A New Energy Efficient Big Data Dissemination Approach Using the Opportunistic D2D Communications

Ambreen Memon(✉), William Liu, and Adnan Al-Anbuky

School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand
ambreen.memon@aut.ac.nz

Abstract. The emerging cyber-physical paradigm endeavours to unite all the physical objects embedded with electronics, software, sensors, and network connectivity to allow more direct interactions and information sharing between the physical and cyber worlds. While these massively connected devices and their associated communications can exponentially increase the data generation, transmission, and processing which consume a huge amount of energy and finally end up with harming the environment seriously. In this paper, we propose a solution for energy efficient data dissemination by using the opportunistic device-to-device (D2D) communications. Each sender can decide either use network infrastructure or through encountering the end-users according to the quality of service (QoS) requirements of each data demand and also the mobility behaviors of the users. These decisions are based on the time and location- traces of daily mobility routines and related activities of users and their social relationship. The case study, based on the similarity analysis of the mobility traces, has confirmed the rich opportunities for encountering among people, thus the proposed approach has great promises to reduce the energy consumption of big data dissemination.

Keywords: Big data · Opportunistic routing
Delay tolerant network · Energy efficient data dissemination
Similarity analysis

1 Introduction

The increase in the use of mobile devices has changed the way that users share data and ultimately leading to mobile traffic expansion exponentially since 2011. As per current expectation, more than 24 ex-bytes of mobile traffic will be navigating operators' networks by 2019, with 72% of this traffic being created by the interactive media [1]. In addition, the growing number of mobile devices and their communications have increased significantly the energy consumption.

Delay tolerant networks (DTN) perform store-carry-forward routing to deliver the data in an end-to-end fashion, although a continuous end-to-end

P. H. J. Chong et al. (Eds.): SmartGIFT 2018, LNICST 245, pp. 155–164, 2018.
https://doi.org/10.1007/978-3-319-94965-9_16

communication path may never exist between sender and destination devices. The integration of the infrastructure-based networks and DTN has shown the benefits since it can boost routing performance and offload traffic from the congested infrastructure networks.

This paper targets to develop a novel eco-friendly and sustainable data transmission approach for offloading the data traffic from the infrastructure to the opportunistic- and social- based device-to-device (D2D) communications. It can be performed by exploring the existing movements and spatial closeness relation among devices. To complement the traditional infrastructure-based data transmission, the new idea is to optimally piggyback data on the moving physical devices for data dissemination to achieve the energy reduction as well as to ensure the QoS requirements. In such way, the proposed approach can fully utilize the users' historical mobility traces to predict the next location. If they are close enough and also satisfy the QoS requirements, the data could be directly transferred using D2D communication which consumes less energy. Otherwise, data could be transferred through the infrastructure-based network.

The rest of the paper is organized as below. Section 2 has reviewed the recent advancements in the areas of D2D communication, delay tolerant network, mobility models and their impact on energy consumption. Section 3 presents the energy consumption model for data dissemination. We introduce the new data dissemination approach by using similarity analysis in Sect. 4. We have conducted a case study in Sect. 5 to validate the new approach and confirm its great promises on substituting the infrastructure-based transmission approach for those delay tolerant data services. Finally, we conclude the main contributions and also layout the future work in Sect. 6.

2 Related Work

In D2D communications, the devices in proximity can interface with each other directly and construct a communication network. Data traffic can be offloaded to the D2D network instead of transmission through the infrastructure based network. For instance, by authorizing D2D communications, some users can download the substances from the cellular base station (BS) while others could get the substances from their associates. Therefore, the D2D communications can significantly reduce the traffic congestions and also energy consumption in networks [2]. In future wireless access networks, balancing the traffic load among base stations can be accomplished through adjusting the user's BS affiliations [3].

