Cellular Traffic Offloading through Opportunistic Communications: A Case Study

Bo Han
Department of Computer Science
University of Maryland
College Park, MD 20742, U.S.

Pan Hui
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin, Germany

V. S. Anil Kumar
Department of Computer Science and Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061, USA

Madhav V. Marathe
Department of Computer Science and Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061, USA

Guanhong Pei
Department of Computer Science and Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061, USA

Aravind Srinivasan
Department of Computer Science and Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742, U.S.

Abstract

Due to the increasing popularity of various applications for smartphones, 3G networks are currently overloaded by mobile data traffic. Offloading cellular traffic through opportunistic communications is a promising solution to partially solve this problem, because there is no monetary cost for it. As a case study, we investigate the target-set selection problem for information delivery in the emerging Mobile Social Networks (MoSoNets). We propose to exploit opportunistic communications to facilitate the information dissemination and thus reduce the amount of cellular traffic. In particular, we study how to select the target set with only k users, such that we can minimize the cellular data traffic.

In this scenario, initially the content service providers deliver information over cellular networks to only users in the target set. Then through opportunistic communications, target-users will further propagate the information among all the subscribed users. Finally, service providers will send the information to users who fail to receive it before the delivery deadline (i.e., delay-tolerance threshold). We propose three algorithms, called `Greedy`, `Heuristic`, and `Random`, for this problem and evaluate their performance through an extensive trace-driven simulation study. The simulation results verify the efficiency of these algorithms for both synthetic and real-world mobility traces. For example, the `Heuristic` algorithm can offload cellular traffic by up to 73.66% for a real-world mobility trace.

CHANTS'10, September 24, 2010, Chicago, Illinois, USA.

Categories and Subject Descriptors

C.2.1 [**Computer Communication Network**]: Network Architecture and Design—*Wireless communication*

General Terms

Algorithms, Design, Performance

Keywords

Cellular traffic offloading, target-set selection, opportunistic communications, mobile social networks.

1. INTRODUCTION

Mobile Social Networks (MoSoNets) have begun to attract increasing attention in recent years, due to the proliferation of smartphones (e.g., Apple's iPhone and Nokia N95), mobile operating systems (e.g., Google's Android and Symbian OS), and online social networking services, such as Facebook and MySpace. The development of MoSoNets has already evolved from the simple extensions of online social networking sites to powerful mobile social software and applications [11, 14, 23, 24]. Currently, a large percentage of mobile data traffic is generated by these mobile social applications and mobile broadband-based PCs [2]. A side effect of the explosion of these applications, along with other mobile applications, is that 3G cellular networks are currently *overloaded*. According to AT&T's media newsroom, its network experienced a 5,000 percent surge of mobile data traffic in the past three years[1]. Thus, it is imperative that novel architectures and protocols be proposed and developed to solve this critical problem.

The information to be delivered in mobile networks may include multimedia newspapers, weather forecasts, and movie trailers etc., generated by content service providers. Benefiting from the delay-tolerant nature of non-real-time applications, these service providers may deliver information to only a small fraction of selected users (i.e., target-users), to

[1] `http://www.att.com/gen/press-room?pid=4800&cdvn=news&newsarticleid=30838`

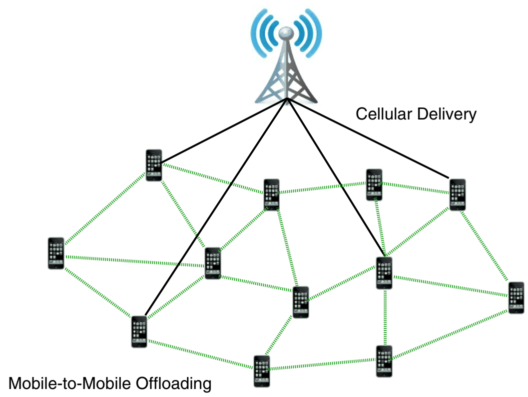

Figure 1: A snapshot of the contact graph for a small group of subscribed mobile users.

reduce cellular data traffic and thus their operation cost. As shown in Figure 1, the target-users can then further propagate the information among all the subscribed users through their social participation, when their mobile phones are within the transmission range of each other and can communicate *opportunistically*. Note that non-target-users can also disseminate the information after they get it from target-users. The major advantage of this cellular traffic offloading approach is that there is no or very little monetary cost associated with opportunistic communications, which are realized through either WiFi or Bluetooth.

In the original delay-tolerant approach, delay is usually caused by intermittent connectivities [13]. For instance, motivated by the fact that the coverage of WLAN hotspots may be very limited, Pitkänen et al. [27] have explored opportunistic web access via WLAN hotspots for mobile phone users. We propose to *intentionally* delay the delivery of information and offload it to the free opportunistic communications, with the goal of reducing cellular data traffic.

In this paper, we study the *target-set selection* problem as the first step towards bootstrapping the cellular traffic offloading for information delivery in MoSoNets. In particular, we investigate how to choose the initial set with only k users, such that we can minimize the amount of cellular data traffic. We can translate this objective into maximizing the expected number of users that can receive the delivered information through opportunistic communications[2]. The larger this number is, the less the cellular traffic will be. To the best of our knowledge, we are the first to apply the target-set selection problem to cellular traffic offloading through opportunistic communications.

