior-aware mobile networking," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2008, pp. 3033–3038.
[38] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "CRAWDAD dataset ncsu/mobilitymodels (v. 2009-07-23)," Jul. 2009. [Online]. Available: http://crawdad.org/ncsu/mobilitymodels/20090723
[39] Magic kingdom - disney world. (2017). [Online]. Available: http://www.wdwinfo.com/maps/magic-kingdom-map.pdf
[40] K. E. Case and R. C. Fair, *Principles of Microeconomics*. London, U. K.: Pearson Education, 2007.
[41] Wikipedia. (2016). [Online]. Available: https://en.wikipedia.org/wiki/Utility
[42] S. Asmussen, *Appl. Probability and Queues*. Berlin, Germany: Springer, 2008, vol. 51.
[43] H. Cramér, *Random Variables and Probability Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 2004, vol. 36.
[44] D. Fudenberg and J. Tirole, *Game theory*, 1991. Cambridge, MA, USA: MIT Press, vol. 393, p. 12, 1991.
[45] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica: J. Econometric Soc.*, vol. 33, pp. 520–534, 1965.
[46] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[48] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.
[49] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
[50] K. Fan, "Fixed-point and minimax theorems in locally convex topological linear spaces," *Proc. National Academy Sci. United States America*, vol. 38, no. 2, 1952, Art. no. 121.

Xiaonan Zhang received the BE degree in communication engineering from the Beijing University of Chemical Technology, China, and the MS degree in electrical and computer engineering from the State University of New York, Binghamton, in 2012 and 2017, respectively, where she is pursuing the PhD degree. She was a research assistant with the Beijing University of Posts and Communications, China, during 2012-2015. Her research interests include resource management in IoT and physical-layer security in the wireless network. She is a student member of the IEEE.



Linke Guo received the BE degree in electronic information science and technology from the Beijing University of Posts and Telecommunications, and the MS and the PhD degrees in electrical and computer engineering from the University of Florida in 2008, 2011, and 2014, respectively. Since August 2014, he has been an assistant professor in the Department of Electrical and Computer Engineering, Binghamton University, State University of New York. His research interests include network security and privacy, social networks, and applied cryptography. He serves as the publication chair of the IEEE Conference on Communications and Network Security (CNS) 2016 and 2017. He was the symposium co-chair of the Network Algorithms and Performance Evaluation Symposium, ICNC 2016. He has served as the technical program committee (TPC) members for several conferences including IEEE INFOCOM, ICC, GLOBECOM, and WCNC. He is the co-recipient of the Best Paper Award of Globecom 2015, Symposium on Communication and Information System Security. He is a member of the IEEE and the ACM.

Ming Li received the BE degree in electrical engineering from Sun Yatsen University, China, the ME degree in electrical engineering from the Beijing University of Posts and Communications, China, and the PhD degree in electrical and computer engineering from Mississippi State University, Starkville, in 2007, 2010, and 2014, respectively. She is currently an assistant professor in the Department of Computer Science and Engineering, University of Nevada, Reno. Her research interests include cybersecurity, privacy-preserving data analysis, resource management and network optimization in cyber-physical systems, cloud computing, mobile computing, wireless networks, smart grid, and big data. She is a member of the IEEE.



Yuguang Fang (F'08) received the MS degree from Qufu Normal University, Shandong, China, the PhD degree from Case Western Reserve University, and the PhD degree from the Boston University, in 1987, 1994, and 1997, respectively. He joined the Department of Electrical and Computer Engineering, University of Florida in 2000 and has been a full professor since 2005. He held a University of Florida Research Foundation (UFRF) Professorship (2017-2020, 2006-2009), University of Florida Term Professorship (2017-2019), a Changjiang Scholar Chair Professorship (Xidian University, Xian, China, 2008-2011; Dalian Maritime University, Dalian, China, 2015-2018), Overseas Academic Master (Dalian University of Technology, Dalian, China, 2016-2018), and a Guest Chair Professorship with Tsinghua University, China (2009-2012). He received the US National Science Foundation Career Award in 2001, the Office of Naval Research Young Investigator Award in 2002, the 2015 IEEE Communications Society CISTC Technical Recognition Award, the 2014 IEEE Communications Society WTC Recognition Award, and the Best Paper Award from IEEE ICNP (2006). He has also received a 2010-2011 UF Doctoral Dissertation Advisor/Mentoring Award, a 2011 Florida Blue Key/UF Homecoming Distinguished Faculty Award, and the 2009 UF College of Engineering Faculty Mentoring Award. He was the editor-in-chief of the *IEEE Transactions on Vehicular Technology* (2013-present), the editor-in-chief of the *IEEE Wireless Communications* (2009-2012), and serves/served on several editorial boards of journals including the *IEEE Transactions on Mobile Computing* (2003-2008, 2011-2016), the *IEEE Transactions on Communications* (2000-2011), and the *IEEE Transactions on Wireless Communications* (2002-2009). He has been actively participating in conference organizations such as serving as the technical program co-chair of the IEEE INFOCOM2014 and the technical program vice-chair of the IEEE INFOCOM'2005. He is a fellow of the IEEE and a fellow of the American Association for the Advancement of Science (AAAS).