A Novel Sensor-Free Location Sampling Mechanism

Panagiota Katsikouli, Diego Madariaga, Marco Fiore, Aline Carneiro Viana, Alberto Tarable

A Novel Sensor-Free Location Sampling Mechanism

Panagiota Katsikouli, Diego Madariaga, Marco Fiore, Aline Carneiro Viana, Alberto Tarable

Abstract—In recent years, mobile device tracking technologies based on various positioning systems have made location data collection an ubiquitous practice. Applications running on smartphones record location samples at different frequencies for varied purposes.The frequency at which location samples are recorded is usually pre-defined and fixed but can differ across applications; this naturally results in big location datasets of various resolutions. What is more, continuous recording of locations results usually in redundant information, as humans tend to spend significant amount of their time either static or in routine trips, and drains the battery of the recording device. In this paper, we aim at answering the question *"at what frequency should one sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*. Our analyses on fine-grained GPS trajectories from users around the world unveil *(i)* seemingly universal spectral properties of human mobility, and *(ii)* a linear scaling law of the localization error with respect to the sampling interval. Building on these results, we challenge the idea of a fixed sampling frequency and present a lightweight, energy efficient, mobility aware adaptive location sampling mechanism. Our mechanism can serve as a standalone application for adaptive location sampling, or as complimentary tool alongside auxiliary sensors (such as accelerometer and gyroscope). In this work, we implemented our mechanism as an application for mobile devices and tested it on mobile users worldwide. The results from our preliminary experiments show that our method adjusts the sampling frequency to the mobility habits of the tracked users, it reliably tracks a mobile user incurring acceptable approximation errors and significantly reduces the energy consumption of the mobile device.

Index Terms—Human Mobility, Location Sampling, Enery Efficiency

1 Introduction

Over the past few years, the pervasive usage of smart devices and location-tracking systems has made it possible to study and understand human mobility at unprecedented scales. An important feature that was found to characterize human mobility is regularity; we tend to follow the same patterns over and over, and we do so in ways that are clearly periodic [5]. Regularity is easily found in the movements of most individuals: as an example, consider Fig. 1, which shows heatmaps of the locations visited by three random users in the dataset presented in Subsec. 3.1. Although these plots convey three weeks of data, a small set of frequently visited places emerges for all users, along with systematic paths connecting them. Likewise, Fig. 2 illustrate the temporal dimension of regularity for the same users: a clear periodicity emerges from the time series of the visited locations.

In this paper we investigate whether the regularity of human mobility entails the possibility of sampling individual movements at reduced frequencies, while allowing for the reconstruction of trajectories that retain a vast portion –if not all– of their original level of detail. Intuitively, periodic visits to a limited set of important places through repeated routes may be captured with a smaller sampling effort than, e.g., a completely random mobility. Identifying suitably reduced frequencies for human mobility sampling would have applications in a number of fields, including but not limited to:

- *(i)* mobile computing, where overly frequent GPS localization unnecessarily reduces the battery life of mobile devices;
- *(ii)* location-based service design, where unwarranted users' position data collection raises significant privacy concerns;
- *(iii)* cellular networks, where active probing of subscribers' positions is a costly task whose rate must be duly optimized;
- *(iv)* trajectory data compression, where information loss must be minimized

Overall, our problem is equivalent to posing the question *"at what frequency should one sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*. To respond, we initially investigate the effect that reduced constant sampling frequencies have on the quality of the tracked movement. We start our quest by adopting a signal processing approach, by considering mobility patterns as signals over time, and carrying out a spectral analysis of human mobility. We find that the spectra of the movements of 119 individuals have very similar, flat shapes that suggest

- *P. Katsikouli is with the DIKU, University of Copenhagen, 2100 Copenhagen, Denmark. E-mail: pk@di.ku.dk.*
- *D. Madariaga is with NIC Chile Research Labs, University of Chile, 8320000 Santiago, Chile. E-mail: diego@niclabs.cl.*
- *M. Fiore is with IMDEA Networks Institute, 28911 LeganÃl's, Spain, E-mail: marco.fiore@imdea.org.*
- *A. C. Viana is with Inria, 91120 Palaiseau, France. E-mail: aline.viana@inria.fr.*
- *A. Tarable is the CNR-IEIIT, 10129 Torino, Italy. E-mail: alberto.tarable@ieiit.cnr.it.*

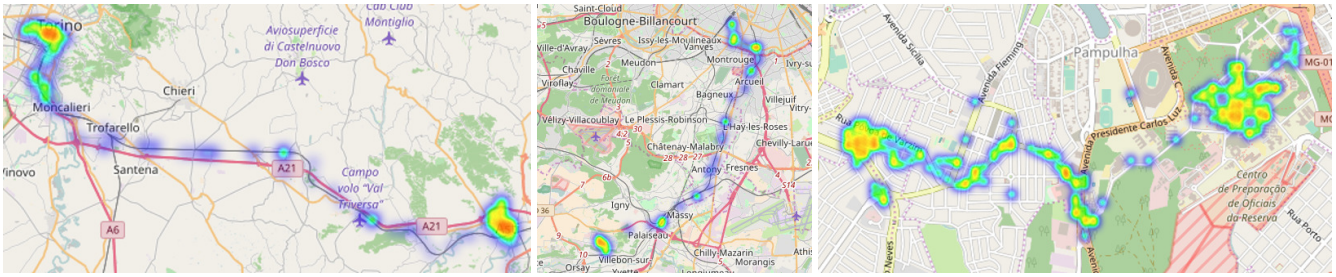


Fig. 1: Heatmaps of locations visited by three distinct users during three weeks: humans tend to commute between a limited set of specific locations. Figures best seen in color.

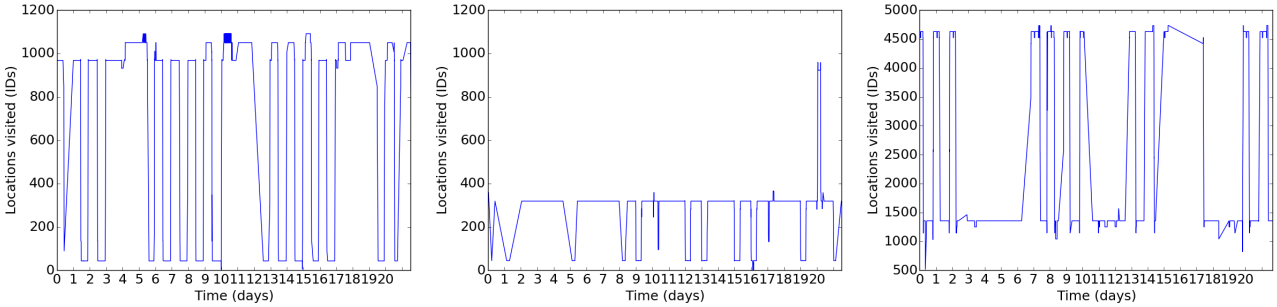


Fig. 2: Location time series for three distinct users during three weeks: humans tend to revisit locations in a periodic fashion. The visited locations are mapped to sequential identifiers, upon discretization on a regular grid of 50 meters step.

the absence of convenient sampling frequency thresholds – even specific to single users – beyond which the error in the reconstructed trajectories drops significantly.

Stimulated by this finding, we carry out a quantitative analysis of the user localization error in movements reconstructed from regular sampling at different periodicities. Our results unveil a linear scaling law of the error with respect to the span of the constant sampling interval. This law corroborates the outcome of the spectral analysis and has significant practical implications, as it controls the trade-off between accuracy and cost of measurements of human mobility.

Building on these results, we set forth to challenge the idea of constant sampling frequency. Our goal is to design a system that adjusts the sampling frequency according to the user's mobility habits. Such a system may act independently or while reinforcing auxiliary sensors or/and alternative positioning systems. Taking advantage of the linear scaling law unveiled in our analysis, we design a lightweight mobility adaptive location sampling mechanism. Our mechanism has been implemented as an app for mobile devices and tested on mobile users worldwide. We find that such a system reliably adjusts the sampling frequency according to the mobility of the user and significantly reduces the energy consumption of the tracking device.

The paper is organized as follows. We discuss related works and ideas in Section 2. In Section 3 we present a reference dataset as well as our spectral and qualitative studies on reduced constant sampling frequencies where the reference dataset is employed. We note here that most of this part of the work corresponds to an earlier work on the sampling frequency of human mobility [3]. Our adaptive sampling mechanism is presented in Section 4 while our implemented app for mobile devices and the preliminary performance results are taken in Section 5. Finally, the significance of the work as well as conclusive remarks are discussed in Section 6.

