Mobile Data Delivery through Opportunistic Communications among Cellular Users: A Case Study for the D4D Challenge*

Ying Zhu   Chao Zhang   Yu Wang
Department of Computer Science, University of North Carolina at Charlotte
9201 University City Blvd., Charlotte, North Carolina, USA
{yzhu17,czhang25,yu.wang}@uncc.edu

ABSTRACT
The appearance of smartphones and increasing popularity of various mobile applications and services have caused the explosion of mobile data traffic. To avoid overloading the cellular networks, different offloading solutions (such as WiFi networks or femtocells) have been proposed and adopted. Recently, offloading cellular traffic through opportunistic communications among mobile phones becomes a new and promising option, due to free cost. In this paper, by using real trace data from the Orange "Data for Development" (D4D) challenge, we investigate the feasibility of delivering data packets among mobile cellular users through opportunistic communications in a large scale network. Our experimental results show that by using social or location properties of mobile users opportunistic routing can indeed complement the traditional cellular network to deliver delay-tolerant data packets among certain portion of cellular users. Such solution is especially cost efficient and beneficial for developing countries, as Ivory Coast.

1. INTRODUCTION
Due to the increasing popularity of mobile applications and services for smartphones, we are currently facing the challenges of mobile data explosion. Based on the most recent Cisco's report [1], mobile data traffic grew 70 percent in 2012 and reached 885 petabytes per month at the end of 2012, which was nearly 12 times the size of the entire Internet in 2000 (75 petabytes per month). Cisco also forecasts that mobile data traffic will surpass 10 exabytes per month in 2017. In addition, the recent advance in machine-to-machine (M2M) communications may potentially add billions of devices into mobile Internet. By the end of 2013, the number of mobile-connected devices will exceed the number of people on earth [1]. However, the current cellular networks do not have enough capacity to support all of the fast-growing mobile data from these devices.

To avoid overloading the cellular networks, different offloading solutions (such as WiFi networks [2–4] or femtocells [5]) have been proposed and adopted. According to Cisco [1], globally, 33 percent of total mobile data traffic was offloaded onto fixed network through WiFi or femtocell in 2012. Without offload, mobile data traffic would have grown 96 percent rather than 70 percent in 2012. Recently, offloading cellular traffic through opportunistic communications [6–8] among mobile phones becomes a new and possible solution. Compared with current WiFi or femtocell solutions, this method uses occasional device-to-device contact opportunities to deliver data rather than using the fixed network infrastructure. The major advantage of this solution is low cost and easy to deploy. However, due to the intermittent connectivity in opportunistic networks, the target data types are limited to delay-tolerant bulk data for non-realtime applications.

In this paper, using real trace data from the Orange "Data for Development" (D4D) challenge [9], we investigate the feasibility of delivering data packets among cellular users through opportunistic communications in a large scale network. Different from the previous studies [6–8] where broadcast traffic from the service provider to all subscriber users are offloaded to opportunistic networks, we focus on data delivery among individual mobile users using opportunistic routing. The released D4D dataset [10] provides anonymized call patterns and mobility data of 5,000 to 50,000 mobile phone users (based on Call Detail Records (CDR) of phone calls and SMS among these users) in Ivory Coast. It is a perfect resource to study the performance of opportunistic routing in a large-scale real network, since the dataset provides fine-quality mobility and location information of a large population of mobile users. Previous study on opportunistic routing [11–19] usually only focus on data delivery in small-scale delay tolerant networks or pocket switch networks with limited number of mobile users in a relevant small region (such as a group of researchers in a conference venue or a group of students on a campus). We believe that this is the first study on data delivery via opportunistic communications in large-scale networks in wide deployment

*This material was prepared for the Data for Development (D4D) Challenge and the third conference on the analysis of mobile phone datasets (NetMob 2013), May 2-3, 2013, MIT, USA. This work was supported in part by the US National Science Foundation (NSF) under Grant No.CNS-0915331 and CNS-1050398.

area. In our experiments, we consider six different opportunistic routing methods and evaluate them under various settings . Our results show that by using social or location properties of mobile users opportunistic routing can indeed achieve certain level of delivery ratios so that it can complement the traditional cellular network to deliver delay-tolerant data packets among active cellular users.

