

Multiple Mobile Data Offloading Through Delay Tolerant Networks

Yong Li[†], Guolong Su[†], Pan Hui[‡], Depeng Jin[†], Li Su[†], Lieguang Zeng[†]

[†] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[‡] Deutsche Telekom Laboratories/TU-Berlin, Ernst-Reuter-Platz 7, Berlin 10587
yong-li07@mails.tsinghua.edu.cn

ABSTRACT

To cope with the explosive traffic demands and limited capacity provided by the current cellular networks, Delay Tolerant Networking (DTN) is used to migrate traffic from the cellular networks to the free and high capacity device-to-device networks. The current DTN-based mobile data offloading models do not address the heterogeneity of mobile traffic and are based on simple network assumptions. In this paper, we establish a mathematical framework to study the problem of multiple mobile data offloading under realistic network assumptions, where 1) mobile data is heterogeneous in terms of size and lifetime, 2) mobile users have different data subscribing interests, and 3) the storage of offloading helpers is limited. We formulate the maximum mobile data offloading as a Submodular Function Maximization problem with multiple linear constraints of limited storage and propose greedy, approximated and optimal algorithms for different offloading scenarios. We show that our algorithms can effectively offload data to DTNs by extensive simulations which employ real traces of both humans and vehicles.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
C.2.1 [Computer Communication Network]: Network Architecture and Design—*Wireless communication*

General Terms

Algorithms, Design, Performance

Keywords

Mobile data offloading, storage allocation, delay tolerant networking.

1. INTRODUCTION

Mobile Internet access is getting increasingly popular today and provides various services and applications including

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHANTS'11, September 23, 2011, Las Vegas, Nevada, USA.
Copyright 2011 ACM 978-1-4503-0XXX-X/11/09 ...\$10.00.

video, audio and images. From the data of the last two years, many researchers (e.g. Cisco) forecast that mobile traffic is growing at an annual rate of 131 % between 2008 and 2014, and will reach over 3.6 exabytes per month in 2014 [1]. Among this traffic, about 66% will be mobile video data [1]. Mobile cellular networks provide the most popular method of mobile access today [9]. With the increase of the mobile services and user demands, cellular networks will, very likely, be overloaded and congested in the near future. Especially during peak time and in urban area, users face extreme performance hits in terms of low network bandwidth, missed voice calls, and unreliable coverage.

To cope with this explosive traffic demands, growth and limited network capacity, it is an urgent and important agenda for cellular providers to provide quick and promising solutions. The most straightforward solution is to increase the cellular network capacity by adding more base stations with smaller cell size such as picocells and femtocells, or by upgrading the cellular network to next generation advanced networks like 4G [16]. However, these activities are at the expense of high financial input but low returns under the current flat pricing model where charges are independent of traffic. Even if the capacity of the networks is enhanced this way, the future demands from the users and applications will deteriorate the situation since they enable more data usage. Consequently, some providers have adopted certain short-term methods, e.g., limiting users' traffic to 5 GB per month, educating the users on responsible access. Obviously, these methods are ineffective and insufficient. In the long-term, the providers should provide different network technologies to offer sufficient bandwidth to the end users.

A more recent approach of using Delay Tolerant Networks (DTNs) to migrate cellular data traffic has been proposed by several previous works [4, 9]. By benefiting from the delay-tolerant nature of non-realtime applications, the service providers can delay and even shift the transmission to DTN. Benefiting from common interests among the users, providers only need to deliver the information to a small fraction of users, and then it will be further disseminated by the selected users through DTN communications. Providers could integrate effective incentive schemes, which stimulate users to participate in mobile data offloading. This kind of offloading should be encouraged by the operators as it is the quickest way, at the smallest cost, to support the exponential growth of mobile data that are still not be able to supported even updating all the infrastructure to 4G [1].

