

CZECH TECHNICAL UNIVERSITY IN PRAGUE



DOCTORAL THESIS STATEMENT

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Telecommunication Engineering

Mgr. Michal Ficek

**TRACKING USERS IN MOBILE NETWORKS:
DATA ACQUISITION METHODS AND THEIR LIMITS**

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Telecommunication Engineering

Doctoral thesis statement for obtaining
the academic title of “Doctor” abbreviated to “Ph.D.”

Prague, June 2013

The doctoral thesis was produced in full-time manner.

Ph.D. study at the department of Telecommunication Engineering of the Faculty of Electrical Engineering of the CTU in Prague.

Ph.D. Candidate: Mgr. Michal Ficek
Establishment Faculty of Electrical Engineering of the CTU in Prague
Address Technická 2, 166 27 Prague 6

Supervisor: Dr. Lukáš Kencl
Department Department of Telecommunication Engineering
Faculty of Electrical Engineering of the CTU in Prague
Technická 2, 166 27 Prague 6

Opponents: _____

The doctoral thesis statement was distributed on

The defence of the doctoral thesis will be held on at a.m./p.m. before the Board for the Defence of the Doctoral Thesis in the branch of study Telecommunication Engineering in the meeting room No. of the Faculty of Electrical Engineering of the CTU in Prague.

Those interested may get acquainted with the doctoral thesis concerned at the Dean Office of the Faculty of Electrical Engineering of the CTU in Prague, at the Department for Science and Research, Technická 2, Prague 6.

.....
Chairman of the Board for the Defence of the Doctoral Thesis
in the branch of study Telecommunication Engineering
Faculty of Electrical Engineering of the CTU in Prague
Technická 2, 166 27 Prague 6.

Contents

1 Introduction and Related Works	1
Motivating Problems	1
Related Work	2
2 Aim of the Dissertation Thesis	4
3 Working Methods and Selected Results	5
Active Tracking in Mobile Networks	5
Extending Utility of Passive Tracking Data	8
Exploring Limits of Crowdsensing	12
4 Conclusion	15
References	20
List of Candidate’s Work Related to the Thesis	20
Journals (Impact)	20
Journals (Reviewed)	20
Patents	20
Web of Science	20
Others	21
List of Candidate’s Work Non-related to the Thesis	22
Journals (Impact)	22
Journals (Reviewed)	22
Patents	22
Web of Science	22
Others	22
Summary	23
Anotace	24

1 Introduction and Related Works

According to the International Telecommunication Union there were 6 billion mobile-phone subscriptions worldwide by the end of 2011, and mobile phone penetration topped 100% of population in many countries [43]. This huge worldwide mobile-phone pervasiveness is increasingly turning the mobile network into a gigantic ubiquitous sensing platform, enabling large-scale analysis and applications of data acquired from the network. In recent years, mobile data-based research reaches important conclusions about various aspects of human characteristics, such as human calling patterns [37], virus spreading [83], social networks [94], human daily activity patterns [70], urban and transport planning [82], network design [92] and others. Particularly, information about movement of network users is of utter interest of many researchers [31, 93] because mobility description and prediction may have a profound effect on various fields of science, including telecommunications [19], human and time geography [5], urban studies [4] and energy-efficient networks [20]. Practical applications of user-tracking using the mobile phones are roaming optimization [21], tracking of criminals [76], traffic-monitoring [13] and targeted mobile advertising [59].

For the purpose of acquiring the data about movement of users, we consider the task of simultaneously tracking a high number of mobile network users or, more specifically, their mobile terminals. By *tracking* we mean collecting continuous information on the user's geographical position by means of various *positioning* techniques. The outcome of this process is a timestamped history of users' positions in the network and their geographical coordinates. The main problem with tracking in mobile network is that suitability of existing methods for large-scale tracking, network-wide application, and their technological limitations are often not discussed or remain unknown.

There are two principal technology options for tracking users in mobile networks:

- *Network-based methods* rely on the mobile-network infrastructure, which performs measurements and calculates the position of a subscriber;
- *Terminal-based methods* refer to the terminal's activity in carrying out the measurements and position calculation.

Network-based tracking methods pose the advantage that they are generally applicable to all network users. Two main approaches to network-based tracking can be used:

- *Active tracking* is based on queries of the network about the tracked device, so the network *actively* gathers information about users' terminal physical coordinates.
- *Passive tracking* uses operating data, such as Call Data Records (CDRs) or network logs, which are generated and stored automatically by the network for *all users* for billing and network troubleshooting reasons, without additional traffic in the network.

Terminal-based solutions are suitable for smartphones and evolving mobile devices, but often require user's cooperation or software installation, which prevents universal coverage.

Motivating Problems

The problem we solve in the thesis is to find and describe particular limits of the three tracking methods above and, when applicable, overcome these limits by special means of utilization of the tracking data.

Network-based active tracking delivers position of mobile-network users with unprecedented temporal granularity, which has proved useful for tracking of criminals and

suspects [76], and studies about mobility patterns [4] and urban dynamics [65]. However, it is a complex process that involves many nodes in mobile network and wastes the resources at the air interface between cell towers and a mobile terminal. Although computational power of network nodes and bandwidth are not usually limiting, the air interface is still a valuable resource that is hard to scale. The main question we try to answer is “How many users can be tracked simultaneously in a mobile network, and how often?” This is becoming a serious issue when thousands of subscribers are to be tracked in order to deliver location-based services and representative studies of human mobility. Knowing such limits would help to adjust the tracking parameters (the number of users or their tracking interval) to a level that would not hamper the usual network traffic.

Network-based passive tracking has become immensely popular as a unique source of large-scale data about individual mobility [31], calling patterns [37], patterns of tourists’ spatial mobility [5] and urban analysis [73], mainly because of the lack of appropriate active-tracking solutions or the cost of active tracking. Posterior interpretation of large-scale CDR datasets is perhaps even more important for purposes such as urban and transport planning [81, 82], network design [92], opportunistic spectrum access [86] and user mobility [35]. Unfortunately, whereas spatial precision of such passive-tracking data is acceptable, the accuracy in temporal dimension is substantially low. The main reason is that the position of a user is recorded *only* at places where the user’s communication event occurs (text message, call, data session), thus depends on communication frequency of an individual. During time of no communication activity it is not clear *where* the user is geographically located. This represents a problem for applications or analyses assuming ubiquitous and continuous user-tracking capability, such as opportunistic data dissemination [36] or epidemiology [39].

Terminal-based tracking remains a viable alternative to network-based methods when active-tracking infrastructure and passive tracking data are not available. It is advantageous in delivering rich data from contemporary sensor-enabled smartphones, and therefore is often sought after in academic research [23], but also in environmental [22], infrastructure [63] and social [74] applications, in which interest groups actively participate on data collection — a method called *crowdsensing*. Crowdsensing proved useful especially in mapping fixed structure of mobile network cells and wireless network access points to reference databases [14, 32], which are subsequently used for geolocation of mobile terminals and various location-based services. The quantitative limits of such approach, however, remain in question: “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?” The answers may help in deciding whether crowdsensing is a viable solution for mapping country-wide networks infrastructure, for example for building a new geo-location business such as [14], or a hint for mobile network providers, whether they should try to monetize their costly network infrastructure and apply mechanisms [17, 85, 8] preventing unauthorized use of such geo-location in their networks.

Related Work

Network-based Active Tracking

In the thesis, we demonstrate an implementation of SMS-based active tracking in mobile networks, a method that utilize Cell-ID based positioning. Positioning of mobile entities in networks is a well-studied problem: work [51] summarize the main approaches. Comparison of network-based positioning techniques is available from sources [26, 51, 78], alternative techniques are discussed in [71, 53].

The limits of Cell-ID positioning in terms of positioning accuracy are discussed in [80, 91]. Authors of [4] provide a short study of Cell-ID+TA-based active-tracking accuracy as a trade-off between tracking interval and tracking costs. The impact on power consumption of a tracked mobile terminal has been measured and discussed in [18]. In the thesis we focus on the properties of active tracking with Cell-ID positioning and its the impact on the network in general.

Recent works describe different methods of mobile terminal activity excitation, a necessary prerequisite for active tracking. These methods are Data oriented approach (simply an ICMP *ping* message), USSD method [9], CAMEL-protocol approach [1], fake handover procedure [3] and an extension of signaling messages [16]. The *SMS-based method* [29] described in the thesis is advantageous in that it can be used in any mobile network technology such as GSM, UMTS and LTE, since the Short Message Service is supported across all types of mobile networks. A similar approach has been presented in [76] under the name “blind SMS”.

There have been significant standardization efforts [2, 67] and corporate initiatives [11, 26] for Location Based Service (LBS). A prominent example of a LBS platform is Ericsson Mobile Positioning System [25], which complies with the latest LBS standards. In comparison with it, tracking solution presented in the thesis is simpler, but lightweight and deployable by adding only a single node into the network.

Network-based Passive Tracking and Available Datasets

Two principal sorts of passive tracking data exist — *location updates* and CDRs [84]. Both methods work with Cell-ID-precision in spatial dimension, but differ in the frequency in temporal dimension, depending on user’s mobility and calling patterns.

Monitoring location updates proved to be helpful in transportation for automatic deriving origin–destination matrices in a studied region [12] and in various applications of intelligent transportation systems [48]. However, location-update data are strongly limited by user mobility since they are recorded when a user crosses borders of location areas.

In recent years, mobile operators shared CDRs with researchers and academia either directly [31, 73], or in data-mining contests [68, 62]. Nevertheless, although many dataset resources exist, for example in [49], a publicly available large-scale CDR dataset is still not present, except for few recordings of individual enthusiasts [10, 28]. For the reasons above we use a substitution for CDRs derived from a real-world trace, the Reality Mining Dataset [23], in the thesis.

Many authors focused on limitations of CDRs from different points of view. Ranjan et al. [72] discuss a potential source of bias in CDRs for human-mobility studies, authors of [31] observed a heavy-tailed distribution of time between user’s communication events, Zang and Bolot [93] analyzed the privacy risks of sharing CDRs. In the thesis we have focused on the principal limitation of CDRs, their poor temporal granularity.

There were several attempts in describing movement of an entity in space and time between two places in general, for example by means of linear weighted interpolation [33], space-time prisms [34] and probabilistic variants of space-time prisms [87]. However, the assumptions behind these models are contradictory to the nature of CDRs. Gonzalez et al. [31] analyze human mobility patterns from CDR, but deliver only a general point of view on the nature of human movement. The idea behind refining some mobility-related data come from [69] — the authors proved that increasing the sampling rate of GPS lowers the localization error. The uncertainty in user’s trajectories has been studied in [50], but with respect to classic time-geography, which hardly applies to CDRs.