Middle East and Africa are the fastest growing regions of mobile phone market. Based on [1], their monthly mobile data traffic will experience the highest compound annual growth rate of 77 percent around the world between 2012 and 2017. Therefore, we strongly believe that the proposed solution is a cost efficient complement to the existing and growing cellular infrastructure for the developing countries in these regions (such as Ivory Coast). Thus, this could contribute to the socio-economic development and well-being of the Ivory Coast population (and fulfill the goal of D4D challenge).

The remainder of this paper is organized as follows: We briefly review related work in Section 2. We then introduce the D4D dataset and how we select and use the dataset for opportunistic communications in Section 3. In Section 4, we present six different opportunistic routing schemes which we test over the D4D dataset. Simulation results are reported in Section 5. Finally, conclusions are presented in Section 6.

2. RELATED WORK

2.1 Opportunistic Networks

Opportunistic network [20,21], where occasional contact opportunities are used to deliver data, is one of the emerging communication paradigms. In opportunistic networks, communication is challenged by sporadic and intermittent contacts as well as frequent disconnections and reconnections. To handle intermittent connectivity, opportunistic routing methods [11–19] share the same principle, "store and forward": If there is no connection available at a particular time, the current node can store and carry the data until it encounters other nodes. When the node has such a forwarding opportunity, all encountered nodes could be the candidates to relay the data. Therefore, relaying selection and forwarding decision need to be made by the current node based on certain forwarding strategy.

The simplest routing method is *epidemic routing* [11], in which a node forwards copies of message to any nodes it encounters. This flooding-based method can guarantee the best delivery ratio, but suffers from huge message overheads. To reduce the overheads, many methods restrict the number of message replicas in the network to a certain constant (such as in *Spray and Wait* [13]) or just one (such as in *SimBet* [15]) or a small one by only replicating the message when certain condition is met (such as in *delegation forwarding* [14]). We call the methods which allow multiple replicas and those which allow a single replica as multi-copy routing and single-copy routing, respectively.

Forwarding decision (or replicating decision) in opportunistic routing usually relies on certain type of quality metric. The message is only forwarded to a node with higher quality metric. During an encounter, if there are multiple nodes with higher quality metric, only the one with highest quality metric is selected as the relay. Examples include *Fresh* [16] (picking the node which has met the destination more recently), *Greedy-Total* [17] (picking the node with a higher encounter frequency to all other nodes), *MobySpace* [18] (picking the node which has more location similarity with the destination), *SimBet* [15] (picking the node with higher social centrality and more common neighbors with the destination), or *Bubble Rap* [12] (picking the node with higher centrality within certain community structure).

All these opportunistic routings are usually evaluated for data delivery in small-scale delay tolerant networks or pocket switch networks with limited number of mobile users (such as a group of researchers in a conference venue or a group of students on a campus).

2.2 Cellular Traffic Offloading

Cellular traffic offloading with complementary network communication technologies has become an emerging topic in recent years due to the dramatic increasing of mobile traffic load. Current solutions mainly rely on either femtocells and WiFi networks for delivering data originally targeted for cellular networks. Femtocells [5] operate on the same licensed spectrum as the macrocells of cellular networks and can offer better indoor voice and data services by offloading traffic from macrocells. WiFi networks on the other hand work on the unlicensed frequency bands and have also been widely used for offloading from cellular networks [2–4]. For example, major cellular operators, such as Orange, AT&T, T-Moblie, all have deployed their own WiFi networks world wide. Recently, Han et al. [6, 7] proposed the third type of solution: offloading cellular traffic to opportunistic networks formed by mobile cellular users. By studying how to select the initial set of users to push the content to all users in the networks, their proposed simple heuristics can improve the delivery efficiency and offload a large fraction of data from the cellular network. Li et al. [8] then studied the problem of multiple mobile data offloading through opportunistic communications among different data subscribers under resource constraints. Our study also uses opportunistic communications to offload traffic from cellular network, but we focus on offloading peer to peer traffic among mobile users instead of broadcasting traffic from the service provider to all subscriber users (as in [6–8]).

