Multiple Mobile Data Offloading Through Disruption Tolerant Networks

Yong Li, *Member, IEEE,* Mengjiong Qian, Depeng Jin, Pan Hui, *Member, IEEE,* Zhaocheng Wang, *Senior Member, IEEE,* and Sheng Chen, *Fellow, IEEE*

Abstract—To cope with explosive traffic demands on current cellular networks of limited capacity, Disruption Tolerant Networking (DTN) is used to offload traffic from cellular networks to high capacity and free device-to-device networks. Current DTN-based mobile data offloading models are based on simple and unrealistic network assumptions which do not take into account the heterogeneity of mobile data and mobile users. We establish a mathematical framework to study the problem of multiple-type mobile data offloading under realistic assumptions, where (i) mobile data are heterogeneous in terms of size and lifetime; (ii) mobile users have different data subscribing interests; and (iii) the storages of offloading helpers are limited. We formulate the objective of achieving maximum mobile data offloading as a submodular function maximization problem with multiple linear constraints of limited storage, and propose three algorithms, suitable for the generic and more specific offloading scenarios, respectively, to solve this challenging optimization problem. We show that the designed algorithms effectively offload data to the DTN by using both the theoretical analysis and simulation investigations which employ both real human and vehicular mobility traces.

Index Terms—Mobile data offloading, storage allocation, Disruption Tolerant Networking

1 Introduction

Mobile Internet access is getting ever-increasingly popular and today provides various services and applications, including audio, images and video. Cisco estimates that mobile traffic is growing at an annual rate of 131% in 2011 and will reach over 6.3 exabytes per month in 2015, while two-thirds of the mobile traffic will be video by 2015 [1]. Mobile cellular networks provide the most popular method of mobile access today [2]. With the explosive increase in mobile services and user demands, cellular networks will very likely be overloaded and congested in the near future. Pessimists already foresee near nightmare scenarios that, during peak time and in urban area, users face extreme performance hits in terms of low or even no network bandwidth, missed calls, and unreliable coverage. Indeed, the limited spectrum and over-the-air bandwidth constrain multimedia applications, such as mobile TV, streaming music or video downloads.

To cope with this explosive growth in traffic demands with the limited current network capacity, it is an urgent agenda for cellular providers to offer quick and promising solutions. A straightforward solution appears to be increasing the cellular network capacity by adding more base stations with smaller cell size, such as picocells and femtocells, or by upgrading cellular networks to next-generation networks, like 4G [3]. However, these activities are very expensive with low financial returns, particularly under the current flat pricing model where charges are independent of traffic. Even if the network capacity is enhanced this way, the future demands from users and new applications will quickly outstrip the network capacity. Consequently, some providers are forced to adopt some drastic short-term “solutions”, like limiting users’ traffic to 5 GB per month and educating users on “responsible” access. Obviously, all these methods are ineffective and insufficient. In the long-term, providers must find different network technologies to offer sufficient bandwidth to end users.

Wireless networks and mobile devices are undergoing a major evolution today. New generation of mobile devices have seen their computation power, storage and communications capabilities enhanced significantly. Up-to-date mobile phones all have multiple network interfaces, such as 3G, WiFi and Bluetooth. These advanced devices help to create an enabling environment, where the network potential bandwidth and communication opportunity can be effectively utilized to enhance users’ experience. For example, direct device-to-device communications can be leveraged to transmit large amounts of data by using the unused bandwidth between mobile devices that are in proximity, over Bluetooth or WiFi, which offloads data from the cellular network. Indeed, many researchers have actively studies offloading mobile data from the overloaded

- *Y. Li, M. Qian, D. Jin, and Z. Wang are with the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: {liyong07, jindp, zcwang}@tsinghua.edu.cn, qianmj@gmail.com.*
- *P. Hui is with Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, Telekom Innovation Laboratories, Berlin, Germany, and Aalto University, Helsinki, Finland. E-mail: panhui@cse.ust.hk.*
- *S. Chen is with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with the Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia. E-mail: sqc@ecs.soton.ac.uk.*

Manuscript received 21 Nov. 2012; revised 30 Apr. 2013; accepted 3 May 2013. Date of publication 16 May 2013; date of current version 2 July 2014. 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 10.1109/TMC.2013.61

cellular networks to WiFi networks [3], [4], and these works have demonstrated that large amounts of mobile data traffics may potentially be offloaded away from 3G to WiFi. Currently, cellular providers are considering mobile offloading as an attractive option to free network capacity.

Huge portion of the mobile data traffics delivered by service providers, such as weather forecasts, multimedia newspapers, stock information and movie trailers, do not have strict realtime constraints. Also these mobile data are typically broadcasted to large number of users. A recent approach utilises these properties to offload cellular data traffics using Disruption Tolerant Networks (DTNs) [2], [5], [6]. Benefiting from the delay-tolerant nature of non-realtime applications, providers can delay and even shift the transmission to DTN. Exploiting common interests among the mobile users for these mobile data, providers may only need to deliver the information to a small fraction of the users, and rely on these selected users to further disseminate the data through DTN. Providers can offer incentive schemes to encourage users to participate in this type of mobile data offloading. This kind of offloading is very attractive to operators as it offers the quickest way, at the lowest cost, to support the exponential growth of mobile data [1], which otherwise operators will be unable to support even they upgrade their infrastructures to 4G.

In this paper, we establish a mathematical framework to study the DTN-based mobile traffic offloading of multiple mobile data items in a realistic mobile environment. This problem is challenging for several reasons. Firstly, mobile data provided by service providers are not single uniform type. Because mobile data have different delay-sensitivities, content sizes, etc, it is difficult for providers to decide how to offload these heterogeneous mobile data. Secondly, users' demands and interests to mobile data are very different, and this must be considered in the design of any offloading scheme. Thirdly, a DTN has very limited network resources in practice, e.g. storage and battery capacity of mobile devices are limited, and communication contacts in a DTN are opportunistic by nature. How to efficiently exploit the limited resources to improve the overall system performance is a very challenging problem. Although several store-carry-and-forwarding schemes have been proposed for DTNs, most of them are end-to-end routing solutions and do not take into account the crucial fact that users' interests are different. Consequently, these schemes cannot be used in mobile traffic offloading directly. For example, for video transmission applications which usually involve large amount of mobile data traffics, end-to-end multi-hop opportunistic forwarding may not be feasible because it would result in huge energy consumption and large transmission delay. Moreover, multi-hop forwarding requires that users in the network are willing and able to cooperate for store-carry-and-forwarding data. However, in practical mobile data offloading, many users may not be willing to or cannot cooperate due to the privacy consideration or resource limitation. Consequently, only a small fractional of users will participate in the offloading. Therefore, new and different schemes are required for efficient mobile traffic offloading.

In order to model a realistic DTN environment, we consider the following network settings: 1) the network contains heterogeneous users in terms of data preference and privacy intention; 2) the offloaded data have different delay sensitivities and sizes; and 3) the offloading helpers' storages are limited in size. These realistic conditions were not considered in the previous analytical works [2], [5] for simplicity reasons but they are explicitly taken into account in our work. Our contributions are now summarized as follows:

- We formulate the optimal DTN-based offloading under the generic consideration of heterogeneous data traffics, users' interests, and limited storage resources as a problem of Submodular Function Maximization (SFM) under multiple linear constraints.
- We prove that this generic optimal DTN-based offloading problem is NP-complete, and provide three suboptimal algorithms to tackle this challenging offloading problem. The first algorithm is designed for the generic offloading scenario considered, and the second algorithm is suitable for the offloading scenario where content lifetimes are short, while the third algorithm provides the optimal solution for the special homogeneous offloading scenario with homogeneous contact rates and data items.
- Through extensive mobility and trace-driven simulations, we show that our designed algorithms achieve good system performance in both human and vehicular networking environments.

The rest of the paper is organized as follows. We describe the system overview and modeling in Section 2, and formulate the associated optimization problem in Section 3. In Section 4, we design three algorithms to obtain the system solution. In Section 5, we introduce the experimental environment for performance evaluation and present the simulation results. In the light of our results, we discuss the related work in Section 6. We conclude the paper in Section 7.

2 System Overview and Modelling

We first provide an overview of the DTN-based mobile data offloading system, and then discuss the underlying heterogeneous network modeling.

2.1 Overview

In our DTN-based mobile data offloading system, some chosen users, referred to as helpers, participate in the offloading. Incentives for these users can be provided by using some micro-payment scheme, or the operator can offer the participants a reduced cost for the service or better Quality of Service (QoS) [7]–[9]. A full analysis of such incentives is not the focus of this paper. Our multiple-type mobile data offloading scheme is depicted in Fig. 1. There are two types of nodes in the system, known as offloading *helper* and mobile data *subscriber*, respectively. The service provider first chooses some users that are willing to participate in data offloading. When it has a set of mobile data items to deliver, the storage allocation decision is made, and it then transmits the mobile data to these chosen helpers through the cellular network

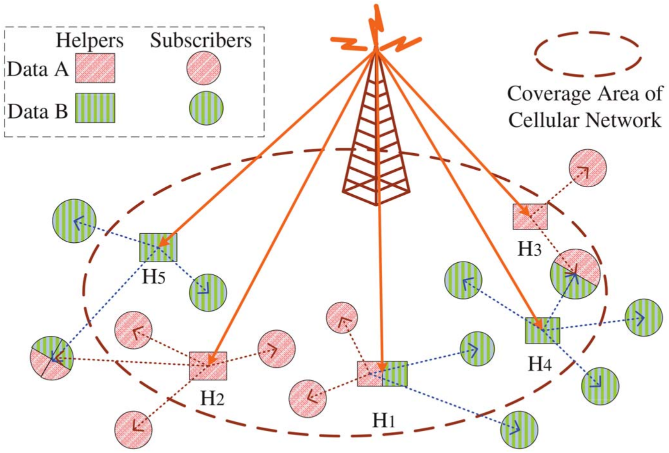


Fig. 1. Multiple-type mobile data offloading through disruption tolerant network.

according to the storage allocation policy obtained. These offloading helpers then further propagate the data to other subscribers that are interested in the data by short range device-to-device communication. However, the subscribers after obtaining the data will not propagate the data further to others that are interested in them. That is, we only consider the two-hop relaying model for the mobile data offloading. If a subscriber could not receive the data from any helper after a specified "tolerable" duration which is related to the data lifetime, it can directly ask to receive the data from the cellular network.

In the case depicted in Fig. 1, there are two different mobile data, denoted by A and B, the service provider first transmits data A to helpers H_1, H_2 and H_3 as well as transmits data B to helpers H_1, H_4 and H_5. Then, subscribers, according to their interests, can obtain the corresponding data from these helpers by the DTN communication paradigm.

2.2 Networking Modeling

In our system, there are $N + H$ mobile users, labelled as $i \in \{1, 2, \cdots, N + H\}$. These users are either vehicles or humans carrying wireless devices, e.g. mobile phones. Since there exist many different types of mobile data, for example, multimedia newspapers, weather forecasts and movie trailers, we model the mobile traffic with C different data items, denoted by $\mathcal{C}$. For any $k \in \mathcal{C}$, its data length is l_k. Some of the users, even when they have additional resource to help with offloading, may not be willing to act as helpers owing to various considerations, including privacy concerns. Hence, it is unrealistic to assume that all the nodes can participate in the offloading activities to pass the data to others unselfishly. On the other hand, even if many users want to become helpers, service providers can only choose a small set of the users that can offload the traffic most efficiently, as they have to pay directly or indirectly to the helpers. Therefore, a service provider will choose an offloading helper carefully by considering many factors [5]. How to select helpers is beyond the scope of this work. The exiting incentive schemes [2], [5] are available to select users in the system to act as offloading helpers. For example, in [5], the helper selection strategies rely on social network properties that some nodes may play a globally important role in the network, which can help offloading a very large amount of the mobile data. Therefore, they may be selected as the offloading helpers.

We use $\mathcal{H}$ to denote the set of chosen helper users that are willing to participate in the offloading and $\mathcal{N}$ to denote the other subscriber users, with $|\mathcal{H}| = H$ and $|\mathcal{N}| = N$. In general, it is possible that some of the chosen helpers are also subscribers as well. However, in our application of offloading mobile data from the cellular network to the DTN, we assume that $\mathcal{H}$ and $\mathcal{N}$ are disjoint. There are good practical reasons to support this assumption. Mobile users are hand held devices, which have very limited resource and capacity. Note that the resources of mobile handsets are primarily used for maintaining their cellular network communication, and these devices can only afford any spare resource and capacity left to form the DTN for assisting data offloading. Therefore, it is reasonable to assume that all the mobile users are rational, and they are able and willing to participate in the data offloading as helpers only when they are not currently retrieving the data for themselves. In other words, we assume that if a mobile user is a subscriber it could not afford additional resource to act as a helper. Therefore, the service provider only selects some of the mobile users that currently are not retrieving mobile data for themselves as the helpers to participate in the offloading to help the data transmission to the relevant subscribers via the DTN communication. We point out that our proposed algorithms can be applied directly to the more generic mobile data offloading problem where a mobile user can be both a subscriber and a helper.

For the helpers, the system requires their storages to buffer some data items. Note that mobile data may include multimedia content like movies that are very large in size, and even the 32 GB of storage available on the current state-of-the-art devices becomes limited in comparison. Also it is unrealistic to ask a helper to contribute all its storage solely for offloading. Thus, we should consider the storage that each helper is willing to share as the constraint, which determines the number of data items that can be stored. Taking this realistic condition into consideration, we assume that helper s, $s \in \mathcal{H}$, can at most buffer L_s size of data items. Since we use device-to-device communication to offload the data to subscribers, nodes can only communicate when they move into transmission range of each other. However, such short-range communications can often provide comparatively high data transmission rates, and nodes may complete very large mobile data transmission during one contact. We assume that the contacts between any two nodes occur at Poisson rates. Therefore, the contact rate between nodes i and j form a homogeneous Poisson process with the contact rate $\gamma_{i,j}$.

