Institut Eurécom
Department of Corporate Communications
2229, route des Crètes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-08-212
Characterizing User Mobility in Second Life
Chi-Anh La, Pietro Michiardi

Tel : (+33) 4 93 00 81 45
Fax : (+33) 4 93 00 26 27
Contact at Eurecom : {Pietro.Michiardi}@eurecom.fr

Abstract

In this work we present a measurement study of user mobility in Second Life. We first discuss on different techniques to collect user traces and then focus on results obtained using a *crawler* that we built. Tempted by the question whether our methodology could provide similar results to those obtained in real-world experiments, we study the statistical distribution of user contacts and show that from a qualitative point of view user mobility in Second Life presents similar traits to those of real humans. We further push our analysis to study the properties of line of sight networks that emerge from user interaction as well as the spatial properties of user movements and observe that users in Second Life revolve around several point of interests traveling in general short distances.

1 Introduction

This work stems form prior studies on human mobility performed in real life. For example, [4–6] conduct several experiments mainly in confined areas and study analytical models of human mobility with the goal of assessing the performance of message forwarding in Delay Tolerant Networks (DTNs). Each user taking part to such experiments is equipped with a wireless device (for example a sensor device, a mobile phone, ...) running a custom software that records *temporal information* about their contacts. Individual measurements are collected, combined and parsed, originating elegant but complex algorithms [5] because the only available information is the temporal distribution of contact times, which are bound to the specific wireless technology used in the experiments.

In this paper we present a novel methodology to capture spatio-temporal dynamics of user mobility that overcomes most of the limitations of previous attempts: it is cheap, it requires no logistic organization, it is not bound to a specific wireless technology and can potentially scale up to a very large number of participants. Our measurement approach exploits the tremendous raise in popularity of Networked Virtual Environments (NVEs), wherein thousands of users connect daily to interact, play, do business and follow university courses just to name a few potential applications. Here we focus on the SecondLife (SL) "metaverse" [3] which has recently gained momentum in the on-line community.

Our primary goal is to perform a temporal, spatial and topological analysis of user interaction in Second Life. Prior works that attempted the difficult task of measuring and collecting traces of human mobility and contact opportunities are restricted by logistic constraints (number of participants to the experiments, duration of the experiments, failures of hardware devices, wireless technology used). In general, position information of mobile users is not available, thus a spatial analysis is difficult to achieve [5]. Some experiments with GPS-enabled devices have been done in the past [7,8], but these experiments are limited to outdoor environments.

In this paper we discuss two monitoring architectures that we tested in our laboratories and focus on the most robust technique, which is based on a custom software module (termed a *crawler*). Our crawler connects to SL and extracts position information of all users concurrently connected to a sub-space of the metaverse: all results presented in this paper have been obtained with this architecture.

One striking evidence of our results is that they qualitatively fit to real life data, raising the legitimate question wether measurements taken in a virtual environment present similar traits to those taken in a realistic setting. Our methodology allows performing large experiments at a very low cost and generate data that can be used for trace-driven simulations of a large variety of applications: the study of epidemics and information diffusion in wireless

3

networks are just some prominent examples.

2   Monitoring architectures

Mining data in a NVE can be approached under different angles. The first architecture we discuss exploit SL and its features to create objects capable of sensing user activities in the metaverse. Although this approach present interesting features, there are several limitations that hinder our ultimate goal, which is to collect a large data set of user mobility patterns. These limitations mostly come from inner design choices made by the developers of SL to protect from external attackers aiming at disrupting the system operation. The limitations incurred by the first approach can be circumvented by building a *crawler* which is not an object of the SL metaverse.

The task of monitoring user activity in the **whole** SL metaverse is very complex: in this work we focus on measurements made on a selected subspace of SL, that is called a land (or island). In the following we use the terminology *target land* to indicate the land we wish to monitor. Lands in SL can be private, public or conceived as sandboxes and different restrictions apply: for example private lands forbid the creation and the deployment of objects without prior authorization.

We now detail the monitoring architectures we investigated in our work.

A sensor network architecture[1]**:** Our first approach has been inspired by current research in the area of wireless sensor networks: it resembles to what one would do in the real world to measure physical data (temperature, movements, etc ...) by deploying sensor devices in the area to be monitored. We built *virtual* sensors using the standard object creation tool accessible from a SL client software. Our sensors collect data and communicate with an external web server that stores the location information of users connected to the target land. The functionality of a sensor is defined using a proprietary scripting language [2].

