CCF Transactions on Networking
https://doi.org/10.1007/s42045-018-0005-y

REGULAR PAPER

Edge-empowered accurate urban vehicle localization with cellular-aware trajectories

Hongzi Zhu[1] · Fan Wu[1] · Siyuan Cao[2] · Shan Chang[3] · Li Lu[4]

Received: 23 July 2018 / Accepted: 24 October 2018

Abstract
Acquiring accurate vehicle location information in urban settings is very challenging due to the complexity of urban environments. In this paper, we propose a novel scheme, called UPS, to tackle urban vehicle localization problem. After extensive empirical study, we find that GSM power spectrogram collected over a distance has ideal temporal–spatial characteristics for fingerprinting. Encouraged by this observation, UPS tries to utilize the geographical trajectory and the associated GSM power spectrogram information of a moving vehicle to identify its location with reference to a map. To this end, two appealing techniques, i.e., *online vehicle localization* and *GSM map construction*, are elegantly integrated. With the former, a vehicle can accurately fix its location under complex urban environments. With the latter, a reliable metropolitan-scale GSM power map can be cost-efficiently built at edges, leveraging the strong power of crowdsourcing. By design, UPS is light-weight, requiring only a minimum hardware deployment. We implement a prototype system to validate the feasibility of the UPS design. Furthermore, we conduct extensive trace-driven simulations and results show that UPS can work stably in various urban settings and achieve an accuracy of 5.3 m with a 90% precision, overwhelming the performance of GPS by five times.

Keywords Urban · Vehicle localization · GSM · Fingerprinting · Trajectory

1 Introduction

Obtaining accurate location information of vehicles, especially in urban environments, is of great importance to many appealing applications. For instance, in the collision avoidance application, one broken car can periodically broadcast its precise location to upcoming vehicles via vehicle-to-vehicle (V2V) communications so that these vehicles can get enough time to take necessary actions (e.g., braking or changing lanes). Besides driving safety, accurate location information is also desired in many location-based services (LBSs). Given the complicated road topology of a big city, navigation is one of the most popular LBSs, where a vehicle needs to know its location before the best route decision can be made. Wrong location information may lead to unnecessary detours and bring unpleasant driving experience to drivers.

This work is partially supported by National Natural Science Foundation of China (Grant No. 61472255, 61420106010, 61772340, 61672348, 61672151,61472068), the Fundamental Research Funds for the Central Universities (Grant No. EG2018028), Shanghai Rising-Star Program (Grant No.17QA1400100), DHU Distinguished Young Professor Program, and the National Key R&D Program of China (2017YFB1003003).

✉ Hongzi Zhu
hongzi@sjtu.edu.cn

Fan Wu
fwu@cs.sjtu.edu.cn

Siyuan Cao
cao208@purdue.edu

Shan Chang
changshan@dhu.edu.cn

Li Lu
luli2009@uestc.edu.cn

[1] Shanghai Jiao Tong University, Shanghai, People's Republic of China

[2] Purdue University, West Lafayette, USA

[3] Donghua University, Shanghai, People's Republic of China

[4] University of Electronic Science and Technology of China, Chengdu, People's Republic of China

To solve the *urban vehicle localization problem*, which refers to obtaining the reliable and accurate location information of vehicles in urban settings, however, is very challenging. The difficulty is four-fold. First, as the resolved location information would be used to build many safety-related applications, the requirement for high accuracy, therefore, is very critical. Large localization errors are the root cause of false alarms in those applications. Second, as the urban environment could be very dynamic and complex (e.g., variant traffic conditions and all types of roads surrounded by various buildings and infrastructures), to achieve stable yet accurate localization performance in all urban scenarios would meet various troubles, such as varying signal availability and non-line-of-sight condition. It is hard for one single solution to deal with all those problems. Third, the process for locating vehicles should be fast as vehicles often move at high speeds. Huge delays result in large deviations between the estimated and the true locations of a vehicle. Last, considering the large area of a metropolis and the vast number of vehicles, it is infeasible to construct new infrastructures or deploy dedicated devices on vehicles for urban vehicle localization, due to the prohibitive cost.

In the literature, there have been quite a number of localization schemes. Range-based localization schemes, such as ToA (Chan et al. 2006), TDoA (Priyantha et al. 2000) and AoA (Elnahrawy et al. 2007), measure the distance or angle from reference points and then perform trilateration or triangulation to obtain the estimated position. All these methods require specialized hardware and dense anchor deployment, which make them less attractive for a large deployment. Global Positioning System (GPS) is the most-widely used outdoor localization system. In metropolises like Toronto or New York, however, it is often the case that GPS may have large errors due to the signal availability problem caused by urban canyons. There are also Received Signal Strength (RSS) modeling-based ranging techniques (Chen et al. 2006) trying to capture the relation between signal strength and distance. In practice, the actual attenuation depends on multipath propagation effects, reflections and noises, which make it hard to build models in urban environments. Besides range-based schemes, range-free localization does not rely on measurement of distance or angles. Cell ID has been used in many schemes (LaMarca et al. 2005; Paek et al. 2011) to get efficient but coarse-grained location information with an error of tens of meters. Fingerprinting techniques (Bahl and Padmanabhan 2000; Varshavsky et al. 2005) can also be introduced and implemented to do localization. Although fingerprinting techniques have better accuracy, the overhead of constructing a reliable and fine fingerprint map at a metropolitan scale is vast. How to cost-efficiently construct such a map is still challenging. As a result, there is no existing solution, to the best of our knowledge, to figuring the urban vehicle localization problem.

