Rethinking Offloading WiFi Access Point Deployment from User Perspective

Eyuphan Bulut
Virginia Commonwealth University
401 E. Main St
Richmond, VA, 23284
ebulut@vcu.edu

Boleslaw K. Szymanski
Rensselaer Polytechnic Institute
110 8th St
Troy, NY 12180
szymansk@cs.rpi.edu

***Abstract*—WiFi offloading has been exploited as a quick and viable solution to decrease the burden on cellular networks. In this paper, we study the problem of deploying new WiFi access points (AP) in a city-wide area for offloading purposes. Different than previous work which look at the problem only from operator's perspective and targets the maximization of offloaded traffic volume, we approach the problem by integrating the user perspective as well. We propose a new AP deployment scheme that aims to increase average individual user satisfaction while still achieving high offloaded total data traffic volume from all users. As the simulation results demonstrate, the proposed approach can achieve more user level satisfaction compared to other algorithms that only target offloaded traffic maximization while keeping operator's benefit from offloading close to others.**

I. Introduction

WiFi offloading has attracted a great deal of attention from academia and industry as it is considered an immediate remedy for taming the mobile data explosion. Several sub-problems emerged with this solution have been studied including the deployment of new WiFi access points [1], [2], [3], [4], managing the seamless control of WiFi and LTE handovers [14] and recruitment of third-party WiFi access points [19], [20], [16].

In this paper, we study the problem of selecting WiFi access point locations in the context of mobile data offloading. Recent work has mainly proposed solutions considering the sole goal of achieving the highest volume of data offloaded from cellular space. However, these solutions look at the problem only from operator's point of view.

Our objective in this paper is to approach the problem from users' perspective as well and find the offloading setting in which user satisfaction is prioritized while also trying to maximize the total volume of traffic offloaded. To this end, we propose a new WiFi access point deployment scheme in which offloading of each individual user's traffic is treated with equal significance. Moreover, as many studies [8], [9], [10] on the energy consumption of network interfaces (WiFi, 3G) have shown in different platforms and devices, cellular access is more costly than WiFi access in terms of the energy spent per byte. For example, it is measured in [9] that the cost of downloading with cellular is twice and uploading is four times expensive than it is with WiFi. Consequently, in the proposed scheme, we also consider the status of users' batteries in the selection process.

The rest of the paper is organized as follows. We first discuss our motivation with some statistics from real network datasets in Section II. In Section III, we define the problem and provide the details of proposed approach. In Section IV, we provide the simulation setting and discuss the evaluation of proposed system using real network traces. Finally, we close by discussing the related work in Section V and conclusion in Section VI.

II. Motivation

Our study is motivated by two observations:

Top region difference: For a mobile operator, the benefit of offloading will be maximized if the WiFi access points are deployed in the regions with maximum aggregate mobile traffic density. However, the regions showing high aggregate mobile data traffic could be different than the highest data usage regions of each individual. This can result in less satisfied users since they are not provided the opportunity of offloading their traffic through WiFi access points (AP).

Battery level diversity: As the per byte cost of data downloading and uploading through cellular connection is more expensive than the cost of the same through WiFi access points, the users with lower battery charge level could be more satisfied if they are provided with WiFi offloading opportunity.

Figure 1 shows some related statistics from real mobile network traces. In Figure 1.a, we compare the most popular regions of all users and individual users in a location-based social network dataset (Gowalla [17]). To get the plot, we first calculated the 300 most dense regions (i.e. grid cell that can hold an AP's coverage area) in terms of total user traffic in San Francisco downtown area. Then, for the users active in these top regions, we found their individual top regions and checked how many of them are included already in the top regions of the total user traffic. If each user is included in more than one of the total traffic top regions, then we checked that many top regions of each individual user. We also considered only users with more than ten data points to eliminate users with some random/irregular behavior. Figure 1.b shows the same data as Figure 1.a but with the number of users per region