In this paper, we establish a mathematical framework to study DTN-based mobile traffic offloading of multiple mobile

data items in a realistic mobile environment. This problem is challenging for several reasons. First, mobile data provided by the service providers is not of just one type, and varies in terms of delay-sensitivity, content size, etc. Therefore, it is difficult for the providers to decide how to offload heterogeneous mobile data with limited network resources. Second, the users' demands and interests to mobile data are different, and this should be efficiently considered in the design of offloading schemes. Third, the DTN network resources are limited in practice (e.g. storage and battery capacity of mobile devices) and the communication contacts are opportunistic in nature. How to efficiently exploit the limited resources to improve the overall system performance is a challenging problem. In order to model a realistic network 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 has different delay sensitivity and size, and 3) the offloading helpers' storage is limited in size. These assumptions are usually not addressed in previous analytical work for simplicity reasons [4, 9]. Our contributions are summarized as follows:

- We formulate the optimal DTN offloading with the consideration of the heterogeneity of traffic, users, and limited storage as a problem of Submodular Function Maximization under multiple linear constraints.
- We prove the problem to be NP-Complete, give three algorithms to the offloading problem with different scenarios, and obtain the optimal solution for the homogeneous system.
- Through extensive real 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 models in Section 2, and then formulate the associated optimization problems in Section 3. In Section 4, we design algorithms to solve the formulated problem. In Section 5, we introduce the experimental environment for performance evaluation and provide simulation results. Finally, we present the related work in Section 6 and conclude the paper in Section 7.

2. SYSTEM OVERVIEW AND MODELS

2.1 Overview

In our DTN-based mobile data offloading system, some chosen users, named helpers, will participate in the offloading. Incentives for these users can be achieved by using some micro-payment scheme, or the operator can offer the participants a reduced cost for the service or better quality of service [10, 20]. A full analysis of such incentives is out of scope of this paper. The service provider chooses some users that are willing to participate in the data offloading, and transmits the mobile data to the chosen users through cellular network, and then these users further propagate the data to other users that are interested in it by short range device-to-device communication. If any users still have not received the data from the helpers after a tolerable duration which is related to the data lifetime, they will directly receive the data from the cellular network.

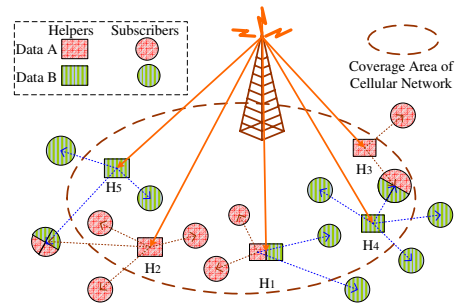


Figure 1: Multiple mobile data offloading through delay tolerant network.

We take two data items as an example to illustrate our multiple mobile data offloading scheme in Fig. 1. There are two types of nodes in the system, one is the offloading *helper*, and the other is the mobile data *subscriber*. The mobile data items are first transmitted to the helpers, and then the helpers transmit the data to some of the users who are interested in these data items. In our system, we consider multiple data offloading, where a helper can store more than one data item depending on their buffer size, and a user may also be interested in different data items. Taking the example shown in Fig. 1, the service provider first transmits Data A to helpers H_1 , H_2 and H_3 , and transmits Data B to helpers H_1 , H_4 and H_5 . Then, the data subscribers obtain the data from these helpers according to their interests by the DTN communication paradigm.

2.2 System Models

In our system, there are $N + H$ mobile users, labeled as $i \in \{1, 2, \dots, N + H\}$. These mobile users are either vehicles or humans carrying wireless devices (e.g. cell phones). Since in reality there are many different types of mobile data, for example, multimedia newspapers, weather forecasts, movie trailers, etc., we model the mobile traffic of C different data items labeled as \mathcal{C} . For any $c \in \mathcal{C}$, its data length is l_c . Users, even when some have additional storage and energy to help with the traffic offloading, may not be willing to act as helpers due to selfish behaviors and privacy concerns. While from the service providers' perspective, even if many users want to be helpers, they will only choose a limited set of them as usually they pay directly or indirectly for such users. Therefore, we use \mathcal{H} to denote the set of helper users that are willing to participate into the offloading, while use \mathcal{N} to denote the other subscriber users, where $|\mathcal{H}| = H$ and $|\mathcal{N}| = N$.