Previous work by Nemhauser, Wolsey, and Fisher [25] shows that if the mapping function from the initial target set to the expected number of infected users is *submodular* (discussed in detail in Section 4), a natural greedy algorithm can achieve a provable approximation guarantee of $(1 - 1/e)$, where e is the base of the natural logarithm. Our first contribution is that we prove the information dissemination function is submodular for the contact graph of mobile users, which changes dynamically over time. Thus, we can apply the greedy algorithm to the target-set selection problem. However, this algorithm requires the knowledge of user mobility in the future, which may not be practical.

[2]We call these users the *infected* users.

Our second contribution is to exploit the *regularity* of human mobility [17, 22] and apply the target set identified using mobility history to information delivery in the future. For example, we determine the target set using the greedy algorithm based on today's user mobility history of a given period, then we use it as the target set for tomorrow's information delivery during the same period. We show through an extensive trace-driven simulation study that this heuristic algorithm always outperforms the simple random selection algorithm (wherein the k target users are chosen randomly), and can offload up to 73.66% of cellular traffic for a real-world mobility trace.

This paper is organized as follows. We give a brief introduction of related work in the next section. We then present the system model and formulate the problem in Section 3. In Section 4, we prove the submodularity of the information dissemination function and propose three algorithms for the target-set selection problem. In Section 5, we evaluate their performance through trace-driven simulation. We discuss several practical issues such as incentives, privacy and energy consumption in Section 6, and then conclude.

2. RELATED WORK

In this section, we briefly review the relate work in two categories: information diffusion/dissemination and cellular traffic offloading.

2.1 Information Diffusion/Dissemination

Social networks have been thought to be the carrier of information flows in communities. Various wireless communication technologies can effectively help the propagation of information among mobile users. As a result, information diffusion/dissemination has been widely studied in traditional social networks [10, 19, 28] and wireless networks [20, 24, 26]. Information diffusion has been extensively investigated through viral marketing [28] and social networks [10, 19]. Domingos and Richardson [10, 28] introduce a fundamental algorithmic problem of information diffusion: what is the initial subset of k users we should target, if we want to propagate the information to the largest fraction of the network? Kempe et al. [19] prove that for the influence maximization problem in social networks, the influence function is submodular for several classes of models. In this paper, we exploit social participation and interaction to offload cellular data traffic.

There are also several existing works for information dissemination in wireless networks. 7DS [26] is a peer-to-peer data dissemination and sharing system for mobile devices, aiming at increasing the data availability for users who have intermittent connectivity. When mobile devices fail to access Internet through their own connections, they can try to query data from peers in their proximity, who either have the data cached, or have Internet access and thus can download and forward the data to them. PeopleNet [24] is designed as a wireless virtual social network that mimics how people seek information in real life. In PeopleNet, queries of a specific type are first propagated through infrastructure networks to bazaars (i.e., geographic locations of users that are related to the query). In a bazaar, these queries are further disseminated through peer-to-peer communications, to find the possible answers. Lindemann and Waldhorst [20] model the epidemic-like information dissemination in mobile ad hoc networks, using four variants of 7DS [26] as examples. They consider the spread of multiple data items by mobile devices with limited buffers and use the least recently used (LRU) approach as their buffer management scheme. Compared to the above works, we focus on the target-set selection problem to reduce cellular data traffic.

2.2 Cellular Traffic Offloading

There are two types of existing solutions to alleviate the traffic load on cellular networks: offloading to femtocells and WiFi networks.

2.2.1 *Femtocell for Indoor Environments*

Originally, the femtocell technique (i.e., access point base stations) was proposed to offer better indoor voice and data services of cellular networks. Femtocells work on the same licensed spectrum as the macrocells of cellular networks and thus do not need special hardware support on mobile phones. But customers may need to install short-range base stations in residential or small-business environment, for which they will provide an Internet connection. Due to their small cell size, femtocells can lower transmission power and achieve higher signal-to-interference-plus-noise ratio (SINR), thus reducing the energy consumption of mobile phones. Cellular operators can reduce the traffic on their core networks when indoor users switch from macrocells to femtocells. A literature review about the technical details and challenges of femtocells can be found in Chandrasekhar et al. [8].

2.2.2 *Cellular Traffic Offloading to WiFi Networks*

Compared to femtocells, WiFi networks work on the unlicensed ISM (Industrial, Scientific and Medical) and U-NII (Unlicensed National Information Infrastructure) frequency bands and thus cause no interference with 3G cellular networks. As a result, cellular network operators, including AT&T, T-Mobile, Vodafone, and Orange, have deployed or acquired WiFi networks worldwide [1]. Meanwhile, there are already several offloading solutions and applications proposed from the industry. For instance, the Line2 iPhone application[3] clones the phone's own software and can initiate voice calls over WiFi networks. Recently, Balasubramanian et al. [3] propose a scheme called Wiffler to augment mobile 3G using WiFi for delay-tolerant applications. Our focus here is on offloading cellular data traffic through opportunistic communications, in metropolitan areas with high population density.

3. SYSTEM MODEL AND PROBLEM STATEMENT

In this section, we describe the system model of MoSoNets and the target-set selection problem we propose to solve.