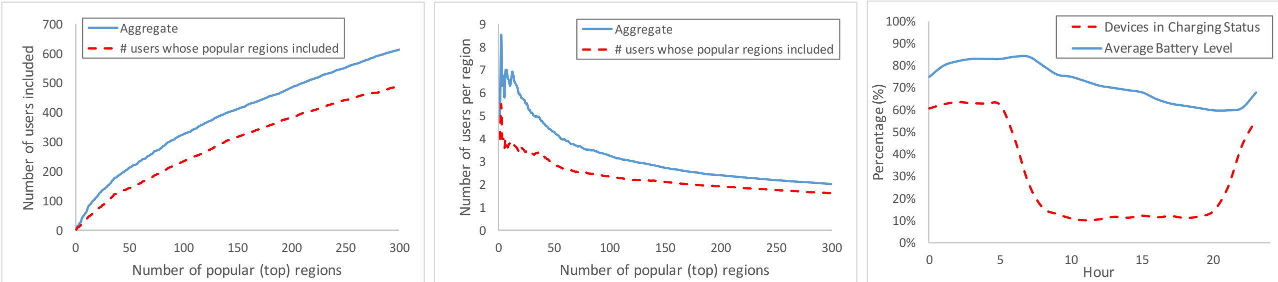


Fig. 1. (a) Top region comparison in aggregate and user based data density, (b) Number of different users per region (c) Average battery level and battery charging status during a day.

information in y-axis. Figure 1.b shows that top regions are mainly dominated by a few users. It starts with 8-9 users then goes down to 2 as the selected top regions increases. Moreover, for a remarkable number of users (starting around 40-45% and stabilizing around 20%) active in these top regions, the individual top regions are different than the top aggregated traffic regions. This is indeed expected because some of the places will be points of interests and will be visited by more people, yielding more aggregated user traffic. However, in case of deploying a WiFi AP in such a region, contribution to average user satisfaction will be limited to average percentage of each user's offloaded traffic amount in that region with respect to their total traffic.

Next, in Figure 1.c, we show the average battery level statistics from a smartphone dataset (Device Analyzer [18]) during a day. Depending on the hour of the day, the average battery charge levels show some significant differences ranging to [60-85]%. Considering that the regions that users visit will be affected by the hour of the day, average battery level of a user visiting different regions will differ. To these differences, we found the average hour of all visits from all users for each region, then compared these means (together with the instances giving the mean) using ANOVA test and observed a significant portion of pairs passing the test. This shows that the average battery levels of users visiting each region can show a different characteristic compared to the averages in other regions.

III. Proposed Solution

Assume that there are n different users and m different locations at which WiFi access points (AP) can be deployed. Let d_{ij} denote the volume of the data requested by user $1 \leq i \leq n$ in the coverage region of a potential AP $1 \leq j \leq m$. We define the utility function on each region as:

$$U_j = \sum_{i=0}^{n} (u_{ij}^d u_{ij}^b)$$

where

$$u_{ij}^d = \frac{d_{ij}}{\sum_{s=0}^{m} d_{is}}$$

and

$$u_{ij}^b = \frac{f(\beta_{avg}^{ij} - \beta_{wifi}(d_{ij}))}{f(\beta_{avg}^{ij} - \beta_{cell}(d_{ij}))}$$

In the last equation, β_{avg}^{ij} denotes the average battery level of user i in region j. The utility of each region is based on two factors: (i) the volume of each user's data that will be offloaded with respect to its total volume of data used in all regions, (ii) the change in the battery level based satisfaction function, $f(...)$, that can be achieved by offloading the traffic to the WiFi AP in that region compared to using cellular data access. Function $f(...)$ is used to account for the user response to the changes in the battery level.

The goal is to maximize the aggregated user satisfaction based on the utility function defined above, after selecting k locations for AP deployment. Assume $K = \{0, 1, 2, \ldots k-1\}$ denote the subset of m locations that are selected for AP deployment. Then the objective is:

$$Max \sum_{j \in K} U_j$$

A. The complexity of the problem

The nature of the problem differs depending on the relation between the coverage areas of potential m AP locations. If these locations are pre-determined depending on several factors including accessibility, availability, and interference-freeness (which could be the most likely case in practice) and there is none or minimal coverage area overlap between them, the problem will reduce to the problem of selecting top k regions in descending order of their U_j values.