In this paper, we propose an innovative scheme, called *Urban Positioning System* (UPS), which utilizes the received GSM wide-band signals and geographical trajectory information of vehicles to tackle the difficulty in solving the urban vehicle localization problem. Based on the intensive analysis on the received signal strength indicator (RSSI) values of a wide band of 194 GSM channels (denoted as GSM *power spectrogram*), we have the observation that GSM power spectrogram has ideal temporal–spatial characteristics for fingerprinting. Encouraged by this observation, UPS can accurately localize a moving vehicle by comparing the GSM power spectrogram information it has collected along its trajectory with a pre-constructed global signal map.

To this end, UPS tackles two following challenges. First, establishing a reliable metropolitan-scale map of GSM power spectrogram is labor-intensive, needing significant data collection. This is also the biggest obstacle which prevents existing fingerprinting indoor localization methods from transplanting to outdoor settings. Second, given a GSM power spectrogram map, how to eliminate the uncertainty in using compound trajectory for localization under dynamic and complex urban environments is nontrivial.

To deal with the first challenge, UPS leverages the extraordinary popularity of smartphones and the power of crowdsourcing to first collect *compound trajectories*, which refers to the combination of a geographical trajectory and the associated GSM power spectrogram measured along the trajectory, from individual vehicles. Then, individual RSSI measures of each vehicle are geographically aligned to a digital map by conducting a dynamic time wrapping (DTW) at edge servers. With the rich set of individual RSSI measures about the GSM channels at each location, an edge server first removes obvious outlier measures and then aggregates all consistent ones. In this way, UPS can cost-effectively retrieve a reliable and accurate GSM fingerprint map at a large scale.

To figure the second challenge, UPS conducts a sliding check on the map to identify the most-likely location with regard to a compound trajectory segment of a vehicle, using Pearson's correlation coefficient. When the segment used for search is sufficiently long, UPS can achieve good localization accuracy in urban settings. The rationale lies in the fact that the influence of faded channels at certain locations in a long trajectory is reduced. Compared with existing work where cell ID and RSS values on a little number of channels (up to 35; Varshavsky et al. 2005) are mostly used, the long trajectory and wide-band signal combine into a high dimension value which can be used to largely enhance the outdoor localization accuracy.

We implement a prototype of UPS and conduct extensive real-world experiments in Shanghai city. With a minimum hardware deployment, it is easy for UPS to gain a large deployment. In addition, when combined with a coarse

GPS- or cell-ID-based localization scheme, the computation cost of UPS can be extremely low as a vehicle only needs to compare with a limited number of roads in a small region on the map. The experiment results demonstrate the efficacy of UPS. On average, UPS can achieve a high localization accuracy of 5.3 m with a precision of 90% and outperform that of GPS by five times.

We highlight our main contributions made in this paper as follows:

- We have conducted intensive analysis on GSM power spectrogram and have the observation that a wide band of GSM signals are ideal for fingerprinting because of the wide availability and the good temporal–spatial characteristics as well.
- We have proposed a cost-efficient GSM fingerprint map construction scheme, leveraging the wide availability of smartphones and the strong power of crowdsourcing and edge computing. Individual GSM RSSI measures can be geographically aligned to a digital map and aggregated with little intervention of outliers.
- We have proposed a robust online vehicle localization scheme based on the correlation calculation using compound trajectories of a vehicle. With this scheme, UPS is resilient to dynamic and complex urban environment.
- We have implemented a prototype system consisting of two vehicles and conducted extensive experiments of over 200 km in Shanghai city. The results show the efficacy of UPS, outwitting that of GPS by five times on average.

The remainder of the paper is organized as follows. In Sect. 2, we survey the related work. We analyze the temporal–spatial characteristics of GSM power spectrogram for fingerprinting in Sect. 3. Section 4 elaborates the UPS design. The detailed implementation of our UPS prototype system is presented in Sect. 5. In Sect. 6, we evaluate the performance of UPS under various conditions and describe the results. We discuss several design issues of UPS in Sect. 7. Finally, we draw our conclusion and give directions for our future work in Sect. 8.

2 Related work

Here we review existing localization techniques, which can be categorized into two classes, i.e., range-based and range-free.

2.1 Range-based

Range-based localization methods measure the distance or angle from reference points and then perform trilateration or triangulation to obtain the estimated position. Through ToA (Chan et al. 2006), distance between sender and receiver of a signal can be determined by using the measured signal propagation time and known signal velocity. Similarly, AoA (Elnahrawy et al. 2005, 2007; Biswas et al. 2005; Li and Lu 2008) (Angle of Arrival) uses measured angle, which is typically achieved using an array of antennas or microphones. Both of these two methods require specialized hardware and highly accurate synchronization, which make them less attractive for a large deployment. Pinpoint (Youssef et al. 2006) improves the idea of TOA which needs no synchronization between devices by increasing the number of exchange times of RF signals. TDoA (e.g., Cricket; Priyantha et al. 2000) is a variance of ToA with no clock synchronization required. There are also RSS modeling-based ranging techniques (Chen et al. 2006) trying to capture the relation between signal strength and distance. In practice, the actual attenuation depends on multipath propagation effects, reflections and noises, which make it hard to build models in urban environments.