Terminal-based tracking: Crowdsensing and Modeling Human Mobility

Crowdsensing [30] represent an important tool for collecting network associations and activity of mobile phones by volunteer individuals. There are publicly available datasets such as [23], other datasets are bound with legal consent [62] or were kept unpublished in a raw form [66]. We use the NRC-Lausanne dataset [62], to obtain an information about mobility of network users.

The limits of crowdsensing have been studied in [88] from the perspective of hardware heterogeneity, burden placed on users and network bandwidth demands. The problem of finding a large-enough user-pool for crowdsensing applications is addressed by various incentive schemes [58, 89]. Energy-efficiency in continuous sensing has been examined in [61, 15] and addressed with various energy-efficient sensing solution [57, 64]. In the thesis we explore the limits of crowdsensing in its ability to map mobile network topology, which may be interesting for various crowdsensing communities [32, 14], mobile network providers [46] and industry [11].

According to recent survey [45], there is no mobility model available to express user-cell association. There were numerous attempts to describe user's movement: Random Way Point, Random Walk [55], Random Direction and Truncated Levy Walk [75], Markovian Way Point [40] and Gaussian-Markov [60] models, Freeway and Manhattan mobility models [7], Obstacle Model [44], and Social models [47]. Human mobility patterns are described in [31], Self-similar Least Action Walk (SLAW) model [56] captures human nature in visiting similar places . Other models are used to predict future places of user's presence [19], or to recognize significant places in the mobility trace [52, 90], but these work usually on short-term outlook. On the contrary, model presented in [79] captures user-place association and the strength of such ties, but it is designed for long-term mobility predictions on the order of months. Closest to our approach in mobility modeling is the methodology recently proposed in [77], which examines human mobility on the basis of motifs from network theory [6]. Similarly to the work in [41, 38] we use a real trace and generate new, synthetic traces, but authors of [41] focus on metropolitan scales and model [38] demands a large number of parameters, which were not derivable from the dataset we used.

2 Aim of the Dissertation Thesis

We aim to deliver convincing results about limits of mobility-related data-collection methods, which may help understand the extent to which a particular technology or tracking method is applicable. The presented work concentrates on addressing the above-mentioned problems and related limits. The primary objectives of the thesis are:

- Limits of network-based active tracking—to show, on an existing tracking solution in a live mobile network, how many users can be tracked simultaneously and how often. To understand the ability of active tracking method to scale in country-wide applications.
- Passive tracking data utilization—to examine the accuracy of network-based passive-tracking data in spatial and temporal dimensions and to propose an extension of the data towards more accurate interpretation.
- Coverage capabilities of crowdsensing—to assess how effective may crowdsensing be in mapping a mobile network infrastructure, measured in the size of a user-pool and the time needed to map critical mass of knowledge about the mobile network.

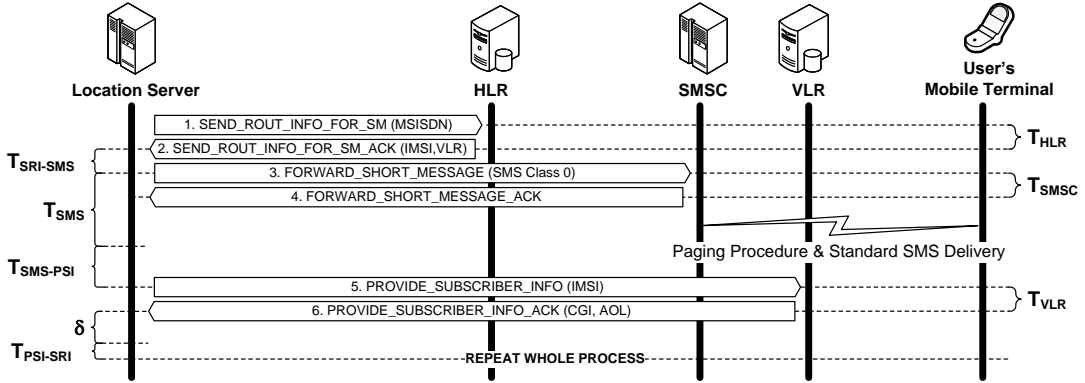


Figure 1: Message flow for obtaining the current Cell-ID of a user. The vertical lines represent time, note the definitions of time intervals T_* between the messages: $T_{\text{SRI-SMS}}$, $T_{\text{SMS-PSI}}$, and $T_{\text{PSI-SRI}}$ represent working times of the Location Server between different types of messages, T_{HLR} , T_{SMSC} , and T_{VLR} , denote service response times of adjacent network nodes, δ stands for the tracking interval.

3 Working Methods and Selected Results

Active Tracking in Mobile Networks

Active tracking is a promising approach in tracking *any* mobile-network user with unprecedented temporal granularity. In the thesis we have presented a particular type of network-based active tracking with Cell-ID positioning — the SMS-based tracking. We have studied the limits of network-based active tracking from various points of view, using an existing active-tracking solution, the SS7Tracker [18, 21], running on a modular signaling platform called SS7Box.

The Cell-ID of a cell currently serving a user can be obtained by sending three MAP¹ primitives in the network. The only input is the Mobile Subscriber ISDN Number (MSISDN) of the user whose Cell-ID is to be retrieved. Figure 1 shows the type of the messages and their ordering, described in detail as follows: First, the Visitors Location Register (VLR) currently maintaining the user record in the network needs to be found, which is done by the Send Routing Info (SRI) request from the Location Server to the user’s Home Location Register (HLR) (message 1). Second, a request to send an “invisible” *SMS Class 0* to the user is sent by the SMS-Center, using the *Forward Short Message* (SMS) (messages 3–4) message. It involves *paging* the user’s mobile terminal which results in updating the location information (Cell Global Identity (CGI) of the cell where the user is located) in the VLR. Finally, an up-to-date Cell-ID and Age Of Location (AOL) is retrieved using a Provide Subscriber Info (PSI) request (messages 5–6).

Based on a measurement on a large-scale tracking, we have described performance statistics of the tracking platform and adjacent network nodes. We have focused particularly on *service response time* distributions of adjacent network nodes (HLR, SMS-Center (SMSC), VLR), SS7Box *working time* and *message lengths variance*. Until explicitly specified, all data described in this section come from a measurement during 6 hours on a sample track with 500 users with a tracking interval of 2 minutes. The total count of sent and received messages is about 700,000 messages.

The tracking process of a set of users is characterized by two main parameters. These

¹Mobile Application Part (MAP) is the topmost part of the Signaling System Number 7 (SS7) stack, it enables applications in the GSM core network

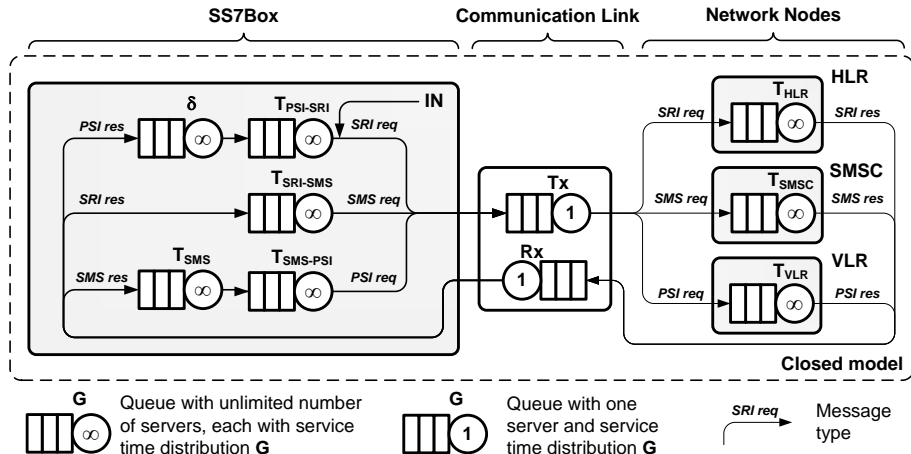


Figure 2: Simulation model. The communication link is modeled as a pair of queues for transmission and reception direction respectively, each with one server that processes messages with time of service equal to the ratio of message length and data rate. Network nodes are modeled as queues with unlimited number of servers thus no waiting time is applied. SS7Box working times, T_{SMS} time and tracking interval δ are modeled in the same way.

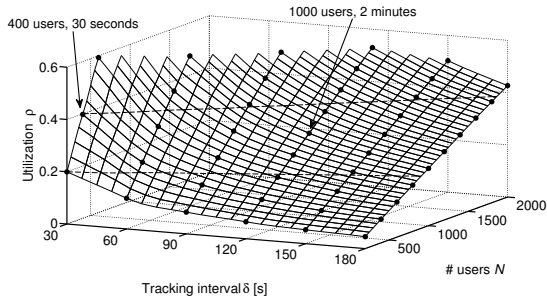


Figure 3: Signaling link utilization. Dependency of signaling-link utilization ρ on the number of tracked users N and the tracking interval δ . As expected, a shorter tracking interval or an increasing number of tracked users cause higher utilization. Dashed lines mark limiting values of allowed utilization maximum ρ_{max} .

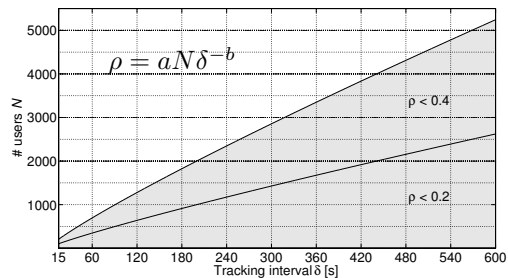


Figure 4: Area of safe operation. Each combination of number of users N and tracking interval δ falling into the area of safe operation (in gray) is allowed with respect to ρ_{max} . The SS7Tracker implementation of SMS-based active tracking yields sufficient performance to track thousands of users with a period of minutes.

are (1) the number of unique users tracked, denoted N , and (2) the tracking interval δ , i.e., the time between consecutive Cell-ID retrievals per single user. We have presented the performance evaluation in the sense of a minimum deployment unit — the SS7Box’s interconnection to the GSM network is realized by only a single E1 line with one timeslot (link) for signaling, which limits the data rate to 64 kbit/s and, moreover, the link utilization policies must be applied². We have modeled the communication process during active tracking in Matlab R2008b and SimEvents Library 2.3 using discrete-time queuing network simulation, schematically depicted in Figure 2. The model is *probabilistic* and *closed*. Signaling link utilization in one direction is interpreted as *server utilization* ρ , i.e., the proportion of the time the server is busy. According to [42], the direction (transmission or reception) with higher utilization stands for the overall signaling link utilization. The main results from the simulation are depicted in Figures 3 and 4.