2 Related Background

Spatial data trajectory compression is a widely addressed research subject, where the objective is to maintain the trajectory shape (see [19], [20] and references therein). Consider the toy example in Fig. 3, where a user leaves home, trains at the gym before work and, later goes to a take-away restaurant. The shape of the spatial trajectory could be well approximated as the sequence of home, junctions B and C, and take-away locations: map-matching based on these cardinal points would allow the description of the whole movement. However, individuals visit places for a purpose, carrying out activities that have different durations. In our example, the size of the circle around each location is proportional to the amount of time spent there. In this work, our purpose is to recreate the complete original mobility of the user, including these temporal features.

The problem we address here is also different from sampling to detect important locations [21], [22], or from simplifying GPS trajectories to preserve semantics of locations [23], [24]. In the example of Fig. 3, important location detection is solved by sampling the trajectory so as to model the original distribution of time spent at home, work, gym, and take-away. However, approaches for the detection of such frequently visited places ignore the time ordering of visits, and do not capture transitions between frequent locations. Instead, our holistic perspective accounts for all these characteristics.

Finally, we are not interested in maintaining locations that would impose a great change in the original direction of the trajectory, if absent [25]; nor we address the similar

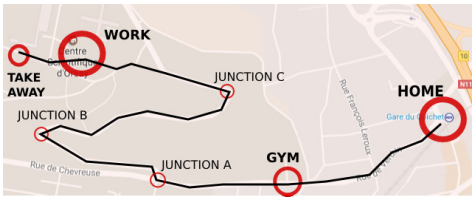


Fig. 3: Toy example of a mobility trajectory. Labels denote important locations or turning points. The size of the circle around each location is proportional to the amount of time spent there by the user.

problem of calculating the current position of a target based on its travelled distance and direction of movement, known as dead reckoning [26], [27]. Indeed, we are not interested in simplifying a pre-recorded GPS trajectory but to find convenient sampling frequencies for human trajectory data.

In terms of dynamically changing the GPS sampling frequency, literature hosts a small collection of works that mainly utilize auxiliary sensors or alternative positioning systems to trigger a GPS recording. Intelligence from the accelerometer of mobile devices has been the main auxiliary system used in adaptive location sensing mechanisms. The authors in [14] look at the problem of minimizing the probability of exceeding a given positional error while maximizing the energy saving when tracking users indoors. In this setting, the main localization system is deployed statically so that it monitors any targets moving in space. Apart from the acceleration information, decisions about the sampling frequency are made also based on the velocity of the moving targets. In [18], the goal of the proposed adaptive location-sensing system is to improve energy efficiency of smartphones. The system builds on four principles, namely substitution, suppression, piggybacking and adaptation. Substitution uses less energy-hungry location-sensing mechanisms than GPS, suppression uses accelerometer during static hours, piggybacking exploits the knowledge of the location from other applications of the phone that have requested a location at the same time and finally, adaptation aggressively adjusts time and distance when the battery is depleted. Experiments of the system show that by reducing GPS records by 98%, battery life can be improved by up to 75%. An adaptive sampling mechanism is presented in [16] as part of a platform based on smart phones that captures the behaviour of users in offices and provides them with information of their own sociability and that of their colleagues. For the adaptive sampling system, the idea is that the sensors need to be mostly sampled during *interesting* events. The sensors sampled in this scenario, apart from the accelerometer, are the bluetooth and the microphone. The mechanism involves an initial learning period, where dense sampling of the sensors is performed, along with event tags in order to later classify the events as stationary or mobile.

The observation that GPS is less accurate in urban areas motivates the work in [17]. The recommended adaptive location sensing system activates the GPS only whenever it is necessary to achieve the *smaller* accuracy tolerated in urban places. Similar to us, the authors use the history of velocities for the activation of the GPS, however, the velocities they consider are the ones that appeared at the same locations and at the same time of the day as at the moment of localization. The authors also present more complex systems that are assisted by sensors, such as the accelerometer and the bluetooth. In [13], the goal is to dynamically use GPS so that the used energy is reduced. The region of interest is divided into geographical zones (which could correspond to GSM coverage cells) and GPS is activated whenever a change of zone is detected.

The effect of human involvement in improving the sensing coverage and data transmission in mobile crowd sensing is investigated in [15]. The paper tackles opportunistic sensing where the goal is to spatiotemporally sample a region, using the presence of humans. Mostly related to our work is an adaptive sampling mechanism for the population of participants. Decisions are made about which user to sample and when, based on the number of times each area in the desired region has been covered during a coverage period. The goal of this work is very different from ours, since here the authors are not interested in accurately recording the mobility of one user but rather using mobile users to spatiotemporally cover the mobility in a region.

In all the works discussed above, the proposed sampling systems are designed with energy efficiency in mind and are not concerned with the accurate reconstruction of an individual's mobility, which is our main motivation. Techniques based on compressive (or compressed) sensing are limited in the field of human mobility, mainly because such techniques offer simple data acquisition at heavier computational cost to reconstruct the sampled data at the processing unit [33]. Preliminary approaches for vehicular networks can be found in [34].

To the best of our knowledge, this is the first work to thoroughly study the problem of investigating what is a good sampling frequency at which to sample human mobility so that users' complete movements can be accurately reconstructed, as well as, the first work to present an energy efficient and lightweight adaptive sampling system that reliably adjusts the sampling frequency according to the tracked user's mobility habits.

3 Spectral and Quantitative Analyses of Mobility Sampling

In this section, we aim at answering the question "*at what frequency should one periodically sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?*". First, we approach the question from a spectral analysis viewpoint, considering human movements as a signal in time and studying its spectrum in frequency (Subsec. 3.2). Next, we conduct extensive quantitative investigation of the exact trade-off between the sampling frequency and the quality of the recorded movement (Subsec. 3.3). Before we discuss our approaches, we present the dataset used in the analyses of this section.

3.1 Reference dataset

Here, we employ a dataset of real-world individual mobility data extracted from three different sources.

- MACACO data was collected between July 2014 and December 2016 as part of the European collaborative project MACACO, funded by the EU CHIST-

Dataset	Users	Weeks
MACACO	19	164
OpenStreetMap	4	7
GeoLife	96	881

TABLE 1: Per-source users and weeks in the reference dataset.

ERA program[1]. A dedicated application, running on smartphones of volunteers in several countries in Europe and South America, recorded GPS positioning information at regular time intervals, typically from one to five minutes. Due to data privacy regulations in France, where the GPS logs are hosted, the data is not publicly available.

- OpenStreetMap (OSM) data was collected by volunteers who recorded and uploaded their trajectories as a contribution to the OSM database[2]. The OSM project is a global crowd-sourcing initiative, aiming at mapping the whole world surface thanks to the activity of a vast user community. GPS traces uploaded by participants typically feature 1-Hz frequency. They are freely available on the official OSM project website.
- Geolife data was collected in Beijing by Microsoft Research Asia from April 2007 to August 2012 [4]. It consists of GPS trajectories recorded through different GPS loggers and smartphone apps. Although sample rates vary significantly across users and time periods, the vast majority of GeoLife positioning data is recorded at intervals from one to five seconds. GeoLife traces are publicly available on the official project website.

The GPS trajectories in our dataset cover sensibly different geographical and temporal spans, even for data coming from the same source. Depending on the user, movements can encompass a single city or multiple continents, over time intervals ranging from days to years. Moreover, the quality of the data for a single user is typically very heterogeneous over time, with periods of days or weeks where GPS logs are erroneous or completely absent. In order to build a consistent reference dataset, we segmented the mobility traces of all users into one-week trajectories[3], and analysed them separately. During each week, we bounded the mobility of each individual to the regions where the activity is concentrated. One-week trajectories and bounded regions avoid biases introduced by singularities, such as international journeys performed once in months. Clearly, bounded regions can create temporal gaps in the weekly traces, whenever users visit areas outside them; however, this effect is marginal, as we found that users spend 85% to 100% of their time within their weekly bounding region. Depending on the user, bounding regions span from 400 to $3,000\ km^2$.

We then filtered the one-week geographically-bounded trajectories based on their quality and retained only the trajectories that contain complete GPS records in at least six out of seven distinct week days. Ultimately, our reference dataset is composed of 1,052 weeks of mobility of 119 different individuals. Table 1 provides a break down of these numbers on a per-source basis. A legitimate question is whether the data is dominated by a few users, i.e., if the majority of weeks refers in fact to a limited set of users, which could bias the analysis. Fig. 4(a)-(c), which portray the distribution of the number of weekly trajectories, show that this is not the case. Indeed, the vast majority of users contribute one to ten weeks of movement data, and the few users who exceed that range provide around 50 weeks of mobility at most. Overall, our reference dataset encompasses a quite diverse base of individuals.