For the helpers, the system requires their storage to buffer some data items. At the same time, mobile data here includes multimedia content like movies that are very large. Even the 32GB of storage available on current devices is limited for large amounts of video content. Furthermore, nobody is willing to contribute all their storage to participate in offloading. Therefore, we should take as our constraint the storage that each helper is willing to share, which influences the number of data items that can be stored. Considering this realistic condition, we assume that helper n , $n \in \mathcal{H}$, can at most buffer L_n size of data items.

Any subscriber in \mathcal{N} may be interested in any data item.

Different subscribers may be interested in different data, and we define a metric named “user interest” to model this. For subscriber m , it is defined as a vector of $w_m = [w_{m,1}, w_{m,2}, \dots, w_{m,C}]$, where $w_{m,j}$ is used to compare the user’s interests in different items, and $w_{m,j} = 0$ means user m is completely not interested in data j . Without loss of generality, we further define $\sum_{j=1}^C w_{m,j} = 1$ for $\forall m$. 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. In this paper, we assume the contacts between any two nodes occurs at Poisson rates; Poisson distributed contacts have been shown to accurately model real-world DTN traces and are widely used to model DTN systems [21, 8, 7, 14, 17].

3. PROBLEM FORMULATION

We first obtain the expectation of the total size of data offloaded in the whole DTN system. Denote $X = (x_{s,k})$ ($s \in \mathcal{H}$, $k \in \mathcal{C}$) as the storage allocation for each helper, in which $x_{s,k} \in \{0, 1\}$ and $x_{s,k} = 1$ means offloading helper s stores item k in its buffer. For the simplicity of derivation, \mathcal{H}_k is further defined as the set of helpers who store copies of item k in their buffers

$$\mathcal{H}_k = \{s \in \mathcal{H} \mid x_{s,k} = 1\}, \quad (1)$$

A lifetime T_k is assigned to each data item k , which means all users will discard this item at time T_k . For a helper $s \in \mathcal{H}$ and a user $i \in \mathcal{N}$, since the contact between them is Poisson process at rate $\gamma_{s,i}$ and user i is interested in item $k \in \mathcal{C}$ with probability $w_{i,k}$, we can model the event that user s forwards item k to user i as a Poisson process with rate $\gamma_{s,i}w_{i,k}$. Then, the *system offloading utility function* $U(x_{s,k})$, which means the expectation of total size of offloaded data before the lifetime, can be derived as

$$\begin{aligned} U(x_{s,k}) &= \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} P(\text{user } i \text{ get item } k \text{ before } T_k) \\ &= \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} \left(1 - \prod_{s \in \mathcal{H}_k} e^{-\gamma_{s,i}w_{i,k}T_k} \right) \\ &= \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k}T_k \sum_{s \in \mathcal{H}} \gamma_{s,i}x_{s,k}} \right). \end{aligned} \quad (2)$$

Based on the total offloaded data size obtained above, we can specify the multiple mobile data offloading problem as the following optimization problem,

$$\begin{aligned} &\text{maximize} && U(x_{s,k}) \\ &\text{over} && x_{s,k} \in \{0, 1\}; \\ &\text{subject to} && \sum_{k \in \mathcal{C}} x_{s,k}l_k \leq L_s \text{ for } \forall s \in \mathcal{H}, \end{aligned} \quad (3)$$

where $\sum_{k \in \mathcal{C}} x_{s,k}l_k \leq L_s$ is the buffer size constraint of offloading helper s . In the formulated problem, we note that the system utility is an increasing and concave function of $x_{s,k}$, and the constraints are linear. Therefore, we can derive the optimal solution by gradient descent algorithm if $x_{s,k}$ is allowed to take the continuous real number. However, in the system, $x_{s,k}$ can either take 1 or 0. Therefore, we should design the corresponding algorithm to obtain the system solution.

4. MULTIPLE DATA OFFLOADING

In this section, we will solve the optimization problem defined in (3). We prove that the formulated problem of (3) is a Submodular Function Maximization under Multiple Linear Constraints, which is NP-hard.

4.1 Utility Function Analysis

Submodularity has long been studied in various problems [15, 18]. A function f defined on subsets of the universe \mathcal{C} is called *submodular*, if and only if $f(\mathcal{A} \cup x) - f(\mathcal{A}) \geq f(\mathcal{B} \cup x) - f(\mathcal{B})$ holds for $\forall \mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$ and $\forall x \in \mathcal{C} \setminus \mathcal{B}$.