²Each signaling link should provide extra capacity and thus its utilization ρ must not exceed a value of maximum utilization ρ_{max} , which usually lies between 0.2 and 0.4 [42].

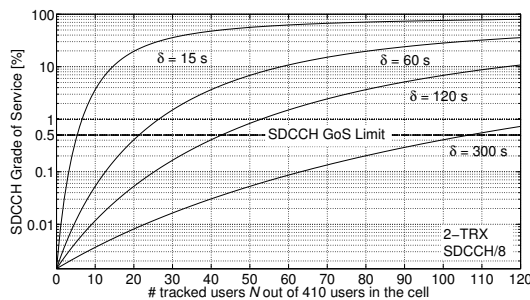


Figure 5: Impact of SMS-based active tracking on SDCCH GoS in a cell. Only 60 tracked users with 60 s tracking interval precludes every 10th user in the cell from being served on average.

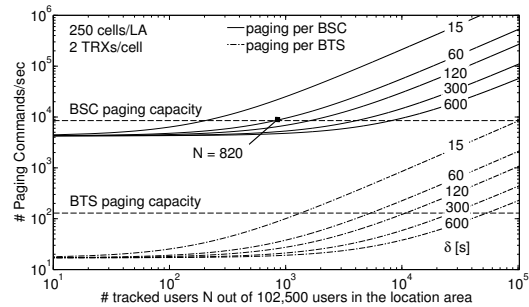


Figure 6: Impact of SMS-based active tracking on paging load. Tracking 820 users in the location area with 60 s tracking interval would disrupt the BSC paging functionality.

We have analyze various limiting factors of active tracking, including basic constraints of the method, constraints of the location server and of the mobile network, finding out that the principal limitation of network-based active tracking is the radio access network.

We have calculated the limitation of the radio access network on different network-infrastructure levels—in a cell and in a location area. We have considered a GSM cell with SDCCH/8 configuration for a 2-Transmitter/Receiver (TRX).

Figure 5 depicts GoS for SDCCHs in a cell as a function of the number of tracked users N and of the tracking interval δ . We assume there are 410 users in the cell, i.e., the maximum TCHs capacity at 2% GoS, and that exactly N of these users are tracked. The graph is calculated using the Erlang B formula for 8 SDCCHs and the offered traffic being a sum of the tracking load and the estimated load from all users in the cell. The impact of the increasing number of tracked users in the cell is significant: only 21 users, tracked every 60 seconds in the cell, suffices to exceed the desired GoS of SDCCHs.

Figure 6 shows the paging load in a location area during tracking as a function of the number of tracked users N and of the tracking interval δ . The number of Paging Commands rises with the number of tracked users, yet longer tracking interval results in slower growth. The graph provides a useful insight: since fewer tracked users suffice to exhaust the BSC paging capacity (8,500 Paging Commands/s [24]) before the BTS capacity is exhausted (17.02 Paging Commands/s per BTS), the bottleneck in the location area is the BSC. Interestingly, under the assumption of positioning an idle mobile terminal in the circuit-switched domain, similar results hold for *all* state-of-the-art network-based positioning methods. Since every positioning method needs to establish a connection with the mobile terminal, the paging procedure is always necessary to locate mobile terminal's cell within a last known location area. We conclude that neither SMS-based active tracking nor any of the state-of-the-art network-based positioning methods can be used for large-scale tracking scenarios, such as tracking all users of a mobile network at the same time.

We have shown how mobility of tracked users, and thus their possible concentration at one place, may affect the performance of the radio access network. We have considered a cell with 2-TRX and the SDCCH/8 configuration, serving 300 users, in which the number or tracked users constantly increases as they arrive into the cell from the neighbor cells.

Figure 7 shows that with more than 16 tracked users arriving in the cell every minute, the desired SDCCH GoS can be exceeded in less than 2 minutes. Such intensity of arrivals can be observed for example before sport events, when tens of thousands of fans meet at a stadium within an hour or two. Dashed lines show, how a leaky-bucket traffic-shaping

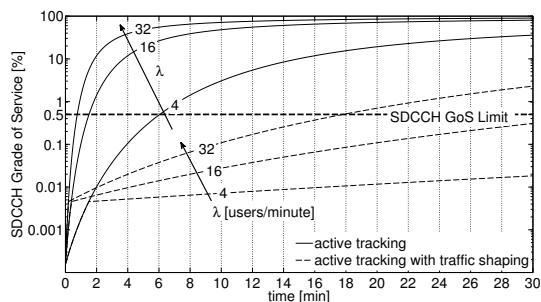


Figure 7: Impact of increasing number of tracked users on SDCCH GoS in a cell. The GoS degrades over time when the tracked users keep concentrating in the cell.

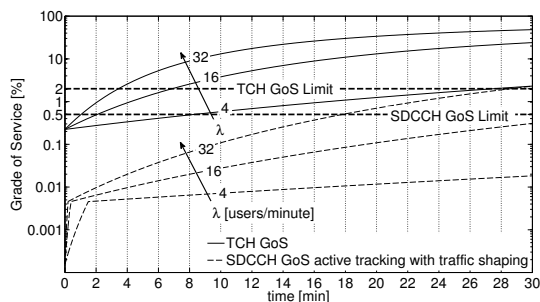


Figure 8: Impact of increasing number of tracked users on TCH GoS in a cell. Traffic shaping can spare SDCCH capacity of the cell, but the capacity of TCHs is exceeded anyway.

mechanism with a fixed rate $r = 6$ positioning requests/min helps to keep GoS under the desired limit. However, because the arriving users bring additional signaling and traffic load, and not only the SDCCH load caused by active tracking, GoS degrades proportionally to the number of users in the cell nonetheless.

Figure 8 depicts the impact of the increasing concentration of tracked users at a cell on GoS of the traffic channels. Since active tracking brings additional load to SDCCHs only, the increase in TCH load and therefore the worse GoS is caused purely by new users in the cells. The most important observation is that the expected SDCCH GoS for a particular λ hits the SDCCH GoS limit long after the TCH GoS limit is reached. As a result, active tracking with the leaky bucket traffic shaping mechanism can spare signaling capacity of the cell, but since arriving users would bring additional voice traffic load, the limiting factor become the capacity of the TCHs anyway.

As a rule of thumb, a mobile network should be dimensioned with respect to all services that the network operator plans to offer to its customers. Should active tracking be a network-safe service, careful re-planning of the mobile network with respect to the expected tracking extent would be necessary.

Extending Utility of Passive Tracking Data

Passive tracking data, mainly the Call Data Records (CDRs), represent a vast amount of easy-to-collect data about *every* mobile network user, for they are automatically generated by telecommunication systems and archived for billing purposes and network troubleshooting. Whereas spatial precision of CDRs is determined by network-cell size, the accuracy in temporal dimension is substantially low. The reason is that the position of a user is recorded *only* at places where the user performs a communication event (SMS, call, data session).

We have addressed the limits of passive-tracking data — their limited accuracy in temporal dimension — by building a probabilistic model of users' position in between communication events. The central idea of our approach was to compare the coarse-grained trajectories from CDRs, and some corresponding ground-truth trajectories representing continuous trace with user's positions. To achieve this, we have spatially extended an existing publicly available dataset, the Reality Mining Dataset (RMD) [23], from which a substitution for both CDRs and a finer trajectory of user-cell association can be derived. We have paired cell identifiers in users' traces with their corresponding geographical coordinates (obtained from the Location API by Google), we have removed spatial outliers

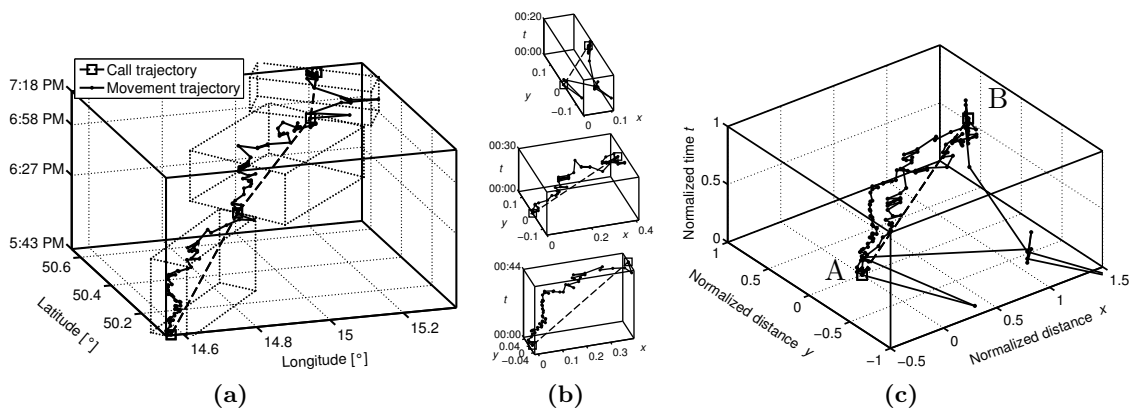


Figure 9: Example of user’s inter-call trajectories aggregation. (a) Four call places divide the call trajectory in three segments, the movement trajectory is divided by call segments in three inter-call trajectories. Each segment determine a separate reference frame (dashed boxes) the orientation of the coordinates is given by the direction of the segment. (b) Each inter-call trajectory is translated to coordinates origin. (c) Inter-call trajectories are scaled into a common reference frame, a normalized space-time cube.

using a novel heuristic approach to agglomerative clustering, and have provided a methodology to extract representative chunks of trajectories [27].

In order to deliver a *general* description of user’s position in between communication records, we have described user’s mobility and detours relative to the inter-call distance—we have studied only pairs of consecutive communication records *at distinct places*. For our purposes, we considered distinct places to be places *more than three kilometers apart*. Conclusions drawn from lower inter-call distance can be affected by the accuracy of Cell-ID positioning method that delivers user’s position only as an approximation of the geographical coordinates within a cell, while the exact position is not known. We have worked with an aggregated view of user movement between calls, derived as depicted in Figure 9. There are 901 inter-call trajectories, made by 56 users out of the 94 user sample in the Reality Mining Dataset. Figure 10 shows these inter-call trajectories aggregated in the common reference frame.