The different techniques employed to collect the GPS positioning information lead to uneven recording intervals across, and even within, the original data sources. In addition to this, weekly trajectories have temporal gaps due to offline GPS receivers, interruptions in the data collection service, or users travelling outside the bounding regions we introduce. Fig. 4(d) shows the cumulative distribution function (CDF) of the sampling intervals observed in all one-week trajectories of our reference dataset. We remark that in almost all cases such intervals are shorter than 15 minutes. More precisely, in trajectories from the lower-granularity MACACO data, 40% and 65% of GPS points are separated by less than 1 and 5 minutes, respectively. In trajectories from OSM and GeoLife data, 90% to 95% of points are less than 10 seconds apart, as highlighted in Fig. 4(e). We argue that the sampling intervals in the one-week trajectories of our reference dataset are sufficient to capture human movements, and we consider them as our ground-truth in the remainder of the study.

3.2 Spectral analysis of human mobility

First, we need to transform individual GPS trajectories into unidimensional time series. Even when ignoring altitude information, points in geographical trajectories are obviously bidimensional. We carried out an extensive evaluation of approaches to reduce bidimensional movements to unidimensional signals, using approximated measures such as velocity or relative displacement from the centre of mass, and transformations such as enumeration of discretized locations in the Hilbert space. However, all of the techniques we tested introduced an exceeding amount of noise in the process, disrupting individual movements or introducing unrealistic jumps in the mobility of users.

We opted for a parallel study of the two dimensions of the geographical space, by considering them in isolation. Instead of using the absolute values of latitude and longitude as unidimensional time series, we replace them with the signed latitude and longitude displacements from the corresponding centre of mass of the one-week trajectory. Formally, the displacements of the i-th point in a trajectory are denoted as $\widetilde{lat}_i$ and $\widetilde{lon}_i$, respectively, and computed as

$$\widetilde{lat}_i = lat_i - \frac{1}{n}\sum_{j=0}^{n} lat_j \ \text{ and } \ \widetilde{lon}_i = lon_i - \frac{1}{n}\sum_{j=0}^{n} lon_j \quad (1)$$

where lat_j, lon_j are the latitude and longitude coordinates of the j-th GPS point, and n is the number of points in the trajectory. Other than making time series more easily

1. https://macaco.inria.fr/macacoapp/
2. https://www.openstreetmap.org
3. The rationale behind our choice is that many human activities have been shown to have a weekly periodicity [28], [29]. Using one-week GPS logs lets us capture both repetitiveness and regularity of human mobility.

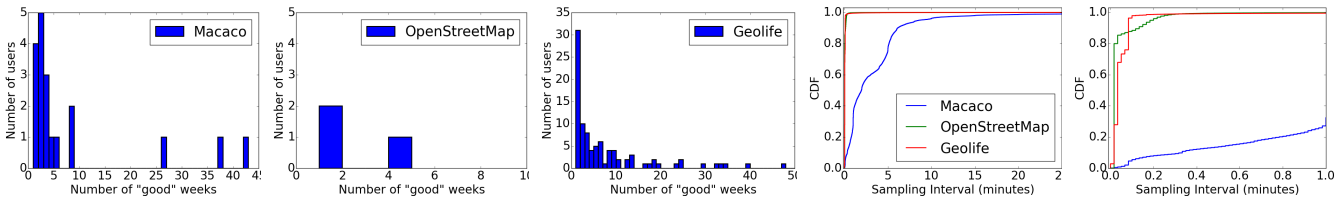


Fig. 4: (a)-(c) User distribution of one-week trajectories in our reference dataset. (d)-(e) Distribution of the sampling interval in one-week trajectories in our reference dataset: full interval span and sub-minute intervals. Figure best seen in color.

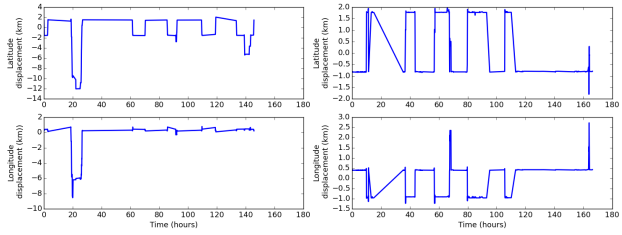


Fig. 5: Unidimensional movement signals of two representative one-week trajectories (left and right). Plots refer to latitude (top) and longitude (bottom) displacements.

Latitude	Longitude
MACACO	
0.55	0.84
OpenStreetMap	
0.53	0.7
GeoLife	
0.57	0.62

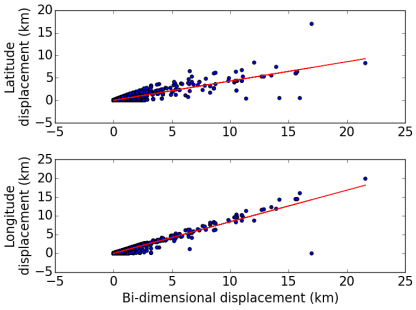


Fig. 6: Correlation between unidimensional displacements and original travelled distance in the bidimensional space. (left) Coefficients separated by data source. (right) Scatterplot of latitude (top) and longitude (bottom) displacemnets with respect to the actual bidimensional movement, for trajectories in the MACACO data.

comparable and readable across users and weeks, the transformations in (1) have the property of generating zero-mean signals whose frequency spectra have no DC components. Note that none of the other tested approaches exhibit such properties. Illustrations of our unidimensional description of individual movements are in Fig. 5, for two one-week trajectories.

By considering the transformation above on the two geographical dimensions in isolation we do not introduce errors; yet, we may lose properties that only emerge when the two dimensions are considered jointly. To verify whether such a problem exists, we analysed the correlation between the isolated latitude or longitude displacements and the actual travelled distance in the bidimensional space. Fig. 6 shows the per-source correlation coefficients, as well as the linear fitting on trajectories from the MACACO data. We observe consistently good correlations in all cases, which lets us conclude that both dimensions, when taken separately, still provide decent approximations of the overall mobility. Interestingly, the correlation is always stronger for longitude than for latitude, indicating that participants to all data sources tend to move along an East-West axis rather than along a South-North one; this may be due to the geographical shape of the cities where the data collection took place.

3.2.1 *Frequency spectra of human mobility*

We apply the Fast Fourier Transform (FFT) in order to compute the spectral representation of the finite-length sequences that represent the one-week latitude and longitude displacement signals. Let $x[n]$, $n = 0, \ldots, N-1$ be a sequence (or signal) of length N. Then, its spectral counterpart,

$$X[k] = \sum_{n=0}^{N-1} x[n] \mathrm{e}^{-j2\pi kn/N}, \quad k = 0, \ldots, N-1 \qquad (2)$$

can be computed with $\mathcal{O}(N \log N)$ operations. It is well known that $X[k]$ is useful to highlight periodic components in the sequence $x[n]$. In particular, if the sequence $x[n]$ is obtained by sampling a continuous-time periodic signal $x(t)$ within a single period, then $X[k]$ will be equal to the sequence of coefficients of the Fourier series of $x(t)$.

The frequency spectrum of a signal yields information about the sampling frequency needed to reconstruct the original time series with a small error. For an ideal signal, whose spectrum drops to zero after some frequency threshold f_s (i.e., the bandwidth of the signal), the Nyquist–Shannon sampling theorem guarantees that a sampling rate $2f_s$ is enough to allow a lossless reconstruction of the original signal from its samples. For practical signals, the spectrum is not strictly limited, but it features limited amounts of noise. In those cases, the spectrum is mostly concentrated within a finite support and shows a negligible amount of power beyond the frequency threshold; again, sampling at a rate twice the threshold, allows reconstructing the original signal with minimum error.