At Massachusetts Institute of Technology (MIT), predicting the user behavior is one of their research themes. Time duration is recorded by mobile phones when adjacent to cell tower IDs and Bluetooth devices. While Bluetooth devices demonstrate different behaviors based on how are the devices related to each other. It has been found that the presence of business students in a similar area is performing the same activities [4]. Bluetooth signals were constructed within individual's house to check the accuracy of data transmission into locations by cell towers [5]. Estimating next location of the objectives by using the dynamic

Bayesian Network can reach a successful rate of 93% to 99%. In prediction, the next cell is a sequence of location that they investigated in communication areas, need to improve those resources for reservation and QoS requirements [6]. Zeibart et al. [7] predict that driving to the destinations, given a partially travelled route by calculating the probabilities of different possible routes. Bauer and Deru [8] suggest the ways of predicting future destination along with previous histories. The work in [9] has used the Naive Bayesian classifier based model, which consists of the time slot of days, weekends and 1–8 h. Everyday time stamp is divided into multiple records, that consist of a list of Bluetooth MAC addresses and locations for mobility prediction.

3 Energy Consumption Model

A network can be represented by a graph G (N, L), where N is the number of nodes and L is the direct links (i, j) or edges between graph nodes. For prototyping the key ideas, We assume a general and simplified energy consumption model for wireless or wired energy dissipation where the transmitter dissipates power to generate the radio or line electronics. The power amplifier then consumes energy to transmit the traffic, and the receiver dissipates energy to receive and process the radio or line electronics, as shown in Fig. 1. Taking radio transmission as an instance, the power control can be used to remedy the signal propagation loss by appropriately setting the power amplifier. For example, if the transmission distance is less than a threshold value d_0, the free space prorogation model with the attenuation parameter of ε_{fx} is used, otherwise the multi-path (mp) propagation model with the attenuation parameter of ε_{mp} is used. For the sender to transmit a volume of k-bits data to the receiver where there is a distance of d away, the energy consumption model can be calculated as below:

$$E_{Tx}(k,d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \tag{1}$$

$$E_{Tx}(k,d) = k \cdot E_{elec} + k \cdot \varepsilon_{fx} \cdot d^2, d < d_0 \quad = k \cdot E_{elec} + k \cdot \varepsilon_{mp} \cdot d^4, d > d_0 \tag{2}$$

where, E_{elec} is the energy consumed by the transmitter and it depends on the factors such as digital coding, modulation and filtering signal processing procedures. As for the energy consumed by amplifier, it depends on the distance to the receiver and the acceptable bit-error rate. Energy consumption for the received data can be calculated by:

$$E_{Rx}(k) = E_{(Rx-elec)}(k) = k \cdot E_{elec} \tag{3}$$

Based on the above general energy consumption model for communications, we can see that the volume of data k and the transmission distance d are two critical and changeable factors which can vary the overall energy consumption, compared to the energy consumed by the electronic components and signal processing mechanism in the transmitter, receiver and also the relay amplifiers. It is possible to reduce the transmission distance d which is being traversed through the infrastructure. In other words, the more transmission distance can

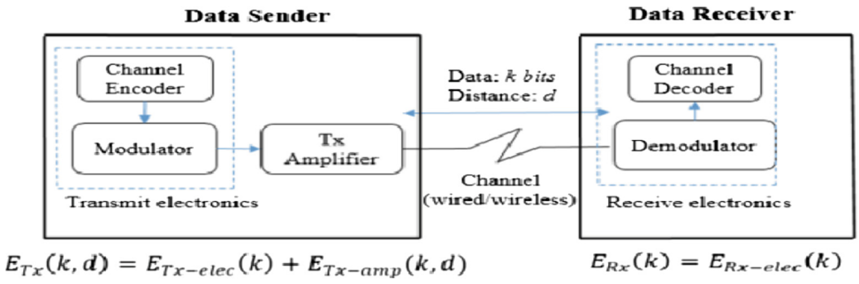

Fig. 1. Energy consumption model

be shortened, the more energy consumption can be reduced. This is the motivation triggering us to propose a new energy efficient data dissemination approach (EEDDA), by fully utilizing the mobility of human (i.e., mobile users) and D2D communications. They carried the data for delivery based on the prediction and users' mobility similarity analysis, especially for those communications services such as file transfer which has delay tolerant characteristic.

