Can Crowdsensing Beat Dynamic Cell-ID?

Michal Ficek, Nathaniel Clark, Lukáš Kencl
Czech Technical University in Prague
Technická 2, 166 27 Prague 6, Czech Republic
{michal.ficek, clarknat, lukas.kencl}@fel.cvut.cz

ABSTRACT

In this work we investigate the limits of crowdsensing in discovering the mapping of mobile network Cell-IDs to geographic locations. We employ original large-scale mobility simulations, derived using the NRC-Lausanne dataset, to determine the fraction of cells visited by a fixed number of users over a time interval. This is vital to judge the ability of crowdsensing to rapidly update an inadequate, malfunctioning or obsolete Cell-ID database, thus preventing mechanisms such as Dynamic Cell-ID from obfuscating the network. We show that crowdsensing is quite a powerful tool, with for example only 25% more users than cells sufficing to scan 99% of the network over a day.

1. INTRODUCTION

Nowadays, communications market players such as Google or Apple utilize crowdsensing techniques to discover the structure of mobile networks. GPS-enabled phones of customers send their current GPS coordinates and Cell Identifier (Cell-ID) to a server that collects, clusters, fingerprints, and stores such data from all customers. Such a Cell-ID database can subsequently lead to the cell geolocations: given a Cell-ID, an approximated position inside the cell is returned. This enables services such as localization or friend proximity lookup, even for mobile phones without GPS receivers. The questions we address in this paper are: What is the required minimal size of a user group needed for obtaining a critical mass of knowledge about the mobile infrastructure? And, how much time is needed to do so?

This is interesting for mobile operators who wish to monetize their costly network infrastructure. Today, user location is delivered for free by third parties (e.g., My Location service by Google) exploiting the fact that the Cell-ID assignment is static and the signal covers a given geographic area. The *Dynamic Cell-ID* mechanism, described in three recent patents [1, 2, 3] and allegedly considered for deployment by China Mobile [4], may alter the situation. The key idea is to mask part of the static Cell Global Identities by providing different, dynamically generated Cell-IDs. With frequent, for example daily changes, third-party Cell-ID databases such as [5, 6] would have difficulty maintaining correct Cell-ID information, enabling network operators to commercialize the mapping of dynamic Cell-IDs to geographical coordinates.

Deploying Dynamic Cell-ID would have consequences. The GPS-less devices, still a majority of all mobile phones (62% in 2011 [7]), rely on network-based (Cell-ID) localization. Third-parties would fail to provide free localization applications, unless they paid operators for the Dynamic Cell-ID mapping, influencing customer's end price. A-GPS-enabled phones would be affected by having a longer time-to-first-fix (order of minutes), as commercial Secure User Plane Location (SUPL) servers would not be able to advise the A-GPS receiver on the approximate satellite positions (based on the current Cell-ID of the user) because SUPL-server databases would become outdated every time the dynamic Cell-IDs are changed. Finally, various cell-fingerprinting methods [8, 9], popular in research and academia as cheap and reliable positioning methods, would be rendered inoperative.

There are two principal ways to deal with Dynamic Cell-ID. Either the mapping may be bought from the mobile operator, which might be costly or the operator might not be willing to sell it. Or, third-parties can assign coordinates to Cell-IDs by wardriving or crowdsensing methods. Wardriving is the act of searching for WiFi hotspots and other information, such as mobile network Cell-IDs, at particular locations by driving around. Crowdsensing [10] refers to a process of collecting data from sensing and computing devices (mobile phones). While wardriving is time-consuming and limited in resources (vehicles, drivers), crowdsensing is advantageous in time and coverage, especially when many mobile users are involved. Nevertheless, two questions arise: How many crowdsensing participants are needed? And, how long does it take them to scan the entire network?

To address these questions, we perform a large-scale simulation of mobile user movements in a mobile network, counting the number of distinct cells users visit over a period of time. First, we build a mobility model, based on the NRC-Lausanne dataset [11], that reflects both temporal properties of human mobility patterns and the number of user-cell associations (Section 3). Such model is necessary for sampling a high number of artificial, yet realistic user mobility patterns to serve as a simulation input. Second, we simulate movement of a different number of users over one day in an approximation of a mobile network (Section 4). This simulation is vital because basic statistics on user movement

PhoneSense'12, November 6, 2012, Toronto, ON, Canada.