In our problem, define the set of pairs of helpers and the items they carry as

$$\mathcal{P} = \{(s, k) \in \mathcal{C} \times \mathcal{H} \mid x_{s,k} = 1\}, \quad (4)$$

the system offloading utility function can be regarded as $U : 2^{\mathcal{C} \times \mathcal{H}} \rightarrow \mathbb{R}$ since (2) is equivalent with

$$U(\mathcal{P}) = \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k}T_k \sum_{s:(s,k) \in \mathcal{P}} \gamma_{s,i}} \right). \quad (5)$$

Then we have the following Theorem.

THEOREM 1. *The system offloading utility function $U : 2^{\mathcal{C} \times \mathcal{H}} \rightarrow \mathbb{R}$ is submodular, and the formulated problem of (3) is NP-hard.*

PROOF. First we show function

$$U_k(\mathcal{A}_k) = l_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k}T_k \sum_{s \in \mathcal{A}_k} \gamma_{s,i}} \right), \quad (6)$$

where $\mathcal{A}_k \in 2^{\mathcal{H}}$, is submodular. Since for $\mathcal{A}_k \subseteq \mathcal{B}_k \subseteq \mathcal{H}$ and $x \in \mathcal{H} \setminus \mathcal{B}_k$

$$\begin{aligned} &(U_k(\mathcal{A}_k \cup \{x\}) - U_k(\mathcal{A}_k)) - (U_k(\mathcal{B}_k \cup \{x\}) - U_k(\mathcal{B}_k)) \\ &= l_k \left(1 - e^{-w_{i,k}T_k \gamma_{x,i}} \right) \\ &\quad \cdot \sum_{i \in \mathcal{N}} e^{-w_{i,k}T_k \sum_{s \in \mathcal{A}_k} \gamma_{s,i}} \left(1 - e^{-w_{i,k}T_k \sum_{s \in \mathcal{B}_k \setminus \mathcal{A}_k} \gamma_{s,i}} \right) \\ &\geq 0, \end{aligned} \quad (7)$$

the submodularity of U_k on sets $2^{\mathcal{H}}$ has been proved. Moreover, U_k can also be regarded as a function defined on $2^{\mathcal{C} \times \mathcal{H}}$ and is also submodular via a similar derivation. Finally, we can observe that the linear combination

$$U = l_k U_k, \quad (8)$$

is a submodular function on $2^{\mathcal{C} \times \mathcal{H}}$. By proper parameter settings, the optimization problem degenerates to the traditional 0-1 Knapsack Problem which is NP-Complete. Therefore, the problem (3) is NP-Complete. \square

4.2 Algorithms

In this section, we design algorithms to approximately solve this optimization problem. First, for the general scenarios, we design a Greedy Algorithm (GA) by heuristic method. Second, for the scenario when the lifetime of data item is short, we design an Approximation Algorithm (AA). Third, we give an Optimal Algorithm (OA) for the case of homogeneous contact rates and homogeneous item and buffer size.

4.2.1 Greedy Algorithm

Consider the problem in the way that each storage allocation is determined one by one. A greedy algorithm can be employed as an approximated solution to the problem. When one or more copies of a certain item is stored in a helper, which is in accordance with the constraints, the gain of objective function will be increased; the gain of the objective function is generally different for different choices of items and helpers. Thus, as our first strategy, we can pick the choice of items and helpers that maximizes the gain on the objective function among all choices at each time. Further consider the length of each data item, we know that although an item may have large gain, it may also have so huge a length that other items cannot be stored. Consequently, our second greedy strategy is to calculate the gain for each choice of item and user per unit length and pick up the maximum one. It is obviously that either one of our above strategies has some drawbacks for some certain scenarios, and thus a combination of the two strategies is used to enhance the overall performance, which is similar with [18]. In the algorithm, the two strategies are both performed and we choose the better result between them.