4 The Proposed EEDDA Approach

The proposed EEDDA approach, as shown in Fig. 2 has four processes including:

1. Data Collection
   In EEDDA approach, the first step is to collect the users' mobility data. In our case study, we reuse the data gathered by the project of Wireless Topology Discovery (WTD) that was handled at UCSD [10]. It has the traces of 300 people's accessibility of PDAs to WiFi. All the traces has two portions of discussion. One portion consisted of trace data. The other file contained the known locations with access points for local coordinates. Eleven-week trace duration started from the 22 September 2002 to 8 December 2002 was the data collection period.
2. Analysis
   The WTD has sampled and recorded the above information for all access points (APs) for every 40 s, which may fill in all its frequencies. The analysis of the collected data was conducted while running on a student's device. During a sample, the above information is given by WTD for all sensed samples. In Fig. 3 below, the three entries were recorded if a device has three APs in one sample (the entries include the IP address, signal strength, and attached flag). To extract basic records that show the user's location, the Associate field is used in this study. The user's device is located near the connected access point (Associated = 1) is based on the assumption. A list of neighboring access points is created for the sensitive access points that were not selected for the Association (Associated = 0). To record that a user should be decided at any time, AC-POWER field can be used for it. The assumption is that

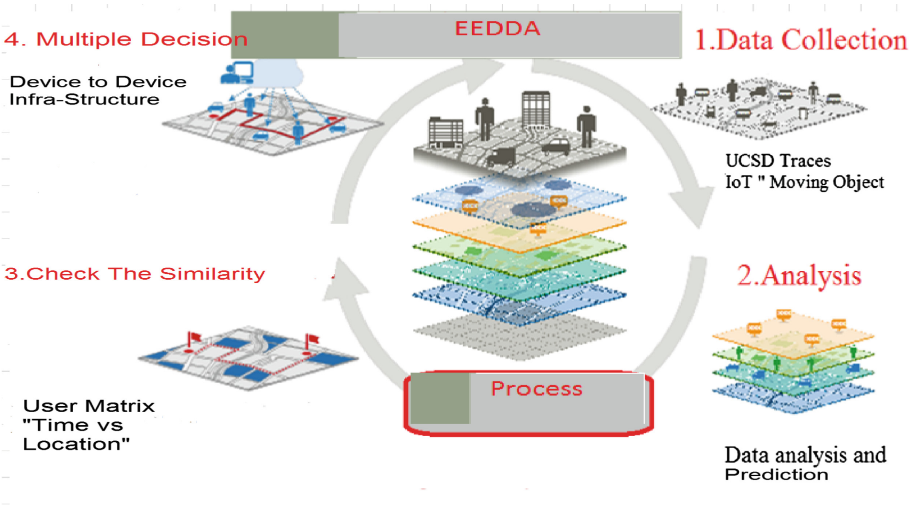

Fig. 2. The Proposed EEDDA approach

USER_ID	SAMPLE_DATE & TIME		AP_ID	SIG_STRENGTH	AC_POWER	ASSOCIATED
123	Sep-22	0:00:00	359	8	0	0
123	Sep-22	0:00:00	363	5	0	0
123	Sep-22	0:00:00	365	11	0	1

Fig. 3. The fields and usage of the database

the user is not mobile and AC-POWER = 1 is a plug-in in the device. In weight gain of individual access points, the given time, signaling power SIG-STRENGTH can be used.

In different location alignment algorithms, the SIG-STRENGTH field can be used. For example, a user is using the trilateration algorithm then it is between all access points. The AP-ID uses only one assigned number to label every access point in the field databases. Here we use these values as a location label, and when evaluating future locations, these values are developed by the model. In a real application, AP-ID field will be backed up on the map or on a map named on a useful location. The SAMPLE-TIME contains the date and time. The recorded state of every 20 s and end of 11 weeks period by their own devices. Throughout the pre-processing, the week fragments at the start and end of during the 11 weeks were rejected so as to provide 10 whole weeks' samples with User ID field for each record to the specific user. This algorithm does not use the user information, but simply uses the user-id field for partitioning the logs into individual user logs. The prediction can be done by individual user but not on the entire group.