On the other hand, if coverage areas could overlap, then the problem can be mapped to a well-known Maximal Covering Location Problem (MCLP) [5] that deals with locating k facilities to an area with demand locations (with different weights) such that the total demand under the coverage area (which is decided by time or travel distance) of all facilities is maximized. The locations of facilities may or may not overlap with the demand locations. The main focus of MCLP problem is to guarantee a worst case performance (by satisfying all demands within distance x or travel time t.).

Considering the data access request demands coming from users as the demands in MCLP problem and the facilities as WiFi APs that will be deployed, our problem of deploying the APs with the range of R (i.e., maximum distance of a demand from a facility at which that facility is able to satisfy

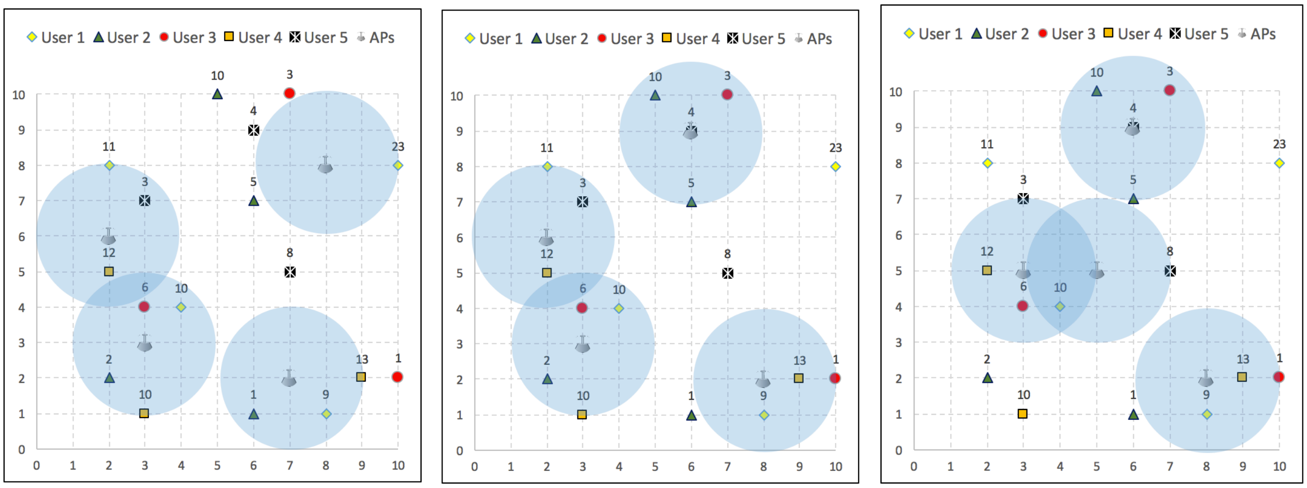


Fig. 2. Example scenario with 17 data points from five users. User requests in each AP region are shown with circles of different colors. (Left) An optimal solution that maximizes the total offloaded traffic from the entire area with 4 APs. (Middle) Optimal locations of APs considering the individual offloading ratios of users. (Right) Greedy heuristic based solution with individual offloading ratios of users.

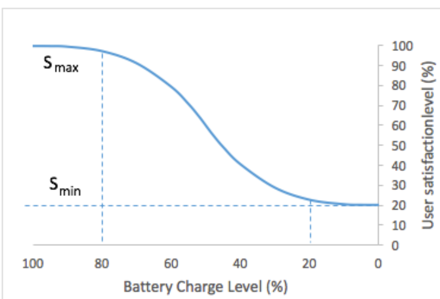


Fig. 3. User satisfaction based on battery level.

this demand) maps to MCLP problem. The MCLP problem is known to be NP-hard as proved by Megiddo et al. in [6].