Fig. 7 shows the spectra of the latitude (top) and longitude (bottom) displacement signals of a representative selection of one-week trajectories. The first two columns refer to the signals in Fig. 5. The original spectra are in light blue, while a moving-average that better displays the overall trends is in dark blue. Vertical orange lines outline the frequencies that correspond to sampling intervals of 10 minutes (farthest from the central frequency), 1 hour and 12 hours (closest to the central frequency). We make two important remarks: *(i)* despite the diversity of the latitude and longitude displacement time series across the different one-week trajectories, all spectra have very similar shapes; *(ii)* the shapes do not show evidence of a bandwidth threshold beyond which the spectra become clearly negligible,

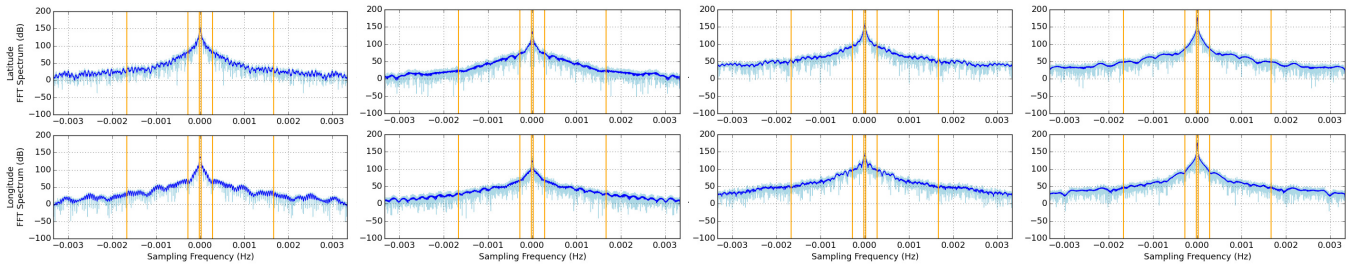


Fig. 7: Frequency spectra of unidimensional movement signals, for a selection of one-week trajectories. Figure best seen in color.

making it impossible to identify an operational point for effective sampling.

Due to space limitations, we cannot show the spectra of all one-week trajectories in our dataset: however, we found the observations above to hold in the overwhelming majority of cases. Considering the heterogeneity of our user base, we hypothesize that such features could be a universal property of human mobility spectra.

We can explain both facts remarked above by considering that the unidimensional movement signals show very steep transitions and deep spikes, see, e.g., Fig. 5: hence, the resulting spectrum shows a slow decay for high frequencies. In other words, although it exhibits some clear periodicity [30], human mobility is in fact a sequence of long periods where individuals are almost static and fast transitions between such important locations. While positions during stationary time intervals contribute to low-frequency spectral components and are hence easily captured by a sparse sampling, travelling causes discontinuities in the mobility signal and is much harder to sample. As a result, considering that sampling at higher frequencies has a cost but leads to a better quality of the reconstructed signal, the spectra do not reveal whether, e.g., collecting samples at every 10 minutes is obviously more efficient than sampling at every hour.

Although disappointing in a sense, this outcome calls forth for an extensive quantitative analysis of the exact trade-off between the quality and cost of sampling in the context of human mobility. We address this aspect next.

3.3 A quantitative analysis of mobility sampling

We perform an experimental analysis, and investigate the impact of different constant sampling frequencies on the quality of the mobility reconstructed from the collected samples. To this end, we create downsampled versions of the one-week trajectories in our reference dataset, using a wide range of sampling intervals, from 10 minutes to 12 hours. We deem longer intervals unreasonable for one-week-long time series. We then create reconstructed versions of the complete trajectories by linearly interpolating the samples, and assess how such reconstructed trajectories compare to the original ones. From the observation that the spectra of human mobility exhibit high similarity among the traces of a particular user, we conduct the analysis on a per user basis. We will show in detail results of a few representative users selected among all considered datasets, however, all users in the analysis exhibit similar trends.

We measure the error in retrieving a complete individual trajectory from sampled data by using the average Haversine distance. Given two points on Earth's surface, $p_a = (lat_a, lon_a)$ and $p_b = (lat_b, lon_b)$, we define $\Delta_{lat} = lat_b - lat_a$, and $\Delta_{lon} = lon_b - lon_a$. The Haversine distance of p_a and b_b is

$$D(p_a, p_b) = R \cdot 2 \cdot atan2\left(\sqrt{\phi}, \sqrt{(1-\phi)}\right), \qquad (3)$$

with $\phi = \sin^2(\Delta_{lat}/2) + \cos(lon_a) \cdot \cos(lat_b) \cdot \sin^2(\Delta_{lon}/2)$, and $R = 6,371km$ is the Earth's radius. The average Haversine distance of a one-week trajectory is the mean of all Haversine distances between the points of the reconstructed and original mobility recorded at the same time instant.

Fig. 8 shows the evolution of the average Haversine error against the sampling interval, for a representative set of eight individuals in our reference dataset. Each plot presents results for all of the one-week trajectories of a specific user: as multiple one-week trajectories are aggregated in every plot, we outline the mean (dots), 25-75% quantiles (dark blue region) and 10-90% quantiles (light blue region) of the error measured over all trajectories of one user.

A surprisingly clear linear relationship characterizes all curves. Fittings on a simple linear model (solid lines in Fig. 8) show an excellent match for all our users. In fact, the linearity of the relationship between the Haversine distance and the sampling interval is a common trait of all individuals in our dataset. Fig. 9(a) portrays the CDF of the Root Mean Square Error (RMSE) between the linear fitting and the mean values (solid lines and dots respectively in Fig. 8) of the average Haversine distance, for all users, over sampling intervals that range from 10 minutes to 12 hours. The probability mass of the distribution is below 250 meters – a very reasonable RMSE for people travelling tens of km per day.

We also highlight that the only parameter of the fitting curve $y = \alpha x$, i.e., the slope α has an important physical meaning: it characterizes the ratio between the average Haversine distance and the sampling interval, or, equivalently, it explains the mean additional error of the reconstructed trajectory when increasing the time that intercurs between samples. Hence, it can be measured in meters per minute (m/min). In other words, our analysis indicates that *adding one minute to the sampling interval used to track one individual leads to an additional positioning error of α meters in her recorded trajectory, irrespective of the absolute span of the sampling interval.*

When looking at the value of α, we remark that it is not identical across users: the plots in Fig. 8 also report the equation of the linear fit, and we can note some diversity there. We study the heterogeneity of α in Fig. 9(b), which portrays the CDF of the distance/interval ratio associated

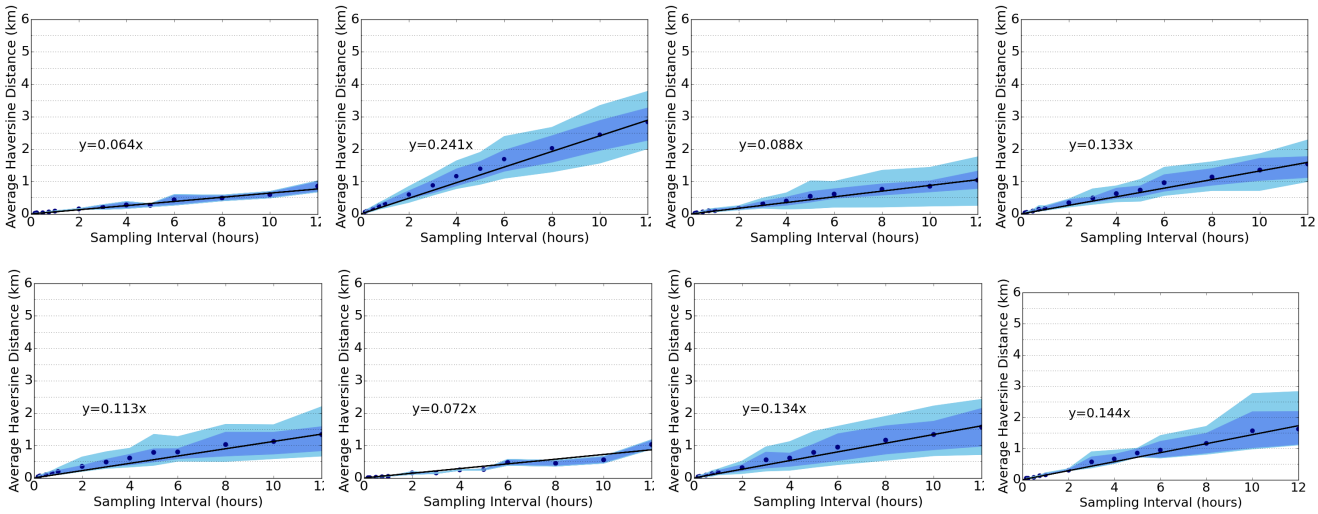
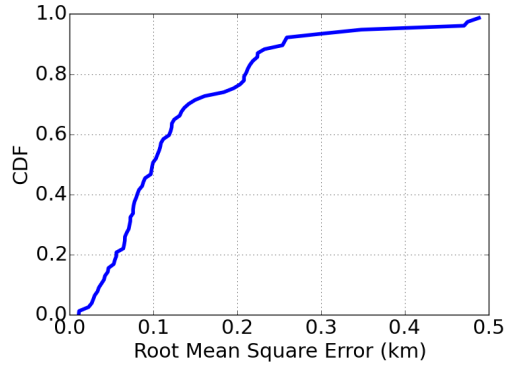
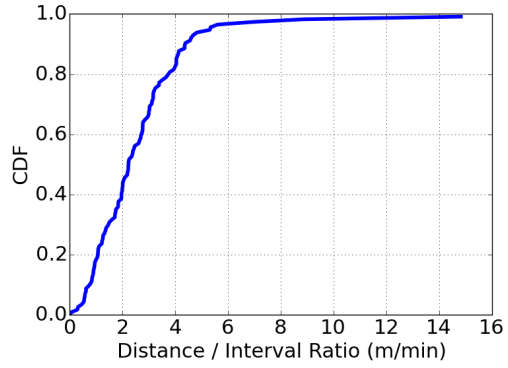