3. Check Similarity

The sequence prediction is to predict the next item in a sequence, which can be considered as a kind of rating. There are potential results for the alphabet used to create a layout. They are known in advance and predict the model in which the next item is in sequence. First signify theorizing symbols S_1, S_2, S_3, S_4, S_n. The n represents the number of symbols in the alphabet. When training sequences are described with the symbol of t:

$$Where\ X_i \in \sum X_1, X_2, X_3 ... X_t \tag{4}$$

this calculation defines the conditional probability

$$Pr\left\{X_{t+1}x_{t+1}|X_t = x_t, x_{t-1} = x_{t-1}\right\} \tag{5}$$

This calculation has been used in stationary Markov chain [11]. In our case, we are considering the stationary one, because the probabilities are not only depend on the same time, even sub-sequences repeat at the same time but with different location in sequences for each repeat or shift S and for all X_i.

$$\begin{aligned} Pr\left\{X_1, X_2 = x_2.....x_n X_1, X_2 = x_2.....x_n\right\} \\ Pr\left\{X_{1+m} = X_1, X_{2+M} = x_2.., X_{n+m} = x_n\right\} \end{aligned} \tag{6}$$

This process is called Markov model because the probability is likely to be on the variable. The number used is the variable for variants, L, model length, or order. The pre-variable sub-division is called history or context. If the contextual length of the context is set continuous, the model is called fixed length Markov chain [12]. The variable length Markov chain with length L, the context length used to vary the maximum number of L on the prediction of the Markov channel. The first order of Markov model makes the basis for prediction model here.

Moreover, the raw data from the WTD experiment combines all logs, from all operators into one file. The entire logs of all the operators are associated with the document by the raw data from the WTD test. The first step of associating all logs from different users into one folder was to divide the raw information into various documentation for an individual user. Records usually here refer to as sensed, non-associated, access points. Records with the same date and time had the maximum amount of indication matrix. Whereas the rest records were rejected. Reports with similar user name, access point and having a starting time within one minute of the preceding record's starting time are termed as contiguous records. The output showed the length of each session collected over a day. The statistics recorded by a full-time active mobile device is regarded as the first movement data. The data representing users' significant locations revealing social interacting applications is called the destination data. In this work, significant locations are determined solely by a length of stay of at least 10 min [12]. The destination data and movement could be considered as movement location and significant location

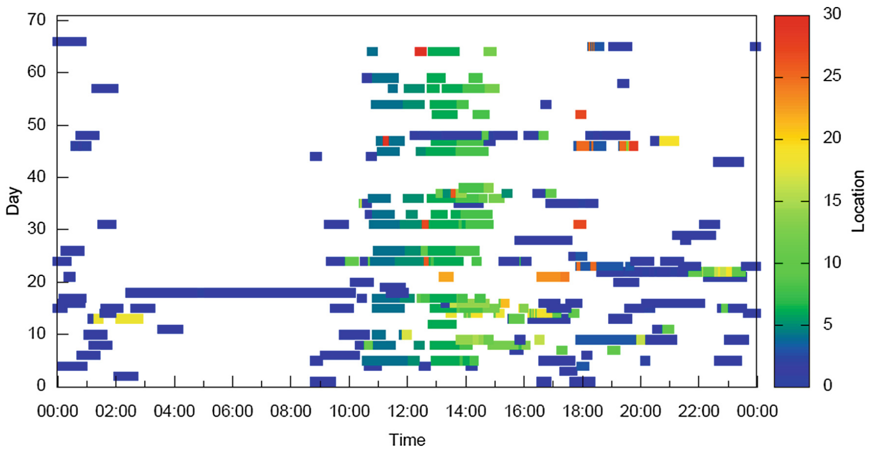

Fig. 4. Data (1-minute or less) for User 003 (Color figure online)