4.2.2 Approximation Algorithm

In this section we propose an approximated solution to the problem for short lifetime scenario. Through the first order Taylor expansion $e^x \approx 1 + x$, the objective function (2) can be approximated as

$$\begin{aligned} U(x_{s,k}) &\approx \sum_{k \in \mathcal{C}} l_k \sum_{i \in \mathcal{N}} w_{i,k} T_k \sum_{s \in \mathcal{H}_k} \gamma_{s,i} \\ &= \sum_{s \in \mathcal{H}} \sum_{k \in \mathcal{C}} \left(l_k T_k \sum_{i \in \mathcal{N}} \gamma_{s,i} w_{i,k} \right) x_{s,k}. \end{aligned} \quad (9)$$

Furthermore, note the constraints in (3) are independent of s , we can break down the problem into

$$\begin{aligned} &\text{maximize} && \sum_{k \in \mathcal{C}} \left(l_k T_k \sum_{i \in \mathcal{N}} \gamma_{s,i} w_{i,k} \right) x_{s,k}; \\ &\text{over} && x_{s,k} \in \{0, 1\}; \\ &\text{subject to} && \sum_{k \in \mathcal{C}} x_{s,k} l_k \leq L_s, \end{aligned} \quad (10)$$

for each $s \in \mathcal{S}$ separately. For each $s \in \mathcal{S}$, the problem in (10) is a Knapsack Problem [5] and has efficiently approximation algorithms. Then the overall optimal solution is the combination of the solution to each problem in (10).

4.2.3 Optimal Storage Allocation Algorithm

We note that obtaining an optimal solution from the defined problem of Submodular Function Maximization under multiple linear constraints is hard. However, we can obtain a stronger result in the case of homogeneous contact rates, homogeneous item length and buffer size. In such homogeneous settings, we can obtain that the system utility of offloading traffic only depends on the number of replicas in all helpers for each data item, and not on the actual subset of nodes that carry it in the buffer.

In the homogeneous settings, we assume that the contact rates of all node pairs are the same, which is γ . Since all nodes' interests are the same, we can use the item popularity to obtain the system utility function. According to

the deviation in Section 3, we can rewrite the system utility function along with the maximization problem as follows,

$$\begin{aligned} &\text{maximize} && \sum_{k \in \mathcal{C}} l_k F_k \left(\sum_{s \in \mathcal{H}} x_{s,k} \right) \\ &\text{over} && x_{s,k} \in \{0, 1\}; \\ &\text{subject to} && \sum_{k \in \mathcal{C}} x_{s,k} \leq L_s, \end{aligned} \quad (11)$$

Now, we give a greedy algorithm described in Algorithm 1 for the formulated problem. The result obtained by Algorithm 1 is the optimal solution—this is proved in [19]. The algorithm repeatedly chooses items for the helpers to store: in each step, we try to select one item that brings the maximum system utility for a helper that still has available storage. Therefore, our algorithm is likely to allocate more helpers to store the items whose corresponding utility gain is larger than others

Algorithm 1 The Greedy Algorithm to Maximize the Objective Function in Homogeneous Settings

```

Set  $x_{s,k} = 0$ ,  $u_k = 0$ ,  $\Delta F_k = 0$  ( $k \in \mathcal{C}$ ,  $s \in \mathcal{H}$ );
Initialize set  $\mathfrak{R} = \{1, 2, \dots, C\}$  and sum = 0;
3: for Every data item  $k$  that  $k \in \mathfrak{R}$  do
     $\Delta F_k \leftarrow l_k (F_k(u_k + 1) - F_k(u_k))$ ;
end for
6: while sum  $\leq \sum_{s \in \mathcal{H}} L_s$  and  $\mathfrak{R} \neq \emptyset$  do
    Select  $i = \arg \max_k \{\Delta F_k | k \in \mathcal{C}\}$ ;
    Select  $q = \arg \max_s \{L_s - l_k \sum_{k \in \mathcal{C}} x_{s,k} | x_{s,i} = 0, s \in \mathcal{H}\}$ ;
9:   Set  $x_{q,i} = 1$ ;
    Update  $u_i \leftarrow u_i + 1$ , sum  $\leftarrow$  sum  $+ l_k$ ;
    Update  $\Delta F_i \leftarrow l_i (F_i(u_i + 1) - F_i(u_i))$ ;
12:  if  $u_i \geq H$  then
     $\mathfrak{R} \leftarrow \mathfrak{R} \setminus \{i\}$ ;
    end if
15: end while