Through the analysis of coarse-grained CDR-based *call trajectories* and corresponding finer trajectories of user-cell associations, *movement trajectories*, we have examined spatio-temporal properties of user’s movement between consecutive calls at distinct places. Figure 11 shows the kernel density estimation of spatio-temporal probability distribution of user’s inter-call movement: as time passes, users move from origin *A* towards destination *B* and take detours from the straight *A – B* direction. The distribution shows three important aspects of inter-call mobility:

1. *Unskewed behavior in spatial dimension.* About half of users closely follow the direct *A – B* linear interpolation line. This is observable in symmetry of 0.5-quantile projection on the *xy*- and *yt*-coordinate planes — it implies unskewed behavior with respect to detours from the straight *A – B* course.
2. *Straight course between calls.* Some users take approximately the shortest path from the origin call-place *A* to the destination place *B*. This is indicated by symmetry in spatial dimension (*xy*-coordinate plane) which does not change in time (*yt*-coordinate) and the fact that 0.5-quantile projection on *xt* plane enclose the direct *A – B* interpolation line.

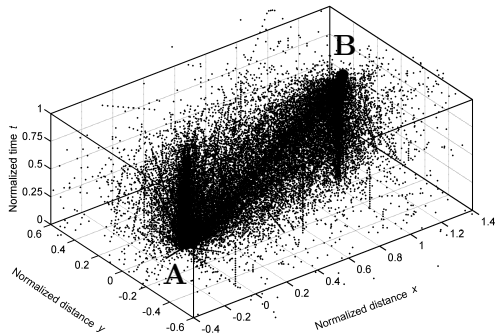


Figure 10: Aggregated inter-call trajectories. Each point represent user’s position between calls (places A and B in normalized common reference frame) after aggregation of inter-call trajectories from the Reality Mining Dataset.

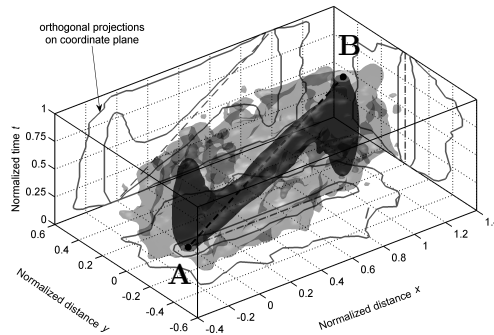


Figure 11: Spatio-temporal characteristics of inter-call movement. Kernel density estimation of the spatio-temporal probability distribution of users’ position between calls. Isosurfaces enclose 0.5- and 0.9-quantiles (dark and light gray).

3. *Staying behavior at call places.* From the shape of the projection on the xt -coordinate plane at points $x_A = 0$ and $x_B = 1$ it follows that some users tend to stay at the origin call-place A before they move to destination B, or they leave the origin soon after the act of communication and stay in the vicinity of the destination place B.

These findings are in strong contrast with assumptions of existing modeling methods, the linear weighted interpolation [33] and the probabilistic extension of space-time prism by Winter [87], and therefore these models cannot be used for precise posterior CDR analysis.

To improve the accuracy of CDR-based deduction of user’s presence at call places over time, we have formulated a new probabilistic Inter-Call Mobility (ICM) model, spatio-temporally fitting the aggregated mobility behavior of the Reality Mining Dataset users. Using a finite Gaussian mixture model, the ICM model approximates the ground-truth of users’ inter-call movement: from a single timestamped call coordinate to a probabilistic distribution of user’s position between calls.

We have represented the aggregated inter-call trajectories from the RMD as a dataset $\mathcal{Z} = \{\mathbf{z}_i\}_{i=1}^N$ of N spatio-temporal records $\mathbf{z}_i \in \mathbb{R}^3$, where each datapoint $\mathbf{z}_i = \{\mathbf{z}_{s,i}, z_{t,i}\}$ comprise spatial coordinates $\mathbf{z}_{s,i} \in \mathbb{R}^2$ and a temporal value $z_{t,i} \in \mathbb{R}$. The initial estimate of mixture parameters θ , needed by the EM algorithm, is given by K -means clustering. Finally, we have run the EM algorithm and retrieve the mixture parameter estimates and the values of model-fit criterions. We have fitted a set of GMMs with $K = 5, \dots, 20$ components to the RMD inter-call trajectories and compared the resulting AIC and the BIC criterions, and have selected the model with 10 components.

We have defined the ICM model as follows. Given the origin position $(x_A, y_A, t_A) = (0, 0, 0)$ and the destination position at $(x_B, y_B, t_B) = (1, 0, 1)$, the probability $\Phi(x, y, t)$ of finding a user at coordinates $(x, y) \in \mathbb{R}^2$ at a time $t \in \mathbb{R}, 0 \leq t \leq 1$, is defined as a Gaussian mixture $p(\mathbf{z} = (x, y, t)^T; \theta) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ of $K = 10$ components with parameters θ (given in the thesis). Figure 12 provides a visual representation of the ICM model, Figure 13 shows an indication of position and shape of the components. As the model approximates the inter-call trajectories of the RMD user-pool, its shape is similar to the estimation in Figure 11.

The ICM model is expressed analytically and thus reusable and general enough for practical application to *any* CDR traces. Moreover, the ICM model allows for description

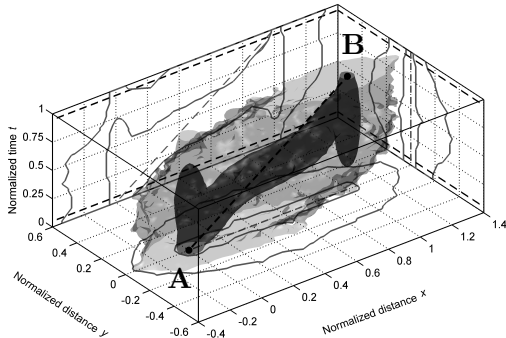


Figure 12: Inter-Call Mobility model. A spatio-temporal probability distribution of user’s position between two communication records at distinct places A and B . Isosurfaces enclose 0.5- and 0.9-quantiles (dark and light gray, respectively).

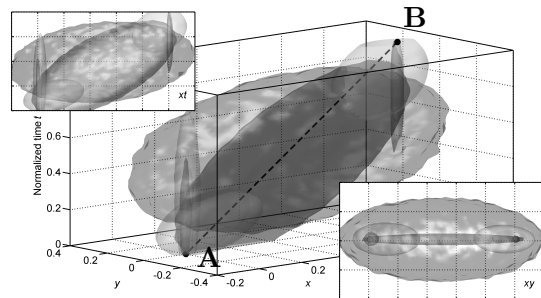
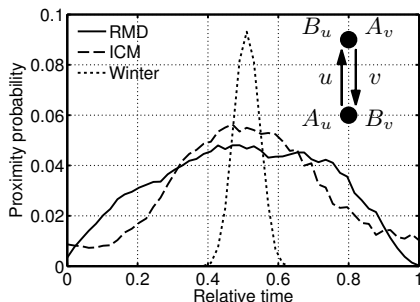
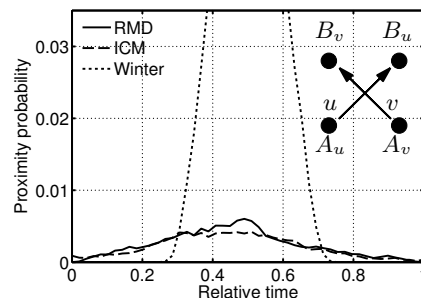


Figure 13: Components of the ICM model. There are 10 three-dimensional Gaussian components whose mixture constitute the ICM model. The orthogonal projections on xt - and xy -coordinate planes are depicted in the upper left and lower right corner, respectively.



(a) *Opposite travel direction*



(b) *Cross travel direction*

Figure 14: Proximity probability of two users. Users u and v travel between their calls at places A_u, B_u and A_v, B_v in two hypothetical scenarios (a) and (b). The ICM model provides more accurate results than the Winter’s probabilistic extension of space-time prisms.

of user’s position at a particular time in between calls and, vice versa, given geographical coordinates, probability of user’s presence at a particular position over time can be derived.

On the example of user proximity probability we have demonstrated that the ICM model outperforms different existing mobility models. By the term *proximity* we understand that two users are closer than x percent of the inter-call distance. Specifically, example in Figure 14 shows 5% proximity probability computed from the ICM model, Winter’s model, and from the estimation of inter-call distribution from the RMD in Figure 11, in two hypothetical scenarios:

Scenario (a), in which *users travel towards each other*, shows that the RMD data demonstrate several times higher proximity probability than the Winter’s model, over majority of time. It is a consequence of user’s staying behavior at call places: both users can simply meet even at the very ends of the time budget, at the origin and destination places.

Scenario (b), with *non-identical user origin and destination places*, shows a significantly lower meeting probability for the ICM model and the RMD in comparison to Scenario (a). It follows from the fact that a meeting chance at either origin or destination places, here impossible, contributes highly to the overall proximity probability.

The ICM model fills the gap between coarse-grained, CDR-based mobility description, limited by the frequency of communication activity of a mobile user, and the highly accu-

rate positioning by GPS, sensor networks or sensing wearables. Broad inter-disciplinary research may benefit from our findings, as CDRs become more readily available to computer scientists, urban planners, geographers, and others.

Exploring Limits of Crowdsensing

Crowdsensing techniques are often utilized by communication market players such as Google and Apple to discover the structure of mobile networks. GPS-enabled phones of customers send their current GPS coordinates and Cell Identifiers (Cell-IDs) to a server that collects, clusters, fingerprints, and stores such data from all customers. Such a Cell-ID database can subsequently lead to the geolocation: 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. In the thesis, we have posed two key questions: *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?* The answers are vital to judge the ability of crowdsensing to build a new Cell-ID database, or rapidly update an inadequate, malfunctioning or obsolete Cell-ID database. We have studied the limits of terminal-based tracking from the point of view of its potential in mapping of mobile network cells to geographic locations.

Since there is no mobility model which counts user-cell associations over a time period (see [45] for a survey), we have proposed a trace-based mobility model to quantify users' ability to detect a number of cells in the network. We have used the NRC-Lausanne dataset [54] to extract the information about users' *places*, i.e., significant locations in terms of time spent at the corresponding cells, and the *transitions* between places during one day. The dataset contains a timestamped sequence of Cell Global Identities (CGIs) per user with one record per every cell change during the campaign period (referred to as *cell trace*). We have processed the data by dividing the cell trace of each user into day-long sequences, each starting at midnight, excluding days where the mobile phone was off, and removing weekends. There is a total of 6,667 day-sequences in the dataset.

To extract mobility patterns from the data, we have processed each day-sequence and find all places in the transitions for that particular day. Figure 15 shows an example of one day-sequence with three transitions T1–T3 of a user from the dataset.

We have focused on extracting the following features from the data: **(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. The statistics above is necessary to describe artificial traces by means of users' daily patterns (visited places and transitions between them), temporal characteristics of varying human mobility during one day, and quantification of daily user-cell associations in terms of unique cells visited.