Poisson distributed contacts have been shown to accurately model real-world DTN traces and are widely used to model DTN systems [10], [11], [13]–[16], where the Inter-Contact Time (ICT) is modeled by the exponential distribution. The recent works for the well-known mobility models, such as random waypoint and random direction [17]–[19] as well as the real-life mobility traces [11], [12], have confirmed that the individual contact time follows an exponential distribution. Although some studies [20]–[22]

suggest that the aggregated ICT follows a truncated power law distribution, the individual ICT with constraints can be better modeled by the exponential distribution with heterogeneous coefficients [11], [23], [24]. Moreover, the recent works [15], [25] for modeling the aggregated ICT of vehicles, which are typical nodes in vehicular DTNs, reveal the exponential distribution of ICT between nodes by analyzing a large amount of car/taxi mobility traces. Therefore, it is reasonable to model the individual pair-wise contact between nodes by the Poisson process with different contact rate of $\gamma_{i,j}$.

2.3 User Interest Modeling

We now characterize the subscriber behaviours in accessing different data items in a mobile data offloading system. In a system with multiple data items, a subscriber has different interests in different items, and different subscribers have different accessing behaviours. Moreover, some data items are popular data that are of interest to many subscribers, while some other data items are not popular data which may only be interesting to a very small number of subscribers. We describe the subscriber's interests to different mobile data by a subscriber profile, and model the popularity of mobile data by an interest distribution. In this way, we model the steady state of the subscribers' interests in different data items.

Specifically, for all the mobile data, the system has M keywords, denoted by the set $\mathcal{M}$, to describe them. Any item $k \in \mathcal{C}$ is described by a subset of keywords, denoted by $\mathcal{M}_k \subseteq \mathcal{M}$, and the associated weights v_{m_k}, which indicates the importance of keyword $m_k \in \mathcal{M}_k$. In this way, we can define the popularity of mobile data items. We assume $\sum_{m_k \in \mathcal{M}_k} v_{m_k} = 1$. To model the interests of different subscribers on different data, we define P_i^m as the degree of subscriber $i \in \mathcal{N}$ been interested in keyword $m \in \mathcal{M}$. In this way, we can compare the interests of subscriber i to two different keywords $m_1, m_2 \in \mathcal{M}$ by $P_i^{m_1}$ and $P_i^{m_2}$. Thus, the interest profile of subscriber i is defined by the set $\mathcal{P}_i = \{P_i^m : m \in \mathcal{M}\}$. We assume $\sum_{m \in \mathcal{M}} P_i^m = 1$. The interest probability of subscriber $i \in \mathcal{N}$ in mobile data $k \in \mathcal{C}$, defined by $w_{i,k}$, can be obtained as follows:

$$w_{i,k} = \sum_{m_k \in \mathcal{M}_k} v_{m_k} P_i^{m_k}. \tag{1}$$

3 Problem Formulation

In this section, the objective of our DTN-based mobile data offloading is formally formulated.

3.1 System Utility

Based on the above characterizations of user interests and system model, we can obtain the expectation of the total data size offloaded in the whole DTN system. Denote $\mathbf{X} = (x_{s,k})$, $s \in \mathcal{H}$ and $k \in \mathcal{C}$, as the storage allocation for helpers, in which $x_{s,k} \in \{0, 1\}$ and $x_{s,k} = 1$ indicates that offloading helper s stores item k in its buffer. A lifetime T_k is assigned to each data item k, which implies that all the helpers will discard this item at time T_k. For assisting the derivation, further define $\mathcal{H}_k$ as the set of the helpers that store copies of item k in their buffers, which is

$$\mathcal{H}_k = \{s \in \mathcal{H} \mid x_{s,k} = 1\}. \tag{2}$$

Since the lifetime T_k is assigned to each data item k, if a subscriber does not receive a required mobile data from helpers after the data lifetime expires, it will directly require this mobile data from the cellular network, which means that this mobile item is not offloaded to the required subscriber. The amount of the not yet offloaded data after the associated lifetimes expire will need to be transmitted by the cellular network directly to the related subscribers. Therefore, the objective of the offloading system is to maximize the expected total size of offloaded data to all the subscribers. Thus, we define the *system offloading utility function* $U(\mathbf{X})$, as the expectation of the total size of the offloaded data before the associated lifetimes expire, which depend on the storage allocation decision $\mathbf{X}$, and this objective function can be expressed as

$$U(\mathbf{X}) = \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} d_{i,k}, \tag{3}$$

where $d_{i,k}$ is the probability that subscriber i receives data k before deadline T_k. The unit of this unity function is the offloaded data size. Since more than one helper may store item k, we consider the *dissemination opportunity* metric for $s \in \mathcal{H}$, $i \in \mathcal{N}$ and $k \in \mathcal{C}$, denoted as $t_{s,i,k}$, which is the probability that subscriber i obtains content k from helper s. Because the contact rate between s and i follows the Poisson distribution with rate $\gamma_{s,i}$ and the contact event is independent of the user interests, we can model the dissemination opportunity as the Poisson process with rate $\gamma_{s,i} w_{i,k}$. If helper s stores content k in its buffer, we can derive the probability that subscriber i obtains content k from s before lifetime T_k as follows:

$$t_{s,i,k} = \int_{T_k}^{\infty} \gamma_{s,i} w_{i,k} e^{-\gamma_{s,i} w_{i,k} y} dy = 1 - e^{-\gamma_{s,i} w_{i,k} T_k}. \tag{4}$$

On the other hand, if helper s does not store content k, this probability is zero. Thus, combining with the definition of $x_{s,k}$, we have $t_{s,i,k} = 1 - e^{-x_{s,k}\gamma_{s,i} w_{i,k} T_k}$. Therefore, the probability that subscriber i cannot obtain content k from all the possible helpers $s \in \mathcal{H}$ can be expressed as $\prod_{s \in \mathcal{H}} (1 - t_{s,i,k})$. It then becomes obvious that the probability of subscriber i receiving data k before deadline T_k, namely $d_{i,k}$, is given by

$$d_{i,k} = 1 - \prod_{s \in \mathcal{H}} (1 - t_{s,i,k}). \tag{5}$$

By substituting $d_{i,k}$ of (5) into (3), the expectation of the total disseminated data size before the related lifetimes expire, $U(\mathbf{X})$, can be written as:

$$U(\mathbf{X}) = \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k} T_k \sum_{s \in \mathcal{H}} \gamma_{s,i} x_{s,k}}\right). \tag{6}$$

3.2 Problem Statement

Based on the total offloaded data size obtained in Subsection 3.1, we can specify the objective of our multiple mobile data offloading problem as to design the storage

TABLE 1
List of Commonly Used Notations and Variables Throughout the Paper

Notation/Variable	*Description*
H, N, C	The numbers of helpers, subscribers and data items in the system, respectively.
$\mathcal{H}, \mathcal{N}, \mathcal{C}$	The sets of helpers, subscribers and data items, respectively.
$\mathcal{H}_k$	The set of helpers that store copies of item k in their buffers, $k \in \mathcal{C}$.
s, i, k	The indexes for helpers, subscribers and data items, respectively.
l_k, T_k	The size and lifetime of data item k, respectively, $k \in \mathcal{C}$.
L_s	The buffer size of helper s, $s \in \mathcal{H}$.
$\gamma_{s,i}$	The contact rate between helper s and subscriber i, $i \in \mathcal{N}$, $s \in \mathcal{H}$.
$w_{i,k}$	The interest probability of subscriber $i \in \mathcal{N}$ in mobile data $k \in \mathcal{C}$.
$\mathbf{X} = (x_{s,k})$	The storage allocation policy of helpers.

allocation for helpers that maximizes the system offloading utility function (6) while satisfying all the helpers' buffer-size constraints. Formally, this is defined as the following optimization problem:

$$\begin{aligned} &\max U(\mathbf{X}) \\ &\text{s.t. } x_{s,k} \in \{0,1\}, \forall s \in \mathcal{H}, \forall k \in \mathcal{C}, \\ &\quad \text{and } \sum_{k \in \mathcal{C}} x_{s,k} l_k \le L_s \text{ for } \forall s \in \mathcal{H}, \end{aligned} \tag{7}$$

where $\sum_{k\in\mathcal{C}} x_{s,k} l_k \le L_s$ is the buffer size constraint of offloading helper s. For the ease of reference, we summarize the commonly used notations and variables throughout the paper in Table 1.

We emphasize that the optimization problem (7), or making the storage allocation decision to maximize the total size of offloaded data before the associated data lifetimes expire, is solved by a centralized algorithm resided in the cellular network operator. This algorithm require the information of the content lengths l_k and content lifetimes T_k, the storage sizes of helpers L_s, and the subscribers' content interests $w_{i,k}$, as well as the contact rates $\gamma_{s,i}$. Note that these parameters are not collected via the DTN. The cellular network operator can obtain the content-related parameters, l_k and T_k, from the content servers via the wired Internet, and the required communication overhead is negligible compared with the bandwidth of the wired Internet. As well as forming a DTN, mobile users are connected to the cellular network. Therefore, they can send the mobile user-related parameters, L_s, $w_{i,k}$, and $\gamma_{s,i}$, via the uplink of the cellular network to the offloading decision-making centralized algorithm, and this signaling overhead is also insignificant compared with the bandwidth of the cellular network. In fact, there already exist certain amount of signaling between a mobile user and the cellular network, and the required mobile-related parameters may take "piggybacking" in these existing signalings. Moreover, l_k and T_k as well as L_s, $w_{i,k}$ and $\gamma_{s,i}$ are often quasi-static parameters, which are known by the content servers or the cellular network operator in advance. More specifically, many mobile users travel on predetermined routes and schedules, and examples include city buses and people traveling by cars to and from work. Therefore, daily mobility patterns exhibit certain regularity, and the contact rates between many users are often quasi-static, which can be obtained by the cellular network operator in advance with high accuracy. Moreover, any new contact information may be collected regularly from mobile users.

4 Multiple Data Offloading

We begin by analyzing the utility function in (7).

4.1 Utility Function Analysis

In the optimization problem (7), the system utility is an increasing and concave function of $x_{s,k}$, which only takes the value of 1 or 0, and the constraints are linear. Therefore, the optimization (7) is a challenging 0-1 programming problem. We note that submodularity has long been studied in various similar problems [26]–[29]. Therefore, we first give a brief introduction to submodularity. We can use submodularity related theory to analyse the optimization problem (7) and to aid the design of algorithms for solving our multiple mobile data offloading problem.

A function f defined on subsets of the universe $\mathcal{C}$ is called *submodular*, if and only if $f(\mathcal{A} \cup x) - f(\mathcal{A}) \ge f(\mathcal{B} \cup x) - f(\mathcal{B})$ holds for $\forall \mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$ and $\forall x \in \mathcal{C} \backslash \mathcal{B}$.

Let us define the set of the pairs of helpers and data items that helpers carry as

$$\mathcal{P}_{\mathcal{H}\times\mathcal{C}} = \{(s,k) \in \mathcal{H} \times \mathcal{C} \mid x_{s,k} = 1\}. \tag{8}$$

The system offloading utility can be defined as $U{:}2^{\mathcal{H}\times\mathcal{C}} \to \mathbb{R}$, as the utility function (6) is equivalent to

$$U(\mathcal{P}_{\mathcal{H}\times\mathcal{C}}) = \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} \left(1 - e^{-w_{i,k} T_k \sum\limits_{s:(s,k)\in\mathcal{P}_{\mathcal{H}\times\mathcal{C}}} \gamma_{s,i}}\right). \tag{9}$$

We have the following theorem, which shows that the problem (7) is a Submodular Function Maximization (SFM) under Multiple Linear Constraints (MLCs). We omit the proof here owing to the space limitation.

Theorem 1. *The system offloading utility function $U(\mathcal{P}_{\mathcal{H}\times\mathcal{C}}){:}2^{\mathcal{H}\times\mathcal{C}} \to \mathbb{R}$ is submodular, and the problem (7) is NP-hard.*

Proof. For $\mathbf{A} \subseteq \mathbf{B} \subseteq 2^{\mathcal{H}\times\mathcal{C}}$ and $(s_0, k_0) \in 2^{\mathcal{H}\times\mathcal{C}}\backslash\mathbf{B}$, we have

$$\begin{aligned}
&U(\mathbf{A}\cup\{(s_0,k_0)\}) - U(\mathbf{A})\\
&= \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} \left(1 - e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{A}\cup\{(s_0,k_0)\}} \gamma_{s,i}}\right)\\
&\quad - \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} \left(1 - e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{A}} \gamma_{s,i}}\right)\\
&= \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} \left(1 - e^{-w_{i,k}T_k x_{s_0,k_0}\gamma_{s,i}}\right) e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{A}} \gamma_{s,i}}. \quad (10)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\Big(U(\mathbf{A}\cup\{(s_0,k_0)\}) - U(\mathbf{A})\Big) - \Big(U(\mathbf{B}\cup\{(s_0,k_0)\}) - U(\mathbf{B})\Big)\\
&= \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} \left(1 - e^{-w_{i,k}T_k x_{s_0,k_0}\gamma_{s,i}}\right)\\
&\quad \left(e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{A}} \gamma_{s,i}} - e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{B}} \gamma_{s,i}}\right)\\
&= \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{A}} \gamma_{s,i}} \left(1 - e^{-w_{i,k}T_k \sum\limits_{s:(s,k)\in\mathbf{B}\backslash\mathbf{A}} \gamma_{s,i}}\right)\\
&\quad \left(1 - e^{-w_{i,k}T_k x_{s_0,k_0}\gamma_{s,i}}\right) \geq 0, \quad (11)
\end{aligned}$$

which proves that $U(\mathcal{P}_{\mathcal{H}\times\mathcal{C}})$ is a submodular function on $2^{\mathcal{H}\times\mathcal{C}}$.