GPS is the most widely used location-sensing system, which can be categorized as ToA-based. Although GPS is ubiquitous among cellphones nowadays, it still suffers from severe power consumption problem. A-Loc (Lin et al. 2010) continually tunes the energy expenditure to meet the changing accuracy requirements using available sensors. RAPS (Paek et al. 2010) turns on the GPS in a rate-adaptive manner to obtain accurate position information while spending minimal energy. LEAP (Ramos et al. 2011) carefully partitions the GPS signal processing pipeline and shifts delay tolerant position calculations to the cloud. Kjærgaard et al. (2011) utilizes the accelerometer and compass of smartphone to track a car based on an initial start location provided by the GPS. Besides huge power consumption, GPS is also restricted by factors such as the availability of signals and atmospheric effects which make GPS-based schemes can hardly achieve high localization accuracy. Differential Global Positioning System (DGPS) is an enhancement to GPS that provides improved location accuracy, from the 15-m nominal GPS accuracy to about 10 cm in case of the best implementation. However, DGPS relies on a network of fixed, ground-based reference stations, which is costly to be widely deployed.

2.2 Range-free

Range-free localization does not rely on measurement of distance or angles. Cell ID has been used in many schemes to get coarse-grained location. Place Lab (LaMarca et al. 2005) is a famous example which evaluates device positioning by listening to radio beacons (such as 802.11 APs, GSM cell towers, and fixed Bluetooth devices) and estimating their own location referenced to the positions of those beacons.

CAPS (Paek et al. 2011) uses a cell ID sequence matching technique to estimate current position based on the history of cell ID and GPS position sequences. There are also commercial systems such as Google's MyLocation (Google 2018) and Skyhook (2018), requiring a data of cell tower locations. Cell ID based localization is efficient but coarse-grained with an error of tens of meters (LaMarca et al. 2005).

Fingerprinting techniques can also be introduced and implemented to do localization. The key idea is to pre-collect the characteristic values at different locations into a map and then match the real-time measurements to this digital map. RADAR (Bahl and Padmanabhan 2000) is a classic indoor localization system based on radio frequency (RF) fingerprinting. Deterministic techniques are proposed in Chen et al. (2006), Varshavsky et al. (2005) for GSM localization while CellSense (Ibrahim and Youssef 2012) is a probabilistic RSSI-based fingerprinting location determination system for GSM phones. Horus (Youssef and Agrawala 2005) and TIX (Gwon and Jain 2004) are other representative work using fingerprinting to localize in indoor environments. Channel State Information (CSI) is also used for indoor localization (Xiao et al. 2013), however CSI is so sensitive to apply to the dynamic and complicated outdoor environment with many moving objects. Note that although fingerprinting techniques can achieve better accuracy, the vast overhead of constructing a reliable and fine fingerprint map at a metropolitan scale still needs a better solution. As a result, fingerprinting is more likely to either be limited in indoor localization or stay unsatisfying in accuracy when extended to outdoor localization, as the best median accuracy of above fingerprinting-based work in outdoor environments is merely 30 m (Ibrahim and Youssef 2012).

In summary, there is no existing successful solution, to the best of our knowledge, which can provide fast and accurate location information for fast moving vehicles in urban settings.

3 Empirical study on GSM power spectrogram

In this section, we first describe the trace of GSM power spectrogram we have collected in our city and then analyze the temporal–spatial characteristics of GSM power spectrogram for fingerprinting.

3.1 GSM primer

Global System for Mobile Communication (GSM) is the most widespread cellular telephony standard in the world, with deployments in more than 210 countries by over 676 network operators (Varshavsky et al. 2005). In China, GSM mainly operates on the 900 MHz frequency band, which is subdivided into 200 kHz wide physical channels with index 955-1023 and 0-124. As GSM is a cellular network, each cell contains an area with the radius of up to 35 km in suburban areas and down to 500 m in downtown areas, which means a higher cell density in downtown. Each cell is allocated a number of physical channels (five at most), according to the needs of network traffic. The indices of physical channels are chose in a way to avoid interference between neighboring cells. Although within a certain cell only a limited number of channels can be used for effective communication, the RSS values of the rest channels still vary between different locations.

Based on the prior knowledge above, we try to utilize the distinguishable characteristic of GSM wide-band signals to identify locations and enhance the stability by using signals along trajectories.

3.2 Collecting GSM power spectrogram

We utilize the OsmocomBB project (Osmocombb 2018) and cheap GSM radios (i.e., Motorola C118 cellphones) to measure GSM power spectrogram. Figure 1 illustrates an R-GSM-900 power spectrogram measured at each meter on an urban road. Although there are 194 channels in the band (indexed from No. 955-1023 and from No.0-124), we only use 125 of them (from No.0-124) since the rest channels seem inactive as shown in the figure.