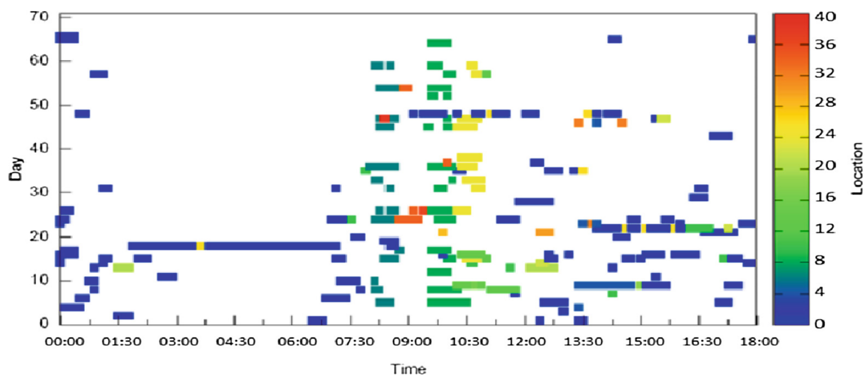

Fig. 5. Data (10-minute) for User 003

individually. The starting time is measured to be on entire number nearest to the last minute dataset of MoveLoc. Later on, all the duration include those, slightly lesser than 20 s, were included of one minute window. When the session started then it was considered as the highest period of one-minute time duration, declining the rest our work prediction on the future location and time. The Fig. 4 shows the MoveLoc data for user 3. The color bar shows that this user visited 40 locations, over the 10-weeks recording period. The x-axis represents the time of day. Anyone can observe that this user had some regular locations between 1:00 pm and 2:00 pm and some of the movement between locations is indicated by the change in colors/shades. Figure 5 depicts that User 3 MoveLoc (1-minute or less) and User 3 SigLoc (10- minute) data and dataset location. Where user consumed the minimum of 10 min, were covered in the SigLoc dataset. Elimination other sessions less than 10 min was the first phase of producing this assessment. The start-time of the other sessions were almost equalized to the bordering 10 min. The SigLoc data for User 3 is indicated in Fig. 4. One can notice that the number of locations fell from 40 to 20 and the transitions between locations were removed. Though the existing point of proof-of-concept on EEDDA, which does not advance towards a far-reaching position, yet the study outcome is relatively stimulating.

4. Multiple Decision
   This is the last process of EEDDA, of which the device will be able to take the decision as per the result of the similarity analysis. The data is either transmitted through the infrastructure based network or D2D communications. The Fig. 6 shows the flow chart of the decision process.

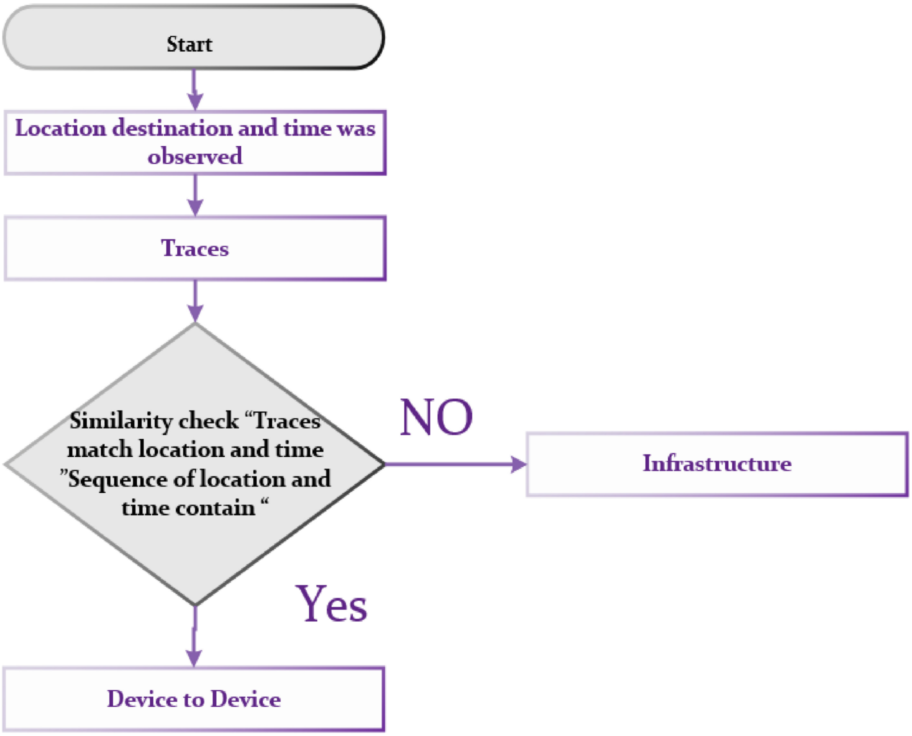

Fig. 6. The process of multiple decision

5 A Case Study

We have identified that the data transmission distance is a key but changeable factor contributing to the overall energy consumption. Here we consider a scenario, of which a professor in university and his students have communicated with each other very frequently. They are all in the same building but on the different floors. We assume that the data is delay tolerantable. In such way, we have two options including option A: send the data through infrastructure-based network, and option B: infrastructure-less approach i.e., D2D communications.