3.1 Model of MoSoNets

There are two kinds of typical connections in MoSoNets, similar to the case of small-world networks [29]:

- Local connections realized by short-range communications, through WiFi or Bluetooth networks. When two mobile phones are within the transmission range of each other, their owners may start to exchange information, although they may not be familiar with each other. This opportunistic communication heavily depends on the mobility pattern of users and usually we can construct *contact graphs* (as shown in Figure 1) for them. Its major advantage is that it needs no infrastructure support and there is no monetary cost.
- Remote connections realized by long-range communications, through cellular networks (e.g., EDGE, EVDO, or HSPA). This communication happens only between friends in real life. It may be used sporadically, compared to the short-range communications. We can construct a *social graph*, as shown in Figure 2, based on the social relationship of mobile users. Of course, users need to pay for such transmissions.

[3] http://www.line2.com/

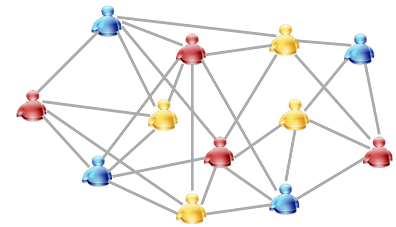

Figure 2: The social graph of mobile users shown in Figure 1. Users connected by an edge are friends of each other. There are three communities depicted by different colors. Users in the same community are friends with each other and form a clique. There are also connections between different communities. The friend relationship within a community is not shown here for clarity.

The study of traditional social networks focuses on the social graph, and the contact graph has been extensively investigated for opportunistic communication. We advocate that *MoSoNets can be viewed as a marriage of traditional social networks with emerging opportunistic networks*. We can exploit both types of communication to facilitate information dissemination in MoSoNets. On the one hand, friends can actively forward (push) information whenever they want. On the other hand, mobile users that are in contact can also pull the information from each other locally. We note that Chierichetti et al. [9] recently study a similar push-pull strategy for rumour spreading.

3.2 Problem Statement

As we mentioned in Section 1, we aim to study how to choose the initial set with only k users, such that we can maximize the expected number of infected users. This number will translate into the decrease of cellular data traffic. If there are totally n subscribed users and m users finally receive the information before the deadline, the amount of reduced cellular data traffic will be $n-(k+(n-m)) = m-k$. For a given mobile user, the delivery delay is defined to be the time between when the service provider delivers the information to the k users until a copy of it is received by that user. The service provider will send the information to a user directly through cellular networks, if he or she fails to receive the information before their delivery deadline.

How the information is propagated is determined by the behavior of mobile users and we exploit a probabilistic dissemination model in this paper. We define the *pull probability* to be the probability that mobile users pull the information from their peers during each of their contacts. The value of pull probability p may not be the same for different types of information and may change as time goes on, which reflects the dynamics of information popularity. After mobile users receive the information from either the service provider or their peers, they may forward it, through cellular networks (e.g., MMS, Multimedia Messaging Service), to their friends with probability q. Usually, $p > q$, because users generally prefer the free opportunistic communications. Moreover, short-range communications usually consume much less energy than long-range ones [4]. We do not consider the push-based approach for opportunistic communications in this paper and leave it as a future work.

The modeling of information dissemination through opportunistic communications can be viewed as a combination of three sub-processes. First, to protect their privacy, mobile users have the control of whether or not to share a piece of information with their geographical neighbors and share it with probability p_1. Second, mobile users may want to explore the information in their proximity only when they are not busy and mobile phones may not always be able to discover each other during the short contact durations. Thus they can find the meta-data of a piece of information with probability p_2. Finally, based on these meta-data, mobile users will decide whether or not to fetch the information from their geographical neighbors and pull it with probability p_3. As a result, $p = p_1 \cdot p_2 \cdot p_3$.

The information dissemination process in MoSoNets is similar to the information diffusion under the independent cascade model [16] of influence maximization [19] in social networks. In the independent cascade model, when a node u becomes active, it has a *single* chance to activate any currently inactive neighbor v with probability $p_{u,v}$. In comparison, mobile users have the chance to pull/exchange information for *every* contact.

4. TARGET-SET SELECTION

We first prove the submodularity of the information dissemination function for the contact graph of mobile users, which leads to the greedy algorithm. The information dissemination function is the function that maps the target set to the expected number of infected users of the information dissemination process. Then we present the details of the greedy algorithm and the proposed heuristic algorithm.

4.1 Submodularity of the Information Dissemination Function

If we can prove that the information dissemination function is submodular, we can then apply the well-known greedy algorithm proposed by Nemhauser et al. [25] to identify the target set. The information dissemination function $g(\cdot)$ is submodular if it satisfies the *diminishing returns* rule. That is, the marginal gain of adding a user, say u, into the target set S is not less than that of adding the same user into a superset S' of S:

$$g(S \cup \{u\}) - g(S) \geq g(S' \cup \{u\}) - g(S'),$$

for all users u and all pairs of sets $S \subseteq S'$. We prove the submodularity of the information dissemination function using an approach very similar to that proposed by Kempe et al. [19]. Generally it is hard to quantify exactly the underlying information dissemination function $g(\cdot)$. However, we can estimate the value of $g(\cdot)$ by Monte Carlo sampling.