A key limitation imposed by the infrastructure of SL is that sensors cannot be arbitrarily deployed on any land. While it is impossible to deploy object on private lands without authorization, on public lands objects expire after a TTL, which is equal to 5 minutes. In our architecture, before a sensor is removed from the land, another sensor will take care of replicating it in the same position.

When a sensor is deployed on the target land, it detects users (a maximum of 16 users can be detected) that fall within the sensing range (96 meters) with a tunable periodicity and stores this information in its local cache (16KB is the maximum storage space). Due to its limited amount of memory, a sensor initiate a connection with the base station (our web server) and flushes its memory using the HTTP protocol as soon as the maximal

---

[1]This approach has been used also by our colleagues in [9].

capacity has been reached. The technical specification of a sensor impose several challenges that hinder the task of covering an entire land. Because of the limitation on the maximum number of users that is possible to monitor with a single sensor, the actual deployment of the monitoring infrastructure becomes a tedious manual exercise.

Moreover, the number of HTTP messages that can be exchanged between sensors and the external web server is restricted by the SL infrastructure to limit the upload bandwidth charged to SL. This inner design choice reduces the quantity of data that can be retrieved from our sensors: a tradeoff exists between the granularity of the sensed data and the duration of a monitoring experiment.

Monitoring using an external crawler: We discuss here an alternative approach that relies on a custom, lightweight SL client program that we designed using `libsecondlife` [1]. Our *crawler* monitors the position of **every** user located on the target land: measurement data is stored in a database that can be queried through an interactive web application[2]. The crawler connects to the SL metaverse as a normal user, thus it does not incur in the limitations imposed by private lands: any accessible land can be monitored in its totality; the maximum number of users that can be tracked is bounded only by the SL architecture (as of today, roughly concurrent 100 users per land); communication between the crawler and the database takes place outside SL hence it is not limited.

During our experiments, we noted that introducing measurement probes in a NVE can cause unexpected effects that perturb the normal behavior of users (hence the measured user mobility patterns). Since our crawler is nothing but a stripped-down version of the legacy SL client and requires a valid login/password to connect to the metaverse, it is perceived in the SL space as an avatar, and as such may attract the attention of other users that try to interact with it: our initial experiments showed a steady convergence of user movements towards our crawler. To mitigate this perturbing effect we designed a crawler that mimics the behavior of a normal user: our crawler randomly moves over the target land and broadcasts chat messages randomly chosen from a small set of pre-defined phrases.

3 Measurement methodology

We create temporal snapshots of *line of sight communication networks* formed by users connected to a target land using their physical coordinates. Given an arbitrary communication range r, it is easy to determine if a communication link exists between any two users. In the following we use the temporal sequence of communication networks extracted from the traces we collected using our *crawler* and analyze contact opportunities between users, their

---

[2]Access to the application can be requested via mail to the authors.

spatial distribution and graph-theoretic properties of their communication network.

Before proceeding any further we discuss the methodology we used to find target lands and the choice of measurement parameters. Choosing a land in the SL metaverse to monitor user behavior is not an easy task: *i)* a large number of lands host very few users; *ii)* lands with a large population are usually built to distribute virtual money: all a user has to do is to sit and wait for a long enough time to earn (for free) money; iii) an automatic synchronization of the crawler to special events supposed to attract many users is very difficult to achieve. While we are currently working on a solution to the latter problem, we manually selected and analyzed the following popular areas: *Apfel Land*, a german-speaking arena for newbies; *Dance Island*, a virtual discotheque; *Island of View*, an open-space land in which an event (St. Valentines) was organized.

We launched the crawler on the selected target lands and set the time granularity (intervals at which we take a snapshot of the users' position) to $\tau = 10$ sec. Although it is possible to collect traces for long periods, in this paper we present results for 24 hours traces. In the following we selected a communication range r to simulate users equipped a bluetooth and a WiFi (802.11a at 54 Mbps) device, respectively $r_b = 10$ meters and $r_w = 80$ meters. In this work we assume an *ideal wireless channel*: line of sight networks extracted from our traces neglect the presence of obstacles such as buildings and trees. We will use $G(t_k) = G(v_i^{t_k}, e_{i,j}^{t_k})$ to identify a snapshot of the communication graph formed by users at measurement time t_k.