To prove that the problem (7) is NP-hard, we use the technique of reduction [30]. We first consider the following 0-1 Knapsack problem which is well-known to be NP-hard [30]

$$\begin{aligned}
&\max \sum_{j=1}^{n} p_j z_j,\\
&\text{s.t. } z_j \in \{0,1\}, \forall 1 \leq j \leq n, \text{ and } \sum_{j=1}^{n} \eta_j z_j \leq W,
\end{aligned} \quad (12)$$

where p_j and η_j are respectively the value and weight of item j, and z_j is the decision valuable which decides whether to choose item j. By defining $x_{1,j} \in \{0,1\}$, $\forall \eta_j$, we have $\sum_{j=1}^{n} p_j z_j = \sum_{j=1}^{n} \eta_j \left(1 - \left(\frac{\eta_j - p_j}{\eta_j}\right)^{x_{1,j}}\right)$. Therefore, the 0-1 Knapsack problem (12) can be rewritten as

$$\begin{aligned}
&\max \sum_{k=1}^{n} \eta_k \left(1 - \left(\frac{\eta_k - p_k}{\eta_k}\right)^{x_{1,k}}\right),\\
&\text{s.t. } x_{1,k} \in \{0,1\}, \forall 1 \leq k \leq n, \text{ and } \sum_{k=1}^{n} \eta_k x_{1,k} \leq W.
\end{aligned} \quad (13)$$

Note that the optimization problem (7) is reduced to the 0-1 Knapsack problem (13) by defining

$$\begin{aligned}
&|\mathcal{N}| = |\mathcal{H}| = 1,\\
&|\mathcal{C}| = n,\ L_1 = W,\ \gamma_{1,1} = n,\\
&w_{1,k} = \frac{1}{n} \text{ and } T_k = \ln\left(\frac{\eta_k}{\eta_k - p_k}\right), 1 \leq k \leq n,
\end{aligned}$$

which is NP-hard. Therefore, the more generic optimization problem (7) must be NP-hard. □

4.2 Algorithms

We now design three suboptimal algorithms to approximately solve the challenging optimization problem (7), suitable for different offloading scenarios.

Algorithm 1 Greedy Algorithm (GA) for generic multiple mobile data offloading.

```
1: Initialise $m = 0$ and $\mathbf{A}_0 = \emptyset$;
2: while $m = 0$ or $U(\mathbf{A}_m) - U(\mathbf{A}_{m-1}) > 0$ do
3:     $m = m + 1$
4:     $(s_m, k_m) = \arg\max_{(s,k)\in\mathbf{P}} (U(\mathbf{A}_{m-1} \cup \{(s,k)\}) - U(\mathbf{A}_{m-1}))$
5:     $\mathbf{A}_m = \mathbf{A}_{m-1} \cup \{(s_m, k_m)\}$
6: end while
7: Initialise $j = 0$ and $\mathbf{B}_0 = \emptyset$;
8: while $j = 0$ or $U(\mathbf{B}_j) - U(\mathbf{B}_{j-1}) > 0$ do
9:     $j = j + 1$
10:    $(s_j, k_j) = \arg\max_{(s,k)\in\mathbf{P}} \frac{U(\mathbf{B}_{j-1}\cup\{(s,k)\}) - U(\mathbf{B}_{j-1})}{l_k}$
11:    $\mathbf{B}_j = \mathbf{B}_{j-1} \cup \{(s_j, k_j)\}$
12: end while
13: if $U(\mathbf{A}_m) > U(\mathbf{B}_j)$ then
14:    $\mathbf{OPT}^* = \mathbf{A}_m$;
15: else
16:    $\mathbf{OPT}^* = \mathbf{B}_j$
17: end if
18: Return $\mathbf{OPT}^*$ and $U(\mathbf{OPT}^*)$
```

4.2.1 Greedy Algorithm

For the generic offloading scenario, we have a Greedy Algorithm (GA). Consider solving the problem (7) in the way that each storage allocation is determined one by one. Similar to [28], greedy algorithm can be employed as an approximate solution to the problem. When one more copy of a certain item is stored in a helper which satisfies the constraints, the objective function will be enhanced. The gain in the objective function is generally different for different choices of item and helper. Thus, we can choose the data item and helper that maximizes the gain on the objective function among all the legitimate choices at each step, as our first strategy. That is, select (s_0, k_0) as

$$(s_0, k_0) = \arg\max_{(s,k)\in\mathbf{P}} \Big(U(\mathbf{A}\cup(s,k)) - U(\mathbf{A})\Big), \quad (14)$$

where $\mathbf{A}$ is the set of the data items and helpers that have been chosen in the previous steps, and $\mathbf{P}$ denotes the set of all the possible solutions that satisfy the storage constraints.

If a selected item has a huge length, then other items cannot be stored. Thus, the length of each data item is also important. Our second greedy strategy selects the data item and helper that maximizes the gain per unit length. That is, select (s_0, k_0) as

$$(s_0, k_0) = \arg\max_{(s,k)\in\mathbf{P}} \frac{U(\mathbf{A}\cup(s,k)) - U(\mathbf{A})}{l_k}. \quad (15)$$

Since each of these two greedy strategies has its own strengths and drawbacks, a combination of these two strategies is used to enhance the overall performance. In our heuristic algorithm presented in Algorithm 1, these two strategies are both performed and we choose the better one from the two solutions. In Line 1 of the algorithm, we initialise the chosen set $\mathbf{A}$ to be empty. Then, from Lines 2 to 6, we choose the data and helper that maximizes the gain on the objective function among all the legitimate choices at each step, and save the results as set $\mathbf{A}$. From Lines 7 to 12, we consider the solution obtained by the greedy strategy that selects the data item and helper to maximize the gain per unit length at each step, and save the results as set **B**. Finally, we choose a better solution by comparing the system utility of **A** with that of **B** in Lines 13 to 17. For this algorithmic procedure, we can show that it is a pseudo-polynomial-time algorithm with the computational complexity $O(H^3C^2N)$. This complexity is acceptable in practice, considering that the number of helpers, H, in real-world DTN-based offloading systems is usually not too large. We now analyze the performance bound of this algorithm in the following theorem. Again the proof is omitted because of space limitation.

Theorem 2. *Denote the optimal solution of the problem (7) by* $\mathbf{OPT} = \arg\max_{\mathbf{A}\in\mathbf{Q}} U(\mathbf{A})$, *where* $\mathbf{Q} \subseteq 2^{\mathcal{H}\times\mathcal{C}}$ *is the feasible solution set. The solution obtained by Algorithm 1,* $\mathbf{OPT}^*$, *satisfies*

$$U(\mathbf{OPT}^*) \geq \frac{1}{2}\left(1 - e^{-\frac{L-\mu}{L}}\right)U(\mathbf{OPT}),$$

where $L = \sum_{s\in\mathcal{H}} L_s$ *and* $\mu = (H-1)\cdot \max_{k\in\mathcal{C}} l_k$.

Proof. We first consider the policy (15). Assume that the algorithm ends at the step J. $\exists j, 1 \leq j \leq J$, we have

$$\begin{aligned}U(\mathbf{OPT}) - U(\mathbf{B}_j) &\leq U(\mathbf{OPT}\cup\mathbf{B}_j) - U(\mathbf{B}_j)\\ &= U(\mathbf{C}_j\cup\mathbf{B}_j) - U(\mathbf{B}_j)\\ &= \sum_{q=1}^{n}\left(U(\mathbf{C}_j^q\cup\mathbf{B}_j) - U(\mathbf{C}_j^{q-1}\cup\mathbf{B}_j)\right),\end{aligned} \tag{16}$$

where $\mathbf{C}_j = \mathbf{OPT}\backslash\mathbf{B}_j = \{(s'_1, k'_1), (s'_2, k'_2), \cdots, (s'_n, k'_n)\}$ and $\mathbf{C}_j^q = \{(s'_1, k'_1), (s'_2, k'_2), \cdots, (s'_q, k'_q)\}$, $0 \leq q \leq n$. According to the submodular property and the greedy policy used,

$$\begin{aligned}\frac{U(\mathbf{C}_j^q\cup\mathbf{B}_j) - U(\mathbf{C}_j^{q-1}\cup\mathbf{B}_j)}{l_{k'_q}} &\leq \frac{U((s'_q, k'_q)\cup\mathbf{B}_j) - U(\mathbf{B}_j)}{l_{k'_q}}\\ &\leq \frac{U((s_{j+1}, k_{j+1})\cup\mathbf{B}_j) - U(\mathbf{B}_j)}{l_{k_{j+1}}}.\end{aligned} \tag{17}$$

Therefore, we have

$$\begin{aligned}U(\mathbf{OPT}) - U(\mathbf{B}_j) &\leq \sum_{q=1}^{n}\frac{l_{k'_q}}{l_{k_{j+1}}}\left(U((s_{j+1}, k_{j+1})\cup\mathbf{B}_j) - U(\mathbf{B}_j)\right)\\ &\leq \frac{L}{l_{k_{j+1}}}\left(U(\mathbf{B}_{j+1}) - U(\mathbf{B}_j)\right),\end{aligned} \tag{18}$$

that is, $U(\mathbf{B}_{j+1}) - U(\mathbf{B}_j) \geq \frac{l_{k_{j+1}}}{L}\left(U(\mathbf{OPT}) - U(\mathbf{B}_j)\right)$. Using induction, we conclude

$$U(\mathbf{B}_{j+1}) \geq \left(1 - \prod_{q=1}^{j+1}\left(1 - \frac{l_{k_q}}{L}\right)\right)U(\mathbf{OPT}).$$

Consider the following $J+1$ step. Note that now we cannot store any data item in any helper. For any possible $\mathbf{B}_{J+1}$, then at least there is one helper whose remaining storage size is 0, and each remaining storage of the others is no more than $\max_{k\in\mathcal{C}} l_k$. Therefore, the total remaining storage is less than μ. That is, $\sum_{q=1}^{J+1} l_{k_q} \geq L - \mu$. Thus, for $\mathbf{B}_{J+1}$, we have

$$\begin{aligned}U(\mathbf{B}_{J+1}) &\geq \left(1 - \prod_{q=1}^{J+1}\left(1 - \frac{l_{k_q}}{L}\right)\right)U(\mathbf{OPT})\\ &\geq \left(1 - \left(1 - \frac{1}{L(J+1)}\sum_{q=1}^{J+1} l_{k_q}\right)^{J+1}\right)U(\mathbf{OPT})\\ &\geq \left(1 - e^{-\frac{L-\mu}{L}}\right)U(\mathbf{OPT}).\end{aligned} \tag{19}$$

Next, let us consider the policy (14). Assume that the algorithm ends at the step M. We can show that $U(\mathbf{A}_M) \geq U(\mathbf{A}_1) \geq U((s_{J+1}, k_{J+1})) \geq U(\mathbf{B}_{J+1}) - U(\mathbf{B}_J)$, that is,

$$U(\mathbf{A}_M) + U(\mathbf{B}_J) \geq U(\mathbf{B}_{J+1}) \geq \left(1 - e^{-\frac{L-\mu}{L}}\right)U(\mathbf{OPT}). \tag{20}$$

Now, we have

$$U(\mathbf{OPT}^*) = \max\left\{U(\mathbf{A}_M), U(\mathbf{B}_J)\right\} \geq \frac{1}{2}\left(1 - e^{-\frac{L-\mu}{L}}\right)U(\mathbf{OPT}),$$

which completes the proof of the Theorem. □

4.2.2 Approximation Algorithm

For the offloading scenario with short data lifetimes, the objective function (6) may be approximated as

$$\begin{aligned}U(\mathbf{X}) &\approx \sum_{k\in\mathcal{C}} l_k \sum_{i\in\mathcal{N}} w_{i,k}T_k \sum_{s\in\mathcal{H}_k}\gamma_{s,i}\\ &= \sum_{s\in\mathcal{H}}\sum_{k\in\mathcal{C}} x_{s,k}\left(l_kT_k\sum_{i\in\mathcal{N}}\gamma_{s,i}w_{i,k}\right),\end{aligned} \tag{21}$$

by using the first order Taylor expansion $e^x \approx 1+x$. We note that the above approximation is valid when

$$T_k \ll \frac{1}{w_{i,k}\sum_{s\in\mathcal{H}} x_{s,k}\gamma_{s,i}}, \forall k\in\mathcal{C}, s\in\mathcal{H}, i\in\mathcal{N}. \tag{22}$$

That is, we consider the scenario of short data lifetimes. By substituting the approximate system utility (21) into the problem (7) and further noting that the constraints in (7) are independent of s, each helper can determine their own policy for data items.