There are many variants of MCLP problem with application specific additional constraints. In our problem of WiFi AP deployment with maximum demand satisfied using a given number of APs, some additional constraints can be considered. For example, there is usually a capacity (i.e., bandwidth) limit of APs. This simply maps to MCLP instance with capacitated facilities [7].

B. Greedy Adding with Substitution (GAS) heuristic

As we also studied the AP deployment in a similar setting but without considering user satisfaction in [4], and it has been shown that greedy heuristic based algorithm results are close to the ILP solution, we will use a similar approach here. This will also allow us to do a fair comparison with [4], which only has the objective of maximizing offloaded user traffic. To extend the algorithm in [4], we use greedy adding with substitution (GAS) heuristic recommended by Church and ReVelle in [5], which attempts to refine the solution set by allowing a selected AP location to be de-selected and a de-selected AP to be re-selected at each iteration. The GAS algorithm performs better than Greedy only algorithm in practice and runs very quickly.

C. Numerical Example

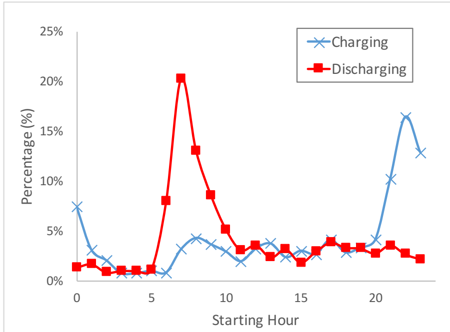
Here, we give a numerical example to show how the results could be different when user satisfaction is also targeted. Figure 2.a illustrates a sample problem with 17 data points from 5 different users. Each data point is tagged with a number representing the weight of mobile data access requests from the user. For this example, we assumed the battery levels of users are similar. The goal is to locate 4 WiFi access points (AP), each having a circular range with 20 meter radius to maximize the offloading benefit. The graph on the left shows an optimal solution that maximizes the total offloaded traffic from the entire area with given number of APs. 100 units out of total 131 data access requests are covered, yielding $O = 76\%$ total offloading ratio. However, this solution gives the optimal solution from operator's point of view and can only achieve $\sum U_j$=59% average user satisfaction in terms of individual data offloading of users:

$$\sum U_j = \left(\sum_{i=1}^{n=5} s_i\right)/5 \text{ where } s_1 = \frac{10+11+23+9}{53}$$

$$s_2 = \frac{2+1}{18},\ s_3 = \frac{6}{10},\ s_4 = \frac{10+12+13}{35} \text{ and } s_5 = \frac{3}{15}$$

The difference is caused by the distribution of data access requests from users in these selected areas. Moreover, total size of data requests from users is highly divergent. Therefore, selecting high weight data points without considering overall size of user's data requests does not work well in terms of user's average individual satisfaction from offloading.

Figure 2.b shows the optimal locations of APs considering the individual offloading ratios of users. Even though the locations of two APs is same as in previous solution (Figure 2.a), the remaining two APs are located around data requests which will overall increase the value of U. This results in O =75.5%



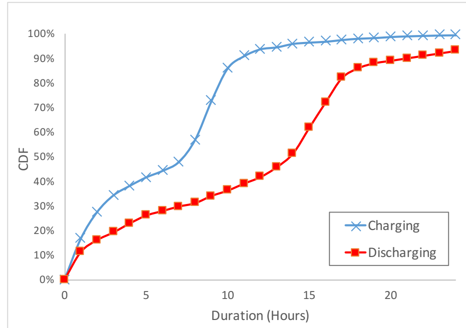


Fig. 4. Battery level and status statistics. (a) initiation of charging and discharging durations (b) CDF of charging and discharging durations.

(close to previous case) and a much higher U=79% with the same s_4 as in previous case but with other s_i values updated to:

$$s_1 = \frac{10+11+9}{53},\ s_2 = \frac{10+5+2}{18},$$
$$s_3 = \frac{6+3+1}{10},\ \text{and}\ s_5 = \frac{3+4}{15}$$

Finally, Figure 2.c shows the greedy-heuristic based solution with individual offloading ratios of users. Greedy approach can achieve very close user satisfaction results (U=78%) to optimal results in Figure 2.b for this specific example, while total offloading ratio achieved decreases to O =68%.