2.3 Cellular Dataset Analysis

The appearance of smartphones equipped with various sensors (especially GPS) and contact/event logs enables pervasive monitoring of mobile user behaviors and mobility. There are several cellular datasets recently collected via smartphone based testbeds: Nokia Data Collection Campaign [22], MIT reality project [23], Nodobo [24], and Context project [25]. These real-life tracing data provide abundant resources to study social, spatial, and temporal characteristics of mobile users in different environments. The D4D datasets [10] are newly released cellular datasets, which complement the existing datasets as the scale of number of users is much larger than those existing ones. This gives us an unique opportunity to study the feasibility of opportunistic routing in large-scale networks.

Table 1: Numbers of users, towers, and contacts in four different settings

Setting	# of users	# of towers	# of encounters
A) subset users within full region	13,436	1,095	617,136
B) subset users within limited region	6,318	496	327,717
C) all users within full region	46,254	1097	6,787,594
D) all users within limited region	21,768	497	3,736,173

(a) traffic load distribution (b) limited region near Abidjan (c) tower distribution in limited region

Figure 1: Illustration of the limited region in Settings B and D: (a) traffic load distribution in the whole nation generated using Geofast site [40] for the first two weeks; (2) the selected region near Abidjan with the heaviest traffic; (3) the detailed cellular tower distribution within the limited region.

Different cellular datasets have been studied by the research community for a wide range of purposes, such as human mobility modeling [26–28], importance place extraction [29–31], mobile recommendation systems [32, 33], urban sensing and planning [34, 35], sociology [36–38], ecology and epidemiology [39]. In this study, we use the D4D to study social or location based opportunistic routing in large-scale cellular networks as a possible offloading solution.

3. D4D DATASETS AND PREPARATION

The released D4D datasets [10] are based on anonymized Call Detail Records (CDR) of phone calls and SMS exchanges between 500,000 Orange mobile users in Ivory Coast between December 1, 2011 and April 28, 2012 (150 days). Among the released four datasets, we mainly use the second one (**SET2**): individual trajectories with high spatial resolution. This dataset contains the access records of antenna (cellular tower) of each mobile user over two-week periods. Such information provides high resolution trajectories for all mobile users. We will use the sequences of visited cellular towers of all users to generate contact encounters among mobile users and location profiles of each mobile user. In the results present in this paper we only use the first two weeks (December 1 to 14, 2011) data for our simulations.

Since D4D datasets do not have direct encounter information between phones via short range communications (such as Bluetooth or WiFi), to support opportunistic communications we assume that two phones can direct communicate to each other if they share the same cellular tower at particular time. Though this assumption may not be true in reality, it gives us an approximated environment for opportunistic communications. All of our experiments (presented in Section 5) are based on the generated encounter databases from SET2.

We will consider four different settings (A-D) for our experiments. Table 1 summaries some statistics of these settings. In term of number of nodes (mobile users), we either use all 500,000 users or a subset of users (around 15,000) in the original SET2. When we pick up the subset of users, we just simply choose the first 15,000 users in our encounter database. Notice that the number of users in our generated encounter database is less than the number of users in original SET2 (such as $46,254 < 50,000$). This shows that there are many mobile users who do not share any cellular towers with other users. The smaller size of user set could accelerate the execution time of our simulations. Notice that the number of encounters is significantly reduced after picking the subset users, though the cellular towers stay the same level. We also have settings where we limit the physical locations of encounters to a small region. As shown in Figure 1(a), the traffic load distribution within Ivory Coast is unbalanced. This figure shows the number of calls (both incoming and outgoing calls) during the first two weeks. Darker color indicates heavier traffic loads. Therefore, when picking up the small region, we choose the region with the heavies traffic load. The longitude and latitude ranges of the region (shown as a tiny blue rectangle in Figure 1(b) around Abidjan) are $[-8.49, -2.69]$ and $[4.41, 10.47]$, respectively. Abidjan is the economic and former official capital of Ivory Coast and the largest city in the nation. From Table 1 we can see that this region holds a large number of cellular towers and mobile users. Figure 1(c) shows the detailed tower distribution in this region.