```

5. PERFORMANCE EVALUATION

In this section, we present the simulation results for our algorithms to demonstrate the designed algorithms for the multiple mobile data offloading problem achieve good performance under the real-world mobility traces. In order to achieve this, we explore a broad set of simulation parameters as follows: a) we use mobility models of both human and vehicular traces, b) we simulate small and large system scale with respect to the number of helpers and data, and c) we compare against various schemes.

In order to show the efficiency of our scheme in real mobility environments, we use real traces for simulation. We use three traces, two are human mobility contact trace from the Reality Project of MIT [12] and *Cambridge* gathered by the Huggle Project [12]. These traces recorded contacts among users carrying Bluetooth devices, which periodically discover peers in communication range and record the contacts. The third one is a taxi GPS trace of *Shanghai* [22]. The used traces cover a large diversity of DTN environments, from sparse university campuses (*Reality*, *Cambridge*) to concentrated road sites (*Shanghai*), with the experiment period ranging from 30 days (*Shanghai*) to 98 days (*Reality*). In the

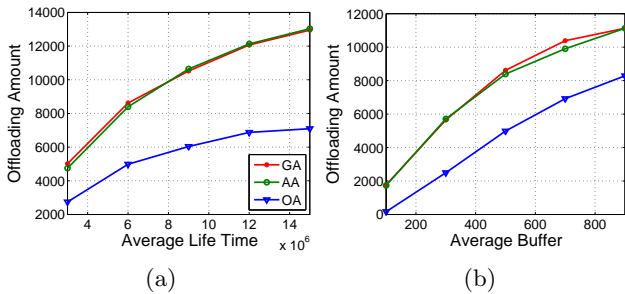


Figure 2: System performance of offloading data amount with trace of *Reality* (a) variable buffer size and average data lifetime is 600000 s; and (b) variable lifetime and average buffer size is 500 kB.

simulation, we vary the number of data items from 10 (*Reality*, *Cambridge*) to 35 (*Shanghai*), the sizes of data items are uniformly generated in the range $[0, 200 \text{ kB}]$ with mean of 100, data deadlines follow a uniform distribution of $[0, 2\beta \text{ s}]$ with mean of β . We randomly choose 10% of nodes as helpers, and they buffer sizes are uniformly generated in $[0, 2\alpha \text{ kB}]$ with mean of α , and user interests to different data items follow uniform distribution with an expectation of 1.

The results under the *Reality* trace are shown in Fig. 2. From the results of Fig. 2 (a), we can see that as we increase the helper's buffer size, the total data amount of offloaded items increases significantly. We find that GA and AA have almost the same performance, and OA has the worst performance. The reason is that in this simulation, settings for the user's buffer and content data length are not homogeneous. Therefore, OA will not obtain the optimal system solution. When we fix the buffer size to 500 and change the mobile data lifetime, we can see that the offloaded data amount increases with the lifetime. The results using *Shanghai* trace are shown in Fig. 3. From the results, we can obtain similar conclusions as the *Reality* trace. However, we can observe that the performance of OA is better in this vehicular trace than the human trace. The reason is that in *Shanghai* trace, the number of nodes is 2000, and the contact rates of any two taxis are almost the same. From Fig. 3 (a), we can see that AA almost achieves the same performance of GA. The reason is that when the lifetime is small, AA will perform well as it is designed for small lifetime case. When the lifetime becomes larger, which is shown in 3 (b), the offloaded amount almost stays the same compared to the shorter lifetime case. The reason is that the contact rate of vehicular is much smaller than the human trace. We also show the results under the *Cambridge* trace in Fig. 4. From the result, we can obtain that gap of OA and GA increases with average lifetime. From all the results, we can conclude that GA almost achieves the best performance under heterogeneous network settings, while AA achieves almost the same performance with GA when the data lifetime is short.