Note that user location in SL is expressed by her coordinates $\{x, y, z\}$ which are relative to the target land whose size is by default 256 m^2. However there is one exception: when a user sits on an object (e.g. a bench) her coordinates are $\{x = 0, y = 0, z = 0\}$. In this paper we neglect users that are sitting.

3.1 Temporal analysis

The metrics we use to analyze mobility patterns are inspired by the work of Chaintreau *et. al.* [4] and allow the analysis of the statistical distribution of contact opportunities between users:

- *Contact time (CT):* defined as the time interval in which two users (v_i, v_j) are in direct communication range, given r;

- *Inter-contact time (ICT):* defined as the time interval which elapses between two contact periods of a pair of users. Let

$$[t^1_{(v_i,v_j)s}, t^1_{(v_i,v_j)e}], [t^2_{(v_i,v_jj)s}, t^2_{(v_i,v_j)e}], ...[t^n_{(v_i,v_j)s}, t^n_{(v_i,v_j)e}]$$

be the successive time intervals at which a contact between user v_i and v_j occurs; then, the inter-contact time between the $k - th$ and the $(k + 1) - th$ contact intervals is:

$$IC_{(v_i,v_j)}^k = t_{(v_i,v_j)s}^{k+1} - t_{(v_i,v_j)e}^k$$

- *First contact time (FT):* defined as the minimum waiting time for a user v_i to contact her first neighbor (ever).

3.2 Spatial analysis

We present here the metrics we used to perform the spatial analysis of our traces:

- *Node degree:* the node degree is defined as the number of neighbors of a given node $n_i \in G(t_k)$ when the communication range is fixed to r and is termed $d_i^{t_k}$;

- *Network diameter:* the network diameter $D(G(t_k))$ is computed as the longest shortest path of the largest connected component of the graph $G(t_k)$. We used the largest component since, for a given r, $G(t_k)$ might be disconnected;

- *Clustering coefficient:* the clustering coefficient of node $v_i \in G(t_k)$ is defined as
$$C_{v_i} = \frac{2|e_{j,k}^{t_k}|}{d_i^{t_k}(d_i^{t_k} - 1)}$$
We compute the clustering coefficient of a graph [10] $G(t_k)$ as $C_{G(t_k} = 1/n \sum_{i=1}^n C_{v_i}$;

- *Travel length:* for every user v_i we compute the distance covered from its login to its logout to SL. We compute the CDF of the travel length for the whole duration of the measurement;

- *Effective Travel time:* for every user v_i we compute the total time spent while moving; hence, this metric neglects *pause* times;

- *Travel time:* for every user v_i we compute the total connection time to the SL land we monitor with the crawler;

- *Zone occupation:* we divided lands in several square sub-cells of size LxL and computed the number of users in every sub-cell, when $L = 20$ meters.
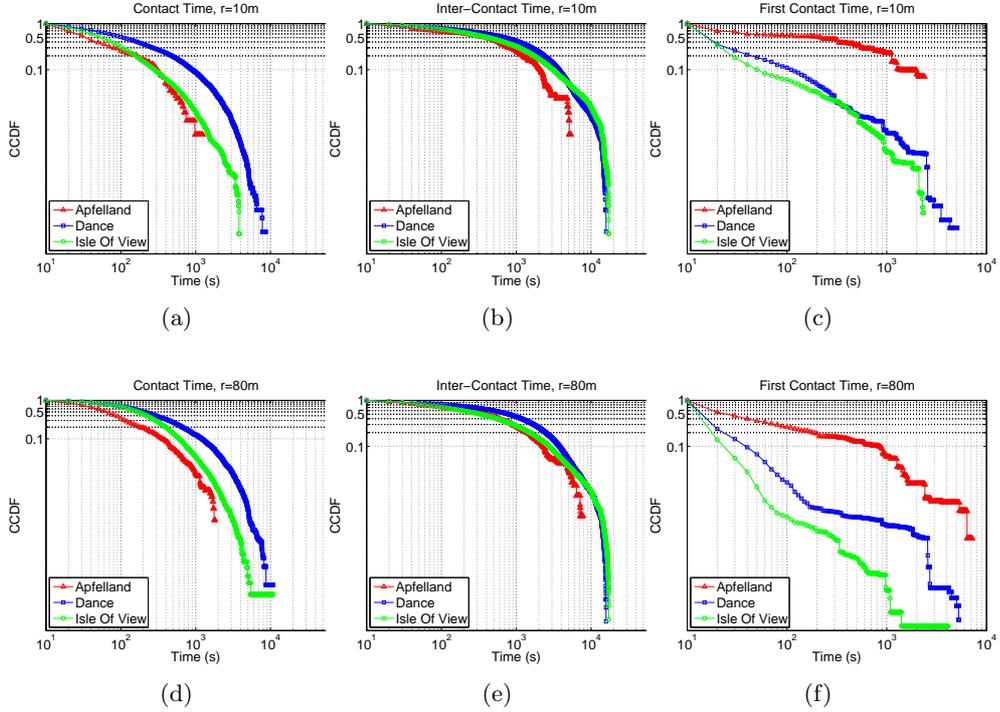
7