As we mentioned in Section 3, compared to the information diffusion in traditional social networks, the contact graph of MoSoNets changes dynamically and mobile users can pull information from their peers for every contact of them. To solve this problem, we generate a *time-stamped* contact graph. For each pair of users u and v, if they are in contact ℓ times during the information dissemination process, there will be ℓ time-stamped edges in the graph, one for each contact. Suppose u's pull probability for v during a given contact t is $p_{u,v,t}$. [4] We can view this random event as flipping a coin of bias $p_{u,v,t}$. Note that whether we flip the coin at the very beginning of information dissemination or when u and v are in contact t, will not affect the final results. Thus, we can assume that for every contact t of each pair of users u and v, we flip a coin of bias $p_{u,v,t}$ at the beginning of the process and save the result to check later.

After we get all the results of coin flips, we mark the edges with successful pulling of information as *active* and the remaining edges as inactive. Since we already know the results of the coin flips and the initial target set, we can calculate the number of infected users at the end of the information dissemination process. In fact, one possible set of results of the coin flips stands for a sample point in the probability space. Suppose z is a sample point and define $g_z(S)$ to be the number of infected users when S is the initial target set. Then $g_z(S)$ is a deterministic quantity. Further define $I(u, z)$ to be the set of users that have a path from u, for which all the edges on it are active and their time-stamps satisfy the monotonically *increasing* requirement[5]. We have

$$g_z(S) = \cup_{u \in S} I(u, z).$$

We now prove that the function $g_z(S)$ is submodular for a given z. Consider two sets S and S', $S \subseteq S'$. $g_z(S \cup \{u\}) - g_z(S)$ is the number of users in $I(u, z)$ that are not in $\cup_{v \in S} I(v, z)$. Note that $\cup_{v \in S'} I(v, z)$ is at least as large as $\cup_{v \in S} I(v, z)$. We then have

$$g_z(S \cup \{u\}) - g_z(S) \geq g_z(S' \cup \{u\}) - g_z(S').$$

Since $g(S) = \sum_z \text{Prob}(z) \cdot g_z(S)$, we thus obtain that $g(\cdot)$ is submodular, because it is a non-negative linear combination of a family of submodular functions.

4.2 Greedy, Heuristic, and Random Algorithms

We present three algorithms for the target-set selection problem, called `Greedy`, `Heuristic`, and `Random`. Target-set selection in general is a NP-hard problem. Nemhauser et al. [25] show that if at each time we select a user that gives the maximum marginal gain of $g(\cdot)$ which is submodular, this greedy algorithm approximates the optimum solution to within a factor of $(1 - 1/e)$, which is the best known approximation ratio so far. For the `Greedy` algorithm, initially the target set is empty, we evaluate the information dissemination function $g(\{u\})$ for every user u, and select the most active user into the target set. Then we repeat this process, in each round selecting the next user from the rest with the maximum increase of $g(\cdot)$ into the target set, until we get the k users. However, we note that the limitation of the `Greedy` algorithm is that it requires the knowledge of user mobility during the dissemination process, which may not be available at the very beginning.

[4] We can define the pull probability $p_{v,u,t}$ accordingly.

[5] This requirement reflects the temporal evolution of the information dissemination process along the path.

To make the `Greedy` algorithm practical, we propose to exploit the 24-hour regularity of human mobility [17, 22], which leads to the `Heuristic` algorithm. Based on a six-month trace of the locations of 100,000 anonymized mobile phone users, Gonzalez et al. [17] identify that human mobilities show a very high degree of temporal and spatial regularity, and that each individual returns to a few highly frequented locations with a significant probability. Benefiting from the regularity of human mobility, the `Heuristic` algorithm identifies the target set using the history of user mobility, and then use this set for the information delivery in the future. That is, for a given period $[s, t]$, we apply the `Greedy` algorithm to determine the target set S of a history period $[s - c \times 24, t - c \times 24]$, where c is a small integer (usually 1 or 2), and then for the information delivery of $[s, t]$, service providers send the information to the mobile users in S at the very beginning to bootstrap the dissemination process. To enable the `Greedy` algorithm, the information dissemination protocol can collect the contact information of the subscribed users. At the end of a day, users can upload the information to the service provider through either their PCs or the WiFi interfaces on their phones.

Finally, for the `Random` algorithm, the service providers select k target users randomly from all the subscribed users. As we will show in the next section, although this algorithm is simple, it is still effective in the offloading process.

5. SIMULATION

We now introduce the mobility traces that we use for performance evaluation, and then present the results from a trace-driven simulator developed in C.

5.1 Mobility Traces

5.1.1 Synthetic Mobility Trace

We use the SIGMA-SPECTRUM simulator [5] to generate a synthetic mobility trace in the region of Portland, Oregon. The simulator combines different real-world data sources and realistic models, including an urban mobility model, synthetic population (according to U.S. Census data) and road-network data of Portland. The trace records the location of mobile users every 30 seconds. We randomly choose 10,000 people from the entire population of the city (around 1,600,000 people) as the subscribed users. The information dissemination periods start from 7:00AM with different durations. Note that the duration of the information dissemination period is, in fact, also the delay-tolerance threshold for mobile users (i.e., the maximum delay they need to tolerate). We use this trace to evaluate the performance of the `Random` algorithm for different pull probabilities and delay-tolerance thresholds.