(average number of unique cells visited during a day) cannot be used alone to form the mobility patterns, as different users are likely to visit the same cells during the day. Among others, we show that only 25% more users than the number of cells in the network suffices to discover 99% of all cells in one day (Section 5). Indeed, crowdsensing can be thus considered as a fight-back strategy against Dynamic Cell-ID.

Even if Dynamic Cell-ID is not deployed, the limits of Cell-ID crowdsensing are still of interest. Networks evolve constantly and relevant databases need to be verified and updated frequently. They might fail, be maliciously manipulated, or contain errors. For users, dependence on a large enterprise such as Google may represent a risk and a privacy issue. Competing crowdsensing services may thus get established, requiring an unknown critical mass of users to obtain a sufficiently precise view of the network infrastructure.

2. BACKGROUND

In this section, we present relevant background knowledge.

Mobile network. We consider an arbitrary commercial mobile network (GSM, UMTS) consisting of thousands of base stations, each equipped with a number of antennas (typically three). The geographical area under an antenna signal coverage is called a *cell*. Each cell is identified by a unique Cell Global Identity (CGI), which consists of four fields: Mobile Country Code, Mobile Network Code, Location Area Code (LAC), and Cell-ID. While the first two fields are specific to the country and operator, the latter two are assigned by the mobile operator.

Dynamic Cell-ID. Dynamic Cell-ID [1, 2, 3] works on the principle of masking one or both of the LAC and Cell-ID parts of a CGI to mobile devices. A new, *dynamic Cell-ID* is calculated by the base station and is transmitted to the mobile device, while the original Cell-ID remains intact in the core network. From time to time (patent [3] suggests once a day), all dynamic Cell-IDs are permuted among the network cells. This process is achieved by an unspecified, time-dependent, invertible function that maps static Cell-IDs to the dynamic ones and vice versa. Such mapping function can be arbitrarily complicated, or it may even represent a simple random permutation. Thus, we assume that the mapping function can not be discovered by simply observing the changes of dynamic Cell-IDs over time.

Dataset. We use part of the NRC-Lausanne dataset, which consists of information about 38 out of 168 users who participated in the Lausanne Data Collection Campaign [11]. It contains a timestamped sequence of CGIs per user with one record per every cell change during the campaign period (referred to as *cell trace*). Also, there is a timestamped GPS log for each user (we call it a *GPS trace*). The dataset covers a large part of Switzerland, including major cities and the countryside (see Figure 1).

Assumptions. We make three key assumptions regarding the principles of Dynamic Cell-ID and crowdsensing. **A1)** The dynamic Cell-ID renumbering occurs only once a day and for all cells in the network at the same time. **A2)** GPS receivers serve as the only tool for reporting geographical positions of the user. **A3)** Mobile phones only report the Cell-ID of the currently attached cell.

3. DATA-DRIVEN MOBILITY MODEL

The NRC-Lausanne dataset gives away the coverage capabilities of the user pool for one year in an area. However, the traces of 38 users cannot cover the whole mobile network (which typically consists of thousand of cells), so several-times more users are necessary in a study of this kind. Straightforward sampling from the traces would lead to the *reuse* of particular user's movement, which we consider harmful. Therefore, we try to infer a general coverage capabilities of the population. We build a model that captures general features of users' movement within the network to generate a high number of synthetic, yet realistic traces.

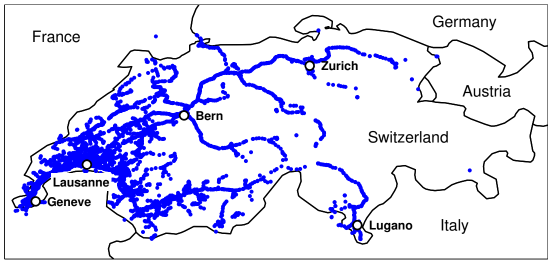

Figure 1: Places visited by the users in the dataset.

We aim to model user mobility patterns in terms of the number of unique cells visited *during one day*, starting from midnight. A common denominator of users' daily mobility is their presence at some *places*, such as home, work, the cinema, etc. for a substantial amount of time. Each of these places (which are not many in one day) is covered by a cell or a set of cells. A key observation here is that the *transitions* between places account for most of the total number of unique cells visited by a user during a day. Thus, the goal of our mobility model is to capture **F1)** the total number and ordering of places as they are visited by a user during a day, **F2)** the start time of all user's transitions between places in the day, and **F3)** the duration of transitions and their length, measured in the number of unique cells visited during the transition. Our model differs from the large body of similar work (see [12] for a survey) in that it incorporates all of the features F1–F3 concurrently, thus it describes the users' *daily patterns* (F1), captures *fine-grained* temporal characteristics of human movement *during one day* (F2), and quantifies daily *user-cell associations* (F3).