Fig. 8: Average Haversine distance between the original and the reconstructed trajectories of eight users, versus sampling intervals between 10 minutes and 12 hours. Dots represent mean values. Dark and light shaded regions depict the 25-75% and 10-90% quantiles. Solid lines are the linear fittings on average points. Figure best seen in color.



(a) Quality of fitting



(b) Fitting slope across users

Fig. 9: (a) Distribution of the RMSE due to the linear approximation of the relation between the average Haversine distance and the sampling interval. (b) Distribution of the ratio between the average Haversine distance and the sampling interval. Plots are for all users.

Feature	Correlation	$p-$value
Radius of gyration	0.73	3.46e-85
Avg. Speed	0.31	0.028
Std. Speed	0.57	0.0001
Fraction being mobile	-0.047	0.28
Fraction being static	0.047	0.28
Visited locations	0.47	3.39e-29
Top popular location	0.08	0.0485
Second popular location	0.11	0.0076
Travelled distance	0.27	8.45e-10
Displacement	0.22	6.55e-07
Entropy	-0.1	0.022
Area	0.47	3.39e-29
Regularity	0.18	2.36e-05

TABLE 2: Pearson's correlation between each mobility feature and the α parameter

to all individuals in our reference dataset. Over 90% of users have slopes that are uniformly distributed between 1 and 4 m/min. Hence, *for the vast majority of individuals, the inaccuracy of their recorded trajectory grows of 1 to 4 meters for each minute added to their movement sampling interval.*

3.3.1 Relation between α and mobility features

To better understand the reasons behind the linear relationship and the heterogeneity of the α parameter, we computed its correlation with a number of human mobility features, computed on the weekly trajectories. The human mobility features considered are shown in Table 2, along with the Pearson's correlation and the corresponding p-value between each feature and α.

We find the strongest positive correlation to exist between the radius of gyration and the parameter α, suggesting that the furthest a user is travelling in her mobility the largest the average error she imposes for sparser sampling intervals. To validate the correlation results, we take a second approach, using a tree classifier. For this, we used a standard sklearn library that implements a meta estimator that fits a number of randomized decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. We used

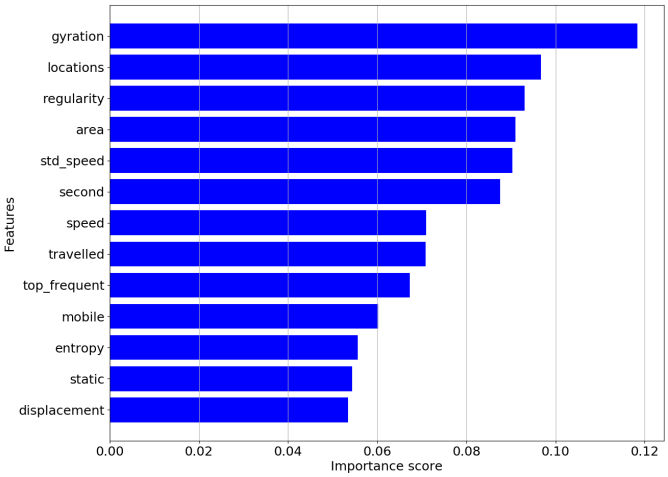


Fig. 10: Feature importance scores from a tree classifier, averaged over 100 independent runs.

the tree classifier on the mobility features, using ranges of the α values as our target variables. From each test we got an importance score for each feature. Fig. 10 shows the features in descending order of their importance scores, averaged over 100 independent runs. We remark that in all tests, radius of gyration is by far the most important feature, according to the tree classifier.

Although the results show some relationship between the α parameter and the radius of gyration, they cannot solidify a strong link between the parameter and any mobility feature. We note here that apart from testing the mobility features individually against the α parameter, we also applied Multiple Linear Regression and Principal Component Analysis in search for combination of features that would exhibit a strong relationship with α; however, with no success. These results suggest that the better choice in learning a user's mobility habits is to apply a training phase of recording their movements. We apply this in our mobility-aware adaptive sampling mechanism, described in the following section.

4 Mobility-aware Adaptive Sampling

One of the key findings in the analyses of the previous section is that dense sampling of human locations can be avoided at small information loss. In this section we challenge the idea of constant sampling intervals that all GPS sampling systems employ. We argue that the way humans move can drive the choice of the sampling interval, so that more samples are acquired when a user is more mobile and less when the user is less mobile. For instance, intuitively we would increase the sampling interval during the night when a user is perhaps at her home location but decrease it while the user is travelling or she commutes between places. In the first case, sparse sampling would sufficiently capture the user's stationary mode at home, whereas in the second, a denser sample is needed to more accurately reconstruct her movement.

Another requirement of ours is to design a complementary system for alternative positioning systems that use auxiliary sensors to trigger the activation of GPS, but could also operate independently, as a standalone application. Our system would be lightweight so that all operations can be performed on small devices with limited computation and storage capabilities, such as a smart-phone.

We are proposing a system that uses the α parameter and the instant speed of the tracked user. We utilize the α parameter in our system because it quantifies the error induced by a sampling interval and it can estimate the added error caused when increasing the sampling interval. As we will see in the details of our system, this is paramount for deciding when to take a next GPS sample so as not to cause high approximation errors.

Our system has two phases. In the first (training) phase, we calculate the α parameter for a specific user and in the second phase we apply the adaptive sampling system while tracking a user.

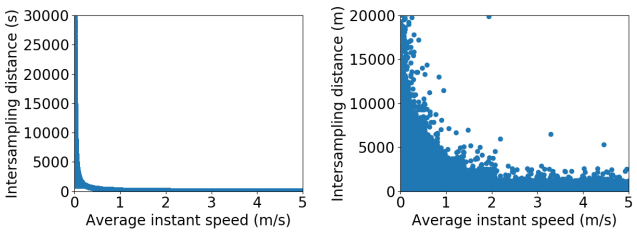
First Phase - α computation: In order to compute the parameter, we require the user to be densely tracked for a period of time T. We do this because we need to capture the ground truth of the user's movement. Next, we downsample the recorded trace for intervals of few minutes upto a few hours (here we consider the intervals used in Subsec. 3.3), and compute the approximation error between the linearly interpolated downsampled trace and the ground truth one. We can then compute the slope of the average approximation error with respect to the sampling interval, which is the α parameter. We note that this step is user-specific, since, as we showed in Fig. 8, different users have different α values.

Second Phase - adaptive sampling system: We make the following assumptions:

- There is a user (or application)-specified error of approximation, er. This represents how much of error a user is willing to allow in their sampled trajectory.
- There is a user (or application)-specified displacement, S, which provides a recommendation for the maximum distance the tracked user could travel unrecorded. That is, we would like to have a location in our sampled trajectory at most every S units of distance, but this is not a hard threshold.
- There is a window of size k for estimating the exponentially weighted moving average of the instant speed of the $k+1$ most recently recorded locations.