In order to obtain a large trace of real GSM power spectrograms for analysis, we mount 16 radios (for scanning channel in parallel) on top of the roof of an experiment vehicle and war drive a typical urban area in Shanghai city within an area of about 6 km^2 and a route about 30 km long. With 16 radios, the 125-channel GSM power spectrogram can be sampled at a rate of 10 Hz. By deliberately driving at a speed lower than 35 km/h, we can obtain a *power vector*, which refers to the vector of RSSI values over all 125 channels measured at one location, at every 1 m. Meanwhile, we also record the geographical trajectory information of the vehicle (including the heading direction and distance

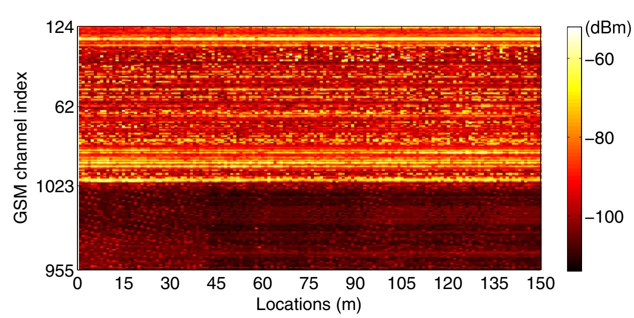

Fig. 1 Illustration of GSM power spectrogram of 194 channels collected over 150 m

information, see Sect. 4.2 for details). We war drive for 16 days from May 4th to May 19th in the year of 2014 and collect the trace. Note that trace collected on different days are geographically aligned using the technique described in Sect. 4.3.

3.3 Temporary stability

For fingerprinting, it is essential for GSM power spectrogram to have the feature of *temporal stability*, which refers to that a power spectrogram measured at a fixed location tends to be invariant during a *long* period of time. Otherwise, the cost for constructing and updating a map of GSM power spectrogram for localization will be enormous, which makes using GSM for localization try in vain. To investigate the temporal stability of GSM power spectrogram, we first measure the similarity between the RSSI values of the same channel over a distance measured at different time, using Pearson's correlation coefficient, defined as

$$r_{X_i^{t_1,l_1,m}X_i^{t_2,l_1,m}} = \frac{\sum_{j=1}^{m}(x_i^{t_1,l_1,j} - \overline{x_i^{t_1,l_1}})(x_i^{t_2,l_1,j} - \overline{x_i^{t_2,l_1}})}{\sqrt{\sum_{j=1}^{m}(x_i^{t_1,l_1,j} - \overline{x_i^{t_1,l_1}})^2}\sqrt{\sum_{j=1}^{m}(x_i^{t_2,l_1,j} - \overline{x_i^{t_2,l_1}})^2}}, \tag{1}$$

where $X_i^{t,l,m} = (x_i^{t,l,1}, x_i^{t,l,2}, \dots, x_i^{t,l,m})$ (referred to as a *power-strip*) denotes the RSSI values of channel i measured at time t from location l over a distance of m meters; $\overline{x_i^{t,l}}$ demotes the average of the power strip, i.e., the average of $x_i^{t,l,j}$ for all $j \in [1,m]$. We further calculate the *spectrogram correlation coefficient* (SCC) to measure the linear dependence between two GSM power spectrogram segments collected at different time, defined as follows,

$$r_{\mathbb{S}^{t_1,l_1,m}\mathbb{S}^{t_2,l_1,m}} = \frac{1}{n}\sum_{i=1}^{n} r_{X_i^{t_1,l_1,m}X_i^{t_2,l_1,m}}, \tag{2}$$

where $\mathbb{S}^{t_1,l_1,m} = (X_1^{t_1,l_1,m}; X_2^{t_1,l_1,m}; \dots; X_n^{t_1,l_1,m})$ and $\mathbb{S}^{t_2,l_1,m} = (X_1^{t_2,l_1,m}; X_2^{t_2,l_1,m}; \dots; X_n^{t_2,l_1,m})$ denote two GSM power spectrogram segments of m meters long consisting of n channels collected from location l_1 and at time t_1 and t_2, respectively. We randomly choose 2000 segments of 100 and 200 m long, respectively, from the trace collected on May 18th which is cloudy. For each segment, we calculate the SCC between this segment and the corresponding segment collected at the same location on May 17th which is rainy, May 11th which is sunny, and May 4th which is cloudy (i.e., with a time interval of 1 day, 7 days, and 15 days), respectively.

Figure 2 plots the cumulative distribution functions (CDFs) of the results. Here we have four observations. First, in general, the GSM power spectrogram has good temporal stability. For example, about 90% of the correlation coefficient values are larger than 0.1 even when the duration is larger than 2 weeks. Second, the temporal stability may slightly vary on different dates. Furthermore, it is not obvious that the temporal stability decreases as time goes by. For example, the coefficient values calculated with an interval of 7 days (on May 11th) is even smaller than that calculated with an interval of 15 days (on May 4th). Third, increasing the length of segments can greatly increase the temporal stability. For instance, about 90% of the correlation coefficient values are larger than 0.2 when the length of segments increases from 100 to 200 m. The reason is that increasing the length of segments can reduce the influence of faded channels at certain locations. The signal wide band and trajectory combine into high signal dimension, which enhances the temporal stability of GSM signals. Last, weather conditions have little impact on temporal stability of the GSM power spectrogram. It confirms to results (Usman et al. 2015) that suggest location as having a dominating impact on the received signal strength among other environmental factors such as weather and time of the day.