We have expressed the model features F1 and F2 by mining the transition probabilities between the places, depending on the time of the day. We have simplified the structure of time by quantizing the day into $T = 288$ 5-minute equidistant time slots, and expressed the probability that a transition between places $i, j \in L$ starts during the time period \tilde{t} , $\tilde{t} \in \{1, \dots, T\}$, by a matrix $p_{i,j}^{\tilde{t}}$. The duration of a transition (model feature F3) has been estimated from the transition sequences 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. We have observed that the transition

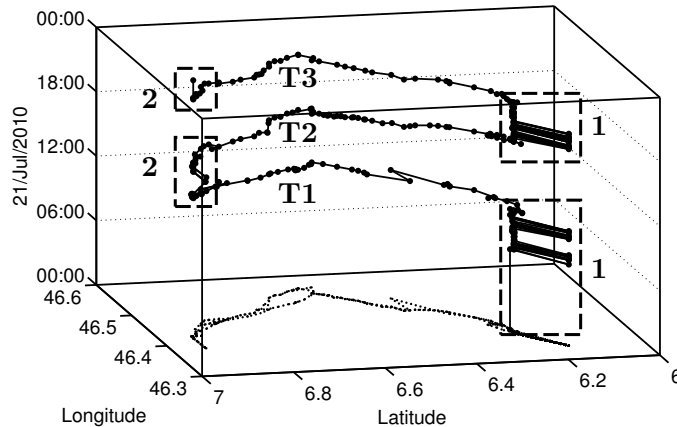


Figure 15: Example of a transition sequence. Rectangles with labels 1 and 2 enclose sets of cells which represent *places* — the user visited two places (1 and 2) during the day. There are in total three transitions between the places: T1 and T3 between places 1 and 2, and T2 between places 2 and 1.

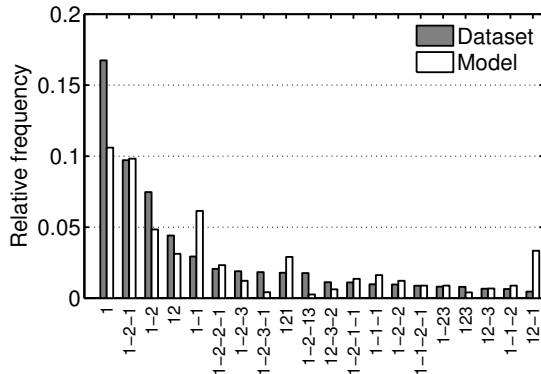


Figure 16: The example of the 20 most frequent daily patterns. The most frequent pattern represents staying at a place for the whole day.

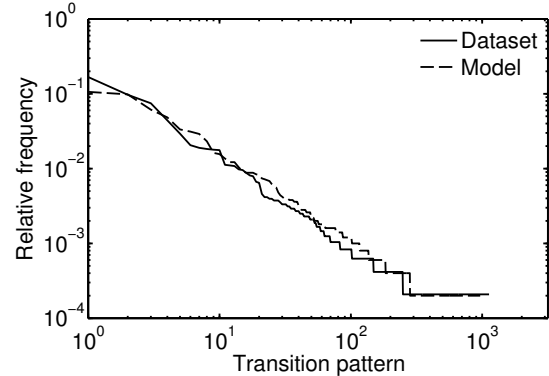


Figure 17: Heavy-tail distribution of different daily patterns. A small number of patterns occur often while numerous patterns are rare.

lengths can be handily approximated by Normal distribution $\mathcal{N}(\mu, \sigma)$ with the mean and standard deviation parameters dependent on the duration of the transition δ and its class. We denote these distributions by $g_{\text{new}}(\delta)$, $g_{\text{same}}(\delta)$, and $g_{\text{old}}(\delta)$. Parameters $p_{i,j}^t$, f_* , and g_* constitute our model. The generation of a new, synthetic transition sequence from the above parameters is algorithmically described in the thesis.

We have validated our model by comparing the features F1–F3 from synthetic traces, generated by the model, and the original traces from 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. Figure 16 compares the most frequent patterns in the dataset with the synthetic traces generated from the model, showing a high correspondence in the frequency. Figure 17 shows that the distribution of users’ daily patterns is heavy-tailed. The fine-grained temporal characteristics of human movement during one day, the model feature F2, are depicted in Figure 18. It compares transition probabilities $p_{i,j}^t$ during a day per each of the three transition classes. Daily user-cell associations, the model feature F3, are depicted in Figure 19a. Clearly, the model well quantifies the total number

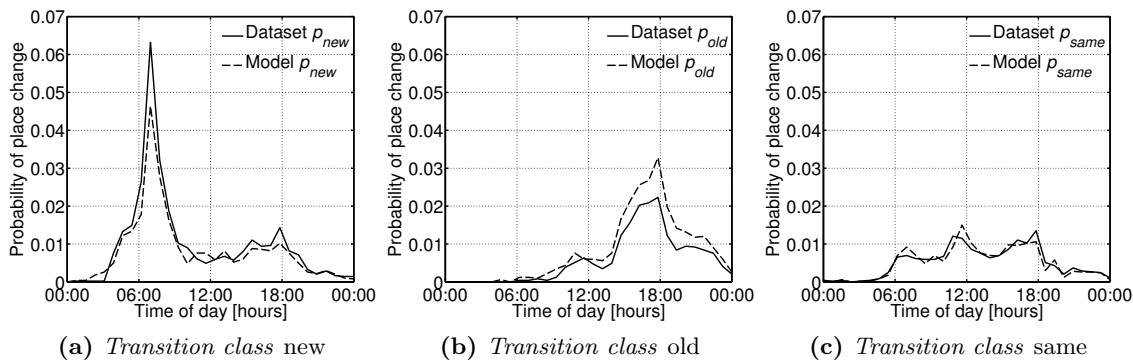


Figure 18: Probability of transitions during day. (a) In the morning users commute to new, previously not visited places, (b) in the afternoon they return to previously visited places, while during the day (c) they tend to leave the place and return to the same place later.

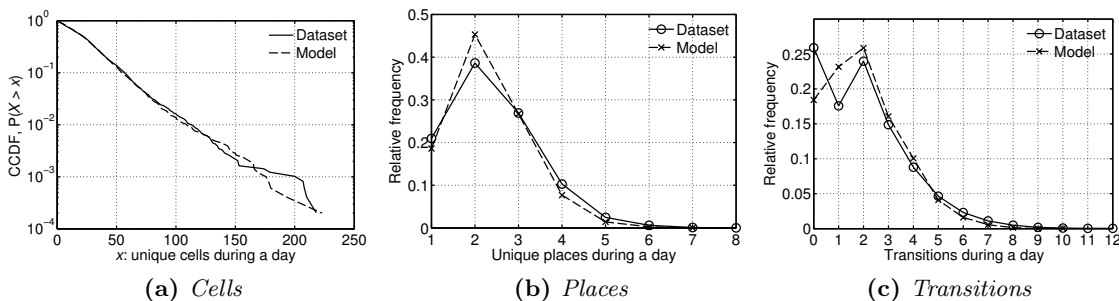


Figure 19: Probability of transitions during day. (a) Distribution of the total unique cells during a day. (b) Distribution of the total places during a day. Almost half of the transition sequences contains only two places. (c) Distribution of the total transitions during a day.

of cells visited during a day. This is achieved by the fact that the synthetic transition sequences follow similar distributions of total number of the places visited during a day (Figure 19b) and the total number of transitions between the places (Figure 19c).

We have applied the model-based traces of user movement by a large-scale simulation to an approximation of a mobile network topology. The simulation is simply a pseudo-random walk on a graph, whose nodes are cells and the neighboring structure is given by Delaunay triangulation. We have simulated the movement of users in a network with $c = 5,000$ cells and number of users $n = \{0.25c, 0.5c, \dots, 5c\}$. Figure 20 shows a relationship between the ratio of cells observed during a day and the number of users involved in crowdsensing.

We have discussed the applicability of crowdsensing as a fight-back method against obfuscation of the network topology by a Dynamic Cell-ID method [8, 17, 85], which is based on periodical changes of Cell-IDs in the network to prevent unauthorized geolocation services offered by third-parties. Figure 21 shows the results for $c = \{5,000, 15,000\}$ users in a network that consists of 5,000 cells. By comparing Figures 21a–21b we see, unsurprisingly, 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 around noon, as mobility of users is low.

Crowdsensing as a fight-back method against the Dynamic Cell-ID method is quite a powerful tool, but as our results show it significantly depends on the user-pool size.

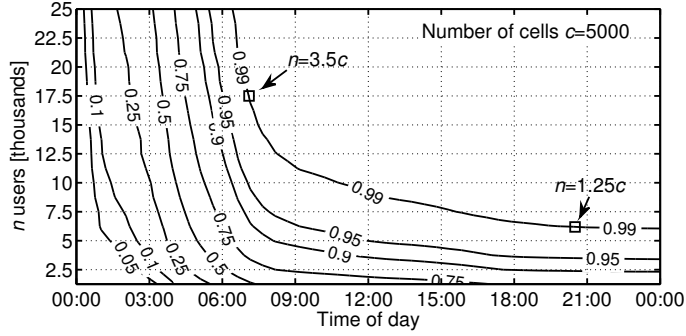


Figure 20: Ratio of cells observed during a day. At least $n = 1.25c = 6,250$ users are needed to observe at least 99% of all cells by the end of the day. Because of a low number of users who travel during the early morning, about $n = 3.5c$ users are needed to visit 99% of the cells by 07:00. Markedly, having more than $3.5c$ users yields only minor improvements.