We define as $t_{track} = \frac{er}{\alpha}$ the maximum period a user is allowed to be untracked so that her maximum error of approximation is not exceeded. The adaptive sampling system operates as follows. Given a user u and her error-to-interval α_u parameter, do the following, for as long as u is tracked:

- We compute the exponentially weighted moving average of the instant speed, $\hat{v}$, for the $k+1$ most recently recorded locations. The moving average assigns larger weight to the most recent locations.
- We compute $t_i = \frac{S}{\hat{v}}$, which gives us a suggestion of when to take the next location sample, given the most recent mobility history of the user (expressed through $\hat{v}$ and her maximum allowed displacement).
- We take the next sample in $min\{t_i, t_{track}\}$ units of time. We take the smaller of the two in order to guarantee that the recorded trajectory won't cause



(a) (b)

Fig. 11: Intersampling distance ((a) in seconds and (b) in meters) versus average instant speed of the 50 most recently sampled locations.

the user to exceed her maximum allowed error of approximation. We note here that t_{track} is a constant value through the tracking period of the user and can only change if the user chooses to recompute and update her α parameter or if she wishes to update her maximum approximation error tolerance. On the other hand, t_i changes based on the recent mobility of the user. Because this can end up being infinitely large, when the user is static for example, we control the recording of the next sample by reverting to the t_{track} value.

- We compute the instant speed between the most recent and the new recorded location and update the instant speed buffer accordingly.

System performance. We conducted a preliminary evaluation of our system on the dataset presented in Subsec. 3.1, by using the first period T of the traces of a user for computing her α parameter. Afterwards, we applied the adaptive sampling system and calculated the correlation between the achieved intersampling distance (in seconds and in meters) between any two consecutive samples s_i and s_{i+1}, against the exponentially weighted moving average of the instant speed for a buffer of the 50 most recently recorded locations, where the most recent location in the buffer is sample s_i. We repeated the experiment for $T = \{24, 48, 72\}$ hours and show the aggregated results in Fig. 11(a) and 11(b). Our system can reliably track a user's mobility and adjust the sampling frequency based on the user's changing mobility habits, since, as we observe in the figures, the faster a user is moving,the denser the samples are in space and in time, whereas, the samples become sparser when the tracked user becomes static.

5 Energy-Efficient Adaptive Sampling App for Mobile Devices

In order to test the adaptive sampling system in the real world, we implemented it as a self-contained Android library that can be embedded in any Android location sampling app. Thus, we also developed a dedicated Android app that uses this library and a centralized server for collecting the data from the test users.

Adaptive Sampling System Overview The system's app is built around four main components: (i) Sampling App, (ii) Adaptive Sampling System Library, (iii) Android System, and (iv) External Server, as shown in Fig. 12. The Sampling App receives as initial parameters the values er, S, k and α (this latter we hypothesize is previously computed as described in Subsec. 3.3 and provided to the app). The Sampling App requests location updates using the Adaptive Sampling System Library, which internally uses the Google Fused Location Provider to obtain location samples from the Android System. Once the Adaptive Sampling System obtains a new location, it saves both the location and the current instant speed (calculated using the previously collected sample) in an internal database. Next, the Adaptive Sampling System retrieves from the internal database the k most recently computed instant speed values and uses those, along with the given values for er, S, and α, to calculate the time interval for collecting the next sample. The Sampling App receives the location samples from the Adaptive Sampling System and saves them in its internal App database. Inside the Sampling App, a Synchronization Service compresses all collected records and stores the compressed files in the mobile device's internal storage until those are sent to the External Server when Wi-Fi is available.

For evaluating our system, we developed three distinct Android apps that run independently and in parallel in users' devices:

1) **Adaptive Sampling App**: Location sampling app that implements the Adaptive Sampling System as described above.
2) **Sensor Sampling App**: Accelerometer-based location sampling that serves as a baseline adaptive sampling method for our proposed mechanism. This methodology relies on the fact that low-power accelerometer sensors can identify user movements [35], [36], and therefore, this information can be used to avoid unnecessary GPS sampling in stationary locations [37]. The Sensor Sampling Apps collects every 3 minutes a short burst (2 seconds) of measurements from the in-built mobile phone accelerometer. After these 2 seconds, the variance of the burst is calculated. If the variance is greater than a given threshold, then the user is classified as in movement, and a GPS sample is recorded.
3) **Fixed Sampling App**: Location sampling at fixed dense intervals (1-minute) that serves as ground truth for the two previous applications. The locations collected with this app serve to reconstruct the ground truth trajectory of users.

5.1 Collected Data

The three apps we developed collect the following data from the devices of the users[4]:

- Device related information, including a randomly generated identifier, the device's industrial design, Android version, the manufacturer, the model, and the release.
- Location samples, including the latitude, longitude, horizontal accuracy, and timestamp of the location.
- Energy consumption information, including the estimated power use in mAh.
- CPU related information, including the total CPU time in ms.

4. All test users were presented with a consent form detailing the purpose of the apps and the collected information.

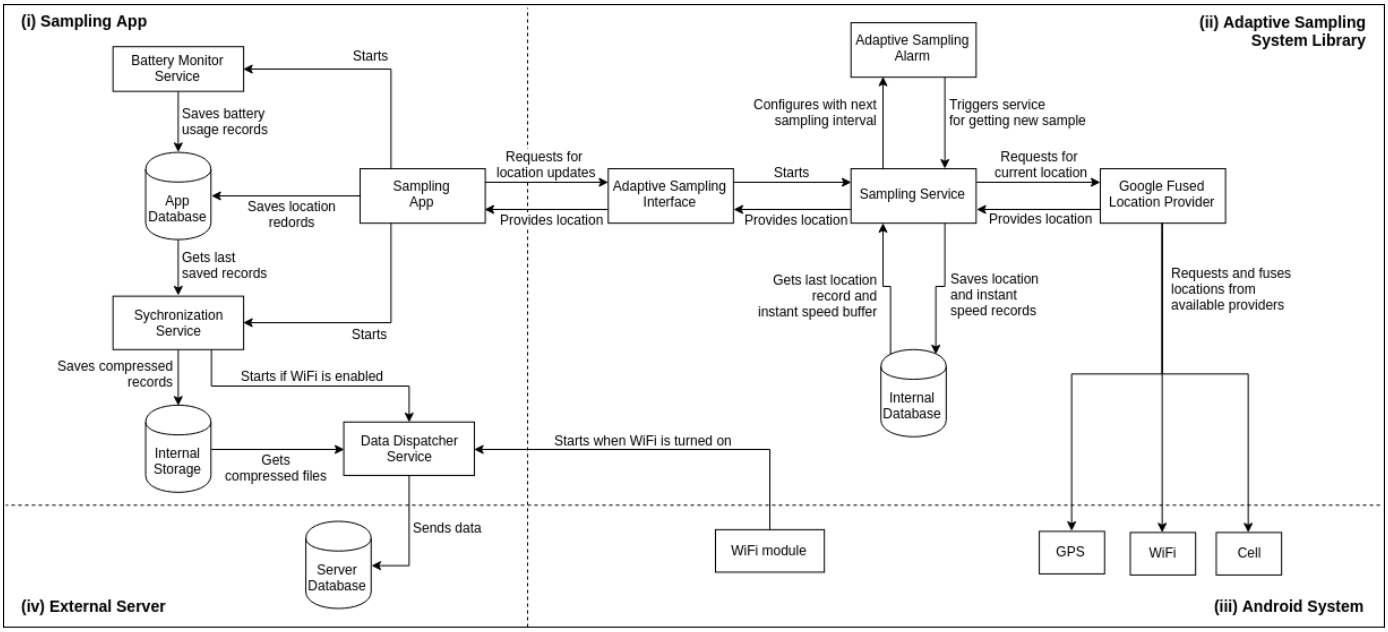


Fig. 12: Adaptive Sampling System Diagram

The collected data allow us to compare the three sampling methods in terms of their energy consumption and CPU usage. Additionally, the collected data allow us to evaluate the error incurred by the trajectories reconstructed from our Adaptive Sampling system and the baseline Accelerometer-based methodology. Our database comprised of 6 test users, collecting data for a total of 7 weeks, in distinct locations around the globe.

5.2 Performance Results on Collected Data

In this section, we present the performance results of our adaptive sampling system on the data we collected from the test users. Our test users were initially asked to install the app that densely samples their locations at a 1-minute frequency (Fixed Sampling App) and use it for two weeks. At the end of this period, the collected traces were used to compute their characteristic α parameter, using the methodology described in Subsec. 3.3. Fig. 13 illustrates the variety of α values in our test users, and therefore, the variety of their mobility patterns.

Next, we asked the users to continue using the Fixed Sampling App and, in parallel, to install the Adaptive Sampling App and the Sensor Sampling App. The users' devices run the three applications in parallel for 5 weeks, where the Adaptive Sampling App used the following configuration:

- The previously computed user-specific α parameter.
- An instant speed buffer size $k = 50$.
- A recommended maximum unrecorded displacement S equal to the allowed approximation error er. These two parameters changed every week to a different value from the following list: $S = er \in \{50, 100, 250, 500, 1000\}$m.