A sample satisfaction function for users depending on their battery levels (with same data request) could be similar to the one in Figure 3. The graph is simply showing that at the very high and low battery charge levels, the change in battery level will not change user's satisfaction much. However, in between these end points, the user's satisfaction will be affected.

IV. SIMULATIONS

In this section, we present the results of simulations performed on a real user data set. To this end, we used an online location-based social network dataset to capture the user data access requests. Specifically, we used the Gowalla dataset and considered the check-ins as the representative weight of data requests in each location. For these simulations, we analyzed the check-ins coming from users in San Francisco area.

For the simulation of battery usage characteristics accurately, we did use the patterns that we extracted from Device Analyzer dataset [18]. These are shown in Figure 4 and Figure 1.c. In other words, we generated a charging pattern for each device by setting a starting time of charging with probabilities observed in Figure 4.a. Then, we assigned a charging duration with distribution given in Figure 4.b. Given these conditions, we also verified the expected device charging status shown in Figure 1.c.

The total data usage distribution among all users and the distribution of each request weight of every user within its total usage are the two significant parameters that will affect the results. As the analysis in many studies [21], [22], [23], [24] on characterization of mobile data traffic shows, these distributions mostly fit to power-law distribution ($Cx^{-\alpha}$). We also observed such distribution within location based social network data set. Finally, the distribution of user data requests to the potential AP locations (or grids representing the coverage area) is also significant. This distribution depends on the geographic location analyzed. For some city-wide regions such as San Francisco downtown on which we perform our simulations, power-law distribution also fits well.

In Figure 5, we show the results that compare the proposed user-based greedy based approach with the aggregated greedy based one in [4]. We first measured the total offloaded user data ratios from all users. As Figure 5.a shows, user-based approach can achieve closer offloading ratio to aggregated greedy approach for different number of APs deployed. On the other hand, as Figure 5.b shows, user-based approach can show better average offloading ratio per user (changes in range of 15-20% for different AP counts). This is simply due to selection process of users in user-based approach which gives preference to users with data points which may not have large volume within all user data points but could be covering a large portion of the user's total data points. In Figure 5.c, we show the CDF of the distribution of offloading ratios per user in the network (with 400 AP deployment). As it is expected, user-based approach can achieve a balanced distribution among all users, thus, the CDF is very linear. Finally, Figure 5.d shows the improvement achieved (in range of 30-40%) in battery based satisfaction ratio in users with user-based approach compared to aggregated approach. As the former gives preference to regions with lower average battery level (from the users in that region) compared to others, the selection process result in deploying APs to such regions when data request weights are similar. As a result, users given offloading opportunity with low level batteries become more satisfied compared to the case where this difference is not considered in selection process (aggregated approach).

V. RELATED WORK

Deployment of WiFi APs has been studied for different goals in the literature [11], [12], [13]. Liao et al. [1] propose an algorithm to deploy minimum number of APs that simultaneously provides full communication coverage and can locate a mobile device with a given accuracy parameter. There