5.1.2 Traces From Real-World Experiments

To evaluate the performance of the `Heuristic` algorithm, we need the mobility traces of different days, which is not available in the SIGMA-SPECTRUM simulator. To this end, we exploit two real-world mobility traces from the Haggle project [7] and the Reality Mining project [12].

We use the INFOCMO06 trace collected by the Haggle project for 4 days (from 2006-04-24 to 2006-04-27) during INFOCOM 2006 in Barcelona, Spain. This trace recorded the mobility of students and researchers attending the student workshop, using 78 iMotes which had a communication range of around 30 meters. We select 3 pairs of 1-hour periods from the trace as shown in Table 1. To exploit the 24-hour regularity of human mobility and evaluate the performance of the `Heuristic` algorithm, we use the target set identified by the `Greedy` algorithm for the periods in the second column ("History") to predict the mobility of users for the periods in the third column ("Delivery") of the same row. We define active users as those who have at least 1 contact with others during the delivery periods. As a result, the numbers of active users for these periods are 70, 66 and 66. We can also use other thresholds instead of 1. But they may exclude some inactive users for the simulation and thus reduce the (already small) number of simulated users.

	History	Delivery
#1	2006-04-24 11:00AM	2006-04-25 11:00AM
#2	2006-04-25 11:00AM	2006-04-26 11:00AM
#3	2006-04-25 12:00PM	2006-04-26 12:00PM

Table 1: The start time of three selected 1-hour periods from INFOCOM06 trace.

	History	Delivery
#1	2004-10-25 12:00PM	2004-10-28 12:00PM
#2	2004-11-15 12:00PM	2004-11-22 12:00PM
#3	2004-12-06 12:00PM	2004-12-07 12:00PM

Table 2: The start time of three selected 6-hour periods from Reality Mining trace.

The Reality Mining trace was collected using 100 Nokia 6600 smartphones carried by people from the MIT Media Laboratory and Sloan Business School, from 2004-07-26 to 2005-05-05. The information in this trace includes call logs, neighboring Bluetooth devices, and associated cell-tower IDs, etc. The contact trace of these users identified by the Bluetooth scanning is very sparse and thus is not suitable for the simulation. As in Ioannidis et al. [18], we instead consider that two mobile users are in contact of each other if their phones are associated with the same cell tower. Even this cell-tower based contact trace is sparse: this is the reason that we use 6-hour periods for the simulation. Similar to Table 1, we show the 3 pairs of 6-hour periods from the trace in Table 2. Benefiting from the long duration of the Reality Mining project, we can also exploit the 3-day (#1 of Table 2) and 1-week (#2 of Table 2) regularity of human mobility. The numbers of active users for these three periods are 61, 71 and 53 for the Reality Mining trace.

5.2 Simulation Results

In this section, we present the simulation results of the `Random`, `Heuristic`, and `Greedy` algorithms. In the simulation, we emulate the information delivery of multimedia newspapers (with size around several MB). Each direct cellular delivery consumes one message containing the newspaper. The goal is to determine the target set which leads to the most efficient cellular traffic offloading.

5.2.1 Pull Probability

We first evaluate the performance of `Random` algorithm for different pull probabilities using the Portland trace. We show the cellular traffic load for different sizes of target set, from 5 to 3,000, and pull probabilities, 0.01, 0.05 and 0.1,

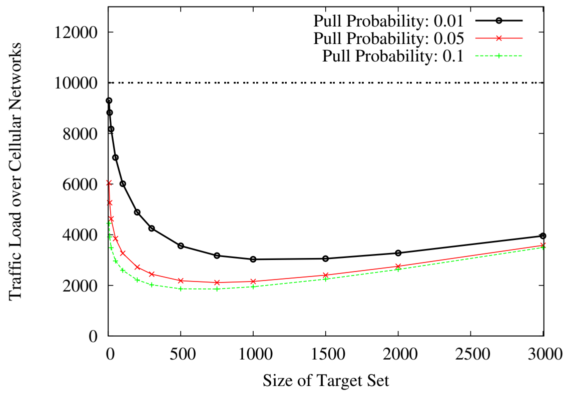

Figure 3: Performance of `Random` algorithm for different pull probabilities (Portland city data set).

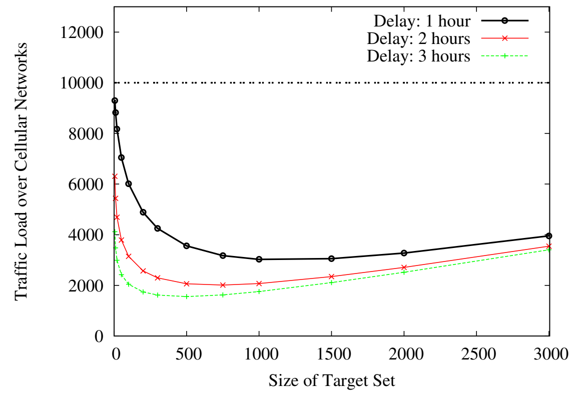

Figure 4: Performance of `Random` algorithm for different delay-tolerance thresholds (Portland city data set).

in Figure 3. The x-axis is the size of target set and the y-axis show the cellular traffic load, in terms of the number of cellular messages. Every user who fails to receive the information before the delivery deadline will consume a cellular message. Moreover, each user in the target set will also consume a cellular message. The delivery deadline is 1 hour. For each combination of the size of target set and pull probability, we run the simulation 10,000 times and report the average value. The horizontal dotted line shows the amount of cellular messages without offloading, which is the same as the total number of subscribed users. As we can see from this figure, even for this very simple random algorithm, it can reduce the amount of cellular traffic by up to 81.42% when the pull probability is 0.1. When we reduce the pull probability to 0.01, it can still offload cellular traffic by up to 69.73%.