In Fig. 14, we show the average sampling interval for all test users during the 5 weeks running the three sampling apps in parallel. The average sampling interval

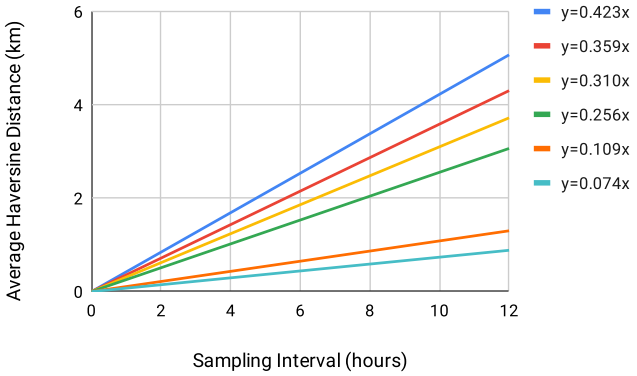


Fig. 13: Calculated α value for each test user.

for the Adaptive Sampling clearly grows from near to 15 minutes (for $er = S = 50$m) to more than 4 hours (for $er = S = 1000$m) as the accuracy requirements are loosened. In the case of the Sensor Sampling, its average sampling interval remains stable between 1 and 2 hours during the 5 weeks.

We used the collected data from the Adaptive Sampling App and the Sensor Sampling App to reconstruct each user's trajectories. These reconstructed trajectories were compared with the ground-truth (constructed from the Fixed Sampling's data) to evaluate the error incurred by both apps. In Fig. 15, we show the approximation error measured using the average Haversine distance for all test users during the last 5 weeks of the experiments. For the Adaptive Sampling App, the approximation error grows almost linearly to the allowed error of approximation, which is coherent with the analysis presented in Section 4. For the Sensor Sampling App, the approximation error varies from

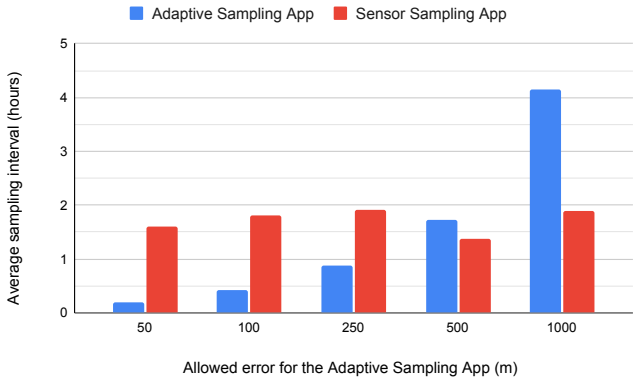


Fig. 14: Average mean sampling interval for all test users using the Adaptive Sampling App and the Sensor Sampling App during the 5 experimental weeks.

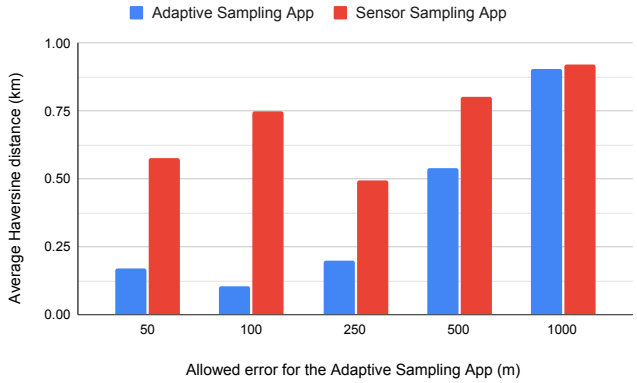


Fig. 15: Average mean haversine error for all test users using the Adaptive Sampling App and the Sensor Sampling App during the 5 experimental weeks.

500m to 900m throughout the 5 weeks. We highlight that, for all possible approximation error values considered, the Adaptive Sampling App yields smaller approximation error compared to the Sensor Sampling app.

5.3 Energy Efficiency

A major reason for reducing the frequency of using the GPS to sample the locations of tracked objects is the energy consumption that dense sampling incurs. Apart from sampling spatio-temporal information, the three apps we created also collect statistics about their battery and CPU usage using the Android dumpsys tool. We show the reduction in energy consumption of the Adaptive Sampling and the Sensor Sampling compared to the Fixed Sampling in Fig. 16. We observe that the Sensor Sampling App reduces the Fixed Sampling App's power consumption by almost 90% during the 5 weeks. In the case of the Adaptive Sampling App, it also generates a significant reduction in battery use of more than 60%, which grows up to 98% as the accuracy requirements are loosened.

Similarly to the above, we show in Fig. 17 the reduction in CPU usage of the Adaptive Sampling and the Sensor

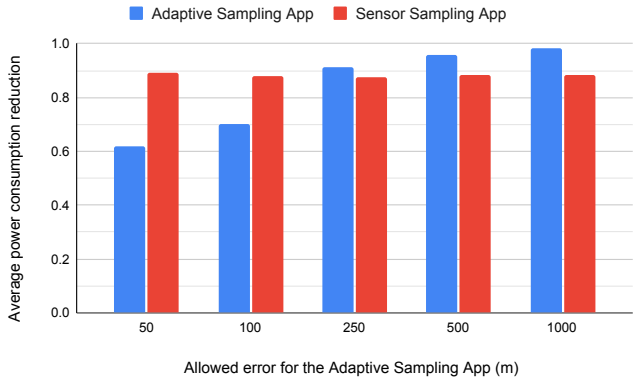


Fig. 16: Average power consumption reduction over the Fixed Sampling App for all test users using the Adaptive Sampling App and the Sensor Sampling App, during the 5 experimental weeks.

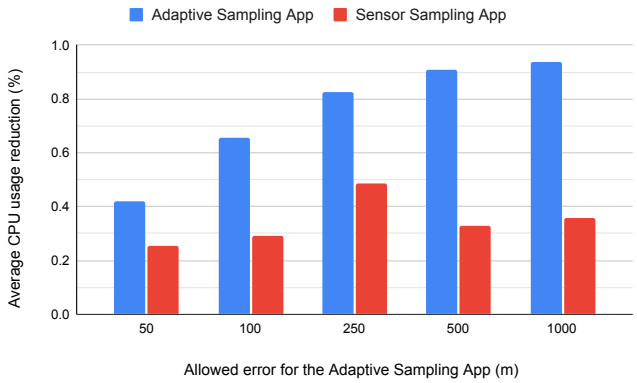


Fig. 17: Average CPU usage reduction over the Fixed Sampling App for all test users using the Adaptive Sampling App and the Sensor Sampling App, during the 5 experimental weeks.

Sampling compared to the Fixed Sampling. We observe that the Adaptive Sampling App generates a significant reduction in CPU usage of more than 40%, which grows up to 94% as the accuracy requirements are loosened. In the case of the Sensor Sampling App, it reduces the CPU usage of the Fixed Sampling App by 25% up to 49%. The Sensor Sampling App generates a lower CPU usage reduction as it triggers a burst of accelerometer measurements every 3 minutes.

6 Discussion and Conclusions

Summarizing our findings, we assert that *the average error incurred by trajectories reconstructed from periodic samples scales linearly with the constant sampling interval*. The linearity of the relationship between error and sampling interval explains the absence of an operational point for the effective sampling of human movements, found by our spectral analysis in Subsec. 3.2. In addition, we find that *the linear scaling law is characterized by a comparable parameter, i.e., the error-to-interval parameter α, across all our user base*. The error-to-interval ration quantifies the error induced by a sampling

interval and it can estimate the added error caused when increasing the sampling interval. Building on this result, we design a mobility-aware adaptive sampling mechanism that is lightweight, it reliably adjusts the sampling frequency to the mobility habits of the tracked user and significantly decreases the battery and CPU usage of the mobile device. The seemingly general linear scaling law and our adaptive sampling system provide a very useful tool and reference for practitioners: we discuss a few especially promising usages.

Energy-efficient mobile computing. It has been shown that frequent sampling of GPS data tends to quickly drain the battery of a mobile device [1], [2]. A natural solution is to sample the device position at a reduced frequency. However, deciding which frequency should be employed is not trivial. The linear scaling law we identified would allow practitioners to control the tradeoff between energy consumption and localization accuracy. What is more, the application of our adaptive sampling system would provably reduce the energy consumption by maintaining high standards on the accuracy of the tracked target.

Location-based service operation. Location-based services (LBS) rely on positioning information about their users. Yet, an excessively frequent collection of user locations is expensive from both energy and communication perspectives, it raises privacy concerns, and it can ultimately bother customers. Our results may help taking informed decisions, reducing the frequency of localization according to the approximation in the user's position that can be tolerated by the service.