4. OPPORTUNISTIC ROUTING PROTOCOLS

To test the feasibility of opportunistic routing among large scale mobile users, we implement six different routing methods which are listed below.

- **Epidemic** [11]: during any encounter, a copy of the message is forwarded to all encountered nodes and the current node still hold a copy of the message.
- **Naive**: during any encounter, the message is always forwarded to the encounter node and the current node will not hold the message after forwarding. If there are multiple nodes during the same encounter, the next hop is randomly picked. It can be treated as a single-copy version of Spray and Wait [13].
- **Fresh** [16]: the message is only forwarded from the current node v_i to the encountered node v_j if v_j has met the destination more recently than v_i does. If there are multiple nodes satisfying such a condition during the same encounter, v_i forwards the message to the one who has met the destination most recently.
- **Destination Frequency** [14]: the message is only forwarded from v_i to v_j if v_j has met the destination more often than v_i does. If there are multiple nodes satisfying such a condition during the encounters, v_i forwards the message to the one who has met the destination most often.
- **Centrality-Based**: the message is only forwarded from v_i to v_j if v_j has higher centrality than v_i does. Here, we simply consider the degree centrality of each node, i.e., how many nodes it has encountered. A node with higher degree centrality is more popular in the network. If there are multiple nodes satisfying the condition during the encounter, v_i forwards the message to the one who has the highest centrality. Similar idea has been used in Greedy-Total [17], SimBet [15] and Bubble Rap [12].
- **Location-Based**: the message is only forwarded from v_i to v_j if v_j has higher location similarity to the destination than v_i does. If there are multiple nodes satisfying the condition during the encounter, v_i forwards the message to the one with the highest similarity. This idea has been used in MobySpace [18] and is based on the obversion that people with similar location profiles (places visited) are likely to meet each other at their common places.

Since most of these methods are quite standard and straightforward to implement, we only introduce the detail about our implementation of location-based method (how to estimate the location similarity between two mobile users).

We treat every cellular tower t_k as one place, where $k = 1, \cdots, N$ and N is the total number of towers. For each mobile user v_i, we first extract its total visit duration and frequency of each tower t_k, represented by $d(v_i, t_k)$ and $f(v_i, t_k)$. Both are normalized between 0 and 1. Then we can estimate the probability that user v_i visits tower t_k using $p(v_i, t_k) = d(v_i, t_k) \times f(v_i, t_k)$. Therefore, for each user v_i, we can have a vector of $P(v_i) = \{p(v_i, t_k)|k = 1, \cdots, N\}$ to represent his *location profile*. Given two mobile users v_i and v_j, their location similarity can be defined as

$$S(v_i, v_j) = |P(v_i) \cdot P(v_j)|_1 = \sum_{k=1}^{N} p(v_i, t_k) \times p(v_j, t_k).$$

Notice that here we only consider the total visit duration and frequency of each user for a particular tower. It is possible to consider more detailed visit pattern, such as time-dependent visit patterns: visit duration and frequency in certain time period (morning, afternoon or night). However, our simulation results show that such refinement does not lead to sufficient improvement.

5. PERFORMANCE EVALUATIONS

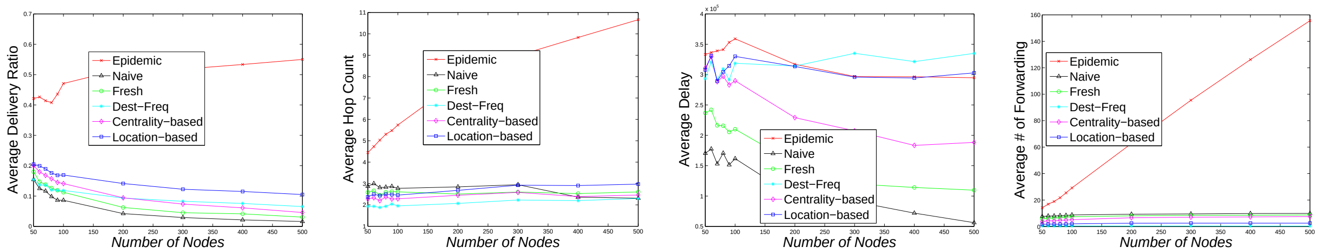
In this section, we present our experimental results on performances of all opportunistic routing methods over the contact database we generated from the D4D dataset as described in Section 3.