6. RELATED WORK

Some of the latest works [2, 3, 9, 16] focus on offloading mobile data from the overloaded cellular networks to other networks to provide better service. These works can be partitioned into three types according to their offloading destination networks. One type named broadcast offloading uses

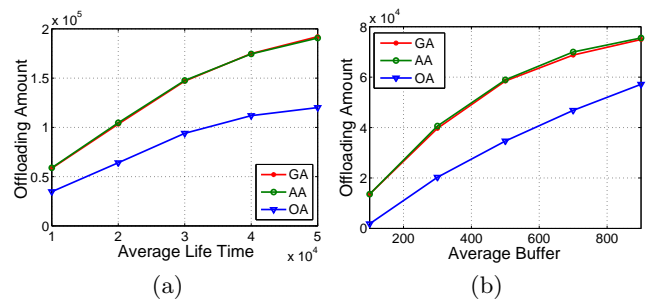


Figure 3: System performance of offloading data amount with trace of *Shanghai* (a) variable buffer size and average data lifetime is 10000 s; and (b) variable lifetime and average buffer size is 500 kB.

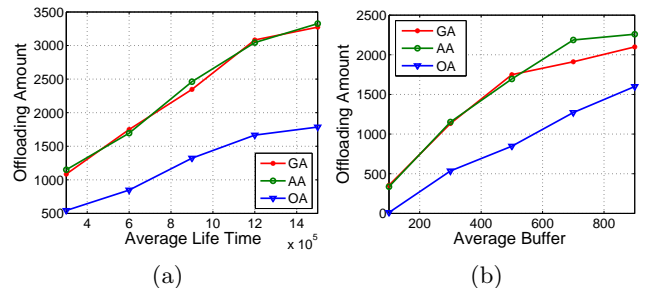


Figure 4: System performance of offloading data amount with trace of *Cambridge* with homogeneous networking settings (a) variable buffer size and average data lifetime is 600000 s; and (b) variable lifetime and average buffer size is 500 kB.

mobile broadcasting networks to offload the cellular traffic [3]. Another popular method named WiFi offloading uses the free available WiFi networks [2, 16, 11]. In our work, we study another type of offloading, which transmits the cellular traffic by opportunistic communications between users.

Opportunistic offloading uses opportunistic communications [9]. For example, Han et al. [9] exploited the opportunistic communications among devices to facilitate offloading by peer-to-peer sharing after one user obtained the content. This paper uses multi-hop opportunistic forwarding, in which a set of targeted users are selected to disseminate the traffic to all other users in the network. This multi-hop forwarding needs all the users to cooperate in the traffic offloading by using their own resources, which is unrealistic in practice. At the same time, it is not suitable to offloading large data items. In our solution, we rely on a crucial small set of users who are willing to participate in the offloading.

Storage allocation problems are also addressed in the area of DTN content sharing [6, 13]. However, there are some big differences between the offloading and content sharing problems. First, in the application of mobile data offloading, the mobile data originates from the Internet. Most of the data consists of large files with very different sizes. Content size does not matter in DTN content sharing. In this work, we consider the different content sizes and storage constraints. Second, in mobile data offloading, the latency of data matters since it impacts the user experience. Third, in offloading problem, the system should focus on how much data is offloaded from the cellular network, and how many capacity

can be saved. Finally, in offloading, the nodes are usually mobile users, which can use the 3G network to communication and to control the system, for example, collecting and obtaining the overall system metrics of node contact rates. Therefore, centralized algorithms work here, and there is no need to design distributed algorithms. However, in the DTN content sharing, the distribution algorithm should be designed.

7. CONCLUSION

In this paper, we investigated the multiple mobile data offloading problem—how to offload mobile traffic from the overloaded cellular networks. We studied this problem in a realistic environment where the network is heterogeneous: offloaded data consists of multiple kinds of content with different delay tolerances and sizes. Furthermore, the storage resources of helpers is limited. By formulating this problem as a Submodular Function Maximization problem, we designed three algorithms to approximately solve it.

In our system, we assume that all helpers cooperate in the offloading scheme. However, in practice the helpers may be too selfish to offload data for others and we need to design appropriate incentive mechanisms. At the same time, to encourage participation of mobile users to act as helpers, service providers can also exploit other efficient incentive schemes. In the future, we plan to integrate such incentive schemes into our mobile data offloading framework.