Important places for a user are recognized from mobile data mainly by clustering methods [13, 14]. Clustering is vital because a place is typically covered by overlapping cells, and the user's mobile phone connects to them even when the user is not moving (so called *cell jitter*). We use time-based clustering to recognize a user's places, because the GPS trace covers only 32% of all cells in the cell trace which precludes spatio-temporal clustering. In this work we define a *place* as a set of neighboring cells in which the user cumulatively spends more than 60 minutes anytime during a day. A *transition* between places is the act of leaving a place and visiting another, or the same place subsequently at least 4 minutes later. Without going into details (for the limited space), the algorithm finds all places and transitions while also detecting and removing cell jitter.

We process the data by dividing the cell trace of each user into day-long sequences, each starting at midnight. Then, we remove days where the mobile phone was off and weekends. Because the daily routine of users and their mobility significantly differ between weekdays and weekends, these must be handled separately. Mainly due to space limitation in this paper we restrict the dataset to weekdays only, however, the weekends can be modeled in a similar manner.

Algorithm 1 Generation of a Transition

Input: $p^{\widetilde{t}}_{i,j}$ for all $i, j \in L$ and $\widetilde{t} \in \{1, \ldots, T\}$,
$f_{\text{new}}, f_{\text{same}}, f_{\text{old}}, g_{\text{new}}(\delta), g_{\text{same}}(\delta), g_{\text{old}}(\delta)$
Output: $TS = \{(O_i, D_i, t_i, \delta_i, l_i)\}_{i=1}^{N}$
$\widetilde{t} \leftarrow 1,\ D_0 \leftarrow 1$
while $\widetilde{t} < T$ **do**
  if no transition (prob. $1 - \sum_{j \in L} p^{\widetilde{t}}_{D_{i-1},j}$) **then**
    $\widetilde{t} \leftarrow \widetilde{t} + 1$
  else
    $O_i \leftarrow D_{i-1}$,
    $D_i \leftarrow d \in L$ with prob. $p^{\widetilde{t}}_{O_i,d}$
    $t_i \leftarrow$ uniformly sampled from interval $\widetilde{t}$
    $\delta_i \leftarrow$ sampled from f_* ($*$ ... transition class)
    $l_i \leftarrow$ sampled from $g_*(\delta_i)$ ($*$ ... transition class)
    $\widetilde{t} \leftarrow$ nearest time period after $t_i + \delta_i$
  end if
end while

Finally, we consider all of the day-long weekday sequences to be independent, even when belonging to the same user. Handling the data in such way is viable because the mobility model is to describe *only* a period of one day and so weekday correlations need not be reflected.

To extract mobility patterns from the data, we process each day-sequence and find all places in the transitions for that particular day. We enumerate places in each day-sequence according to the time of the first visit with numbers $L = \{1, 2, \ldots\}$. Then, we describe all N transitions during one day in a *transition sequence* $TS = \{(O_i, D_i, t_i, \delta_i, l_i)\}_{i=1}^{N}$, where $O_i, D_i \in L$ are the origin and destination places of the i-th transition, $t_i, \delta_i \in [0, 1] \in \mathbb{R}$ represent the time (by a fraction of the day) of the transition start and its duration, and $l_i \in \mathbb{N}$ is the length of the transition expressed in the number of unique cells visited during the transition. The TS is empty for a user who spends the entire day at one place and makes no transitions.