Figure 1: *Temporal Analysis:* Complementary CDF of contact opportunity metrics for three target lands.

4   Results

We now discuss on the results of our measurements for the three selected target lands. We study the influence of one parameter, *i.e.* the communication range (r_b or r_w), and three typologies of user mobility corresponding to the lands we monitored: one (Dance island) that reflects users in a confined space and two (Apfel Land and Isle of View) that are representative of users in an open-space.

Temporal Analysis: Fig. 1 illustrates the distribution of the temporal metrics we used in this work for $r_b = 10$ meters and $r_w = 80$ meters. A glance at the complementary CDF (CCDF) of the contact time CT indicates that the *median* contact time is roughly 30, 60 and 100 seconds respectively for Apfel Land, Isle of View and Dance Island when $r = r_b$, and about 70, 200 and 300 seconds for the same set of islands when $r = r_w$. Besides the intuitive result which indicates larger transfer opportunities for larger r, we observe that the distribution of CT has two phases: a first power-law phase and an exponential cut-off phase that limits the CT to a few hundreds seconds.

Similar observations can be done for the CCDF of the inter contact time

ICT: for the three target lands we analyzed, the distribution follows a first power-law phase, followed by an exponential cut-off phase. The median *ICT* is around 400 seconds for the two open-space lands and between 700 and 800 seconds for the Dance Island. Interestingly enough, the shape of the distribution of *ICT* appears to be only slightly affected by r. We computed the exponent of the power-law phase and obtained: $\alpha(r_b) = 0.1975$, $\alpha(r_w) = 0.2189$ for Apfel Land; $\alpha(r_b) = 0.3275$, $\alpha(r_w) = 0.2842$ for Dance Island; and $\alpha(r_b) = 0.4318$, $\alpha(r_w) = 0.4123$ for Island of View.

Although the distribution of contact opportunities appears to be similar for the two open-space lands, the CCDF of the first contact time *FT* illustrates a great disparity between these lands: in Apfel Land users have to wait for a long time before meeting their first neighbor. The median *FT* is around 300 seconds for Apfel Land, while it is less than 20 seconds for the other two lands when $r = r_b$. The *FT* improves a lot when increasing r: the median is around 30 seconds for Apfel Land and less than 5 seconds for the other lands.

These results are quite surprising: from a *qualitative* point of view, we obtained a statistical distribution of contact opportunities that mimics what has been obtained for experiments in the real world [5, 7, 8]. Obviously, human activity roughly spans the 12 hours interval, while even the more assiduous user we were able to track spent less than 4 consecutive hours on SL, hence a quantitative comparison is not immediate. In our future work we will try and address the following key question from a quantitative point of view: *is mobility of users in SL representative of real human mobility?*

Line of sight networks: We now delve into a detailed analysis of the communication networks that emerge from user interaction when we assume them to be equipped with a wireless communication device covering a range $r \in \{r_b, r_w\}$. A temporal line of sight network is constructed every measurement interval: in Fig. 2 we show the aggregate (over the whole measurement period) CCDF of the node degree and the aggregate CDF of the network diameter and the clustering coefficient.