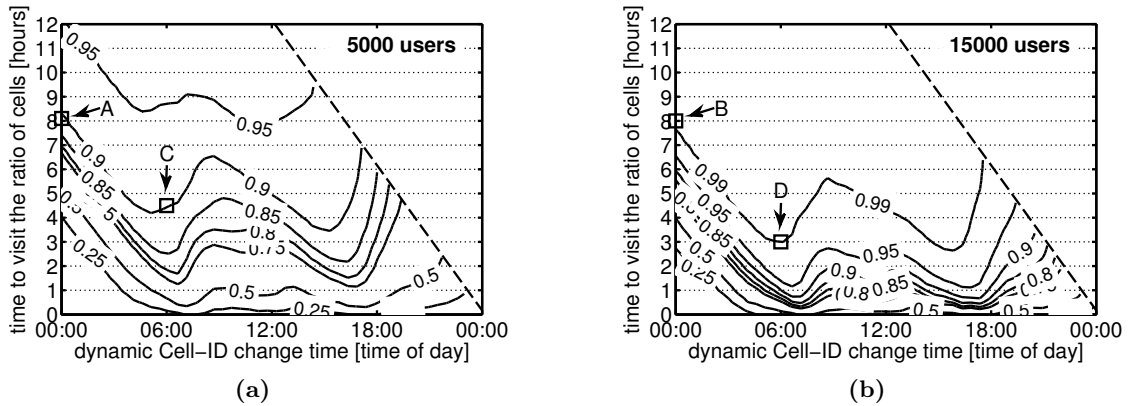


Figure 21: Impact of dynamic Cell-ID renumbering time on cell discovery in a network with 5,000 cells. Example A shows that in case the dynamic Cell-ID renumbering occurs at midnight, 5,000 users visits 90% of all cells at 08:00 (after 8 hours from midnight). Higher number of users results in shorter discovery time and more cells discovered (example B). Examples C and D show the Cell-ID-crowdsensing performance in the time of the day with high user mobility. If the dynamic Cell-ID renumbering occurs at 06:00, then 5,000 users discovers 90% of all cells in 4 hours (at 10:00), whereas with 15,000 users 99% of all cells are discovered in 3 hours (at 09:00).

Apparently, having the network scanned within couple of hours anytime during a day is possible, but with an almost unrealistic number of users. As a result, with Dynamic Cell-ID adopted, a third-party service relying on a crowdsensed Cell-ID database may suffer from bad localization performance for hours-long period. Our study on crowdsensing limits may be particularly interesting for contemporary and new Cell-ID-based location-services providers, but, at the same time, for mobile network providers as a hint whether they should try to monetize their costly network infrastructure and apply network-infrastructure obfuscating mechanisms.

4 Conclusion

In the thesis, we have investigated principal limits of tracking methods in mobile networks. Tracking data, simply a timestamped history of mobile users' positions in the network, is a sought-after and still scarce source of information for research studies in telecommunications, transportation, urban studies, network design, cloud computing, and other fields of science.

The main problem with tracking methods in mobile network is that their suitability for large-scale tracking, network-wide application, and their technological limitations are often not discussed or remain unknown. We have addressed the following topics related to the limits of tracking methods in mobile networks:

(1) Technological limits of networks-based active tracking: A model describing SMS-based active tracking in mobile networks has been presented, its parameters come from a large-scale measurement on a real tracking platform. Limits of the implementation have been simulated using a discrete-time simulation of the tracking process. Limits of SMS-based active tracking has been deduced and described: basic constraints of the method, constraints of the location server and the mobile network. Principal limitations of the GSM radio access network has been derived.

(2) Extensions beyond low temporal granularity of passive tracking data: Detailed analysis of users' mobility behavior in between their communication places (calls, text messages) has been presented. It showed that the nature of human mobility between communication events is in strong contrast with assumptions of existing modeling methods. A probabilistic spatio-temporal refinement of Call Data Records, the *Inter-Call Mobility* (ICM) model, has been presented. The model is analytically tractable and can be used for practical application to any CDR traces to describe spatio-temporal mobility behavior of network users between their communication events.

(3) Scaling limits of cooperative terminal-based tracking: A data-driven mobility model has been proposed to express the number of unique mobile-network cells a user is capable of visiting during one day. The model describes users' daily patterns, capture fine-grained temporal characteristics of human movement during a day, and quantifies daily user-cell associations. A large-scale simulation of users' movement in an approximation of a mobile network has been conducted to assess the coverage capabilities of a user pool for the time period in an area. Limits of crowdsensing in discovering the mapping of mobile-network cell identifiers to geographic locations as a method against mobile-network-topology obfuscation has been presented.

References

- [1] 3GPP, "TS 09.78: Customised Applications for Mobile network Enhanced Logic (CAMEL); CAMEL Application Part (CAP) specification," 2000, v7.1.0.
- [2] 3GPP, "TS 23.271: Functional stage 2 description of Location Services (LCS)," 2010, v10.0.1.
- [3] 3GPP, "TS 45.811: Feasibility study on Uplink TDOA in GSM and GPRS," 2002, v6.0.0.
- [4] R. Ahas *et al.*, "Mobile positioning in space-time behaviour studies: Social positioning method experiments in estonia," *Cartography and Geographic Information Science*, vol. 34, no. 34, pp. 259–273, 2007.
- [5] R. Ahas *et al.*, "Evaluating passive mobile positioning data for tourism surveys: An Estonian case study," *Tourism Management*, vol. 29, no. 3, pp. 469–486, 2008.
- [6] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450–461, 2007.
- [7] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proc. of INFOCOM*. IEEE, 2003, pp. 825–835.
- [8] M. Bensimon *et al.*, "Method to mask cell identifiers or location area codes of a mobile network with respect to a mobile terminal," U.S. Patent 20 090 170 477, 2009.
- [9] A. Berg, "Method for providing location information," U.S. Patent 20 020 160 789, 2002.
- [10] K. Biermann, "Betrayed by our own data," *Zeit Online*, 2011. [Online]. Available: <<http://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte>>

- spitz> [Accessed: 2013.04.07]
- [11] Broadcom Corporation, “AGPS Server and Worldwide Reference Network,” 2007.
 - [12] N. Caceres, J. P. Wideberg, and F. G. Benitez, “Deriving origin destination data from a mobile phone network,” *IET Intelligent Transport Systems*, vol. 1, no. 1, pp. 15–26, 2007.
 - [13] California Center for Innovative Transportation, “Mobile millenium,” 2013. [Online]. Available: <<http://traffic.berkeley.edu/>> [Accessed: 2013.04.30]
 - [14] Combain Mobile AB, “Location-API,” 2009. [Online]. Available: <<http://location-api.com/>> [Accessed: 2011.06.15]
 - [15] I. Constandache, R. Choudhury, and I. Rhee, “Towards mobile phone localization without war-driving,” in *Proc. of INFOCOM*. IEEE, 2010, pp. 1–9.
 - [16] N. Deligiannis, S. Louvros, and S. Kotsopoulos, “Mobile positioning based on existing signalling messages in GSM networks,” 2007. [Online]. Available: <<http://sfhmy.ntua.gr/2007/papers/paper10.pdf>> [Accessed: 2013.04.07]
 - [17] Z. Duan, X. Ren, and J. Qu, “Method and device for configuring a cell identity,” Chinese Patent WO/2009/140 914, 2009.
 - [18] K. Dufková *et al.*, “Can active tracking of inroamer location optimise a live GSM network?” in *Proc. of CoNEXT Student Workshop*. ACM, 2007, pp. 1–2.
 - [19] K. Dufkova *et al.*, “Predicting user-cell association in cellular networks from tracked data,” in *Proc. of MELT*. Springer-Verlag, 2009, pp. 19–33.
 - [20] K. Dufkova *et al.*, “Energy savings for cellular network with evaluation of impact on data traffic performance,” in *Proc. of EW*, 2010, pp. 1–9.
 - [21] K. Dufková *et al.*, “Active GSM cell-id tracking: ”Where did you disappear?”,” in *Proc. of MELT*. ACM, 2008, pp. 7–12.
 - [22] P. Dutta *et al.*, “Common sense: participatory urban sensing using a network of handheld air quality monitors,” in *Proc. of SenSys*. ACM, 2009, pp. 349–350.
 - [23] N. Eagle, A. S. Pentland, and D. Lazer, “Inferring friendship network structure by using mobile phone data,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15 274–15 278, 2009.
 - [24] Ericsson, “Location area dimensioning guideline,” 2002, 19/10056-HSC 103 12/4 Rev C 2005-06-09.
 - [25] Ericsson, “Mobile Positioning System,” 2007. [Online]. Available: <http://www.ericsson.com/ourportfolio/products/mobile-positioning-system?nav=fgb_101_743|fgb_101_746> [Accessed: 2013.04.02]
 - [26] Ericsson, “Positioning with LTE – Maximizing performance through integrated solutions,” 2011.
 - [27] M. Ficek and L. Kencl, “Spatial extension of the Reality Mining Dataset,” in *Proc. of MASS*, 2010, pp. 666–673.
 - [28] M. Ficek, “CRAWDAD data set ctu/personal (v. 2012-01-11),” 2012. [Online]. Available: <<http://crawdad.org/ctu/personal>> [Accessed: 2013.04.07]
 - [29] M. Ficek, T. Pop, and L. Kencl, “Active tracking in mobile networks: An in-depth view,” *Computer Networks*, 2013, *in press*. [Online]. Available: <<http://www.sciencedirect.com/science/article/pii/S1389128613000996>> [Accessed: 2013.04.03]
 - [30] R. Ganti *et al.*, “Mobile Crowdsensing: Current State and Future Challenges,” *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.
 - [31] M. C. González, C. A. Hidalgo, and A.-L. Barabasi, “Understanding individual human mobility patterns,” *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
 - [32] GSMweb, “GSMweb,” 2010. [Online]. Available: <<http://gsmweb.cz/>> [Accessed: 2013.03.19]
 - [33] R. H. Güting and M. Schneider, *Moving Objects Databases*. Morgan Kaufmann, 2005, ISBN 0-12-088799-1.
 - [34] T. Hägerstrand, “What about people in regional science?” *Papers in Regional Science*, vol. 24, no. 1, pp. 6–21, 1970.