We express the model features F1 and F2 by mining the transition probabilities between the places, depending on the time of the day. We simplify the structure of time by quantizing the day into $T = 288$ 5-minute equidistant time slots $\widetilde{t} \in \{1, \ldots, T\}$. So, for example $\widetilde{t} = 2$ represents a time period from 12:05 a.m. to 12:10 a.m. Then, $p^{\widetilde{t}}_{i,j}$ denotes the probability that a transition between places $i, j \in L$ *starts* during the time period $\widetilde{t}$. The duration of a transition (model feature F3) is estimated from the data with probability density functions $f_{\text{new}}, f_{\text{same}}, f_{\text{old}}$. These describe different transition classes, depending on the relationship between the origin O_i and destination D_i places during the day: A transition is classified as *new* if it ends at a new, previously not visited place, *same* if it starts and ends at the same place, and *old* if it is between places already visited. Finally, we found that the transition lengths (number of unique cells) follow a Normal distribution with the mean and standard deviation parameters linearly dependent on the duration of the transition and its class. We denote these distributions by $g_{\text{new}}(\delta)$, $g_{\text{same}}(\delta)$, and $g_{\text{old}}(\delta)$. Parameters $p^{\widetilde{t}}_{i,j}$, f_*, and g_* constitute our model. The generation of a new, synthetic transition sequence from the above parameters works according to Algorithm 1.

In the rest of this section we show by comparing the features F1—F3 that the synthetic traces from the model correspond to the NRC-Lausanne dataset. Users' daily patterns, the model feature F1, represent the number and ordering of different places visited by a user during a day. For example, a typical daily pattern for the majority of users would be *1—2—1*, where *1* could represent "Home" and *2* could stand for "Work". Figure 2 (Left) compares the most frequent patterns in the dataset with the synthetic traces generated from the model, showing a high correspondence in the frequency. Figure 2 (Right) shows that the distribution of users' daily patterns is heavy-tailed, i.e., a small number of patterns occur often while numerous patterns are rare. As depicted, the model faithfully captures this daily pattern distribution.

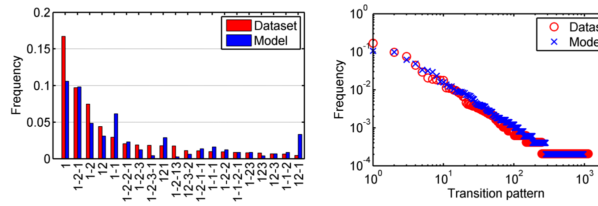

Figure 2: (Left) The example of the 20 most frequent daily patterns. (Right) Heavy-tail distribution of different daily patterns.

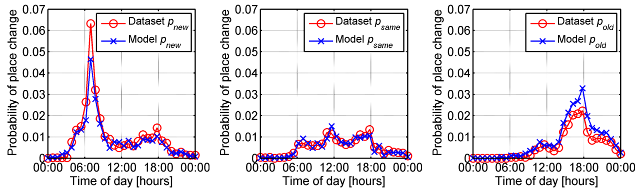

Figure 3: Probability of transitions during a day.

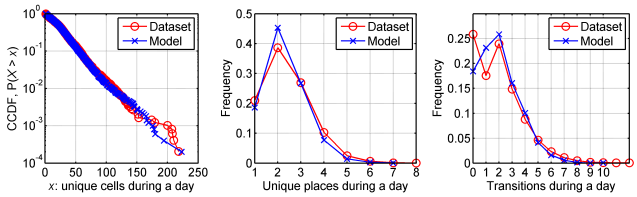

Figure 4: Distribution of the total unique cells, places, and transitions during a day.

The fine-grained temporal characteristics of human movement during one day, the model feature F2, are depicted in Figure 3. It shows that in the morning users commute to new, previously not visited places (p_{new}), during the day they tend to leave the place and return to the same place later (p_{same}), and in the afternoon they return to previously visited places (p_{old}).

Daily user-cell associations, the model feature F3, are depicted in Figure 4 (Left). Clearly, the model well quantifies the total number of cells visited during a day. This is achieved by following the distributions of total number of the places visited during a day (Figure 4 (Middle)) and the total number of transitions between the places (Figure 4 (Right)).

4. SIMULATING USER-CELL ASSOCIATION

We model a mobile network with c cells as a Voronoi tessellation [15] of a unit square simulation area where a spatial Poisson process of constant intensity represents the