8. ACKNOWLEDGMENT

This work was supported by National Major Scientific and Technological Specialized Project (No. 2010ZX03004-002-02), National Basic Research Program (No. 2007CB310701), National High Technology Research and Development Program (No. 2008AA01Z107, No. 2008AA01A331 and No. 2009AA011205), NSFC (No. 61021001), PCSIRT and TNLIST.

9. REFERENCES

- [1] Cisco visual networking index: Global mobile data traffic forecast update, 2009-2014, February 2010. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/whitepaper_c11-520862.html . .
- [2] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting mobile 3G using WiFi. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 209–222. ACM, 2010.
- [3] R. Bhatia, G. Narlikar, I. Rimaq, and A. Beck. UNAP: User-Centric Network-Aware Push for Mobile Content Delivery. In *INFOCOM 2009, IEEE*, 2009.
- [4] A. Boc. Vip delegation: Enabling vips to offload data in wireless social mobile networks. 2011.
- [5] T. Cormen. *Introduction to algorithms*. The MIT press, 2001.
- [6] P. Costa, C. Mascolo, M. Musolesi, and G. Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 26(5):748–760, 2008.
- [7] W. Gao and G. Cao. User-centric data dissemination in disruption tolerant networks. In *Proceedings of INFOCOM*, 2011.
- [8] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proc. of ACM MobiHoc*, pages 299–308, 2009.
- [9] B. Han, P. Hui, V. Kumar, M. Marathe, G. Pei, and A. Srinivasan. Cellular traffic offloading through opportunistic communications: a case study. In *Proceedings of the 5th ACM workshop on Challenged networks*, pages 31–38. ACM.
- [10] B. Han, P. Hui, and A. Srinivasan. Mobile data offloading in metropolitan area networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 14(4):28–30, 2011.
- [11] B. Higgins, A. Reda, T. Alperovich, J. Flinn, T. Giuli, B. Noble, and D. Watson. Intentional networking: opportunistic exploitation of mobile network diversity. In *Proceedings of MobiCom 2010*, pages 73–84. ACM.
- [12] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay tolerant networks. In *Proc. of ACM MobiHoc*, 2008.
- [13] S. Ioannidis, A. Chaintreau, and L. Massoulié. Optimal and scalable distribution of content updates over a mobile social network. In *INFOCOM 2009, IEEE*, pages 1422–1430. IEEE, 2009.
- [14] M. Khouzani, S. Sarkar, and E. Altman. Maximum Damage Malware Attack in Mobile Wireless Networks. In *Proc. of IEEE INFOCOM*, 2010.
- [15] A. Kulik, H. Shachnai, and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 545–554. Society for Industrial and Applied Mathematics, 2009.
- [16] K. Lee, I. Rhee, J. Lee, Y. Yi, and S. Chong. Mobile data offloading: how much can WiFi deliver? *ACM SIGCOMM Computer Communication Review*, 40(4):425–426, 2010.
- [17] K. Lee, Y. Yi, J. Jeong, H. Won, I. Rhee, and S. Chong. Max-contribution: On optimal resource allocation in delay tolerant networks. In *Proc. of 29th IEEE INFOCOM*, pages 1–9, March 14-19, 2010.
- [18] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.
- [19] Y. Li, P. Hui, L. Su, D. Jin, and L. Zeng. An optimal distributed malware defense system for mobile networks with heterogeneous devices. In *SECON 2011, IEEE*, pages 1–9. IEEE.
- [20] H. Luo, X. Meng, R. Ramjee, P. Sinha, and L. Li. The design and evaluation of unified cellular and ad hoc networks. *IEEE Transactions on Mobile Computing*, pages 1060–1074, 2007.
- [21] J. Reich and A. Chaintreau. The age of impatience: optimal replication schemes for opportunistic networks. In *Proc. of ACM CoNext*, pages 85–96, 2009.
- [22] SJU Traffic Information Grid Team, Grid Computing Center. Shanghai taxi trace data. <http://wirelesslab.sjtu.edu.cn/>.