In all experiments, we compare each algorithm using the following routing metrics.

- **Delivery ratio:** the average percentage of successfully delivered messages from the sources to the destinations.
- **Hop count:** the average number of hops during each successful delivery from the sources to the destinations.
- **Delay:** the average time duration of successfully delivered messages from the sources to the destinations.
- **Number of forwarding:** the average number of messages forwarding in the network during the whole period.

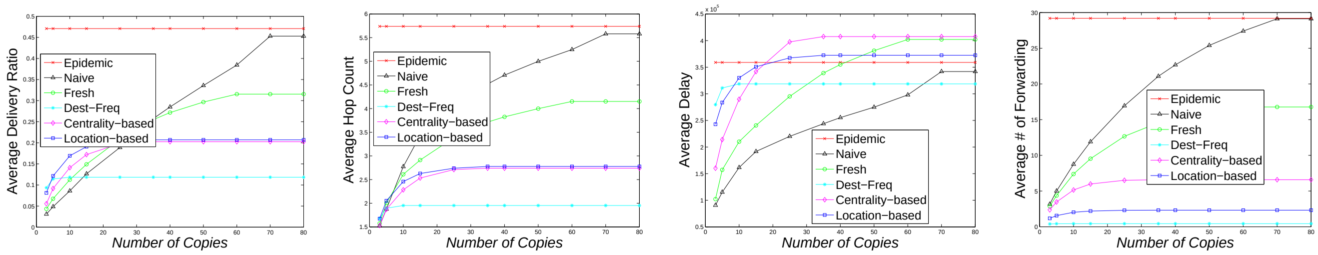
For all experiments, we perform $5,000$ random routing tasks among the selected participators. All results reported here are the average over these tasks.

For each experiment, we pick different number of nodes to participate the opportunistic communications, ranging from 50 to 500. Here we always pick the most active nodes (based on overall centrality) in the user set, since they are better candidates for opportunistic forwarding. We believe that these active users are the major target customers for our proposed traffic offloading scheme via opportunistic communication. For those users who are not active or even isolated in the opportunistic network, the only choice is using the cellular or fixed networks. We did perform some experiments over random chosen users, however the delivery ratios of all routing methods (even Epidemic) are very low (worse than 1 percent).



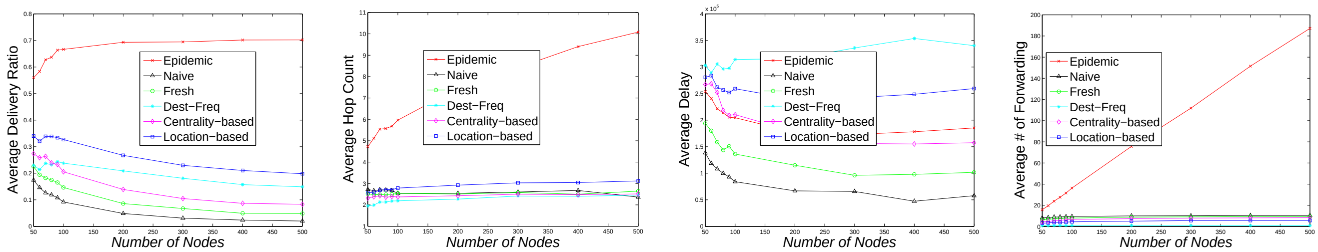
(a) delivery ratio (b) hop count (c) delay (d) # of forwarding

Figure 2: Performance results over Setting A (the number of copies is fixed at 10).