- [35] E. Halepovic and C. Williamson, "Characterizing and modeling user mobility in a cellular data network," in *Proc. of PE-WASUN*. ACM, 2005, pp. 71–78.
- [36] A. Heinemann, J. Kangasharju, and M. Muehlhaeuser, "Opportunistic data dissemination using real-world user mobility traces," in *Proc. of AINAW*. IEEE, 2008, pp. 1715–1720.
- [37] H. Hohwald, E. Frías-Martínez, and N. Oliver, "User modeling for telecommunication applications: Experiences and practical implications," in *Proc. of UMAP*. Springer-Verlag, 2010, pp. 327–338.
- [38] W.-J. Hsu *et al.*, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1564–1577, 2009.
- [39] R. Huerta and L. S. Tsimring, "Contact tracing and epidemics control in social networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 66, no. 5, p. 056115, 2002.
- [40] E. Hyttia, P. Lassila, and J. Virtamo, "A markovian waypoint mobility model with application to hotspot modeling," in *Proc. of ICC*. IEEE, 2006, pp. 979–986.
- [41] S. Isaacman *et al.*, "Human mobility modeling at metropolitan scales," in *Proc. of MobiSys*. ACM, 2012, pp. 239–252.
- [42] ITU-T, "Methods for dimensioning resources in Signalling System No.7 networks," ITU-T Recommendation E.733, 1998.
- [43] ITU-T, "Measuring the information society 2012," 2012. [Online]. Available: http://www.itu.int/ITU-D/ict/publications/idi/material/2012/MIS2012_without_Annex_4.pdf [Accessed: 2013.04.09]
- [44] A. Jardosh *et al.*, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. of MobiCom*. ACM, 2003, pp. 217–229.
- [45] D. Karamshuk *et al.*, "Human Mobility Models for Opportunistic Networks," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 157–165, 2011.
- [46] F. Kattan, "Dynamic cell-ID: Clever way to block Google, but will it backfire?" 2010. [Online]. Available: <http://franciscokattan.com/2010/02/06/dynamic-cell-id-clever-way-to-block-google-but-will-it-backfire/> [Accessed: 2011.08.08]
- [47] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. of INFOCOM*. IEEE, 2006, pp. 1–13.
- [48] Z. Koppanyi *et al.*, "Tracking vehicle in gsm network to support intelligent transportation systems," in *Proc. of ISPRS*, 2012, pp. XXXIX–B2:139–144.
- [49] D. Kotz and T. Henderson, "Crawdad," 2010. [Online]. Available: <http://www.crawdad.org/index.php> [Accessed: 2013.01.07]
- [50] B. Kuijpers and W. Othman, "Trajectory databases: Data models, uncertainty and complete query languages," *Journal of Computer and System Sciences*, vol. 76, no. 7, pp. 538–560, 2010.
- [51] A. Kupper, *Location-Based Services: Fundamentals and Operation*. Wiley, 2005, ISBN 0470092319.
- [52] K. Laasonen, "Mining cell transition data," Ph.D. dissertation, University of Helsinki, Finland, 2009.
- [53] A. LaMarca *et al.*, "Place Lab: Device Positioning Using Radio Beacons in the Wild," in *Proc. of Pervasive*. Springer-Verlag, 2005, pp. 116–133.
- [54] J. K. Laurila *et al.*, "The Mobile Data Challenge: Big Data for Mobile Computing Research," in *Proc. of Pervasive*. Springer-Verlag, 2012.
- [55] J.-Y. Le Boudec and M. Vojnovic, "The random trip model: Stability, stationary regime, and perfect simulation," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1153–1166, 2006.
- [56] K. Lee *et al.*, "SLAW: A mobility model for human walks," in *Proc. of INFOCOM*. IEEE, 2009, pp. 855–863.
- [57] Y. Lee *et al.*, "Comon: cooperative ambience monitoring platform with continuity and benefit

- awareness,” in *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys'12)*. ACM, 2012, pp. 43–56.
- [58] J. J. Li and B. Faltings, “Incentive Schemes for Community Sensing,” in *Proc. of CompSust*, 2012.
- [59] K. Li and T. C. Du, “Building a targeted mobile advertising system for location-based services,” *Decision Support Systems*, vol. 54, no. 1, pp. 1–8, 2012.
- [60] B. Liang and Z. J. Haas, “Predictive distance-based mobility management for multidimensional PCS networks,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 718–732, 2003.
- [61] K. Lin *et al.*, “Energy-accuracy trade-off for continuous mobile device location,” in *Proc. of MobiSys*. ACM, 2010, pp. 285–298.
- [62] B. Ly, “Mobile Data Challenge 2012: Unlocking the secrets of smartphone data,” Conversations by Nokia, 2012. [Online]. Available: <<http://conversations.nokia.com/2012/06/20/mobile-data-challenge-2012-unlocking-the-secrets-of-smartphone-data/>> [Accessed: 2013.03.28]
- [63] S. Mathur *et al.*, “Parknet: drive-by sensing of road-side parking statistics,” in *Proc. of MobiSys*. ACM, 2010, pp. 123–136.
- [64] S. Nath, “Ace: exploiting correlation for energy-efficient and continuous context sensing,” in *Proc. of MobiSys*. ACM, 2012, pp. 29–42.
- [65] J. Novak, “Lokalizacni data mobilnich telefonu: Moznosti vyuziti v geografickem vyzkumu,” Ph.D. dissertation, Charles University in Prague, 2010.
- [66] E. Oliver and S. Keshav, “Data driven smartphone energy level prediction,” 2010. [Online]. Available: <<https://cs.uwaterloo.ca/research/tr/2010/CS-2010-06.pdf>> [Accessed: 2011.08.08]
- [67] Open Mobile Alliance (OMA), “OMA Mobile Location Service V1.2,” 2011. [Online]. Available: <http://technical.openmobilealliance.org/Technical/release_program/mls_v1_2.aspx> [Accessed: 2013.04.04]
- [68] Orange, “D4D Challenge,” 2012. [Online]. Available: <<http://www.d4d.orange.com/home>> [Accessed: 2013.04.07]
- [69] D. Pfoser and C. Jensen, “Capturing the uncertainty of moving-object representations,” in *Advances in Spatial Databases*. Springer, 1999, vol. 1651, pp. 111–131.
- [70] S. Phithakkitnukoon *et al.*, “Activity-aware map: identifying human daily activity pattern using mobile phone data,” in *Proc. of HBU*. Springer-Verlag, 2010, pp. 14–25.
- [71] A. Quigley *et al.*, “BlueStar, a Privacy Centric Location Aware System,” in *Proc. of PLANS*. IEEE, 2004, pp. 684–689.
- [72] G. Ranjan *et al.*, “Are call detail records biased for sampling human mobility?” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, pp. 33–44, 2012.
- [73] C. Ratti *et al.*, “Mobile landscapes: Graz in real time.” in *Location Based Services and TeleCartography*. Springer-Verlag, 2007, pp. 433–444, ISBN 978-3-540-36727-7.
- [74] S. Reddy *et al.*, “Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype,” in *Proc. of EmNets*. ACM, 2007, pp. 13–17.
- [75] I. Rhee *et al.*, “On the Levy-walk nature of human mobility,” in *Proc. of INFOCOM*. IEEE, 2008, pp. 924–932.
- [76] P. Schmitz and A. Cooper, “Using mobile phone data records to determine criminal activity space,” in *Proc. of IQPC*, 2007, pp. 1–29.
- [77] C. Schneider *et al.*, “Unraveling daily human mobility motifs,” *Journal of the Royal Society Interface*, 2013, under review. [Online]. Available: <http://humnet.scripts.mit.edu/wordpress/wp-content/uploads/2011/04/Interface_paper_2013.pdf> [Accessed: 2013.04.07]
- [78] L. Shi and T. Wigren, “AECID fingerprinting positioning performance,” in *Proc. of GLOBECOM*. IEEE, 2009, pp. 1–6.
- [79] C. Song *et al.*, “Modelling the scaling properties of human mobility,” *Nature Physics*, vol. 6,

- no. 10, pp. 818–823, 2010.
- [80] E. Trevisani and A. Vitaletti, “Cell-ID location technique, limits and benefits: an experimental study,” in *Proc. of WMCSA*. IEEE, 2004, pp. 51–60.
- [81] M. Vieira *et al.*, “Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics,” in *Proc. of SocialCom*. IEEE, 2010, pp. 241–248.
- [82] H. Wang *et al.*, “Transportation mode inference from anonymized and aggregated mobile phone call detail records,” in *Proc. of ITSC*. IEEE, 2010, pp. 318–323.
- [83] P. Wang *et al.*, “Understanding the spreading patterns of mobile phone viruses,” *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.
- [84] S. Weixiong *et al.*, “Collecting and analyzing mobility data from mobile network,” in *Proc. of IC-BNMT*. IEEE, 2009, pp. 810–815.
- [85] P. Willars, J. Bolin, and T. Wigren, “Methods and systems for obscuring network topologies,” U.S. Patent 20 110 116 632, 2011.
- [86] D. Willkomm *et al.*, “Primary users in cellular networks: A large-scale measurement study,” in *Proc. of DySPAN*. IEEE, 2008, pp. 1–11.
- [87] S. Winter and Z.-C. Yin, “Directed movements in probabilistic time geography,” *International Journal of Geographical Information Science*, vol. 24, no. 24, pp. 1349–1365, 2010.
- [88] Y. Xiao *et al.*, “Lowering the barriers to large-scale mobile crowdsensing,” in *Proc. of Hot-Mobile*. ACM, 2013, pp. 9:1–9:6.
- [89] D. Yang *et al.*, “Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing,” in *Proc. of Mobicom*. ACM, 2012, pp. 173–184.
- [90] G. Yang, “Discovering significant places from mobile phones: a mass market solution,” in *Proc. of MELT*. Springer-Verlag, 2009, pp. 34–49.
- [91] J. Yang *et al.*, “Accuracy characterization of cell tower localization,” in *Proc. of Ubicomp*. ACM, 2010, pp. 223–226.
- [92] H. Zang and J. C. Bolot, “Mining call and mobility data to improve paging efficiency in cellular networks,” in *Proc. of MobiCom*. ACM, 2007, pp. 123–134.
- [93] H. Zang, F. Baccelli, and J. Bolot, “Bayesian inference for localization in cellular networks,” in *Proc. of INFOCOM*. IEEE, 2010, pp. 1963–1971.
- [94] H. Zhang and R. Dantu, “Predicting social ties in mobile phone networks,” in *Proc. of ISI*, 2010, pp. 25–30.

List of Candidate's Work Related to the Thesis

Journals (Impact)

- [1] M. Ficek, T. Pop, and L. Kencl, “Active tracking in mobile networks: An in-depth view,” *Computer Networks*, 2013, *in press*. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128613000996> (50%)

Journals (Reviewed)

None.

Patents

None.

Web of Science

- [1] M. Ficek and L. Kencl, “Inter-Call mobility model: A spatio-temporal refinement of call data records using a gaussian mixture model,” in *Proceedings of The 31st*

Annual IEEE International Conference on Computer Communications (IEEE INFO-COM '12). IEEE, 2012, pp. 469–477. (70%)

Others

- [1] M. Ficek, N. Clark, and L. Kencl, “Can crowdsensing beat Dynamic Cell-ID?” in *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones (PhoneSense '12)*. ACM, 2012, pp. 10:1–10:5. (50%)
- [2] M. Ficek, “CRAWDAD data set ctu/personal (v. 2012-01-11),” 2012. [Online]. Available: <http://crawdad.org/ctu/personal> (100%)

Cumulative download statistics (for the period from 2012-03-24 to 2013-06-04):

- Number of users who downloaded this dataset: 117

- [3] M. Ficek and L. Kencl, “Spatial extension of the Reality Mining Dataset,” in *IEEE 7th International Conference on Mobile Adhoc and Sensor Systems (MASS '10)*, 2010, pp. 666–673. (70%)

Best Paper Presentation Award (San Francisco, CA, USA)

The paper has been cited in:

- S. Nath, “Ace: Exploiting correlation for energy-efficient and continuous context sensing,” in *Proc. of MobiSys*. ACM, 2012, pp. 29–42.
- L. Gao *et al.*, “Multi-dimensional routing protocol in human associated delay-tolerant networks,” *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2012.
- K. Yadav *et al.*, “Poster: cloud-enabled content search and sharing system for mobile phones (MobiShare),” in *Proc. of MobiSys*. ACM, 2012, pp. 521–522.
- S. A. Shad and E. Chen, “Spatial outlier detection from GSM mobility data,” *International Journal of Advanced Research in Computer Science*, vol. 3, no. 3, pp. 68–74, 2012.
- S. A. Shad and E. Chen, “Precise location acquisition of mobility data using Cell-id,” *International Journal of Computer Science Issues*, vol. 9, no. 3, pp. 222–231, 2012.