Active probing of device position in mobile networks. Precise knowledge of subscribers' locations is a valuable information for mobile operators, for both network management and value-added service development [31]. Actively probing mobile devices for their position in a country-scale mobile network is a computationally expensive task, which has traditionally pushed operators to favor less controllable passive measurements [32]. In fact, subscribers are localized based on their associated antenna sector, and those sectors cover hundreds of m^2 in the best case. In this context, our results suggest that running active probing at, e.g., every hour, would not decrease significantly the measurement accuracy, as the incurred error would be typically comparable to the antenna sector coverage.

Trajectory data compression. A straightforward application of our results is data compression. If large amounts of trajectories must be stored, and memory becomes an issue, one could sample the original movement data at some reduced fixed frequency, or apply mobility adaptive sampling solutions, and store only those samples. This is a lossy operation, yet our results confirm that the information loss is controlled and non prohibitive.

In the light of the results presented in this work, a laconic answer to the original question posed in the Introduction is *"it depends on the error one can afford"*. Fortunately, our findings are more informative than that, and provide both a simple scaling law for the user positioning error, with general validity and limited parameter diversity across individuals, and a system that permits one to adjust the sampling frequency at the mobility habits of the tracked user.

REFERENCES

[1] Y. Qi, C. Yu, Y.-J. Suh, and S. Y. Jang, "Gps tethering for energy conservation," in *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*. IEEE, 2015, pp. 1320–1325.

[2] S. Vanini, F. Faraci, A. Ferrari, and S. Giordano, "Using barometric pressure data to recognize vertical displacement activities on smartphones," *Computer Communications*, vol. 87, pp. 37–48, 2016.

[3] P. Katsikouli, A. C. Viana, M. Fiore, and A. Tarable, "On the sampling frequency of human mobility," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–6.

[4] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 2008, pp. 312–321.

[5] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature*, vol. 453, no. 7196, p. 779, 2008.

[6] T. S. Azevedo, R. L. Bezerra, C. A. V. Campos, and L. F. M. de Moraes, "An analysis of human mobility using real traces," in *2009 IEEE Wireless Communications and Networking Conference*, April 2009, pp. 1–6.

[7] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky, "Human mobility characterization from cellular network data," *Commun. ACM*, vol. 56, no. 1, pp. 74–82, Jan. 2013. [Online]. Available: http://doi.acm.org/10.1145/2398356.2398375

[8] G. Chen, A. Carneiro, and M. Fiore, "Takeaways in large-scale human mobility data mining," in *IEEE International Symposium on Local and Metropolitan Area Networks*, 2018.

[9] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, Apr 2006. [Online]. Available: https://doi.org/10.1007/s10994-006-6226-1

[10] H. Senaratne, M. Mueller, M. Behrisch, F. Lalanne, J. Bustos-Jiménez, J. Schneidewind, D. Keim, and T. Schreck, "Urban mobility analysis with mobile network data: A visual analytics approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1537–1546, 2018.

[11] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, 2013.

[12] K. SiÅĆa-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. DemÅąar, and A. S. Fotheringham, "Analysis of human mobility patterns from gps trajectories and contextual information," *International Journal of Geographical Information Science*, vol. 30, no. 5, pp. 881–906, 2016. [Online]. Available: https://doi.org/10.1080/13658816.2015.1100731

[13] N. Deblauwe and G. Treu, "Hybrid gps and gsm localization âĂŤ energy-efficient detection of spatial triggers," in *2008 5th Workshop on Positioning, Navigation and Communication*, March 2008, pp. 181–189.

[14] C.-W. You, P. Huang, H. hua Chu, Y.-C. Chen, J.-R. Chiang, and S.-Y. Lau, "Impact of sensor-enhanced mobility prediction on the design of energy-efficient localization," *Ad Hoc Networks*, vol. 6, no. 8, pp. 1221 – 1237, 2008, energy Efficient Design in Wireless Ad Hoc and Sensor Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870507001667

[15] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, Aug 2014.

[16] K. K. Rachuri, C. Mascolo, M. Musolesi, and P. J. Rentfrow, "Sociablesense: Exploring the trade-offs of adaptive sampling and computation offloading for social sensing," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '11. New York, NY, USA: ACM, 2011, pp. 73–84. [Online]. Available: http://doi.acm.org/10.1145/2030613.2030623

[17] J. Paek, J. Kim, and R. Govindan, "Energy-efficient rate-adaptive gps-based positioning for smartphones," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '10. New York, NY, USA: ACM, 2010, pp. 299–314. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814463

[18] Z. Zhuang, K.-H. Kim, and J. P. Singh, "Improving energy efficiency of location sensing on smartphones," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '10. New York, NY, USA: ACM, 2010, pp. 315–330. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814464

[19] P. Katsikouli, R. Sarkar, and J. Gao, "Persistence based online signal and trajectory simplification for mobile devices," in *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2014, pp. 371–380.

[20] J. Muckell, J.-H. Hwang, C. T. Lawson, and S. S. Ravi, "Algorithms for compressing gps trajectory data: An empirical evaluation," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2010, pp. 402–405.

[21] B. Thierry, B. Chaix, and Y. Kestens, "Detecting activity locations from raw gps data: a novel kernel-based algorithm," *International Journal of Health Geographics*, vol. 12, no. 1, p. 14, 2013.

[22] B. C. Csáji, A. Browet, V. Traag, J.-C. Delvenne, E. Huens, P. V. Dooren, Z. Smoreda, and V. D. Blondel, "Exploring the mobility of mobile phone users," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 6, pp. 1459 – 1473, 2013.

[23] Y. Chen, K. Jiang, Y. Zheng, C. Li, and N. Yu, "Trajectory simplification method for location-based social networking services," in *Proceedings of the 2009 International Workshop on Location Based Social Networks (LBSN)*, ACM, 2009, pp. 33–40.

[24] S. Khetarpaul, R. Chauhan, S. K. Gupta, L. V. Subramaniam, and U. Nambiar, "Mining gps data to determine interesting locations," in *Proceedings of the 8th International Workshop on Information Integration on the Web: In Conjunction with WWW 2011 (IIWeb)*, ACM, 2011, pp. 8:1–8:6.

[25] C. Long, R. C.-W. Wong, and H. V. Jagadish, "Direction-preserving trajectory simplification," *Proc. VLDB Endow.*, vol. 6, no. 10, pp. 949–960, Aug. 2013.

[26] G. Trajcevski, H. Cao, P. Scheuermanny, O. Wolfsonz, and D. Vaccaro, "On-line data reduction and the quality of history in moving objects databases," in *Proceedings of the 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE)*, ACM, 2006, pp. 19–26.

[27] R. Lange, F. Dürr, and K. Rothermel, "Online trajectory data reduction using connection-preserving dead reckoning," in *Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (Mobiquitous)*, ICST, 2008, pp. 52:1–52:10.

[28] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking (MobiCom)*, ACM, 2010, pp. 341–352.

[29] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, ACM, 2011, pp. 305–316.

[30] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[31] A. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger, "Is there a case for mobile phone content pre-staging?" in *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, ACM, 2013, pp. 321–326.

[32] 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)*, ACM, 2010, pp. 239–254.

[33] L. Ang, K. P. Seng, A. M. Zungeru and G. K. Ijemaru, *Big Sensor Data Systems for Smart Cities*,IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1259-1271, Oct. 2017.

[34] Yu, X., Zhao, H., Zhang, L., Wu, S., Krishnamachari, B., and Li, V. O. *Cooperative sensing and compression in vehicular sensor networks for urban monitoring*. IEEE International Conference on Communications pp. 1-5. May, 2010.

[35] J. Kwapisz, G. Weiss, and S. Moore, "Activity recognition using cell phone accelerometers," in *ACM SigKDD Explorations Newsletter*, 2011, vol. 12, no 2, p. 74–82.

[36] M. Puyau, A. Adolph, F. Vohra, I. Zakeri, and N. Butte, "Prediction of activity energy expenditure using accelerometers in children." in *Medicine & Science in Sports & Exercise*, 2004, vol. 36, no 9, p. 1625-1631.

[37] D. Brown, S. LaPoint, R. Kays, W. Heidrich, F. Kümmeth, and M. Wikelski, "Accelerometer-informed GPS telemetry: Reducing the trade-off between resolution and longevitys." in *Wildlife Society Bulletin*, 2012, vol. 36, no 1, p. 139-146.