Specifically, let us define

$$U(\mathbf{X}; s) = \sum_{k\in\mathcal{C}} x_{s,k}\left(l_kT_k\sum_{i\in\mathcal{N}}\left(\gamma_{s,i}w_{i,k}\right)\right), \ s\in\mathcal{H}. \tag{23}$$

Then, we have

$$U(\mathbf{X}) = \sum_{s\in\mathcal{H}} U(\mathbf{X}; s). \tag{24}$$

Clearly, the system utility is given as a summation of the individual utilities of all the helpers. Furthermore, the buffer constraints in (7) are independent for individual helpers. Consequently, we can decompose the problem (7) into the following sub-problems

$$\begin{aligned}&\max\ U(\mathbf{X}; s),\\ &\text{s.t.}\quad x_{s,k}\in\{0,1\},\ \forall k\in\mathcal{C},\ \text{and}\ \sum_{k\in\mathcal{C}} l_kx_{s,k} \leq L_s,\end{aligned} \tag{25}$$

for each $s\in\mathcal{H}$ separately. Therefore, the overall approximately optimal solution is obtained by combining all the

Algorithm 2 Approximation Algorithm (AA) for multiple mobile data offloading under the scenario (22) of short lifetimes.

```
1: Initialize $\mathbf{OPT}^* = \emptyset$, $U(\mathbf{OPT}^*) = 0$. Initialize the approximation parameter $\varepsilon$ as required.
2: for every helper $s$ that $s \in \mathcal{H}$ do
3:     Compute the weight for each data item, denoted by $p_k$, $p_k = l_k T_k \sum_{i \in \mathcal{N}} (\gamma_{s,i} w_{i,k}), \forall k \in \mathcal{C}$
4:     Compute the quantisation precision $r = \left\lfloor \log\left( \frac{\varepsilon \cdot \max_{k \in \mathcal{C}}\{p_k\}}{C} \right) \right\rfloor$
5:     Update the weights for all data items, denoted by $p'_k = \left\lfloor \frac{p_k}{10^r} \right\rfloor, \forall k \in \mathcal{C}$
6:     Initialize $\mathbf{T} = \emptyset$, $P_0 = 0$, $\mathbf{Q}_0 = \{(\mathbf{T}, P_0)\}$
7:     for $\forall k \in \mathcal{C}$, $i = 1{:}C$ do
8:         Obtain $\mathbf{Q}_i = \mathbf{Q}_{i-1} \cup \{(\mathbf{T} \cup \{k\}, P_{i-1} + p'_k) | \sum_{j \in \mathbf{T}} l_j + l_k \leq L_s$ and $(\mathbf{T}, P_{i-1}) \in \mathbf{Q}_{i-1}\}$
9:         if $\exists\ (\mathbf{T}^1, P^1), (\mathbf{T}^2, P^2) \in \mathbf{Q}_i$, satisfies $P^1 = P^2$ and $\sum_{j \in \mathbf{T}^1} l_j > \sum_{j \in \mathbf{T}^2} l_j$ then
10:            Update $\mathbf{Q}_i = \mathbf{Q}_i \backslash \{(\mathbf{T}^1, P^1)\}$
11:        end if
12:    end for
13:    Select $\{(\mathbf{T}^*, P^*)\} \in \mathbf{Q}_C$, where $P^* = \max P \in \{(\mathbf{T}, P)\} \in \mathbf{Q}_C$
14:    Update $x_{s,k} = 1, \forall k \in \mathbf{T}^*$
15:    $\mathbf{OPT}^* = \mathbf{OPT}^* \cup \{(s,k) | x_{s,k} = 1, \forall k \in \mathcal{C}\}$
16: end for
17: Return $\mathbf{OPT}^*$ and $U(\mathbf{OPT}^*)$
```

optimal solutions for all the sub-problems (25), $\forall s \in \mathcal{H}$. For each $s \in \mathcal{H}$, the sub-problem (25) is a 0-1 knapsack problem. Thus, we can exploit an established method for solving 0-1 knapsack problems, known as the Fully Polynomial-Time Approximation Scheme (FPTAS) [30], which obtains a $(1 - \varepsilon)$ approximation solution to each sub-problem of (25) with the computational complexity of $O(C^4/\varepsilon)$, given any positive value ε. We present our Approximation Algorithm (AA) in Algorithm 2 which has the complexity of $O(HNC + HC^4/\varepsilon)$ in solving the problem (7) by approximation. In Algorithm 2, $\mathbf{OPT}^*$ is the solution obtained, $\lfloor \bullet \rfloor$ denotes the integer floor operator, and ε is the given algorithm precision. We loop for all the helpers beginning from Line 2. For each helper, we first prepare and initialize the parameters in Lines 3 to 6, and then use the FPTAS algorithm from Lines 7 to 12 to solve each sub-problem of (25). Finally, we combine the obtained results of the individual sub-problems as the final solution.

4.2.3 Homogeneous Algorithm

Obtaining an optimal solution for the SFM problem with MLCs is extremely hard, and we have so far presented two suboptimal algorithms to solve the problem (7). However, we can obtain a stronger result for the special scenario of homogeneous content sizes and contact rates, in which all node pairs meet with each other at the same contact rate and all the content sizes are identical. Under such homogeneous settings, the system offloading utility depends only on the number of replicas in all the helpers for each data item, but not on the actual subset of nodes that carry it in buffers. We now present our third algorithm, which is referred to as a Homogeneous Algorithm (HA), as it is for the special offloading scenario with homogeneous contact rates and data items.

Under the homogeneous setting considered, the contact rates of all node pairs are the same, which is denoted by γ_0, and all the content sizes are identical to l. Therefore, we can use the popularity of data items, the data lifetimes and the helpers' buffer sizes alone to obtain the system utility function. Specifically, we can rewrite the system utility function as

$$U(\mathbf{X}) = \sum_{k \in \mathcal{C}} l \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k} T_k \gamma_0 \sum_{s \in \mathcal{H}} x_{s,k}} \right). \tag{26}$$

Let us define the individual utility F_k for $k \in \mathcal{C}$ as

$$F_k\left(\sum_{s \in \mathcal{H}} x_{s,k} \right) = \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k} T_k \gamma_0 \sum_{s \in \mathcal{H}} x_{s,k}} \right). \tag{27}$$

Under the special scenario of homogeneous content size and contact rate, the problem (7) is reduced to:

$$\begin{aligned} &\max \sum_{k \in \mathcal{C}} l F_k\Big(\sum_{s \in \mathcal{H}} x_{s,k} \Big), \\ &\text{s.t. } x_{s,k} \in \{0,1\} \text{ and} \\ &\quad \sum_{k \in \mathcal{C}} x_{s,k} l \leq L_s, \ \forall s \in \mathcal{H}, \forall k \in \mathcal{C}. \end{aligned} \tag{28}$$

We are now ready to describe our HA for solving the problem (28), summarized in Algorithm 3 where

$$u_k = \sum_{s \in \mathcal{H}} x_{s,k}, \ k \in \mathcal{C}. \tag{29}$$

The algorithm repeatedly chooses items for the helpers to store, and in each step, it attempts to select one item that results in the maximum system utility for a helper who still has available storage. Specifically, in Lines 1 and 2 of the algorithm, the related parameters are initialized. From Lines 3 to 5, the algorithm computes the incremental utility. Then from Lines 6 to 14, it repeatedly chooses items that results in the maximum system utility for a helper which still has available storage until all the buffers are full. Therefore, this algorithm is likely to allocate the items to those helpers to store whose corresponding utility gains are larger than other helpers. In fact, Algorithm 3 provides the optimal solution for the problem (28), and this is proved in the following.

With the definition of u_k given in (29), the optimization problem (28) can be rewritten as

$$\begin{aligned} &\max \sum_{k \in \mathcal{C}} l F_k(u_k), \\ &\text{s.t. } \sum_{k \in \mathcal{C}} u_k \leq \sum_{s \in \mathcal{H}} \lfloor L_s / l \rfloor. \end{aligned} \tag{30}$$

Noting that in $F_k(u_k)$ we have $0 \leq u_k \leq H$, we can define $F_k(i) = F_k(u_k = i)$ for $0 \leq i \leq H$. Further define

$$\Delta_k(i) = l\big(F_k(i+1) - F_k(i)\big), 0 \leq i \leq H-1, k \in \mathcal{C}. \tag{31}$$

Sort $\Delta_k(i)$ for $0 \leq i \leq H-1$ and $k \in \mathcal{C}$ in the order

$$\widehat{\Delta}_1 \geq \widehat{\Delta}_2 \geq \widehat{\Delta}_3 \geq \cdots \geq \widehat{\Delta}_{HC}, \tag{32}$$

Algorithm 3 Homogeneous Algorithm (HA) for multiple mobile data offloading under the homogeneous scenario.

```
Set $x_{s,k} = 0$, $u_k = 0$, $\triangle F_k = 0$, $k \in \mathcal{C}$, $s \in \mathcal{H}$;
Initialize set $\mathcal{I}_\mathcal{C} = \{1, 2, \cdots, C\}$ and sum $= 0$;
for every data item $k$ that $k \in \mathcal{I}_\mathcal{C}$ do
    $\triangle F_k \leftarrow l\,(F_k\,(u_k + 1) - F_k\,(u_k))$;
end for
while sum $\le \sum_{s \in \mathcal{H}} \lfloor L_s/l \rfloor$ and $\mathcal{I}_\mathcal{C} \neq \emptyset$ do
    Select $i = \arg\max_k \{\triangle F_k | k \in \mathcal{C}\}$;
    Select $q = \arg\max_s \left\{\lfloor L_s/l \rfloor - \sum_{k \in \mathcal{C}} x_{s,k} | x_{s,i} = 1,\ s \in \mathcal{H}\right\}$;
    Set $x_{q,i} = 1$;
    Update $u_i \leftarrow u_i + 1$, sum $\leftarrow$ sum $+ 1$;
    Update $\triangle F_i \leftarrow l\,(F_i\,(u_i + 1) - F_i\,(u_i))$;
    if $u_i = H$ then
        $\mathcal{I}_\mathcal{C} \leftarrow \mathcal{I}_\mathcal{C} \backslash \{i\}$;
    end if
end while
```

where $\widehat{\Delta}_1$ and $\widehat{\Delta}_2$ are the largest and the second largest of all the $\Delta_k(i)$, respectively, and so on. At each greedy step of maximizing the system utility in Algorithm 3, we calculate

$$\Delta F_k = l\big(F_k(u_k + 1) - F_k(u_k)\big) = \Delta_k(u_k), k \in \mathcal{C}, \tag{33}$$

choose $i = \arg\max_k \{\Delta F_k \mid k \in \mathcal{C}\}$, update $u_i \leftarrow u_i + 1$ and add ΔF_i to the objective function. Now denote

$$\widehat{\delta}_j = \max_k \{\Delta F_k \mid k \in \mathcal{C}, \text{at } j\text{th step of Algorithm 3}\}. \tag{34}$$

For notational simplicity, we also denote $\Delta_k(i)$s that are related to $\widehat{\Delta}_j$ and $\widehat{\delta}_j$, respectively, by

$$\widehat{\Delta}_j = \Delta_{K_1(j)}(U_1(j)), \tag{35}$$

$$\widehat{\delta}_j = \Delta_{K_2(j)}(u_{K_2(j)}), \tag{36}$$

where $K_1(j)$, $U_1(j)$, and $K_2(j)$ are the corresponding indexes.

Lemma 1. *The set of $\{\widehat{\delta}_j\}$ obtained by Algorithm 3 matches the set of $\{\widehat{\Delta}_j\}$, namely, $\widehat{\delta}_j = \widehat{\Delta}_j$, $\forall j$. That is, the increment of the objective function obtained at the jth step of Algorithm 3 is exactly the jth largest element among all the $\Delta_k(i)$, $0 \le i \le H - 1, k \in \mathcal{C}$.*

Proof. First, we prove $\widehat{\delta}_j \ge \widehat{\delta}_{j+1}$. If $K_2(j) = K_2(j+1)$, it can be deduced that $u_{K_2(j+1)} = u_{K_2(j)} + 1$ since the update of $u_{K_2(j)}$ is performed at the jth step. From the concavity of $F_k(u_k)$, we know $F_k(u_k+1) - F_k(u_k) \ge F_k(u_k+2) - F_k(u_k+1)$, and thus $\Delta_k(u_k) \ge \Delta_k(u_k + 1)$. Substituting k and u_k with $K_2(j)$ and $u_{K_2(j)}$, respectively, we have $\widehat{\delta}_j \ge \widehat{\delta}_{j+1}$. If $K_2(j) \neq K_2(j+1)$, we know that $u_{K_2(j+1)}$ remains the same between the jth and $(j+1)$th steps of Algorithm 3. Thus $\Delta F_{K_2(j+1)}$ is the same for the jth and $(j+1)$th steps. Since $K_2(j) = \arg\max_k \{\Delta F_k \mid k \in \mathcal{C}, \text{at the } j\text{th step}\}$, we know that $\Delta F_{K_2(j)}$ is no less than $\Delta F_{K_2(j+1)}$ at the jth step. Consequently,

$$\widehat{\delta}_j = \Delta F_{K_2(j)} \ge \Delta F_{K_2(j+1)} = \widehat{\delta}_{j+1}. \tag{37}$$

This proves that

$$\widehat{\delta}_1 \ge \widehat{\delta}_2 \ge \cdots \ge \widehat{\delta}_n \ge \widehat{\delta}_{n+1} \ge \cdots. \tag{38}$$

From the definitions of $\widehat{\delta}_j$ and $\widehat{\Delta}_j$, it is obvious that $\widehat{\delta}_j \le \widehat{\Delta}_j$. We obtain $\widehat{\delta}_j = \widehat{\Delta}_j$ by proving the contradiction otherwise. If it is not true, we can denote

$$j_0 = \min\{j \mid \widehat{\delta}_j \neq \widehat{\Delta}_j\}. \tag{39}$$

Since $\widehat{\delta}_j \le \widehat{\Delta}_j$, it must be $\widehat{\delta}_{j_0} < \widehat{\Delta}_{j_0}$. Without loss of generality, we can assume that if $\widehat{\Delta}_j = \widehat{\Delta}_{j+1}$, the indexes will satisfy $K_1(j) < K_1(j+1)$ or $K_1(j) = K_1(j+1)$ and $U_1(j) < U_1(j+1)$. Furthermore, we have $\widehat{\Delta}_{j_0} = \Delta_{K_1(j_0)}(U_1(j_0))$. From our proof in the previous paragraph, it can be attained that $\Delta_{K_1(j_0)}(q) \ge \Delta_{K_1(j_0)}(U_1(j_0))$ for all $q < U_1(j_0)$. From the above assumption on indexes and the definition of j_0 in (39), we know that $\Delta_{K_1(j_0)}(q)$, $1 \le q < U_1(j_0)$ all appear in $\widehat{\Delta}_j$, $j < j_0$, and thus they appear in $\widehat{\delta}_j$, $j < j_0$. Now considering the j_0th step of the algorithm, it can be deduced that $u_{K_1(j_0)} = U_1(j_0)$ from above analysis and, therefore, $K_1(j_0) = \arg\max_k \{\Delta F_k | k \in \mathcal{C}, \text{ at the } j_0\text{th step}\}$ because $\Delta F_{K_1(j_0)} = \widehat{\Delta}_{j_0}$ and $\widehat{\Delta}_{j_0} \ge \widehat{\Delta}_{j'}$ when $j' > j_0$. Consequently, from the algorithm we have $\widehat{\delta}_{j_0} = \widehat{\Delta}_{j_0}$, which leads a contradiction.