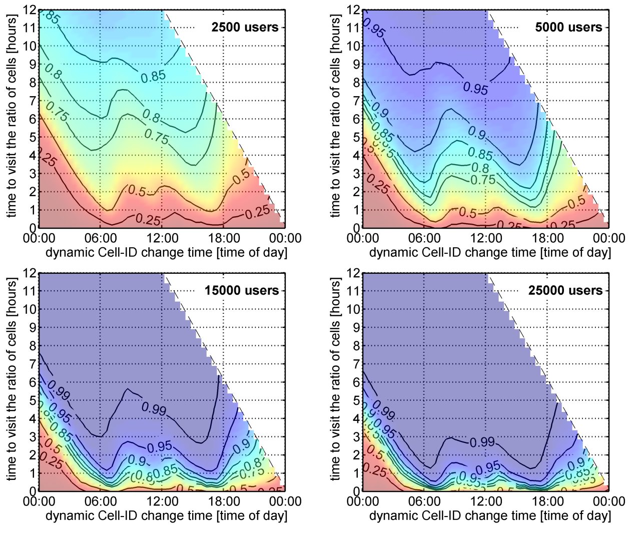

Figure 5: Impact of dynamic Cell-ID renumbering time on cell discovery in a network with 5,000 cells.

cells' positions[1]. Connectivity between cells (and thus possible handovers) is captured by the Delaunay triangulation $DT = (V, E)$ [15], the dual of Voronoi tessellation: if cells $u, v \in V$ are neighbors, there exists an edge $(u, v) \in E$. Let $p_{u,v,k} = \{u = c_0, c_1, c_2, ..., c_{k-1}, c_k = v\}$ denote a simple path (without cycles) between nodes $u, v \in V$ of length $k \in \mathbb{N}$ (i.e., having k edges and $k-1$ unique nodes in the path excluding u and v). And, let $S \in \mathbb{N}^{|V|\times|V|}$ be an all-pairs shortest path matrix for the DT graph with unit edge-weights, computed by the Floyd-Warshall algorithm.

We simulate the movement of a set of users $U = \{1, \ldots, n\}$ in the network as a discrete-time pseudo-random walk on the DT graph, according to the transition sequences $TS_u = \{O_i, D_i, t_i, \delta_i, l_i\}_{i=1}^{N}$. These are sampled independently from the mobility model for each user $u \in U$. As the users traverse the graph, the number of distinct nodes visited corresponds to the number of cells they have been associated with. The simulation consists of three steps, explained below by the example of a single user.

First, the user's place labels are mapped to DT nodes by a randomly selected one-to-one function $m : L \to v$, where $L = \{O_1\} \cup \{D_i | i = 1, \ldots, N\}$ and $v \subset V$. The function m is found with respect to the lengths of the paths between places in a user's TS, such that $\forall i \exists p_{u,v,k} : u = m(O_i) \wedge v = m(D_i) \wedge k = \max(S_{u,v}; l_i + 1)$. The selection of a path length k guarantees triangle inequality of path lengths.

Second, a randomly selected path $p^i_{u,v,k}$, such that $u = m(O_i)$, $v = m(D_i)$ and $k = l_i + 1$, is considered to be a sequence of nodes (cells) the user visits between places O_i and D_i during the i-th transition. However, finding a simple, k-length path $p_{u,v,k}$ is known to be NP-hard (can be reduced to the Hamiltonian Cycle problem), and even probabilistic algorithms [16] are too slow on large graphs. Therefore, we simplify this task and look for a path $\widetilde{p}^i_{u,v,k} = \{u, \ldots, w, \ldots, v\}$ with a maximal number of unique nodes and the desired length, i.e., $w \in \operatorname*{argmin}_{x \in V}(p_{u,x,S_{u,x}} \cap p_{x,v,S_{x,v}})$ and $S_{u,w} + S_{w,v} = k$. Using the matrix S to find a node w is fast, although it can result in a non-simple path.

Third, we express the user-cell association during the i-th transition $(O_i, D_i, t_i, \delta_i, l_i)$ as a sequence $A^i = \{(\tau_j, v_j)\}_{j=0}^{l_i+1}$, where τ_j denotes the time of the change of association to the j-th cell v_j. Assuming that the speed of the user is constant on the path between O_i and D_i, then the user-cell association in time changes proportionally to the distance between the nodes in the path. Further, we assume that a user-cell association change happens on the boundary between cells. Since the shortest distance between nuclei of two adjacent Voronoi cells to their common boundary is equal (by definition of the Voronoi tessellation), the cell-association changes when the user is in the middle of the Delaunay triangulation edge between the cells' nuclei. Thus, the elements of the sequence A^i are as follows: $\tau_0 = t_i$, $\tau_j = t_i + \delta_i(\frac{\sum_{l=1}^{j-1} d_l + d_j/2}{\sum_{l=1}^{n} d_l})$, and $v_j = c_j$, where d_l is the length of the edge (c_{l-1}, c_l) in terms of Euclidean distance.