(a) delivery ratio (b) hop count (c) delay (d) # of forwarding

Figure 3: Performance results over Setting A (the number of copies is ranged from 3 to 80, and the number of nodes is fixed at 100).



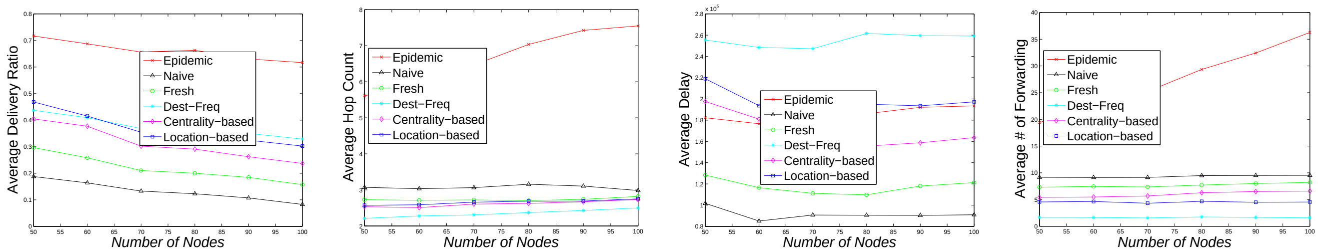
(a) delivery ratio (b) hop count (c) delay (d) # of forwarding

Figure 4: Performance results over Setting B (the number of copies is fixed at 10).

For all opportunistic routing methods except for Epidemic, we allow multiple copies of the same message but limit the number of copies by a small constant. In the default setting, we use 10 as the constant bound.

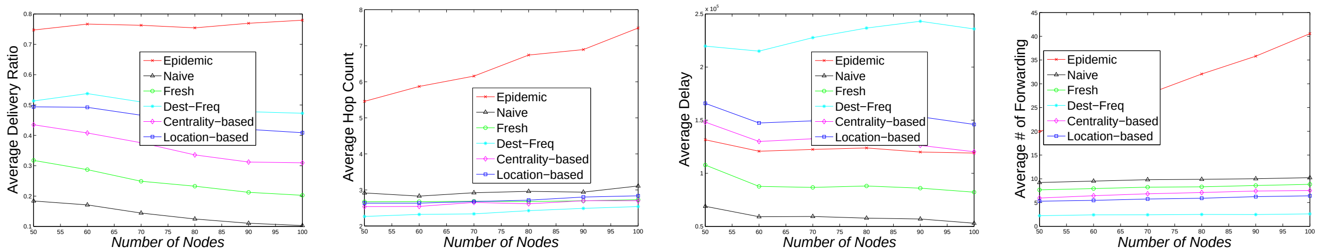
In the first set of simulations, we use Setting A (with around 15,000 selected users and within the full region). Figure 2 illustrate the results. From Figure 2(a), we can see that location-based and centrality-based methods can achieves better delivery ratio than Naive and Fresh methods. This confirms that the understanding and usage of social or location relationships among mobile users is beneficial for making smarter forwarding decision. Notice that that even though Epidemic routing has the best delivery ratio, it costs extremely large amount of forwarding as shown in Figure 2(d). It is also noticeable that the delivery ratio is decreasing as the number of nodes increases. This is reasonable since we always choose the most active nodes as the participators. With more nodes included, more routing tasks are among less active nodes. This again shows that our proposed solution mainly benefits the active mobile users in the network. In terms of hop count and number of forwarding, all opportunistic routing methods are at the similar level except for Epidemic. Notice that for delay since we only consider the successful routes, thus Epidemic usually has the largest delay.

For the same Setting A, we then test the effect of the number of copies in multi-copy opportunistic routing. We fixed with 100 nodes and change the number of copies from 3 to 80. Figure 3 shows the results. It is obvious that with more message copies all methods can achieve higher delivery ratio but increase the number of forwarding too. There is clearly a trade-off between number of copies and forwarding overhead. When the number of copies reaches certain value, the delivery ratio will be stable. Further adding more copies does not help. For different methods, such critical value of copy number may vary.