This completes the proof of Lemma 1. □

We are ready to provide the following theorem.

Theorem 3. *The optimal solution of the problem (28) is obtained by Algorithm 3.*

Proof. From the definition of $\Delta_k(i)$, the objective function becomes

$$\sum_{k \in \mathcal{C}} lF_k(u_k) = \Big(\sum_{k \in \mathcal{C}} \sum_{i=0}^{u_k - 1} \Delta_k(i)\Big) + \sum_{k \in \mathcal{C}} lF_k(0), \quad \text{subject to } \sum_{k \in \mathcal{C}} u_k \le Q, \tag{40}$$

where Q is a constant. Since $\sum_{k \in \mathcal{C}} lF_k(0)$ is a constant and $\Big(\sum_{k \in \mathcal{C}} \sum_{i=0}^{u_k - 1} \Delta_k(i)\Big)$ is the sum of at most HC terms of $\Delta_k(i)$s, it can be deduced that

$$\Big(\sum_{k \in \mathcal{C}} \sum_{i=0}^{u_k - 1} \Delta_k(i)\Big) \le \sum_{j=1}^{HC} \widehat{\Delta}_j = \sum_{j=1}^{HC} \widehat{\delta}_j, \tag{41}$$

according to the definition of $\widehat{\Delta}_j$ and Lemma 1. As the objective function cannot exceed the value obtained by Algorithm 3, the solution obtained by Algorithm 3 is the optimal solution of the problem (28). □

The computational complexity of this HA can be shown to be $O\Big(C + \sum_{s \in \mathcal{H}} \lfloor L_s/l \rfloor \log(C)\Big)$. Like the GA and AA schemes, it is also a centralized algorithm as it requires the content-related parameters T_k and the network-related parameters $w_{i,k}$ and L_s.

5 Performance Evaluation

We present the extensive simulation results obtained for our three algorithms: i) **Greedy Algorithm** (GA) shown in Algorithm 1, ii) **Approximation Algorithm** (AA) shown in Algorithm 2, and iii) **Homogeneous Algorithm** (HA) shown in Algorithm 3. We note that these three algorithms are designed for different scenarios of multiple mobile data offloading. Therefore, we first used Poisson process to generate different traces for these three algorithms to verify their efficiency. Specifically, we would like to confirm that the GA scheme can achieve good performance under the generic multiple mobile data offloading scenario, and the AA scheme can provide good performance for the scenario of small data lifetimes, while the HA scheme indeed offers the optimal solution for the scenario of homogeneous content sizes and contact rates.

Furthermore, we used mobility models as well as realistic human and vehicular mobility traces to compare the performance of our GA, AA and HA schemes with three other existing schemes:

1) **SFM Algorithm** (SFM) [26], which uses some approximation algorithms to maximise a submodular set function subject to MLCs. This algorithm is the best existing scheme to solve the optimization problem (7) in terms of performance, which yields a $(1 - 1/e - \Theta(\varepsilon))$ approximation solution, but it has a very high computational complexity of $O(H! + C!)$.

2) **Random Selection** (RS), in which each helper chooses the data items randomly to fill its buffer until no more items can be stored.

3) **Equal Selection** (ES), where the system allocates the buffer for each data equally. That is, each data is assumed to have the same priority to occupy the buffers of all the helpers.

It is worth emphasising again that our main contribution is to consider the realistic heterogeneous conditions of the DTN-based offloading system, which were not taken into account in the previous works [2], [5], [8], and the benchmarks adopted reflect the whole spectrum of the currently available techniques for multiple mobile data offloading in DTNs. Specifically, the RS scheme is extremely simple and can be implemented distributively. The SFM scheme [26] offers the best existing solution in terms of performance at the cost of very high computational complexity. It is also a centralized algorithm when applying to our DTN based multiple mobile data offloading system, requiring the exactly the same set of the content-related and network-related parameters as our GA and AA schemes. We further point out that the most widely approach for DTN-based offloading systems adopts a homogeneous network environment and, therefore, our HA scheme, which allocates the buffer based on the assumption of homogeneous contact rates and content lengths, represents the most widely used scheme for DTN-based offloading.

5.1 Mobility Traces and Simulation Settings

In the simulation, we used the first half of the mobility traces to estimate the initial contact rates for all the node pairs, and then continuously updated them during the data dissemination process. Among all the simulated nodes, we randomly chose 10% as the helpers and the rest as the subscribers. For the AA scheme, the algorithm precision ε was set to 0.2.

For the purpose of comparing our three algorithms, we simulated the network with the size of 200 nodes and the number of data items $C = 10$, where the Poisson process with the parameter $\gamma = 0.01\,\mathrm{s}^{-1}$ was used to generate node contact events. The contact rates were homogeneous in this case. The data lifetimes followed the uniform distribution in $[0,\ 2T_a]$ s, where T_a was the average data lifetime. The helpers' buffer sizes were randomly and uniformly generated in $[0,\ 2l_a]$ MB, where l_a was the average buffer size. Three different settings of content sizes were simulated. In *Setting 1*, all the message sizes were equal to 100 MB, and in *Setting 2*, the messages were randomly and uniformly generated in [50, 150] MB, while in *Setting 3*, the message sizes followed the uniform distribution of [0, 200] MB. Note that *Setting 1* simulated the scenario of homogeneous content sizes.

To demonstrate the efficiency of our proposed GA and AA schemes in realistic mobility environments, we used the synthetic mobility model, Random WayPoint (RWP), and real-world MAP-driven DTN model (MAP) with pedestrians and transportation systems, in the Opportunistic Network Environment (ONE) simulator [31]. RWP is a mobility model while MAP in the ONE simulator represents the realistic DTN environments, where the contact rate is not governed by Poisson process. In the MAP mobility, the nodes moved along the streets on an imported map of downtown Helsinki, Finland. The size of the map was $4500 \times 3400\,\mathrm{m}^2$, and the nodes' transmission range was 20 m. About 15% of the nodes were configured to follow pre-defined routes with the speed uniformly distributed in the range of [1, 3] m/s, which imitated the regularity of certain type of mobility existing in human or vehicles, such as public buses that have very regularity mobility trajectories and patterns every day. Other nodes were divided into three groups. For each group, there were Points-Of-Interest (POIs), and we assigned a different probability for picking the next node from a particular group of POIs to simulate the phenomenon that people visit certain areas of a place more frequently than other areas based on features such as age and profession. The nodes' walking speed was chosen from a uniform distribution with the range of [0.8, 1.2] m/s. Other settings were given by the default values in the ONE simulator. We also used two real-world traces in the simulation. The first one was the human mobility contact trace of *Cambridge* gathered by *Haggle* Project [32], which recorded contacts among users carrying Bluetooth devices that periodically discover peers in communication range and record the contacts. The second one was the taxi GPS trace of *Shanghai* [33]. The mobility models and real-world traces used are listed in Table 2, where it can be seen that they cover diverse DTN environments, from sparse university campuses (*Cambridge*) to concentrated city road sites (*Shanghai*), with experimental periods ranging from 1 days (MAP, RWP) to 30 days (*Shanghai*). In the simulation, we varied the number of data items from $C = 200$ (for RWP, MAP and *Cambridge*) to $C = 350$ (for *Shanghai*), and the sizes of data items were randomly and uniformly generated in the range of [50, 150] MB. It is clear that the simulated network environments under these conditions were highly heterogeneous in terms of contact rates and content sizes.

As the subscribers' interests are highly heterogeneous in practice, we used the Zipf distribution to describe the interest distribution. The Zipf distribution is widely used in content population modeling [13], in which most of the subscribers' interests concentrate on the popular data. To generate the Zipf interest distribution of realistic data sets,

TABLE 2
Heterogeneous Mobility Model and Trace Summary

Trace	RWP	MAP	Cambridge	Shanghai
Network Type	The ONE Simulator	The ONE Simulator	Bluetooth	GPS
Device	Simulation Node	Simulation Node	iMote	GPS
Number of devices	200	200	54	2000
Duration (days)	1	1	11	30

we set the number of keywords to $M = 205$ for RWP, MAP and *Cambridge* while setting $M = 355$ for *Shanghai*, and we assumed that the keywords m_1 to m_M were ranked according to their popularity. For keyword m_j, we assumed that the average interest of all the subscribers was I_j, and used the Zipf distribution with exponent 2 to generate I_j, which is defined by $I_j = \frac{1/j^2}{\sum_{q=1}^{M} 1/q^2}$. The normal distribution was employed to generate the interest profile $\mathcal{P}_i$ for subscriber i. For each data item k, $k \in \{1, 2, \ldots, C\}$, its describing subset of keywords was given by $\mathcal{M}_k = \{m_k, m_{k+1}, \ldots, m_{k+4}\}$ with equal weighting $v_{m_j} = 1/5$ for each keyword m_j. Then, by using (1), we obtained the interest probability of subscriber $i \in \mathcal{N}$ in data item $k \in \mathcal{C}$.

To illustrate the property of the Zipf interest distribution, we plot the interest probability of each keyword and the average interest probability of all the subscribers on each data item in Fig. 2(a) and (b), respectively, for a simple case of 30 keywords and 10 data items, where the results of the uniform and exponential distributions are also shown for comparison. From the results depicted in Fig. 2, we observe that for the exponential distribution, most of the user interests concentrate on the popular data items, while for the Zipf distribution, the difference between high popular data and low popular ones is smaller than the case of exponential distribution. Thus, the Zipf interest probability distribution lies between the uniform and exponential distributions. This type of user interest is also demonstrated to fit the realistic mobile data downloading applications well [34].

In a most generic DTN based content dissemination system, the most important performance metrics include the amount of the contents disseminated and the content dissemination delay. For our mobile data offloading system, however, the content dissemination delay becomes a secondary issue. In our investigation, we consider the delay as the system constraint by imposing the requirement of delivering each mobile data item k before its deadline T_k. Therefore, we mainly focus on the amount of the offloaded data items, which is clearly the optimization goal in our problem (7). Although the mobile data for offloading do not have strict real-time requirements and can tolerate a certain amount of delay, the latency in the offloading does affect the QoS. It would be highly desired to simultaneously maximize the total amount of the offloaded data items and minimize the expected offloading delay. However, this is an open multi-objective optimization. In our simulation, we will also evaluate the proposed algorithms by observing the resulting offloading delays. Besides the amount of the offloaded data and the data offloading delay, the data offloading ratio also influence the QoS. The data offloading ratio, defined as

$$\mathrm{OR} = \sum_{\forall i \in \mathcal{N}, \forall k \in \mathcal{C}} w_{i,k} p_{i,k}, \tag{42}$$

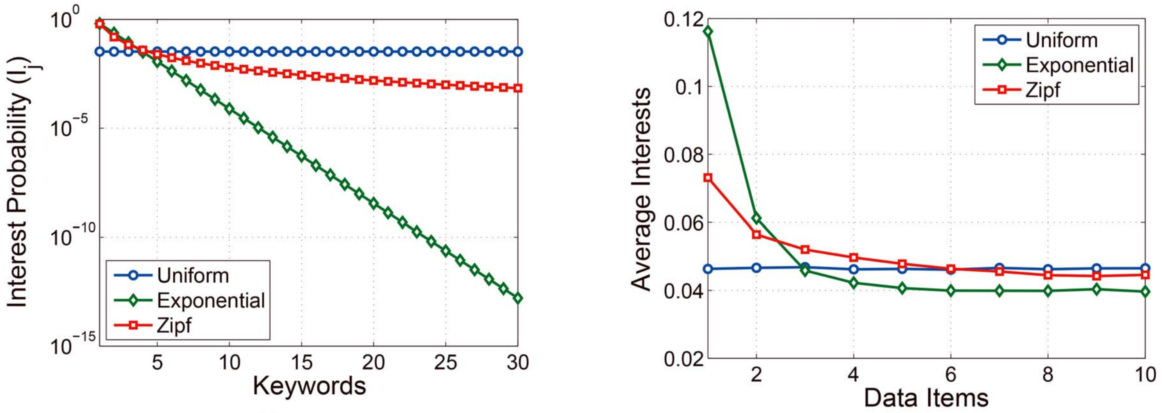
where

$$p_{i,k} = \begin{cases} 1, \text{ subscriber } i \text{ receives data } k \text{ before } T_k, \\ 0, \text{ otherwise}, \end{cases} \tag{43}$$

specifies the average probability of all the subscribers receiving their requested data. We also evaluated this data offloading ratio in the simulation.