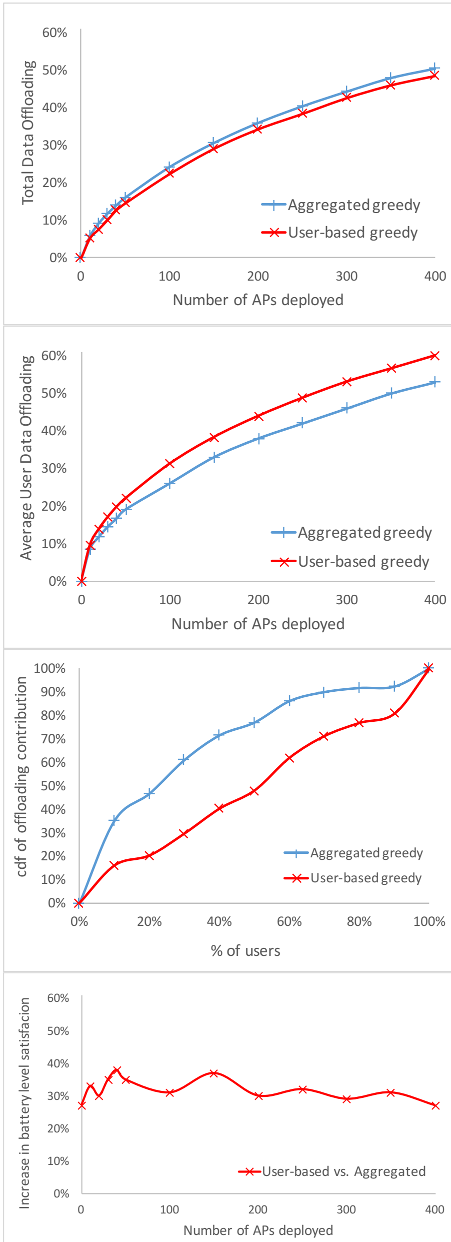


Fig. 5. (a) Total offloaded aggregated data ratio and (b) average offloaded user data ratios in aggregated greedy and user-based greedy algorithms.(c) The CDF of distribution of user offloading ratios and (d) improvement in battery-level satisfaction ratio.

are also a few studies that propose WiFi AP deployment algorithms with the goal of maximizing cellular offloaded data. In [2], AP locations are decided in a sequential manner without considering the efficiency of deployment. Similarly, in the HotZones algorithm proposed in [3], APs are deployed to cover the areas of most used cell towers. However, the user traffic distribution inside each macrocell coverage area has not been considered. Thus, APs are not efficiently deployed. A more granular deployment algorithm in a city-wide scenario is studied in [4]. There are also some works that study the recruitment of third-party WiFi access points via incentives and auctions [16], [19], [20].

All of these works mainly approach the problem from operator's point of view without considering the user satisfaction in WiFi offloading domain. There have been some work which considered user satisfaction in the context of delayed offloading [16]. But these studies still consider overall user satisfaction not individual user satisfaction. In contrast, we do consider user satisfaction based on average individual data offloading ratios and battery level changes. The proposed idea could be extended to application of offloading (not just to deployment of APs) and could be utilized in giving priority to users which otherwise would not be benefiting from offloading opportunity.

VI. Conclusion

In this paper, we study the WiFi AP deployment within the context of offloading cellular networks. Our objective is to increase average user satisfaction and give the opportunity of offloading to each user equally while still trying to maximize the overall offloading ratio as much as possible. Simulation results on real user data set show that the proposed approach can help increasing average user satisfaction (in terms of average user offloading ratios and battery level based satisfaction) while keeping the aggregate offloading ratio close to maximum possible values. In our future work, to see the impact of several factors on results (such as the coefficients of data usage distribution among all users and among each user's different data requests), we will evaluate the proposed approach in different real and generated datasets. Moreover, we will work on analytical derivation of these gains and loses depending on the parameters defined and try to find their upper and lower bounds.

References

[1] L. Liao, W. Chen, C. Zhang, L. Zhang, D. Xuan, and W. Jia, Two Birds With One Stone: Wireless Access Point Deployment for Both Coverage and Localization, IEEE Transactions on Vehicular Technology, Vol. 60, No. 5, June 2011.

[2] S. Dimatteo, P. Hui, B. Han and V. O. K. Li, *Cellular Traffic Offloading through WiFi Networks*, in Proc. of IEEE MASS, 2011.

[3] N. Ristanovic, J.-Y. L. Boudec, A. Chaintreau, V. Erramilli, Energy Efficient Offloading of 3G Networks, in Proc. of IEEE MASS, 2011.

[4] E. Bulut, and B. Szymanski, *WiFi Access Point Deployment for Efficient Mobile Data Offloading*, in Proceedings of ACM International Workshop on Practical Issues and Applications in Next Generation Wireless Networks (PINGEN 2012) at Mobicom 2012., Istanbul, Turkey, August 26, 2012, pp. 45-50.