The node degree CCDF illustrates a diverse user behavior in each target land: for Apfel Land we observe that 60% of users have no neighbors, for the Dance Island only 10% of users have no neighbors while in the Island of View, all users have at least one neighbor when $r = r_b$. When the communication range is set to $r = r_w$ all users have at least one neighbor in all lands. The maximum degree and the whole distribution varies a lot between target lands: the main reason lies in the physical distribution of users on a land. In Apfel Land users are relatively sparse while in the Dance Island, for example, most of the users spend most of the time in a tiny portion of the land: this observation is corroborated[3] by our study on the spatial distribution of

---

[3]There is an intuitive reason for this phenomenon: in a discotheque, for example, users spend most of their time on the dance floor or by the bar, while in an open space user are
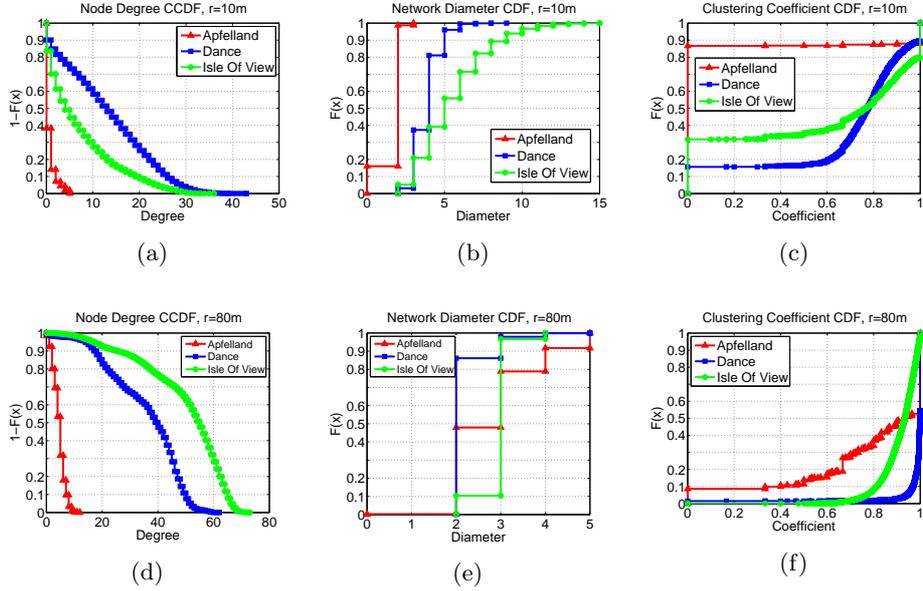
Figure 2: *Line of sight networks:* graph theoretic properties for three selected target lands.
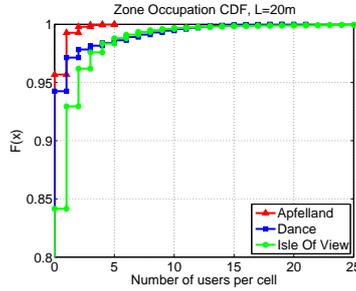

Figure 3: *Spatial distribution* of users.

users as shown in Fig. 3. Although the general trend for all target lands we inspected is that a large fraction of the land has no users, some lands (e.g., Dance Island) are characterized by hot-spots with several tens of users.

The CDF of the network diameter illustrates the impact of different transmission ranges: it is clear that the diameter shrinks for $r = r_w$. We note, however, that for Apfel Land there is an apparent contradiction: for $r = r_b$ the maximum diameter is smaller than for $r = r_w$. This phenomenon is due to the fact we compute the diameter of the largest connected component of the temporal graph formed by users: when the radio range is small
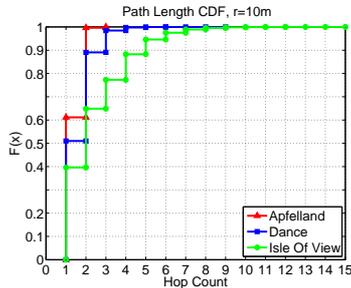
generally more sparse.

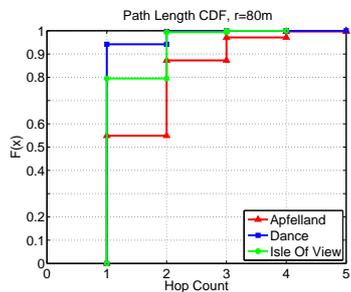Figure 4: *Path length distribution.*

Figure 5: *Path length distribution.*

(and users are scattered through the target land) we observe the emergence of relatively small connected components, whereas for larger ranges the connected component is large (eventually it includes all users), hence a larger diameter.

In Fig. 3 we also plot the CDF of the ratio between $C_{G(t_k)}$ and the *equivalent random graph* which has the same number of vertex and edges that are placed uniformly at random, for the whole measurement period. Our results clearly point to a ratio larger than one, indicating that our temporal line of sight networks do not share features common to random graphs. We are tempted to claim that the graphs that emerge from user interaction under our definition of contact (two users that fall within the communication range r) have small world characteristics, but this claim cannot be supported due to the relatively small scale of our graphs and to our results on network diameter and average path length (Fig. 4 and Fig. 5). We defer for our future work the analysis of other kind of contacts (e.g. friendships) that exists in SL which allow the collection of a larger data set.