There are two main observations from this figure. First, the amount of cellular traffic decreases as the pull probability increases. It is because when mobile users are all active in information propagation, a large number of users can get the delivered information through opportunistic communications, and thus avoid the data transmissions over cellular networks. Hence, active *social participation* is a key enabling factor of efficient information delivery. Second, as the size of target set increases, the amount of cellular traffic first decreases and then increases. The reasons are: (1). when the size of target set is small, the expected number of users that can receive the information through opportunistic communications is also small and thus a large number of users need to get the information through cellular networks; (2). when the size of target set is large, although it can make more users receive the information through opportunistic communications, the users in the target set will directly generate a large amount of cellular traffic.

For the three curves in Figure 3, the pull probability is fixed for all the contacts of these mobile users. We also tried different probabilities for different contacts, uniformly and randomly selected between 0.01 and 0.1. The result looks very similar to the curve with pull probability 0.05. Thus, we omit that result for clarity.

5.2.2 *Delay-Tolerance Threshold*

We then evaluate the performance of `Random` algorithm for different delay-tolerance thresholds for the Portland trace. We show the traffic load over cellular networks for three delay-tolerance thresholds, 1, 2 and 3 hours, in Figure 4. The pull probability is 0.01. We also run the simulation 10,000 times for a point in that plot and report the average value. As we can see from this figure, if mobile users are willing to tolerate longer delay we may be able to offload more traffic from cellular networks. However, the benefit of increasing the delay-tolerance threshold from 2 hours to 3 hours is not very significant, compared to that from 1 hour to 2 hours. One possible reason is that when we increase the threshold to 2 hours, most of the active users can receive the delivered information through opportunistic communications and thus the improvement of increasing it to 3 hours is limited.

Start at	No. 1	No. 2	No. 3	No. 4	No. 5
2006-04-25 11:00AM	**43** (31.18)	53 (31.17)	40 (30.77)	73 (29.46)	78 (29.31)
2006-04-26 11:00AM	68 (18.08)	**43** (16.67)	69 (15.78)	60 (14.98)	30 (14.86)
2004-12-06 12:00PM	**94** (34.07)	15 (34.03)	80 (34.01)	97 (33.61)	7 (33.57)
2004-12-07 12:00PM	**94** (26.22)	95 (26.07)	15 (25.97)	92 (25.79)	7 (25.31)

Table 3: The top 5 most active users for different periods and the expected number users that they can infect.

5.2.3 *Comparing Random, Heuristic, and Greedy*

We finally compare the performance of `Random`, `Heuristic` and `Greedy` algorithms using the two real-world traces. To verify the regularity of human mobility, we show in Table 3 the IDs of the top 5 most active users for 2 pairs of selected periods, one for the INFOCOM06 trace and another for the Reality Mining trace. The numbers in the parentheses are the expected number of infected users when each of the active users is selected as the single user in the target set. From this table, we can see that the most active user (with ID 43) for the period 2006-04-25 11:00AM-12:00PM is the second most active user for the period 2006-04-26 11:00AM-12:00PM for the INFOCOM06 trace. For the Reality Mining trace, the most active user for the period 2004-12-06 12:00PM-06:00PM is also the most active one for the period

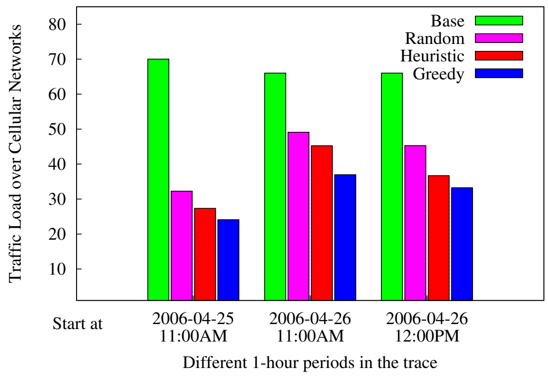

Figure 5: Performance comparison of `Random`, `Heuristic`, and `Greedy` algorithms for the INFOCOM06 data set.

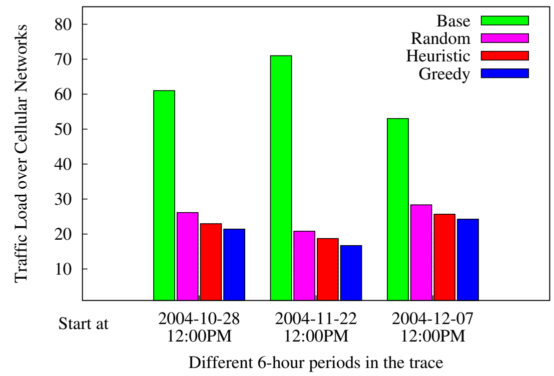

Figure 6: Performance comparison of `Random`, `Heuristic`, and `Greedy` algorithms for the Reality Mining data set.