5.2 Comparison of the Proposed GA, AA and HA

We first show the results obtained by our three algorithms, the GA, AA and HA, under the network simulated with the homogeneous contact rate and three different settings of content sizes in Figs. 3 to 5, respectively. The amount of offloaded data achieved by each algorithm was obtained by simulating the system with the buffer allocation strategy of each algorithm, which was observed to agree with



(a) (b)

Fig. 2. (a) Interest probability of each keyword. (b) Average interest probability of each data item, where the number of keywords $M = 30$ and the number of data items $C = 10$.

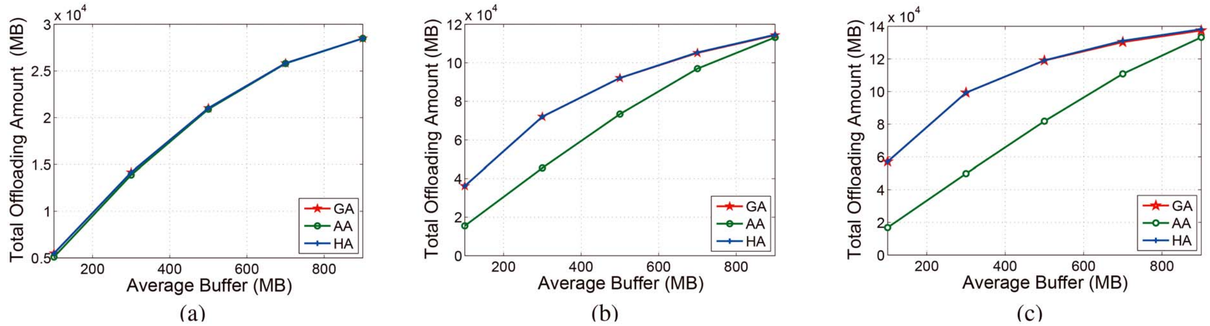


Fig. 3. Performance comparison of three proposed algorithms in terms of total amount of offloaded data for the simulated network of homogeneous contact rate under *Setting 1* of content sizes with variable average buffer size l_a and average data lifetime. (a) $T_a = 10$ s. (b) $T_a = 100$ s. (c) $T_a = 200$ s.

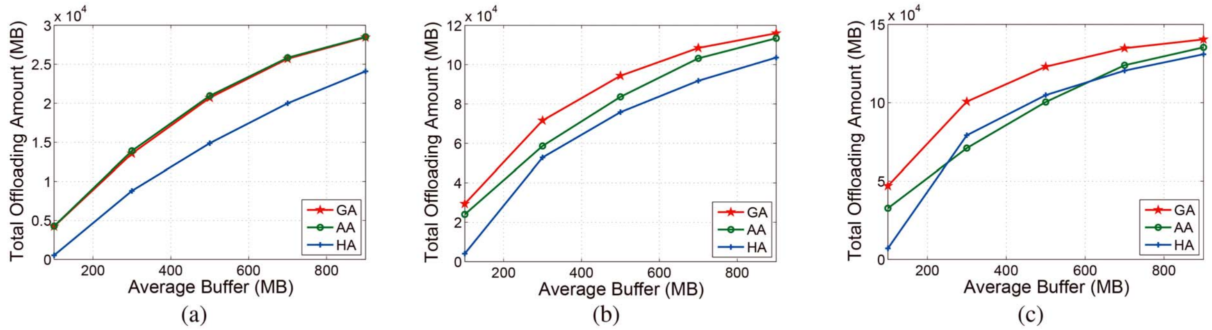


Fig. 4. Performance comparison of three proposed algorithms in terms of total amount of offloaded data for the simulated network of homogeneous contact rate under *Setting 2* of content sizes with variable average buffer size l_a and average data lifetime. (a) $T_a = 10$ s. (b) $T_a = 100$ s. (c) $T_a = 200$ s.

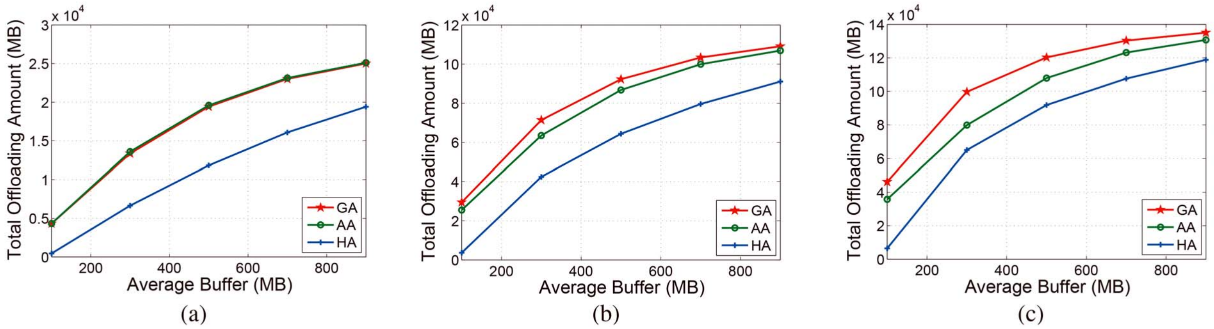


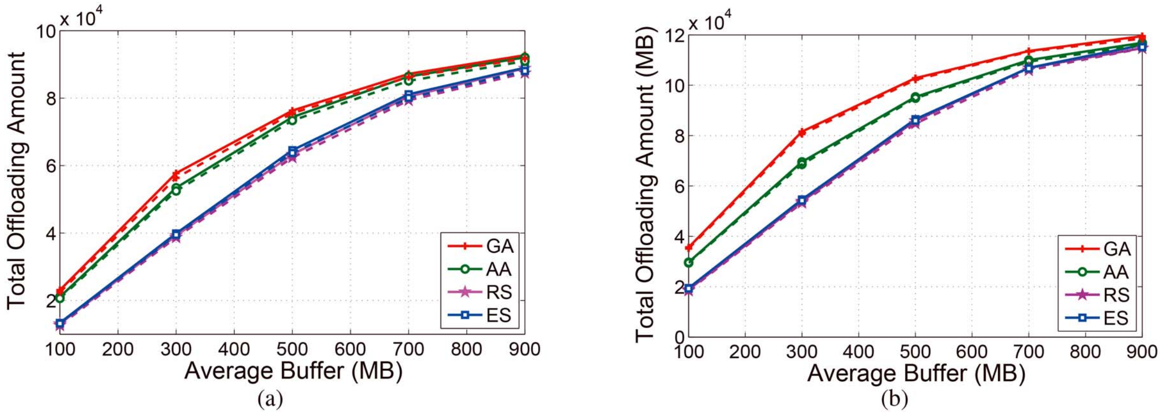
Fig. 5. Performance comparison of three proposed algorithms in terms of total amount of offloaded data for the simulated network of of homogeneous contact rate under *Setting 3* of content sizes with variable average buffer size l_a and average data lifetime. (a) $T_a = 10$ s. (b) $T_a = 100$ s. (c) $T_a = 200$ s.

the theoretical calculation using (6) as expected. The results under *Setting 1* depicted in Fig. 3 show that the HA and GA achieved the same best performance for different average data lifetimes T_a. Under *Setting 1*, the contact rates are homogeneous and so are the content sizes, and the HA attains the optimal performance which confirms Theorem 3. With small data lifetimes of no more than $2l_a = 20\,\text{s}$, the AA is observed to achieve the same optimal performance as that of the HA and GA in Fig. 3(a). This confirms that the AA scheme is capable of achieving good performance under the scenario of small data lifetimes. With the increase of data lifetimes, the performance of the AA deteriorates considerably, as can be seen in Fig. 3(b) and (c).

With *Setting 2* and *Setting 3*, the system becomes increasingly heterogeneous in terms of data sizes, and the performance of the HA degrades significantly. For example, the GA outperforms the HA by about 36% in Fig. 4(b) and by about 54% in Fig. 5(b). Similar to the *Setting 1* case, the AA obtains the same performance as the GA when the data lifetimes are short, but its performance degrades with the increase of data lifetimes, as can be seen in Figs. 4 and 5. From the results of Figs. 3 to 5, we can draw the conclusion that the GA scheme always achieves the best result among the three designed schemes. However, the computational complexity of the AA scheme is lower than that of the GA scheme. Therefore, for the scenario of short data lifetimes, the AA offers a viable choice. Although the HA scheme has the lowest computational complex, its application is limited to homogeneous systems in terms of content sizes and contact rates.

5.3 Amount of Offloaded Data

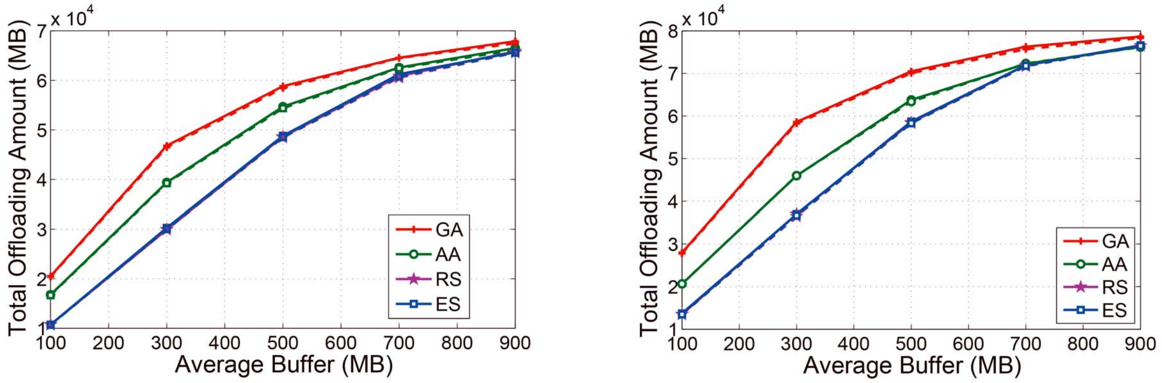
The performance comparison of our GA and AA schemes with the RS and ES benchmarks, in terms of amount



(a)

(b)

Fig. 6. Performance comparison of four algorithms in terms of total amount of offloaded data for the RWP model with the variable average buffer size, the fixed average data lifetime and the average data lifetime. (a) 110000 s. (b) 210000 s, where the dashed and solid curves indicate the theoretical and simulation results, respectively.



(a)

(b)

Fig. 7. Performance comparison of four algorithms in terms of total amount of offloaded data for the MAP model with the variable average buffer size, the fixed average data lifetime and the average data lifetime. (a) 120000 s. (b) 200000 s, where the dashed and solid curves indicate the theoretical and simulation results, respectively.

of offloaded data under the RWP network simulation, is shown in Fig. 6, where the dashed curves indicate the theoretical amount of offloaded data calculated by each algorithm using the utility function (6) while the solid curves are the real system performance obtained by simulating the system with the buffer allocation strategy of each algorithm. It is clearly seen from Fig. 6 that the simulated real-system performance agrees with the theoretical calculation, which validates the accuracy of our formulated problem. From Fig. 6, we observe that our GA achieves the best performance among the four algorithms compared, and it outperforms the RS and ES algorithms considerably. Comparing Fig. 6(a) with (b), we again observe that the performance of the AA scheme degrades as the data lifetimes increase.

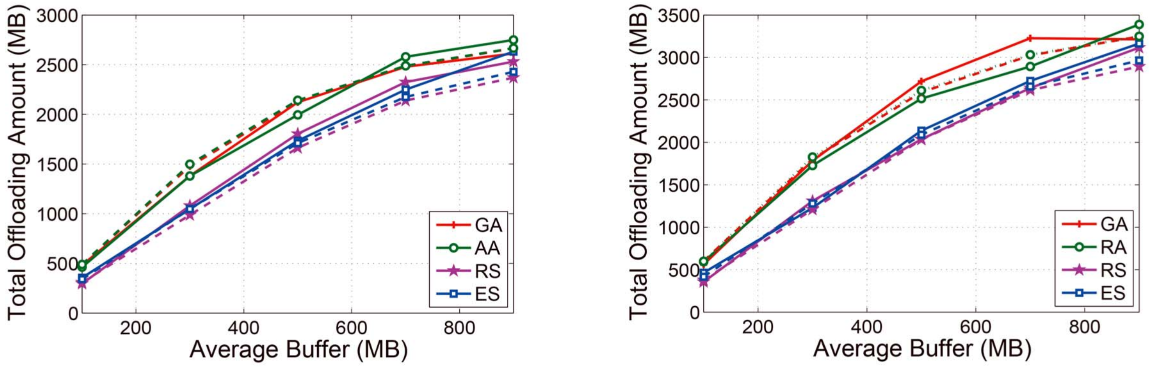
The results obtained under the MAP network simulation are shown in Fig. 7. Again, the simulation results agree with the theoretical results, and similar observations to those drawn for the RWP results can be made. For example, increasing the helper's buffer size leads to a significant increase in the total amount of offloaded items as expected. Our GA scheme achieves the best performance, and the RS and ES algorithms have the worst performance. Also, the data lifetime has a significant influence on the achievable performance of the AA scheme.

The results obtained using the *Cambridge* human mobility trace are shown in Fig. 8. Unlike the cases of the RWP and MAP models, we note that there exists a slight derivation between the simulation and theoretical results, as the *Cambridge* trace is a real-world human mobility trace. In this realistic experiment, the GA and AA achieve almost the same performance, and they outperform the ES and RS considerably.

The results obtained using the *Shanghai* vehicular mobility trace are depicted in Fig. 9. We observe that the AA achieves almost the same performance as the GA does, even when data lifetimes are large, as in the case of Fig. 9(b). The reason is that the contact rates of vehicular traces are much smaller than human traces. The results of Fig. 9 also show that our GA and AA considerably outperform the RS and ES.