The number of unique cells a user u visits by the time T is the cardinality of the set $U_u = \{v \in V | \exists i, j, A^i_u = \{(\tau_j, v_j)\} : \tau_j \leq T \wedge v = v_j\}$. A higher number of users in the simulation at one time is handled *independently*, so the total number of cells visited by n users is $|\cup_{u=1}^{n} U_u|$.

5. RESULTS

Assume we have a mobile network that consists of c cells, and there are n users involved in crowdsensing. How long does it take them to visit x% of all cells in the network? We simulated the movement of users in a network for all combinations of network sizes $c = \{1{,}000, 2{,}000, \ldots, 5{,}000\}$ cells and number of users $n = \{0.25c, 0.5c, \ldots, 5c\}$.

Figure 5 shows that collecting network Cell-IDs takes a shorter time when more users are involved. However, users' mobility during a day significantly affects the duration of the network scan: in the morning and afternoon users commute and travel more, resulting in a shorter scan time. On the contrary, it takes longer to scan the network during the night and while the users are at work, as their mobility is low. This may be a clue for *when* operators should renumber the dynamic Cell-IDs to strike at the heart of third-party Cell-ID databases the most.

We focus in detail on the case where dynamic Cell-IDs are renumbered at midnight: Figure 6 shows a relationship between the ratio of cells observed during a day and the number of users in a network with $c = 5{,}000$ cells. We can see that at least $n = 1.25c = 6{,}250$ users are needed to observe at least 99% of all cells by the end of the day. Additionally, there is a significant difference between the times to visit all cells as the number of users increases from $n = 1.25c$ to $n = 3.5c$. Because of a low number of users who travel during the early morning, about $n = 3.5c$ users are needed to visit 99% of the cells by 7 a.m. Markedly, having more than $3.5c$ users yields only minor improvements.

6. DISCUSSION

The assumptions we made in Section 2 stem from following reasons. Naturally, renumbering all cells at the same time (assumption A1) causes the highest havoc because it affects each user of a Cell-ID database provider — each user would obtain a false position. The renumbering period of one day is suggested in [3], but we admit that a shorter interval or a different renumbering scheme could be even-

[1]To obtain a more realistic network, we can use a non-homogeneous Poisson process in which higher density of cells corresponds to the cities, or even apply any relevant background knowledge, such as population density or transportation networks (roads, railways) in the studied region.

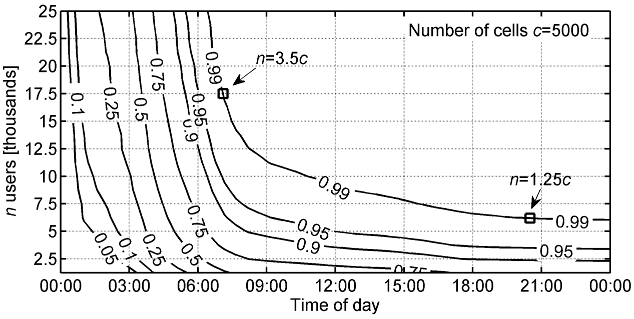

Figure 6: Ratio of cells observed during the day.

tually possible. We leave this for future experimentation. The assumption A2 of having only GPS coordinates from the user's mobile phone represents the simplest positioning solution. Although other positioning techniques or phone-sensed data may be available (such as WiFi beacons, signal strength, etc.), we assume this easiest and most accurate method. Finally, assumption A3 of reporting only the currently attached cell comes from the fact that the knowledge of neighboring Cell-IDs would not immediately bring any advantage. However, it may be used in some future extensions to deliver the approximate position of a cell.

Our approach to the simulation is not limited by the model we used. Any model that quantifies user-cell association during a day can be applied. Nevertheless, to the best of our knowledge, there is no such model publicly available.

The model presented in this study may seem limited by the lack of any spatial relation to a real geographic background. However, because it captures user's movement in a mobile network in terms of cell transitions without conditioning on the real-world cell tower locations, the model is area-independent. As such, it is not limited to the area covered by the dataset and can be be applied to any arbitrary cellular network topology — either a real one or an artificial one. Nevertheless, the parameters of the model may change with different network technology or with a larger and more representative user-pool.