- [4] M. Ficek, T. Pop, P. Vláčil, K. Dufková, L. Kencl, and M. Tomek, “Performance study of active tracking in a cellular network using a modular signaling platform,” in *Proceedings of the 8th international conference on Mobile systems, applications, and services (MobiSys '10)*. ACM, 2010, pp. 239–254. (20%)

The paper has been cited in:

- Q. Xu *et al.*, “Identifying diverse usage behaviors of smartphone apps,” in *Proc. of IMC*. ACM, 2011, pp. 329–344.
- C. Neuberg, “A mobile world of security,” Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2011.

- [5] M. Ficek and L. Kencl, “Improving roamer retention by exposing weak locations in GSM networks,” in *Proceedings of the 5th international student workshop on Emerging networking experiments and technologies (Co-Next Student Workshop '09)*. ACM, 2009, pp. 17–18. (70%)

- [6] M. Ficek, K. Dufková, P. Vláčil, T. Pop, J. Danihelka, J. Kouba, and L. Kencl, “SS7Tracker,” Funkční vzorek, Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Telecommunication Engineering 2009. (14%)
- [7] K. Dufková, M. Ficek, L. Kencl, J. Novak, J. Kouba, I. Gregor, and J. Danihelka, “Active GSM cell-id tracking: ”Where did you disappear?“,” in *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments (MELT '08)*. ACM, 2008, pp. 7–12. (20%)

The paper has been cited in:

- G. Yang, “Discovering significant places from mobile phones: a mass market solution,” in *Proc. of MELT*. Springer-Verlag, 2009, pp. 34–49.
 - K. Laasonen, “Mining cell transition data,” Ph.D. dissertation, University of Helsinki, Finland, 2009.
 - T. Mantoro *et al.*, “HajjLocator: A hajj pilgrimage tracking framework in crowded ubiquitous environment,” in *Proc. of ICMCS*. IEEE, 2011, pp. 1–6.
 - S. Weixiong *et al.*, “Collecting and analyzing mobility data from mobile network,” in *Proc. of IC-BNMT*. IEEE, 2009, pp. 810–815.
 - M. Kwan, “Visualization and analysis of mobile phone location data,” Ph.D. dissertation, Mathematical and Geospatial Sciences, RMIT University, 2012.
 - B. Srivastava *et al.*, “Towards a sustainable services ecosystem for traffic management,” in *Proc. of SRII*. IEEE, 2011, pp. 392–400.
- [8] K. Dufková, J. Danihelka, M. Ficek, I. Gregor, and J. Kouba, “Can active tracking of inroamer location optimise a live GSM network?” in *Proceedings of the 2007 ACM CoNEXT conference (CoNEXT '07)*, no. 42. ACM, 2007, pp. 1–2. (20%)

List of Candidate's Work Non-related to the Thesis

Journals (Impact)

None.

Journals (Reviewed)

None.

Patents

None.

Web of Science

None.

Others

- [1] E. B. Martinez, M. Ficek, and L. Kencl, “Mobility data anonymization by obfuscating the cellular network topology graph,” in *Proceedings of the IEEE International Conference on Communications (ICC) '13*, 2013. (33%)

Summary

In the thesis we investigate tracking methods in mobile networks and their principal limits. We study various methods of tracking (i.e., periodical positioning) of a number of mobile network users. We consider two basic options of tracking methods in mobile networks — terminal-based and network-based. Terminal-based techniques require user’s cooperation and special hardware or software on the side of the localized mobile terminal. Network-based tracking, generally reaching all subscribers, is implemented in the network in either active or passive manner. Active tracking is based on queries of the network about the tracked device, whilst passive tracking uses operating data, which are generated and stored automatically by the network for all users.

Our original contribution to the area of *network-based active tracking* is a detailed study of a particular method of active tracking, the SMS-based one, using our proof-of-concept tracking platform connected to a live mobile network. Based on a large-scale measurement, we build a model of tracking process, and simulate how many users a tracking platform is able to track simultaneously without overloading the mobile network. Finding out that the principal limitations lie in the radio access network, we express the scalability of the method at various network-infrastructure levels and point out some pitfalls of active tracking, such as user mobility.

To the area of *network-based passive tracking* we significantly contribute by deriving a novel, probabilistic Inter-Call Mobility model, which overcomes the main limitation of passive tracking data — the poor temporal granularity of Call Data Records (CDRs). Our Inter-Call Mobility model spatio-temporally fits the aggregated mobility behavior of a large user-pool and significantly improves the CDR-based deduction of user’s presence at some place in time: from the timestamped cell coordinates of mobile phone communication records, to a probabilistic distribution of user’s position in between consecutive communication records, over time. On the example of user-proximity probability we demonstrate large disproportions in expected user’s position in space and time among different mobility models, concluding that the Inter-Call Mobility model outperforms existing modeling techniques.

Finally, we investigate the limits of cooperative *terminal-based tracking* (crowdsensing) in discovering the mapping of mobile-network cell identifiers to geographic locations. Based on a real-world trace, we propose a novel data-driven mobility model to express the number of unique mobile-network cells a user is capable of visiting during one day. The model describes users’ daily patterns, captures the fine-grained temporal characteristics of human movement during a day, and quantifies daily user-cell associations. Synthetic traces of user mobility from the model serve as an input for a large-scale simulation in an approximation of a mobile network. We show how crowdsensing may serve as a fight-back solution against a particular mobile-network-topology obfuscation method.

These three topics, studied in the thesis, illustrate the extent to which a particular technology or tracking method is applicable. Individually, they present relevant information about exploring and modeling human mobility based on mobile-network data, and the proposed mobility models can be used in future research.

Anotace

V dizertační práci zkoumáme metody sledování uživatelů v mobilních sítích a limity těchto metod. Pojmem „sledování“ rozumíme opakované zjišťování polohy účastníka mobilní sítě. Podle místa, kde probíhá zjištění a vyhodnocení pozice uživatele, rozlišujeme dvě základní metody sledování: terminálové a síťové. Terminálové metody využívají mobilního telefonu (terminálu) uživatele, k čemuž je obvykle zapotřebí spolupráce uživatele, nebo speciální vybavení sledovaného zařízení. Síťové metody jsou realizovány uvnitř mobilní sítě, využívají jejich specifických možností a umožňují zjištění polohy obecně libovolného uživatele sítě. Způsob sledování pomocí síťových metod je dvojitý: aktivní a pasivní. Během aktivního sledování se síť aktivně dotazuje na polohu účastníka, pasivní sledování využívá provozních dat, která jsou v mobilní síti automaticky vytvářena a ukládána.

Náš původní přínos do oblasti aktivního síťového sledování spočívá v detailním zmapování konkrétní metody založené na posílání textových zpráv (SMS), kterou zkoumáme na existující implementaci připojené do mobilní sítě. Průběh zjištění polohy účastníka modelujeme pomocí diskrétní simulace s parametry získanými na základě rozsáhlého měření. Simulace slouží ke zjištění kritického množství uživatelů, které by bylo možné sledovat v síti najednou, bez negativních důsledků pro mobilní síť. Podrobně se věnujeme stěžejním limitům aktivního sledování, které pramení z omezené kapacity rádiového rozhraní přístupové sítě, a popisujeme, do jaké míry je nasazení aktivního sledování použitelné v aplikacích náročných na velký počet uživatelů.

K výzkumu v oblasti pasivního síťového sledování významně přispíváme vytvořením pravděpodobnostního Inter-Call Mobility (ICM) modelu, jenž popisuje pohyb uživatele mobilní sítě mezi místy, kde komunikoval. Tento model překonává základní nedostatek provozních dat mobilních sítí — záznamů volání (Call Data Records) — jejichž velké časové rozestupy znemožňují přesné určení polohy uživatele kdykoli během dne. ICM model je založen na agregovaných datech pohybu mnoha uživatelů; rozšiřuje možnosti použití záznamů volání pro zjištění polohy uživatele mobilní sítě: od zaznamenaných poloh ve chvíli komunikace až po pravděpodobnostní rozložení předpokládané polohy mezi dvěma po sobě následujícími záznamy. Použití a výhody modelu ilustrujeme na příkladu odhadu pravděpodobnosti setkání uživatelů a ukazujeme, že ICM model překonává současné způsoby modelování pohybu uživatelů na základě záznamů volání.

V poslední části dizertační práce zkoumáme limity sledování účastníků, konkrétně použití kooperativního sběru dat z mobilních telefonů, tzv. crowdsensing. Crowdsensing je často používán ke zmapování topologie mobilní sítě, tj. přiřazení geografických souřadnic jednotlivým buňkám sítě, a tím pádem k vytvoření databáze buněk pro lokalizační služby. Na základě rozsáhlého záznamu pohybu uživatelů v mobilní síti odvozujeme model, který umožňuje odhadnout počet buněk, které uživatel navštíví během jednoho dne. Tento model popisuje vzorce chování uživatele z pohledu jeho přesunů mezi buňkami, zachycuje tendenci k přesunům v průběhu dne, a vyčísluje očekávaný počet buněk, ke kterým je terminál uživatele během dne připojen. Model generuje charakteristické chování uživatele během dne a tento výstup simulujeme v uměle vytvořené topologii mobilní sítě. Výsledkem je odhad za jak dlouho může určitý počet uživatelů navštívit předem stanovené množství buněk v síti. Závěrem ukazujeme úspěšnost crowdsensingu v případě, že je v mobilní síti použita konkrétní metoda zabraňující mapování poloh buněk.

Výše uvedená tři témata ukazují, pro které aplikace a do jaké míry je konkrétní technologie sledování uživatele v mobilní síti vhodná. Jednotlivě prezentují řešené problémy významný přínos do oblasti zkoumání a modelování lidské mobility na základě dat z mobilní sítě. Představené modely mobility mohou být použity v další výzkumné práci.