(a) delivery ratio (b) hop count (c) delay (d) # of forwarding

Figure 5: Performance results over Setting C (the number of copies is fixed at 10).



(a) delivery ratio (b) hop count (c) delay (d) # of forwarding

Figure 6: Performance results over Setting D (the number of copies is fixed at 10).

To test the performance of all methods in a small and dense region, we then test our methods on Setting B, which limits the region around a rectangle region near Abidjan. Compared with the results in the full region (Setting A), all methods can achieve better performances in this setting. This is reasonable since a limited dense network provides more close opportunities for message delivery among mobile users than a larger and sparser network does.

Last, we also perform simulation over the full population of D4D dataset (Settings C and D). Figure 5 and Figure 6 show the results, respectively. Compared with previous results, all methods can achieve better performances too. The reason is still the same that within larger population the selected participants are more active thus lead to better chances for mobile delivery. Once again, better performance can also be achieved in a smaller and denser area.

In summary, via the above simulations over the D4D dataset, we can have the following overall conclusions.

- Epidemic can achieve the highest delivery ratio since it takes every forwarding opportunities and does not have limitation on the number of copies. However, it suffers from the large number of forwarding, especially when the number of nodes is large.
- Location-based, Centrality-based, and Destination Frequency can achieve relevant high delivery ratios while still use reasonable number of forwarding.
- Compared with different settings, all opportunistic routing can achieve better performance when the participants are active users and the physical region is small and dense. This can be shown in Figure 7 which summarizes the average delivery ratios over four different settings under the same parameters.

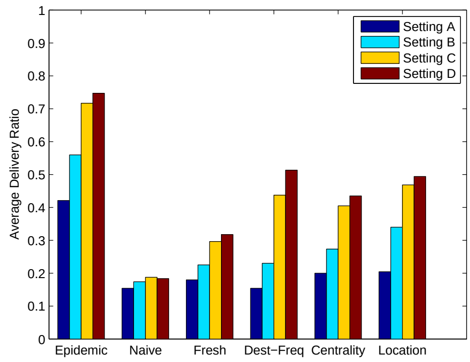


Figure 7: Average deliver ratios over Settings A to D (the number of nodes and the number of copies are 50 and 10, respectively).

6. CONCLUSIONS

In this paper, by leveraging the rich trace data from the Orange D4D challenge, we investigate the feasibility of delivering data packets among mobile cellular users through opportunistic communications. Our experimental results show that by using social or location properties of mobile users opportunistic routing can indeed achieve certain level of deliver ratio, especially among the active mobile users and within dense region. Therefore, it is possible that such solution can complement the traditional fixed network to deliver delay-tolerant data packets. On the other hand, there are still spaces to further improve the deliver ratio of opportunistic routing in such large scale networks. We will continue investigate new techniques to enhance the performance of opportunistic communications. Finally, we would like to thank the Orange D4D challenge organizers to provide such a great opportunity for us to participate in this event. We hope that Orange and other cellular companies can further release high quality real life datasets to research community.

7. REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017 (February 6, 2013). http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html.

[2] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting mobile 3G using WiFi. In *ACM MobiSys 2010*, 2010.

[3] K. Lee, I, Rhee, J. Lee, S. Chong, and Y. Yi. Mobile data offloading: how much can WiFi deliver? In *ACM Co-NEXT 10*, 2010.

[4] S. Dimatteo, P. Hui, B. Han, and V.O.K. Li. Cellular traffic offloading through WiFi networks. In *Proceeding of IEEE 8th International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2011*, 2011.

[5] V. Chandrasekhar, J.G. Andrews, and A. Gatherer. Femtocell networks: A survey. *IEEE Communications Magazine*, 46(9):59-67, 2008.

[6] B. Han, P. Hui, V.S.A. Kumar, M.V. 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 (CHANTS '10)*, 2010.

[7] B. Han, P. Hui, V.S.A. Kumar, M.V. Marathe, J. Shao, and A. Srinivasan. Mobile data offloading through opportunistic communications and social participation. *IEEE Transactions on Mobile Computing*, 11(5):821-834, 2012.