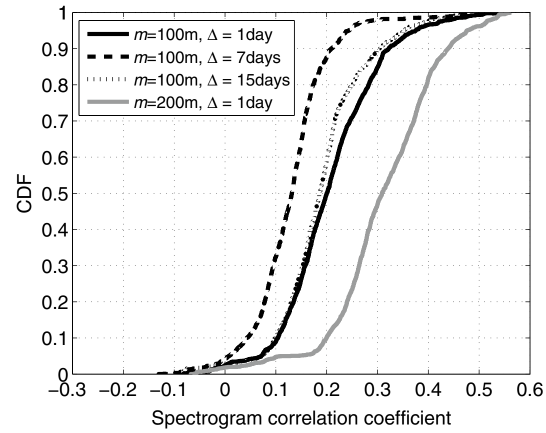

Fig. 2 CDFs of SCC calculated with segments collected at different time

3.4 Geographical uniqueness

In addition to temporal stability, it is also desired for GSM power spectrogram to have the feature of *geographical uniqueness*, which refers to that power spectrograms collected on different locations within a sufficiently large area should be distinctive. To examine the geographical uniqueness of GSM power spectrogram, we calculate the SCC between $\mathbb{S}^{t_1,l_1,m}$ and $\mathbb{S}^{t_2,l_2,m}$, which denote two power spectrogram segments of m meters long collected from location l_1 and l_2 and at time t_1 and t_2, respectively. Moreover, we consider the Euclidean distance between l_1 and l_2 is less than a given distance d, i.e., l_2 is within a disk area $R^{l_1,d}$ with l_1 located at the disk center and a radius of d.

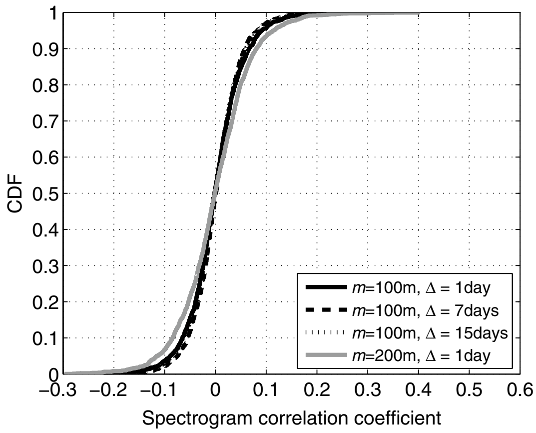

Fig. 3 CDFs of SCC calculated with segments collected at different locations

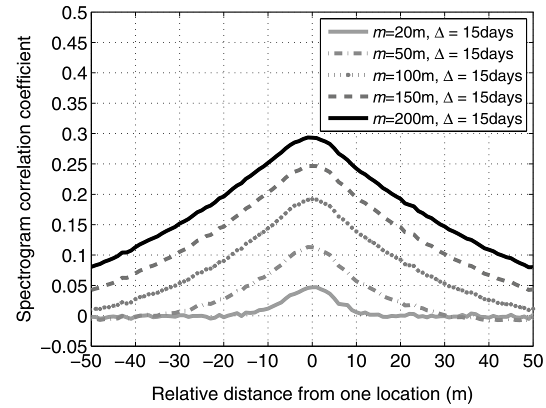

Fig. 4 Spectrogram correlation coefficient calculated with segments with a distance offset of *m* meters

We randomly choose 500 segments of 100 and 200 m long, respectively, from the trace collected on May 18th. For each segment $\mathbb{S}^{t_1,l_1,m}$, we randomly select 100 segments of the same length within the region $R^{l_1,1\,\mathrm{km}}$ from the trace collected on May 17th, May 11th, and May 4th, respectively, and calculate the SCC.

Figure 3 plots the cumulative distribution functions (CDFs) of the results. We have two main observations. First, the GSM power spectrogram has excellent geographical uniqueness. For example, about 95% of the correlation coefficient values are less than 0.1 over all time. Second, increasing the length of segments has only negligible impact on the geographical uniqueness. For example in the figure, the correlation coefficient of using 200-m segments is almost the same as that of using 100-m segments. It is because that for two segments along different locations, the measurements on each corresponding points are supposed to have poor similarity with each other, which would not be improved by aggregating more points with poor similarity together. Combing with the observations about temporal stability, we are highly encouraged to use GSM power spectrograms for fingerprinting.

3.5 Distinctive resolution

Besides the capability for fingerprinting, we also care about the resolution of GSM power spectrogram for localization as high accuracy is required in localizing urban vehicles. We check the *distinctive resolution* of GSM power spectrogram, which refers to the minimum distance over which two power spectrogram can be distinguished. We randomly choose 500 segments of 20, 50, 100, 150, and 200 m, respectively, from the trace collected on May 18th. For each segment $\mathbb{S}^{t_1,l_1,m}$, we calculate the SCC using $\mathbb{S}^{t_1,l_1,m}$ and $\mathbb{S}^{t_2,l_2,m}$, where t_2 refers to May 4th and location l_2 is taken m meters away from location l_1 for $m \in [-50, 50]$.