[5] R. L. Church, C. S. ReVelle, *The maximal covering location problem*, Papers of the Regional Science Association, 32 (1974), pp. 101-118.

[6] N. Megiddo, E. Zemel, ans S. L. Hakimi, *The maximum coverage location problem*, SIAM J. Alg. Disc. Meth., Vol. 4, No.2, June 1983.

[7] H. Pirkul, and D. Schilling, *The maximal covering location problem with capacities on total workload*, Management Science, Vol. 37, No. 2, pp 233-248, 1991.

[8] N. Balasubramanian , A. Balasubramanian , A. Venkataramani, *Energy consumption in mobile phones: a measurement study and implications for network applications*, Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, November 04-06, 2009, Chicago, Illinois, USA.

[9] G. Kalic, I. Bojic, and M. Kusek, *Energy consumption in android phones when using wireless communication technologies*, in Proc. of 35th Int. Convention on Information and Communication Technology, Electronics and Microelectronics, 2012.
[10] A. Carroll, and G. Heiser, *An analysis of power consumption in a smartphone*, In USENIX, Boston, MA, USA, June 2010.
[11] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, *Mobile data offloading: how much can wifi deliver?* in Proc. of Co-NEXT 10, pp. 26:1-26:12, 2010.
[12] S. Dimatteo, P. Hui, B. Han and V. O. K. Li, *Cellular Traffic Offloading through WiFi Networks*, in Proceedings of 8th IEEE Int. Conf. Mobile Ad Hoc and Sensor Systems (IEEE MASS 11), 2011.
[13] B. Han, P. Hui, and A. Srinivasan, *Mobile data offloading in metropolitan area networks*, SIGMOBILE Mob. Comput. Commun. Rev., vol. 14, pp. 28-30, November 2010.
[14] A. Balasubramanian, R. Mahajan, and A. Venkataramani, *Augmenting Mobile 3G Using WiFi*, in ACM MobiSys'10, Jun. 2010, pp. 209-222, 2010.
[15] Balasubramanian, N., Balasubramanian A. and Venkataramani, *A. Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications*, in Proceedings of the 9th Internet measurement conference, November 2009, pp. 280-293.
[16] X. Zhuo, W. Gao, G. Cao, and Y. Dai, *WinCoupon: An Incentive Framework for 3G Traffic Offloading*, in Proc. of IEEE ICNP, 2011.
[17] E. Cho, S. A. Myers, J. Leskovec, *Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks*, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011.
[18] D. Wagner, A. Rice and A. Beresford, *Device Analyzer: Understanding smartphone usage*, 10th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Tokyo, Japan, December 2013
[19] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, *A double auction mechanism for mobile data offloading markets*, IEEE/ACM Transactions on Networking (Sept. 2014).
[20] W. Dong, S. Rallapalli, L. Jana, R.and Qiu, K. Ramakrishnan, L. Razoumov, Y. Zhang, and T. W. Cho, *Ideal: Incentivized dynamic cellular offloading via auctions*, in Proc. IEEE INFOCOM, 2013.
[21] X. Zhou, Z. Zhao, R. Li, Y. Zhou, J. Palicot, and H. Zhang, *Human mobility patterns in cellular networks*, IEEE Communication Letters, vol. 17, no. 10, pp. 1877-1880, Oct. 2013.
[22] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, *Understanding traffic dynamics in cellular data networks*, in Proc. IEEE INFOCOM, Apr. 2011, pp. 882-890.
[23] M. Z. Shaq, L. Ji, A. X. Liu, J. Pang, and J. Wang, *Characterizing geospatial dynamics of application usage in a 3G cellular data network*, in Proc. IEEE INFOCOM, Mar. 2012, pp. 1341-1349.
[24] E. Hazan, P. Pont, and K. Roche, *Averages lie: Using smart segmentation to find growth*, McKinsey iConsumer Research, April, 2014.
[25] G.P. Perrucci, F.H.P Fitzek, Q. Sasso, W. Kellerer, J. Widmer ,*On the impact of 2G and 3G network usage for mobile phones' battery life*, European Wireless Conference, pp. 255-259, May. 2009.