[8] Y. Li, G. Su, P. Hui, D. Jin, L. Su, and L. Zeng. Multiple mobile data offloading through delay tolerant networks. In *Proceedings of the 6th ACM workshop on Challenged networks (CHANTS '11)*, 2011.

[9] The Data for Development (D4D) Challenge. http://www.d4d.orange.com.

[10] V.D. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda, and C. Ziemlicki. Data for development: The D4D challenge on mobile phone data. arXiv.1210.0137v2, Jan 2013.

[11] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. Tech. Rep. CS-200006, Duke University, 2000.

[12] P. Hui, J. Crowcroft, and E. Yonek. Bubble Rap: Social-based forwarding in delay tolerant networks. In *Proc. ACM MobiHoc*, 2008.

[13] K T. Spyropoulos, Psounis, and C. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *Proc. of ACM SIGCOMM workshop on DTN (WDTN)*, 2005.

[14] V. Erranmilli, M. Crovella, A. Chaintreau, and C. Diot. Delegation forwarding. In *Proc. of ACM MobiHoc*, 2008.

[15] E.M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. In *Proc. ACM MobiHoc*, 2007.

[16] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *Proc. of ACM MobiHoc*, 2003.

[17] V. Erramilli and M. Crovella. Diversity of forwarding paths in pocket switched networks. In *Proc. of ACM IMC*, 2007.

[18] J. Leguay, T. Friedman, and V. Conan. DTN routing in a mobility pattern space. In *Proc. of ACM SIGCOMM workshop on DTN*, 2005.

[19] L. Zhao, F. Li, C. Zhang, and Y. Wang. Routing with multi-level social groups in mobile opportunistic networks. In *Proc. of IEEE Globecom*, 2012.

[20] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, New York, NY, USA, 2005. ACM.

[21] Y. Zhu, B. Xu, X. Shi, and Y. Wang. A survey of social-based routing in delay tolerant networks: Positive and negative social effects. *IEEE Communication Survey and Tutorials*, 15(1):387-401, 2013.

[22] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Mobile Data Challenge 2012 (by Nokia) Workshop, in conjunction with Int. Conf.. on Pervasive Computing*, Newcastle, June 2012.

[23] MIT Media Lab. Reality mining project. http://reality.media.mit.edu/.

[24] University of Strathclyde. Nodobo data release. http://nodobo.com/release.html.

[25] University of Helsinki. The context project. http://www.cs.helsinki.fi/group/context/#data.

[26] F. Calabrese, G.D. Lorenzo, L. Liu, and C. Ratti. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4):36-44, 2011.

[27] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi. Limits of predictability in human mobility. *Science*, 327:1018–1021, 2010.

[28] M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.

[29] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying important places in people's lives from cellular network data. In *Proceedings of the 9th international conference on Pervasive computing*, Pervasive'11, pages 133–151, Berlin, Heidelberg, 2011. Springer-Verlag.

[30] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of

locations. In *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, WMASH '04, pages 110–118, New York, NY, USA, 2004. ACM.

[31] Y. Ye, Y. Zheng, Y. Chen, J. Feng, and X. Xie. Mining individual life pattern based on location history. In *Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, MDM '09, pages 1–10, Washington, DC, USA, 2009. IEEE Computer Society.

[32] V. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with GPS history data. In *Proc. of WWW' 10*, pages 1029–1038, 2010.

[33] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS tragectories. In *Proc. of WWW' 09*, pages 791–800, 2009.

[34] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing*, 10:18–26, 2011.

[35] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3):30–38, July 2007.

[36] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 2012.

[37] N. Eagle, A. Pentland, and D. Lazer. Inferring Friendship Network Structure by Using Mobile Phone Data. *PNAS*, 106(36), 2009.

[38] W. Dong, B. Lepri, and S. Pentland. Tracking co-evolution of behavior and relationships with mobile phone. *Tsinghua Science and Technology*, 17(2):136-151, 2012.

[39] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modeling disease outbreaks in realistic urban social networks. *Nature*, 2004.

[40] The Geofast Web Interface. http://www.geofast.net.