Figure 4 shows the SCC as a function of the offset distance m. It can be seen that longer segments would have larger SCC. Moreover, the correlation coefficient can always achieve maxima when $m = 0$, i.e., two segments are aligned, and monotonically and steeply decrease as the offset distance increases. As urban environment has many obstructions due to large buildings, these obstructions actually prove advantageous because they help form unique fingerprints with satisfactory resolution to allow the algorithm to differentiate between nearby locations. The blocking effect as well as the intense cell density in urban areas results in the unique GSM signal spectrogram, thereby enhances the resolution of GSM wide-band signals. It is also clear to tell that spectrogram correlation coefficient is an ideal metric to represent the fine distinctive of GSM power spectrogram as it varies even on a distance difference of 1 m. Given these observations, we are also encouraged to localize a vehicle with an appealing accuracy of 1 m by conducting a hill-climbing search for the maximum of the spectrogram correlation coefficient instead of using a threshold-based scheme.

As illustrated in Fig. 1, many GSM channels seem idle and therefore may affect the overall temporal–spatial characteristics of GSM power spectrogram for fingerprinting. A channel is *effective* if it has good temporal stability, geographical uniqueness and distinctive resolution. To check the effectiveness of each channel, for a power strip $X_i^{t_1,l_1,m}$ of channel i, we first calculate the utility of temporal stability of this power strip as

$$u^{stability}_{X_i^{t_1,l_1,m}} = r_{X_i^{t_1,l_1,m} X_i^{t_2,l_1,m}}. \tag{3}$$

The utility of geographical uniqueness of this power strip is defined as

$$u^{uniqueness}_{X_i^{t_1,l_1,m}} = \frac{1}{K} \sum_{k=2}^{K} r_{X_i^{t_1,l_1,m} X_i^{t_2,l_k,m}}, \tag{4}$$

where location l_k is within the disk area $R^{l_1,d}$. The utility of distinctive resolution of this power strip is calculated as

$$u^{resolution}_{X_i^{t_1,l_1,m}} = r_{X_i^{t_1,l_1,m} X_i^{t_2,l_1,m}} - r_{X_i^{t_1,l_1,m} X_i^{t_2,l_1+\delta,m}}, \tag{5}$$

where δ denotes the target resolution that we would like to achieve with this channel. By averaging the utility of temporal stability, geographical uniqueness and distinctive resolution over all power strips of channel i, we get the utility of temporal stability, geographical uniqueness and distinctive resolution of channel i, denoted as $u_i^{stability}$, $u_i^{uniqueness}$ and $u_i^{resolution}$, respectively. Finally, we define the *utility* of channel i as follows

$$u_i = u_i^{stability} - u_i^{uniqueness} + u_i^{resolution}. \tag{6}$$

3.6 Identifying effective GSM channels

We randomly choose 1000 segments of 300 m long from trace collected on May 18th. For each channel, we calculate the utility of that channel with d and δ set to 1 km and 10 m, respectively, using the trace collected on May 4th. Figure 5 plots the utility of temporal stability, geographical uniqueness and distinctive resolution and the corresponding channel utility of each channel. It can be seen that the channel utility varies with different channels. We select 41 channels with the utility higher than 0.2 (as shown by the dotted rectangles in the figure) and use these channels for fingerprinting.

It should be noted that the selected effective channels would vary between different locations, which is determined by the channel allocation in current and neighboring cells. Based on our observation, the effective channels seldom change within a large area of several square kilometers. This suggests that we construct a correspondence table between cell IDs and the sets of effective channels and dynamically choose the proper channel set according to current coarse location information getting from cell ID or GPS.

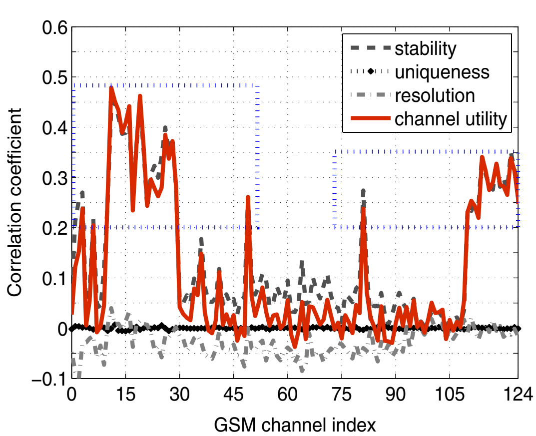

Fig. 5 Selecting effective channels based on channel utility

4 Design of UPS system

4.1 Overview

With the observation that GSM power spectrogram has ideal temporal–spatial characteristics for fingerprinting, we are highly encouraged to localize a moving vehicle by comparing the GSM power spectrogram information it has collected along its trajectory with a pre-constructed global signal map. To this end, there are two main challenges, i.e., *reliable metropolitan-scale map construction* and *robust online vehicle localization*. To tackle those challenges, UPS elegantly integrates two key components in the system, i.e., *online vehicle localization* at the vehicle side and *GSM map construction* at edge servers. The system architecture of UPS is shown in Fig. 6.