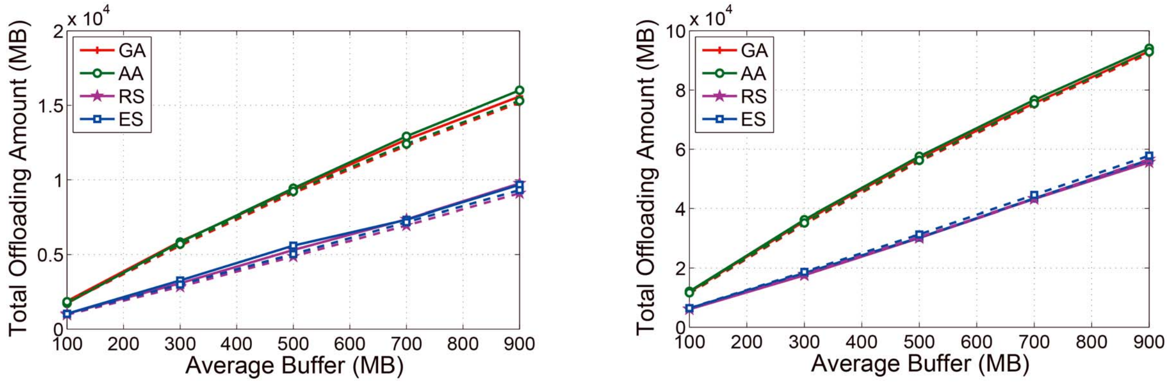
5.4 Data Offloading Latency and Ratio

We used the *Shanghai* trace to evaluate the achievable mobile data offloading delay and the data offloading ratio by the GA and HA schemes in comparison with the SFM and RS benchmarks. Fig. 10(a) shows the results of the average data offloading latency obtained by these four algorithms with the fixed average data lifetime of 10000 s and the variable average buffer size. Observe that the RS and



(a) (b)

Fig. 8. Performance comparison of four algorithms in terms of total amount of offloaded data for the *Cambridge* trace with the variable average buffer size, the fixed average data lifetime and the average data lifetime. (a) 800000 s. (b) 1000000 s, where the dashed and solid curves indicate the theoretical and simulation results, respectively.



(a) (b)

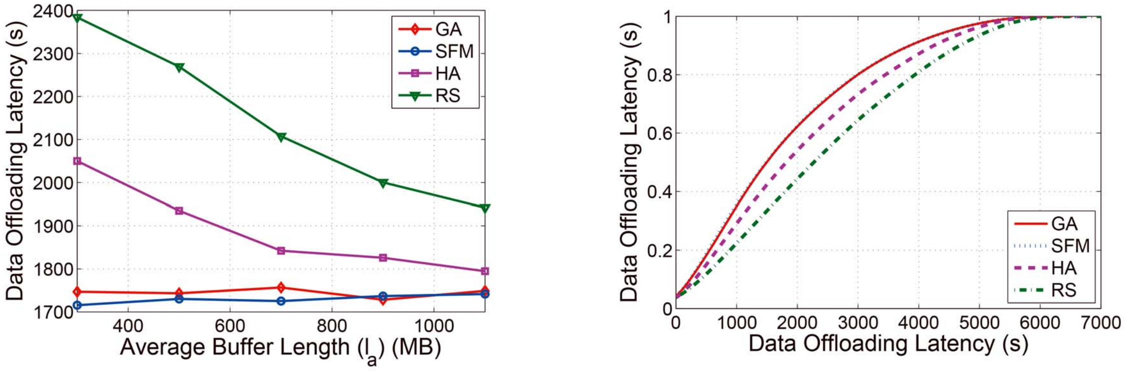
Fig. 9. Performance comparison of four algorithms in terms of total amount of offloaded data for the *Shanghai* trace with the variable average buffer size, the fixed average data lifetime and the average data lifetime. (a) 3000 s. (b) 20000 s, where the dashed and solid curves indicate the theoretical and simulation results, respectively.

HA require much larger average data offloading delays than the GA and SFM. Compared with the SFM algorithm, our GA needs a slightly larger offloading delay than the SFM algorithm when the average buffer size is relatively small, but this difference disappears when the average buffer size is sufficient large. Thus, under the network condition of relatively abundant buffer resources, our GA can obtain the same offloading latency performance as the SFM algorithm, while benefiting from a significant saving in computational complexity.

To provide the distribution of the average data offloading latency, we set the average data lifetime to 10000 s as well as the average buffer length to 500 MB, and Fig. 10(b) plots the Cumulative Distribution Function (CDF) of the data offloading latency obtained by each algorithm in the simulation. Observe from Fig. 10(b) that the SFM and our GA schemes have practically the same CDF, under the simulated network senario. For both the SFM and GA schemes, 80% of the offloading latency is under 3000 s, while for the HA and RS algorithms, they are about 3500 s and 4000 *s*, respectively. The 80% of the offloading latency achieved by each algorithm is much smaller than the imposed average data lifetime.

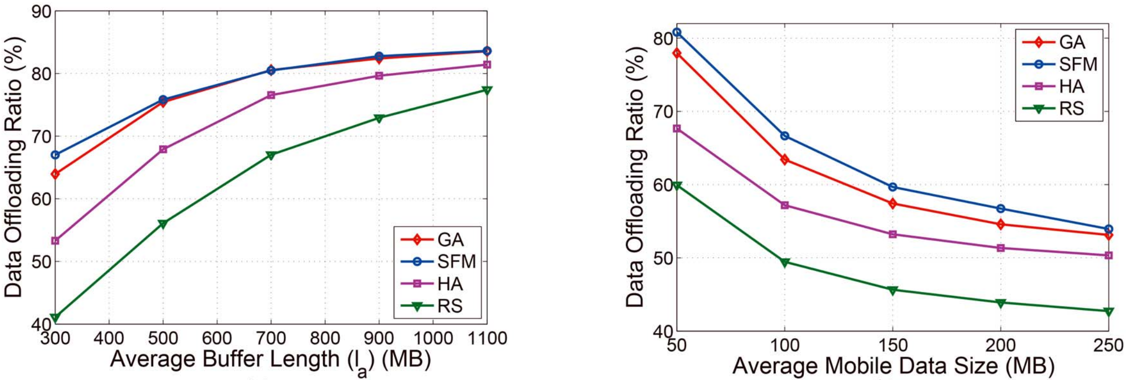
We next investigated the achievable average probability of all the subscribers receiving their interested data before deadlines, i.e. the data offloading ratio defined in (42). Fig. 11(a) depicts the data offloading ratios obtained in the simulation by the four algorithms, the GA, HA, SFM and RS, with the fixed average data lifetime of 3000 s and the variable average buffer size. Observe from Fig. 11(a) that compared with the SFM scheme our GA scheme is slightly inferior when the average buffer size $l_a < 500\,\text{MB}$, but both the algorithms attain the identical offloading ratio for $l_a \geq 500\,\text{MB}$. Moreover, the SFM and GA attain much higher data offloading ratio than the HA and RS. It can also be seen from Fig. 11(a) that our HA scheme outperforms the RS benchmark considerably.

In all the above simulation experiments, the sizes of the mobile data were randomly and uniformly generated in [50, 150] MB. To further investigate the impact of data sizes on the achievable performance, we designed an additional experiment by letting the mobile data sizes to follow the uniform distribution in $\left[(d_s - 50), \quad (d_s + 50)\right]$ MB, while varying the "average mobile data size" d_s from 50 MB to 250 MB. This created various experimental conditions, where the ranges of the uniform distribution for the data sizes changed from as small as [0, 100] MB to as large as [200, 300] MB. Fig. 11(b) shows the data offloading ratios obtained by the four algorithms under the different average mobile data sizes d_s, with the average data lifetime set to 3000 s and the average buffer size to 500 MB. From Fig. 11(b), we can see that compared with the SFM our GA



(a) (b)

Fig. 10. Performance comparison of four algorithms in terms of average data offloading latency for the *Shanghai* trace. (a) Average data offloading latency given the fixed average data lifetime of 10000 s and the variable average buffer size. (b) Cumulative distribution function of data offloading latency given the average data lifetime of 10000 s and the average buffer length of 500 MB.



(a) (b)

Fig. 11. Performance comparison of four algorithms in terms of data offloading ratio for the *Shanghai* trace. (a) Data offloading ratio given the fixed average data lifetime of 3000 s and the variable average buffer size. (b) Data offloading ratio given the average data lifetime of 3000 s, the average buffer size of 500 MB and the sizes of mobile data randomly and uniformly generated in the range of $\left[(d_S - 50),\ (d_S + 50)\right]$ MB with the given average mobile data size of d_S.

scheme is slightly inferior. Both the SFM and GA schemes significantly outperform the HA and RS algorithms, while our HA achieves a much better performance than the RS scheme.

5.5 Discussions

From the above simulation study, we conclude that among the three designed algorithms the GA scheme achieves the best performance under the generic heterogeneous network setting, and the AA scheme obtains the equally good performance as the GA in the heterogeneous network setting with short data lifetimes, while the HA scheme is only optimal under the special network setting of homogeneous content sizes and contact rates. The results also demonstrate that our GA and AA schemes significantly outperform the simple RS and ES benchmark algorithms. Interestingly, the AA scheme is observed to achieve almost the same performance as the GA under the simulation experiments of real-world mobility traces. This is significant, as the AA has a lower computational complexity than the GA. Thus, the AA scheme does offer a viable alternative to the GA for real-world DTN-based mobile data offloading, particular in the vehicular based DTNs where the contact rates are small owing to fast moving vehicles.

Most significantly, our GA scheme achieves almost the same performance in comparison with the SFM algorithm[1]. The SFM algorithm is the most accurate algorithm currently available to solve the optimisation problem (7) of the heterogeneous mobile data offloading but it is not very practical due to its very high computational complexity. For example, in the above simulated scenarios, the average running time of the SFM algorithm is 0.53 hours, while our GA scheme only takes a few seconds to obtain the results. This clearly demonstrates the effectiveness of our proposed GA algorithm. In the light of these results, we offer the following review of the existing related works, to further highlight our novel contributions.

6 Related Works

The latest works [2]–[4], [35], [36] focus on offloading mobile data from overloaded cellular networks to other networks in order to provide better service. These works can be divided into three types according to their offloading

[1] We also evaluated the achievable performance by the SFM algorithm in terms of the amount of data offloaded under the same experimental conditions given in Section 5.3. We observe that the GA achieves almost the same performance as the SFM algorithm. For the purpose of graphic clarification, we do not include the performance of the SFM in Section 5.3.

destination networks. Broadcast offloading utilizes mobile broadcasting network to offload the cellular traffic, and the work [35] proposes an architecture called UNAP, which makes efficient use of the unicast 3G network while at the same time efficiently and reliably schedules content delivery over the mobile broadcast network. Another popular offloading method utilizes freely available WiFi networks [3], [4], [37], [38], which may be referred to as WiFi offloading. Specifically, the work [37] proposes a mechanism for using network diversity to improve the application-level aggregate bandwidth, which is termed *Intentional Networking*. By using the measurements obtained in a WiFi and cellular 3G network coverage area, it shows that the system latency is enhanced significantly. The problem of augmenting mobile 3G using WiFi in a moving vehicular environment is studied in [4], which designs a system called *Wiffler* to augment the mobile 3G capacity by leveraging delay-tolerance and fast switching. It shows that, for a delay tolerance of 60 seconds, 45% of the traffic can be offloaded to WiFi. The study [3] investigates daily mobility patterns of humans, and finds that WiFi can offload about 65% of the total mobile data traffic from mobile 3G networks and save 55% of battery power without using any delayed transmission. In our work, we also study the problem of WiFi offloading but we focus on another type of mobile social network offloading, which transmits the traffic by *opportunistic communications* between users.

Opportunistic offloading utilises opportunistic communications [2], [5], [36]. The work [2] exploits the opportunistic communications among devices to facilitate offloading by peer-to-peer *sharing* after one user has obtained the content [39], [40]. Specifically, the heuristic algorithm designed in [2] is shown to be capable of offloading cellular traffic by up to 73.66% for the traces studied. But multi-hop opportunistic forwarding is employed in [2], in which a set of targeted users are selected to disseminate the traffic to all other users. This multi-hop forwarding needs all the users to cooperate in the traffic offloading by using their own resources, which can be unrealistic, and it is unsuitable to mobile data offloading. In our solution, we rely on crucially a small set of users who are willing to participate in the offloading. The works [2] and [5] study the offloading helper selection problem. By contrast, we focus on how to efficiency allocate the buffer resource of the helpers after they are selected by the system. In vehicular networks, opportunistic communication is exploited to disseminate content between the moving vehicles [41], [42]. Specially, the work [41] focuses on studying the maximum downloading performance that can be theoretically achieved through opportunistic networks, while the work [42] studies the message forwarding strategy by mining the temporal correlation and the evolution of contact between each pair of vehicles. Different from these works, we focus on how to efficiency allocate the buffer resource for multiple mobile items with the consideration of heterogeneous network settings.

Storage allocation problems are addressed in the area of DTN content sharing [43], [44]. However, there exist significant differences between offloading and content sharing. Firstly, in the application of mobile data offloading, the mobile data originates from the Internet. Most of the data consist of large files with very different sizes. Content size is usually not considered in the works of DTN algorithm and protocol design [43], [44]. In our work, we have to consider different content sizes and storage constraints. Secondly, unlike content sharing, in mobile data offloading, the latency of data matters greatly since it impacts the user experience. Thirdly, in a offloading problem, the system focuses on how much data are offloaded from the cellular network, and how much capacity can be saved. Finally, in offloading, the nodes are usually mobile users, which can use the 3G network for example to communication. More importantly, the cellular system can use the control feedback channel to collect the offloading system's parameters, such as node contact rates. Therefore, centralized algorithms work. By contrast, for DTN content sharing, distribution algorithms are usually required, although the current proposed schemes like [10] still focus on analyzing the centralized algorithms.