To disseminate data through these options, we assume further two possible ways: (a) Human mobility traces get matched according to the movement behavior. In such way, the proposed approach can fully utilize the benefits of mobility traces and check similarity of location and time. When they are matched, the data can transfer through D2D while consuming less energy. (b) In this option, we use the prediction model with human mobility traces that checks predicted location and time of user's movement. When it matches, the data will transfer by D2D communications.

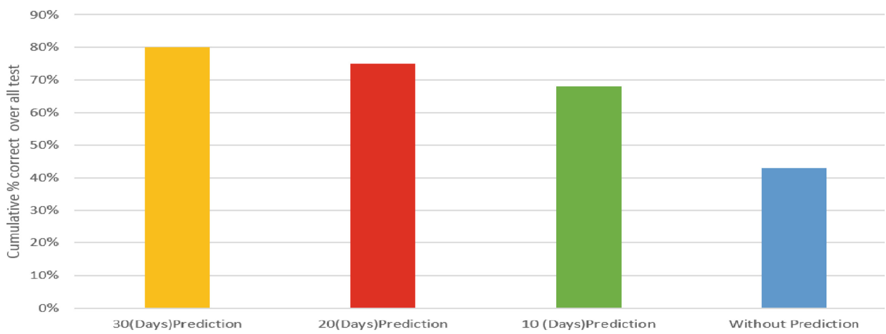

Fig. 7. Test result with 30 days, 20 days, 10 days interval prediction and without prediction

Figure 7 shows the correctness of the results as per time allocated for prediction. If we simply use the traces, the accuracy of the results is 45%, while considering delay for about 10 days interval. While in the prediction option, it gets increased by 70%. However, when we increase the interval of delay for about 20 days or 30 days, the accuracy could be better than the previous. This indicates that the longer tolerated delay and mobility duration, the higher opportunity that two users encounter each other to have D2D communications, which could reduce the energy consumption significantly.

6 Conclusion and Future Work

In this paper, a novel energy efficient data dissemination approach (EEDDA) is introduced, of which a mobile encounter between the communicating pairs is sought for directly exchanging data. The similarity analysis framework reveals the mobile encounter opportunities among communication peers. Peer device similarity depends on the moving ability and adaptability of interacting with the devices. The work is ongoing for developing communication protocol and algorithm but also for peer countering and automation. The greater number of complex mobility models and the unpredictability of devices' movements should be considered too to develop more adaptable EEDDA in future work.

References

1. Micinski, K.: Interaction-based security for mobile apps. Ph.D. dissertation, University of Maryland, College Park (2017)
2. Kjærgaard, M.B., Langdal, J., Godsk, T., Toftkjær, T.: Entracked: energy-efficient robust position tracking for mobile devices. In: Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, pp. 221–234. ACM (2009)
3. Dogar, F.R., Steenkiste, P., Papagiannaki, K.: Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 107–122. ACM (2010)
4. Ziebart, B.D., Maas, A.L., Dey, A.K., Bagnell, J.A.: Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 322–331. ACM (2008)
5. Elnekave, S., Last, M., Maimon, O.: Predicting future locations using clusters' centroids. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, p. 55. ACM (2007)
6. Roddick, J.F., Spiliopoulou, M.: A survey of temporal knowledge discovery paradigms and methods. IEEE Trans. Knowl. Data Eng. **14**(4), 750–767 (2002)
7. Mayrhofer, R., Radi, H., Ferscha, A.: Recognizing and predicting context by learning from user behavior. na (2003)
8. Holbach, H., Bourassa, M.A.: The effects of gap wind induced vorticity, the ITCZ, and monsoon trough on tropical cyclogenesis. In: 2012 International Ocean Vector Winds Science Team Meeting (2012)
9. Bellotti, V., Begole, B., Chi, E.H., Ducheneaut, N., Fang, J., Isaacs, E., King, T., Newman, M.W., Partridge, K., Price, B., et al.: Activity-based serendipitous recommendations with the magitti mobile leisure guide. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1157–1166. ACM (2008)
10. UCSD WTD Trace. http://sysnet.ucsd.edu/wtd/
11. Yang, G.: Discovering significant places from mobile phones – a mass market solution. In: Fuller, R., Koutsoukos, X.D. (eds.) MELT 2009. LNCS, vol. 5801, pp. 34–49. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04385-7_3
12. Vu, L., Do, Q., Nahrstedt, K.: Exploiting joint wifi/bluetooth trace to predict people movement (2010)