7. CONCLUSION

In this work we investigated the limits of crowdsensing in discovering the mapping of mobile network Cell-IDs to geographic locations. Based on the NRC-Lausanne dataset, we build a model which describes user-cell association in a mobile network over a day. Using the model we generated thousands of artificial yet realistic traces of user movement applied by a large-scale simulation to a mobile network topology. The results show that crowdsensing is quite a powerful tool. For example with only 25% more users than cells sufficing to map 99% cells of a mobile network to geographic locations over a day.

Crowdsensing as a fight-back method against the Dynamic Cell-ID method poses several issues. First, it is questionable whether any third-party can persuade a user-pool of at least three-times the number of cells in the network to participate in crowdsensing. Consider the Czech Republic with 10.5 million inhabitants living in approx. 78,000 km^2 $\cong$ 30,500 sq mi. Each of the three biggest mobile operators has about $c = 14{,}000$ cells in the network [17], so about $n = 3c = 42{,}000$ users are needed to visit 99% of the cells by 8 a.m. Let us assume the sensing software is built on the Android platform, and the smartphone penetration (50%) and Android share on the smartphone market (48%) are similar to the U.S. [18]. Then a calculation ($10{,}500{,}000 \times 1/3 \times 0.5 \times 0.48 = 840{,}000$) shows that every *twentieth* user of an Android smartphone (per each operator) should participate in crowdsensing, making it seem viable.

Other issues are related to the localization business models and Dynamic Cell-ID implementation details. Is the hours-long period of bad localization performance acceptable for the users? Apparently, having the network scanned within one hour anytime during a day is possible, but with an unrealistic number of users. And finally, what if the dynamic Cell-ID renumbering interval will be shorter, let us say on the order of hours? The near future may enlighten us on these concerns.

Acknowledgement

This work was supported by Microsoft Research through its PhD Scholarship Programme.

References

[1] M. Bensimon *et al.*, “Method to mask cell identifiers or location area codes of a mobile network with respect to a mobile terminal,” 2009, u.S. Patent No.: 20090170477.

[2] Z. Duan *et al.*, “Method and device for configuring a cell identity,” November 2009, international Patent No.: PCT/CN2009/071861.

[3] P. Willars *et al.*, “Methods and Systems for Obscuring Network Topologies,” 2011, u.S. Patent No.: 20110116632.

[4] G. Lu, “China Mobile Is To Implement Dynamic Cell-ID, LBS Services Need Pay for the Location,” Online, 2011, [Online]. Available: http://technode.com/2011/01/06/china-mobile-to-implement-dynamic-cell-id/.

[5] Location-API.com. (2009) Combain Mobile AB. [Online]. Available: http://location-api.com/

[6] 8motions. (2010) OpenCellId. [Online]. Available: http://www.opencellid.org/

[7] Berg Insights, “Handset Connectivity Technologies 3rd Ed.” 2012, [Online]. Available: http://www.berginsight.com.

[8] T. Wigren, “Adaptive Enhanced Cell-ID Fingerprinting Localization by Clustering of Precise Position Measurements,” *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 5, pp. 3199–3209, 2007.

[9] J. Paek *et al.*, “Energy-efficient Positioning for Smartphones Using Cell-ID Sequence Matching,” in *Proc. ACM MobiSys*, 2011, pp. 293–306.

[10] R. Ganti *et al.*, “Mobile Crowdsensing: Current State and Future Challenges,” *Communications Magazine, IEEE*, vol. 49, no. 11, pp. 32–39, 2011.

[11] 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 *Proc. Pervasive 2012*, 2012.

[12] D. Karamshuk *et al.*, “Human Mobility Models for Opportunistic Networks,” *Communications Magazine, IEEE*, vol. 49, no. 12, pp. 157–165, 2011.

[13] G. Yang, “Discovering Significant Places from Mobile Phones - A Mass Market Solution,” in *Proc. MELT*, 2009, pp. 34–49.

[14] R. Montoliu and D. Gatica-Perez, “Discovering Human Places of Interest from Multimodal Mobile Phone Data,” in *Proc. MUM*, 2010, pp. 12:1–12:10.

[15] F. Aurenhammer, “Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure,” *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, 1991.

[16] N. Alon *et al.*, “Color-coding,” *Journal of the ACM*, vol. 42, no. 4, pp. 844–856, 1995.

[17] GSMweb, 2010, [Online]. Available: http://gsmweb.cz/.

[18] Nielsen, “Smartphones Account for Half of all Mobile Phones,” 2012, [Online]. Available: http://blog.nielsen.com/nielsenwire.