7 Conclusion

We have studied the problem of how to offload multiple mobile data traffics from overloaded cellular networks in a realistic heterogeneous DTN-based network environment, where offloaded data consist of multiple contents with different delay tolerances and sizes, and the storage resources of helpers are limited. By formulating this challenging problem as a submodular function maximization problem with multiple linear constraints, we designed three algorithms to approximately solve it. Our first algorithm, the greedy algorithm, achieves the best performance under the generic heterogeneous DTN-based offloading scenario, and our second algorithm, the approximation algorithm, which offers lower complexity than the GA scheme, can achieve the same performance as the GA for the offloading scenario where data life times are short or in the fast-moving vehicular environment where contact rates are short. Our third algorithm, the homogeneous algorithm, has the lowest complexity but its application is limited to the offloading scenario with homogeneous contact rates and content sizes. In this special offloading case, the HA can be proved to provide the optimal solution.

In our system, we assume that all the helpers cooperate in the offloading scheme. However, in practice, helpers may act strategically to offload data for others. For example, some helpers may be more willing to offload certain contents to certain other users. Therefore, designing appropriate and efficient incentive mechanisms has a crucial role in enhancing offloading performance. In our current framework, we consider the opportunistic contacts as atomics, in which the mobile data can be transferred from one mobile user to another during a single contact. However, when the size of a data item is very large, the transmission of the data may not complete in single contact, owing to the limited contact time. Therefore, considering the limited contact duration for the current offloading framework is also needed. Our future work will investigate how to integrate appropriate incentive schemes with our mobile data offloading framework with the consideration of limited contact time.

ACKNOWLEDGMENTS

This work is supported by National Basic Research Program of China (Grant 2013CB329001), National Nature Science Foundation of China (Grants 61301080, 61171065 and 61273214), National High Technology Research and Development Program (Grants 2013AA013501 and 2013AA013505), and Program for Changjiang Scholars and Innovative Research Team in University. A small part of this work was presented at CHANTS'11, Nevada, USA.

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015*, Feb. 1, 2011 [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

[2] B. Han *et al.*, "Cellular traffic offloading through opportunistic communications: A case study," in *Proc. 5th ACM Workshop Challenged Netw.*, Chicago, IL, USA, Sep. 2010, pp. 31–38.

[3] K. Lee, I. Rhee, J. Lee, Y. Yi, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 425–426, Oct. 2010.

[4] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. 8th Int. Conf. Mobile Syst. Appl. Services*, San Francisco, CA, USA, Jun. 2010, pp. 209–222.

[5] M. V. Barbera, J. Stefa, A. C. Viana, M. D. de Amorim, and M. Boc, "VIP delegation: Enabling VIPs to offload data in wireless social mobile networks," in *Proc. 2011 Int. Conf. Distributed Computing Sensor Systems Workshops*, Barcelona, Spain, pp. 1–8.

[6] V. Vukadinović and G. Karlsson, "Spectral efficiency of mobility-assisted podcasting in cellular networks," in *Proc. 2nd Int. Workshop Mobile Opportunistic Networking*, Pisa, Italy, Feb. 2010, pp. 51–57.

[7] H. Luo, X. Meng, R. Ramjee, P. Sinha, and L. Li, "The design and evaluation of unified cellular and ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 9, pp. 1060–1074, Sep. 2007.

[8] B. Han, P. Hui, and A. Srinivasan, "Mobile data offloading in metropolitan area networks," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 14, no. 4, pp. 28–30, Oct. 2011.

[9] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié, "Incentivizing peer-assisted services: A fluid shapley value approach," *ACM SIGMETRICS Perf. Eval. Rev.*, vol. 38, no. 1, pp. 215–226, Jun. 2010.

[10] J. Reich and A. Chaintreau, "The age of impatience: Optimal replication schemes for opportunistic networks," in *Proc. CoNEXT*, Rome, Italy, Dec. 2009, pp. 85–96.

[11] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. MobiHoc*, New Orleans, LA, USA, May 2009, pp. 299–308.

[12] Y. Li *et al.*, "The impact of node selfishness on multicasting in delay tolerant networks," *IEEE Trans. Veh. Tech.*, vol. 60, no. 5, pp. 2224–2238, Jun. 2011.

[13] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. 30th IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 1–9.

[14] M. H. R. Khouzani, S. Sarkar, and E. Altman, "Maximum damage malware attack in mobile wireless networks," in *Proc. 29th IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[15] K. Lee *et al.*, "Max-contribution: On optimal resource allocation in delay tolerant networks," in *Proc. 29th IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[16] Y. Li, D. Jin, P. Hui, L. Su, and L. Zeng, "Revealing contact interval patterns in large scale urban vehicular ad hoc networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 299–300, 2012.

[17] Y. Li *et al.*, "Energy-efficient optimal opportunistic forwarding for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4500–4512, Nov. 2010.

[18] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in *Proc. 7th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Florence, Italy, May 2006, pp. 49–60.

[19] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, "Evaluating the impact of social selfishness on the epidemic routing in delay tolerant networks," *IEEE Commun. Lett.*, vol. 14, no. 11, pp. 1026–1028, Nov. 2010.

[20] A. Chaintreau *et al.*, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.

[21] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[22] P. Hui *et al.*, "Pocket switched networks and human mobility in conference environments," in *Proc. ACM SIGCOMM Workshop Delay Tolerant Networking*, Philadelphia, PA, USA, Aug. 2005, pp. 244–251.

[23] V. Conan, J. Leguay, and T. Friedman, "The heterogeneity of inter-contact time distributions: Its importance for routing in delay tolerant networks," *arXiv Report cs.NI/0609068*, Sep. 2006.

[24] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law intermeeting time in mobile ad-hoc networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1578–1591, Oct. 2009.

[25] H. Zhu *et al.*, "Recognizing exponential inter-contact time in VANETs," in *Proc. 2010 IEEE INFOCOM*, San Diego, CA, USA, pp. 1–5.

[26] A. Kulik, H. Shachnai, and T. Tamir, "Maximizing submodular set functions subject to multiple linear constraints," in *Proc. 20th Annu. ACM-SIAM Symp. Discrete Algorithms*, New York, NY, USA, Jan. 2009, pp. 545–554.

[27] Y. Azar and I. Gamzu, "Efficient submodular function maximization under linear packing constraints," in *Proc. 39th Int. Colloq. Automata, Lang. Program.*, Coventry, U.K., Jul. 2012, pp. 38–50.

[28] J. Leskovec *et al.*, "Cost-effective outbreak detection in networks," in *Proc. 13th ACM SIGKDD Int. Conf. KDD*, San Jose, CA, USA, Aug. 2007, pp. 420–429.

[29] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operat. Res. Lett.*, vol. 32, no. 1, pp. 41–43, Jan. 2004.

[30] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, 2001.

[31] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proc. Int. Conf. Simulation Tools Techniques*, Rome, Italy, Mar. 2009, pp. 1–10.

[32] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. 9th ACM MobiHoc*, Hong Kong, China, May 2008, pp. 241–250.

[33] SJU Traffic Information Grid Team, Grid Computing Center. *Shanghai Taxi Trace Data* [Online]. Available: http://wirelesslab.sjtu.edu.cn/

[34] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Cooperative caching for efficient data access in disruption tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 611–625, Mar. 2014.

[35] R. Bhatia, G. Narlikar, I. Rimac, and A. Beck, "UNAP: User-centric network-aware push for mobile content delivery," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1–9.

[36] Y. Li, Z. Wang, D. Jin, L. Zeng, and S. Chen, "Collaborative vehicular content dissemination with directional antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1301–1306, Apr. 2012.

[37] B. Higgins *et al.*, "Intentional networking: Opportunistic exploitation of mobile network diversity," in *Proc. 16th Annual Int. Conf. Mobile Computing Networking*, Chicago, IL, USA, Sep. 2010, pp. 73–84.

[38] E. Altman, R. El-Azouzi, D. S. Menasche, and Y. Xu, "Forever young: Aging control for smartphones in hybrid networks," *arXiv preprint arXiv:1009.4733*, pp. 1–59, 2010.

[39] T. Stading, P. Maniatis, and M. Baker, "Peer-to-peer caching schemes to address flash crowds," in *Peer-to-Peer Systems*, Berlin, Germany: Springer, 2002, pp. 203–213.

[40] S. Tewari and L. Kleinrock, "Proportional replication in peer-to-peer networks," in *Proc. 25th IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[41] F. Malandrino, C. E. Casetti, C. F. Chiasserini, and M. Fiore, "Content downloading in vehicular networks: What really matters," in *Proc. 30th IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 426–430.

[42] H. Zhu, S. Chang, M. Li, K. Naik, and S. Shen, "Exploiting temporal dependency for opportunistic forwarding in urban vehicular networks," in *Proc. 30th IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 2192–2200.

[43] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 748–760, Jun. 2008.

[44] S. Ioannidis, A. Chaintreau, and L. Massoulié, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1422–1430.

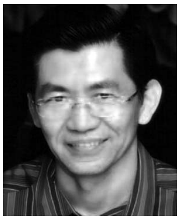
Yong Li (M'09) received the B.S. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. From 2012 to 2013, he worked as a Visiting Research Associate at Telekom Innovation Laboratories and HK University of Science and Technology, respectively. From 2013 to 2014, he visited the University of Miami, FL, USA, as a Visiting Scientist. He is currently a Faculty Member of Electronic Engineering at the Tsinghua University. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet. He received Outstanding Postdoctoral Researcher, Outstanding Ph.D. Graduates and Outstanding Doctoral thesis at Tsinghua University, and his research is granted by the Young Scientist Fund of Natural Science Foundation of China, Postdoctoral Special Find of China, and industry companies of Hitachi and ZET. He has published over 100 research papers and has 10 granted and pending Chinese and International patents. He has served as Technical Program Committee (TPC) Chair for WWW workshop of Simplex 2013, served as the TPC of several international workshops and conferences. He is also a Guest-Editor for the ACM/Springer *Mobile Networks and Applications*, Special Issue on Software-Defined and Virtualized Future Wireless Networks. Now, he is the Associate Editor of the *EURASIP Journal on Wireless Communications and Networking*. He is a member of the IEEE.

Mengjiong Qian received the B.S. and M.D. degrees from Tsinghua University, Beijing, China, in 2010 and 2013, respectively, both in electronic engineering. His current research interests include vehicle networks and mobility modeling.

Depeng Jin received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1995 and 1999, respectively both in electronics engineering. He is an Associate Professor at Tsinghua University and a Vice Chair of Department of Electronic Engineering. He was awarded the National Scientific and Technological Innovation Prize (Second Class) in 2002. His current research interests include telecommunications, high-speed networks, ASIC design, and future Internet architecture.

Pan Hui received B.Eng. and M.Phil. degrees from the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong, and the Ph.D. degree from Computer Laboratory, University of Cambridge, Cambridge, U.K. He is currently a Faculty Member of the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology, Hong Kong, where he directs the HKUST-DT System and Media Lab. He also serves as a Distinguished Scientist of Telekom Innovation Laboratories (T-labs), Germany and an Adjunct Professor of Social Computing and Networking at Aalto University, Aalto, Finland. Before returning to Hong Kong, he spent several years at T-labs and Intel Research Cambridge. He has published over 100 research papers and has some granted and pending European patents. He has founded and chaired several IEEE/ACM conferences/workshops, and served on the organizing and technical program committee of numerous international conferences and workshops including IEEE Infocom, ICNP, SECON, MASS, Globecom, WCNC, ITC, ICWSM, and WWW. He is an Associate Editor for the *IEEE Transactions on Mobile Computing* and *IEEE Transactions on Cloud Computing*. He is a member of the IEEE.

Zhaocheng Wang (M'09-SM-'11) received the B.S., M.S., and Ph.D. degrees from Tsinghua University, Beijing, China, in 1991, 1993, and 1996, respectively. From 1996 to 1997, he was with Nanyang Technological University (NTU), Singapore, as a Post-Doctoral Fellow. From 1997 to 1999, he was with OKI Techno Centre Pte. Ltd., Singapore, firstly as a Research Engineer and then as a Senior Engineer. From 1999 to 2009, he was at SONY Deutschland GmbH, firstly as a Senior Engineer and then as a Principal Engineer. He is currently a Professor in the Department of Electronic Engineering, Tsinghua University. His current research interests include wireless communications, digital broadcasting, and millimeter wave communications. He holds 22 granted U.S./E.U. patents and has published over 60 technical papers. He has served as Technical Program Committee co-chair/member of many international conferences. He is a senior member of the IEEE and a Fellow of IET.



Sheng Chen (M'90-SM'97-F'08) received the B.Eng. degree from the East China Petroleum Institute, Dongying, China, in January 1982, and the Ph.D. degree from the City University, London, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, Doctor of Sciences (D.Sc.), from the University of Southampton, Southampton, U.K. From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in the U.K. Since 1999, he has been with Electronics and Computer Science, the University of Southampton, Southampton, U.K., where he currently holds the post of Professor in Intelligent Systems and Signal Processing. He is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia. He is a Chartered Engineer (C.Eng.) and a Fellow of IET (FIET) and the IEEE. His current research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimisation. He has published over 470 research papers. He is an ISI highly cited Researcher in the engineering category during March 2004.