2004-12-07 12:00PM-06:00PM. For almost all the other periods, the most active user of the *History* period is in the top 5 most active users of the *Delivery* period.

We plot in Figure 5 and Figure 6 the traffic load over cellular networks for the 6 pairs of periods listed in Table 1 and Table 2. Due to the small number of mobile users in the traces, we set the size of target set to be 5. For the `Random` and `Heuristic` algorithms, we simulate the information dissemination process 100,000 times and report the averaged values. For the `Greedy` algorithm, we run the simulation 10,000 times to determine the marginal gain for each user. In these figures, the `Base` shows the amount of cellular traffic without offloading, which is the same as the number of active users during these periods.

The performance of these algorithms depends on the pull probability. The pull probability is 0.01 for the INFOCOM06 trace and 0.001 for the Reality Mining trace. For high pull probabilities, there is no significant difference among them. As we can see from these figures, `Greedy` performs the best, followed by the `Heuristic` algorithm, for all the cases. Compared to the `Base`, the `Random` algorithm can reduce the amount of cellular traffic by up to 53.91% for the INFOCOM06 trace and 70.72% for the Reality Mining trace.

Owing to the regularity of human mobility, `Heuristic` can further reduce the amount of cellular traffic of `Random` by up to 18.95% for the INFOCOM06 trace and 12.25% for the Reality Mining trace. We note that due to the incompleteness of the real-world traces (e.g., caused by hardware errors), some users in the target set of the *History* period may not be active during the *Delivery* period. In these cases, we replace them with randomly selected users.

We have not evaluated how the push-based approach can help the information dissemination among friends, because there is no information about the social graph of mobile users for the above traces. However, we note that it is possible to construct the graph through the analysis of traffic between mobile users [30], or historical data of mobile users, such as proximity and location at a given time [12]. We leave the evaluation of push-based approach as a future work.

6. DISCUSSION

In this section, we discuss several practical issues for the large-scale deployment of our proposed cellular traffic offloading solution.

6.1 Incentives

How to integrate effective incentive schemes into the cellular traffic offloading is a challenging problem. For information service providers, with the offloading solution they can decrease the number of cellular messages and thus reduce their operation cost. As a result, they may reduce the subscription fee for their customers. Moreover, since the offloading solution can reduce the mobile data traffic over cellular networks, cellular operators may provide additional benefit to information service providers, which can in turn further reduce the cost of mobile users. To encourage social participation of mobile users, information service providers can also exploit other incentives: see, e.g., the Coupons approach of Garyfalos and Almeroth [15]. This system appends a sorted list of user IDs to a propagated message, which records the sequence of users who helped to disseminate the message. Similarly, information service providers can ask mobile users to optionally report when they got the delivered information and from where. Then they can offer discounts to mobile users who actively help the information delivery process.

6.2 Energy Consumption

Energy consumption may be the most important issue for the deployment of mobile applications. There are three main phases of opportunistic communications: device discovery, content discovery, and data transfer. Nowadays, more and more smartphones have multiple wireless interfaces (e.g., 3G, WiFi, and Bluetooth). The unique features of various wireless communication techniques make them suitable for different tasks. For example, Bluetooth may be suitable for device discovery due to its low transmission power and WiFi may be a better solution for content transfer due to its high data rate. Moreover, the energy consumption per bit for WiFi is lower than that of Bluetooth. Sticking with one technology may not be the best solution and thus we can combine them to make the opportunistic communications more energy efficient. For example, we can use Bluetooth for device- and content- discovery and WiFi for content transfer. Moreover, since device discovery is a common component for several mobile applications like Social Serendipity [11] and Media Sharing [21], its energy consumption can also be amortized by them.

6.3 Privacy

Unlike some existing protocols which determine whether to exchange information during the contact period [6], we aim to provide a general platform for information dissemination among mobile users. It is the users, not the platform, who make the decision about whether or not to share the information with peers and thus can protect their content privacy. They can opt-in and opt-out of the information dissemination process anytime they want, by turning off the information dissemination application. We finally note that we require only the contact information among the users and there is no need to track mobile users' locations to enable our proposed solution.

7. CONCLUSION

In this paper, we propose to offload cellular traffic through opportunistic communications and investigate the target-set selection problem for information delivery in MoSoNets. We present three algorithms for this problem, `Random`, `Heuristic`, and `Greedy`, and evaluate their performance through trace-driven simulation, using both a large-scale synthetic mobility trace and two real-world mobility traces. The simulation results show that `Greedy` performs the best, followed by `Heuristic`. Although the `Greedy` algorithm may not be practical, it is the basis of the `Heuristic` algorithm which exploits the regularity of human mobility. We are currently implementing a prototype of the information delivery framework using Linux-based smart-phones (e.g., Nokia N900).