Online vehicle localization In order to determine its location, a vehicle first utilizes on-board motion sensors such as the accelerometer, gyroscope, and compass to perceive its motion and *estimate geographical trajectory* information. At the same time, the vehicle also captures the GSM power spectrogram via GSM radios as it moves, and conducts a *compound trajectory generation* procedure to bind the measured GSM spectrogram to the geographical trajectory, forming its compound trajectory. The vehicle contributes its compound trajectory to a nearby edge server and also downloads the up-to-date GSM signal map in return (as illustrated by the dashed arrow lines) once wireless communications are available. With the signal map, the vehicle executes an online *maximum similarity search*, seeking the most-likely location in the map with the help of GPS with its own compound trajectory.

GSM map construction To build a metropolitan-scale GSM signal map, edge servers keep collecting enormous

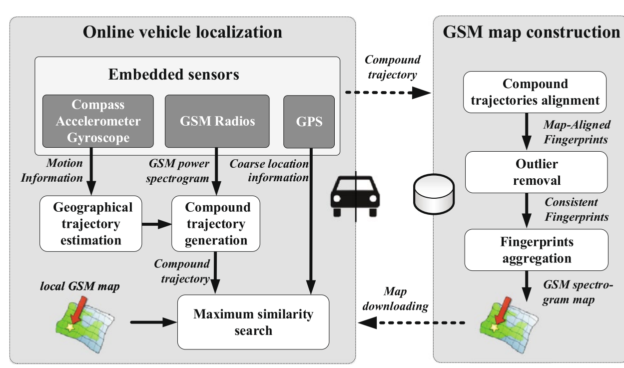

Fig. 6 System architecture of UPS design

amount of compound trajectories from individual vehicles. Because GPS is not trustworthy in many urban scenarios and different vehicles may have variant trajectories even when moving along the same path, edge servers carry out the same *compound trajectory alignment* procedure as described above to align all collected compound trajectories to a digital map so that the location of a GSM power vector measured by a vehicle can be best estimated. After gathering sufficient power vectors assumed to be measured at a location, an *outlier removal* procedure is conducted to eliminate obvious outlier measures before those consistent fingerprints are aggregated by conducting a *fingerprints aggregation* procedure. Leveraging the tremendous power of crowdsourcing, UPS can cost-effectively establish a reliable and accurate GSM power spectrogram map at a metropolitan scale.

As the observations found in temporal stability analysis suggest, GSM power spectrogram should be measured over time and aggregated in order to achieve a reliable signal map. Moreover, GSM power map do not need to be updated frequently because of its good temporal stability. Therefore, the cost for map updating in UPS is low. In addition, combining a coarse localization such as GPS or a cell-ID-based scheme, the response time of UPS for fine localization can be greatly reduced as the searching area is restricted.

4.2 Online localization at a vehicle

In this subsection, we first describe the key techniques adopted at the vehicle side for accurate urban localization.

4.2.1 Physical trajectory estimation

From the temporal–spatial analysis, we have the observation that GSM power spectrogram segments over a long distance is better for fingerprinting than those with a short length. This encourages a vehicle to collect GSM power measures over its geographical trajectory for localization. Therefore, it is fundamental for the vehicle to first obtain its accurate geographical trajectory. In UPS, the geographical trajectory of a vehicle can be estimated with two techniques.

Perceiving the heading direction The vehicle can utilize the embedded compass to get both the strength and the direction of the magnetism of the earth by calculating the sum of magnetization readings along the *x*- and *y*-axis of the compass. By aligning the coordinate system of the compass with that of the vehicle using the scheme proposed in work Xiao et al. (2013), the heading direction of the vehicle can be derived as the angle between the *y*-axis of the compass and the direction of the magnetism of the earth.

Estimating the instant speed The distance traversed along one direction can be calculated as the integral of instant speed values over time. In UPS, a simple way to obtain accurate speed information is to gain access to the onboard Electronic Control Unit (ECU) of the vehicle using an OBD interface. Schemes like SenSpeed (Han et al. 2014), where inertial sensors such as a 3D accelerometer and a gyroscope is used to estimate the instant speed of a vehicle, leveraging the uneven surface of urban roads and variant driving behavior to eliminating accumulative errors, can also be adopted. Due to the page limitation, we omit the details from this paper.

With the heading and distance information, the vehicle can estimate its geographical trajectory of m meters, denoted as a vector $\mathscr{T}^m$ with m elements. An element i in $\mathscr{T}^m$ can be denoted as a tuple (θ_i, t_i) , where θ_i and t_i represent the heading angle and the timestamp calculated for the ith meter on the trajectory.

4.2.2 Compound trajectory generation

A vehicle can scan GSM channels with one or multiple GSM radios while moving. The retrieved power spectrogram, however, is a time-domain signal, which should be bound to the geographical trajectory of the vehicle to form a *compound trajectory* before it can be used for comparison with the GSM signal map.