Trip analysis: using physical coordinates, we were able to study the statistical distribution of the distance travelled by users on the three target
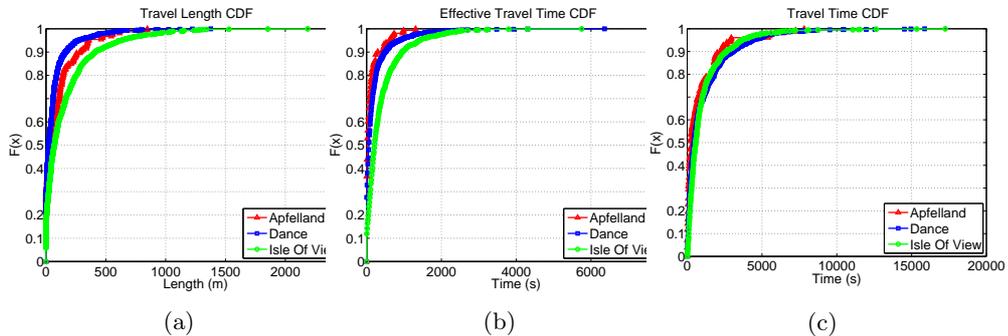
Figure 6: *Trip analysis* for three selected target lands

lands we analyze in this paper. Fig. 6 illustrates the aggregate CDF of the travel length, the travel time and the login time for all users. Fig. 6(c) shows the CDF of the login time: in our measurement we observed that the longest log-in time for a user was around 4 hours while 90% of users are logged in for less than 1 hour.

Fig. 6(a) provides further hints towards a better understanding of user mobility in the selected target lands. For a confined area such as Dance Island, the vast majority of users travel less than 230 meters (90*th* percentile). This observation however applies also for open spaces: for Apfel Land, the 90*th* percentile is around 400 meters while it grows up to 500 meters for Isle of View. There is a small fraction of users who travel a very long distance: for the Isle of View, around 2% of users travel more than 2000 meters. Fig. 6(b) is useful to infer the distribution of the times a user takes to travel from her initial point (the first time our crawler tracked the user) to her final point (the last time the user has been seen on the target land). We defer to our future work the study of the correlation between the travel time and the travel length so as to determine the distribution of the speed at which users move.

5   Conclusion

In this paper we discuss a novel methodology to perform user profiling that exploits the raising popularity of on-line communities emerging from user interaction in Networked Virtual Environments. In this work we study the mobility patterns of users connected to Second Life: we built a crawler that extract the position of users on a target land at regular intervals. Tempted by the question whether any similarity can be found between our results and measurements performed in the real world, we first characterize the statistical distribution of contact opportunities among users. Our analysis indicates that mobility patterns in a virtual environment share common traits, from

a qualitative point of view, with those in the real world. We further pushed our analysis to characterize the spatial distribution of users and their mobility behavior: users are generally concentrated around point of interests and travel small distances in the vast majority of cases. This phenomenon could be the result of the user interface used to interact with SL: users hardly move while chatting because movement commands are situated on the keyboard.

An interesting area of future research would be to build the network of "relationship" among SL users, rather than defining contacts as done in this work. Based on the "relation graph", new question can be addressed such as the frequency and the strength of contacts (in the sense of this paper) between acquaintances.

References

[1] Libsecondlife: www.libsecondlife.org/.

[2] Lsl: http://wiki.secondlife.com/wiki/lsl_portal.

[3] Second life: http://www.secondlife.com.

[4] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, J. Scott, and R. Gass. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, 2007.

[5] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot. The diameter of opportunistic mobile networks. In *Proc. of CoNEXT*, 2007.

[6] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic. Power law and exponential decay of inter contact times between mobile devices. In *Proc. of MOBICOM*, 2007.

[7] J. Krumm and E. Horvitz. The microsoft multiperson location survey. Technical report, Microsoft Research, MSR-TR-2005-13, 2005.

[8] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *Proc. of IEEE INFOCOM*, 2008.

[9] M. Varvello, E. Biersack, and C. Diot. A networked virtual environment over kad. In *Proc. of ACM CoNext*, 2007.

[10] D. J. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440442, 1998.