8. ACKNOWLEDGEMENT

We thank the anonymous reviewers for their insightful comments. We thank Kan-Leung Cheng and Xiaoyu Zhang for useful discussions and valuable inputs. Aravind Srinivasan and Bo Han were supported in part by NSF ITR Award CNS-0426683 and NSF Award CNS-0626636. Madhav V. Marathe, V. S. Anil Kumar, and Guanhong Pei were supported in part by NSF Nets Grant CNS-0626964, NSF HSD Grant SES-0729441, NSF PetaApps Grant OCI-0904844, DTRA R&D Grant HDTRA1-0901-0017, DOE DE-SC0003957, DTRA CNIMS Grant HDTRA1-07-C-0113, NSF CAREER CNS 0845700 and NSF NETS CNS-0831633.

9. REFERENCES

[1] Mobile Data Offload for 3G Networks. White Paper, IntelliNet Technologies, Inc., 2009.

[2] Mobile data traffic surpasses voice. `http://www.cellular-news.com/story/42543.php`, 2010.

[3] A. Balasubramanian *et al.* Augmenting Mobile 3G Using WiFi. In *Proceedings of MobiSys 2010*, June 2010.

[4] N. Balasubramanian *et al.* Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications. In *Proceedings of IMC 2009*, pages 280–293, Nov. 2009.

[5] R. Beckman *et al.* Implications of Dynamic Spectrum Access on the Efficiency of Primary Wireless Market. In *Proceedings of DySPAN 2010*, pages 1–12, Apr. 2010.

[6] C. Boldrini *et al.* Context and resource awareness in opportunistic network data dissemination. In *Proceedings of WoWMoM 2008*, pages 1–6, June 2008.

[7] A. Chaintreau *et al.* Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, June 2007.

[8] V. Chandrasekhar *et al.* Femtocell Networks: A Survey. *IEEE Communications Magazine*, 46(9):59–67, Sept. 2008.

[9] F. Chierichetti *et al.* Rumour spreading and graph conductance. In *Proceedings of SODA 2010*, Jan. 2010.

[10] P. Domingos and M. Richardson. Mining the Network Value of Customers. In *Proceedings of SIGKDD 2001*, pages 57–66, Aug. 2001.

[11] N. Eagle and A. Pentland. Social Serendipity: Mobilizing Social Software. *IEEE Pervasive Computing*, 4(2):28–34, Apr.-June 2005.

[12] N. Eagle *et al.* Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, Sept. 2009.

[13] K. Fall. A Delay-Tolerant Network Architecture for Challenged Internets. In *Proceedings of SIGCOMM 2003*, pages 27–34, Aug. 2003.

[14] S. Gaonkar *et al.* Micro-Blog: Sharing and Querying Content Through Mobile Phones and Social Participation. In *Proceedings of MobiSys 2008*, pages 174–186, June 2008.

[15] A. Garyfalos and K. C. Almeroth. Coupons: A Multilevel Incentive Scheme for Information Dissemination in Mobile Networks. *IEEE Transactions on Mobile Computing*, 7(6):792–804, June 2008.

[16] J. Goldenberg *et al.* Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 12(3):211–223, Aug. 2001.

[17] M. C. González *et al.* Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[18] S. Ioannidis, A. Chaintreau, and L. Massoulié. Optimal and Scalable Distribution of Content Updates over a Mobile Social Network. In *Proceedings of the IEEE INFOCOM 2009*, pages 1422–1430, Apr. 2009.

[19] D. Kempe, J. Kleinberg, and Éva Tardos. Maximizing the Spread of Influence through a Social Network. In *Proceedings of SIGKDD 2003*, pages 137–146, Aug. 2003.

[20] C. Lindemann and O. P. Waldhorst. Modeling Epidemic Information Dissemination on Mobile Devices with Finite Buffers. In *Proceedings of SIGMETRICS 2005*, pages 121–132, June 2005.

[21] L. McNamara, C. Mascolo, and L. Capra. Media Sharing based on Colocation Prediction in Urban Transport. In *Proceedings of MOBICOM 2008*, pages 58–69, Sept. 2008.

[22] A. Mei and J. Stefa. SWIM: A Simple Model to Generate Small Mobile Worlds. In *Proceedings of the IEEE INFOCOM 2009*, pages 2106–2113, Apr. 2009.

[23] E. Miluzzo *et al.* Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. In *Proceedings of SenSys 2008*, pages 337–350, Nov. 2008.

[24] M. Motani *et al.* PeopleNet: Engineering A Wireless Virtual Social Network. In *Proceedings of MOBICOM 2005*, pages 243–257, Aug.-Sept. 2005.

[25] G. L. Nemhauser *et al.* An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, Dec. 1978.

[26] M. Papadopouli and H. Schulzrinne. Effects of power conservation, wireless coverage and cooperation on data dissemination among mobile devices. In *Proceedings of MOBIHOC 2001*, pages 117–127, Oct. 2001.

[27] M. Pitkänen, T. Kärkkäinen, and J. Ott. Opportunistic Web Access via WLAN Hotspots. In *Proceedings of PerCom 2010*, pages 20–30, Mar.-Apr. 2010.

[28] M. Richardson and P. Domingos. Mining Knowledge-Sharing Sites for Viral Marketing. In *Proceedings of SIGKDD 2002*, pages 61–70, July 2002.

[29] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[30] Z. Zhu *et al.* A Social Network Based Patching Scheme for Worm Containment in Cellular Networks. In *Proceedings of the IEEE INFOCOM 2009*, pages 1476–1484, Apr. 2009.