To generate a compound trajectory, the RSSI readings of GSM channels measured at each time instance is associated to the corresponding location traversed by the vehicle at that time in the geographical trajectory. We denote the tuple $\mathbb{C}^m = (\mathscr{T}^m, \mathbb{S}^m)$ as a resolved compound trajectory of m meters, where $\mathscr{T}^m$ and $\mathbb{S}^m$ are the corresponding geographical trajectory and the associated GSM power spectrogram measured over $\mathscr{T}^m$, respectively. An element i in $\mathbb{C}^m$ can be denoted as a triple $(\theta_i, t_i, V^{t_i,l,i})$, where $V^{t_i,l,i} = (x_1^{t_i,l,i}, x_2^{t_i,l,i}, \dots, x_n^{t_i,l,i})$ is the associated power vector. It should be noted that, as it takes time to scan all effective GSM channels, when the vehicle moves fast, it is possible that some channels (referred to as *missing channels*) within a power vector at a particular location are not measured. In UPS, missing channels are filled with the nearest value along the trajectory.

The generated compound trajectories can be used for two purposes. One is for a nearby responsible edge server to construct local GSM map; the other one is to localize the vehicle with a pre-downloaded GSM signal map obtained from nearby edge servers.

4.2.3 Maximum similarity search for localization

With the observations that the SCC can always reach maxima when two spectrogram segments are perfectly aligned, UPS conducts a maximum similarity search to localize a vehicle.

Specifically, to acquire its accurate location information, a vehicle uses a segment of its own compound trajectory to search in a restrained area on the GSM map to identify the most likely location that the segment. Given a compound trajectory segment $(\mathscr{T}^m, \mathbb{S}^m)$ of the vehicle, we want to identify the location of the corresponding power spectrogram segment in the map that has the maximum similarity with $\mathbb{S}^m$, i.e.,

$$\arg\max_{l} r_{\mathbb{S}^m \mathbb{S}^{l,m}}, \quad l \in R^{l_0,d}, \tag{7}$$

where l is an arbitrary location within the disk region $R^{l_0,d}$ on the GSM map; $\mathbb{S}^(l, m)$ is a GSM power spectrogram segment of m meters picked from the map starting from location l; l_0 denotes the coarse starting location of the compound trajectory segment obtained with GPS and d is the accuracy error of GPS. As a result, the solution to (7) l^*, referred to as a *location reference point* (LRP), is treated as the best location where $\mathbb{S}^m$ might start. In UPS, and a hill-climbing procedure using a sliding window of m meters is conducted to search for l^* road by road in $R^{l_0,d}$.

As localizing a moving vehicle calls for short response time, how to minimize the complexity for searching l^* is of great importance. Firstly, we should try the best to restrain the initial searching area by using some kind of coarse-grained localization methods as supplement. In rather open areas where GPS is mainly available, such as on the main roads, an initial area with a radius of tens of meters can be attained by GPS. Whereas when surrounded by tall buildings or shielded by elevated roads, cell ID can be used as an alternative to GPS, though the initial area may expand to a distance of several kilometers. Secondly, note that given the limited area of $R^{l_0,d}$, the number of roads needed to be searched is limited. In order to avoid unnecessary SCC calculation between $\mathbb{S}^m$ and $\mathbb{S}^{l,m}$, the angle information of the corresponding geographical trajectory and the topology of road segments are first compared to filter out non-interest roads.

With a LRP, the location of the vehicle can be easily obtained by adding the distance traveled since that LRP.

4.3 GSM map construction at edge severs

In this subsection, we describe how to cost-efficiently construct a reliable and accurate GSM map at a metropolitan scale.

4.3.1 Compound trajectories alignment

Different vehicles may have slightly different trajectories along the same path. In UPS, an edge server collects compound trajectories of nearby individual vehicles, trying to obtain reliable GSM power measurement from those compound trajectories. To reduce the divergence of individual trajectories, we adopt Dynamic Time Warping (DTW) (Chandrasekaran et al. 2011) to align each individual compound trajectory to a digital map.

Specifically, a large urban area is divided into small regions according to the deployment of edge servers. Each edge server collects compound trajectories from individual vehicles within its region via 3G/4G mobile networks to construct a GSM map of this region. For an edge server, collected compound trajectories are first partitioned into segments at turns. With the GPS location information of the starting and ending point of such a segment $\mathbb{C}^{m_1}$, the corresponding intersections can be located in the digital map. Given the two intersections, all paths on the map with a similar length with $\mathbb{C}^{m_1}$ are selected. To generate a warping path between $\mathbb{C}^{m_1}$ and one such path $\mathcal{P}^{m_2}$ (with a length of m_2 meters), DTW constructs a distance matrix $D[m_1 \times m_2]$ which represents the minimum distance to reach any point (i, j) in the matrix from (0, 0) using a dynamic programming formulation shown as follows,

$$D(i,j) = d(i,j) + min(C(i-1,j-1), C(i-1,j), C(i,j-1)), \tag{8}$$

where $d(i, j)$ refers to the Euclidean distance between the ith position in $\mathcal{P}^{m_2}$ and the jth position in $\mathbb{C}^{m_1}$. The algorithm is illustrated in Fig. 7.

We apply DTW to all paths and select the path with the minimum $D(m_1 - 1, m_2 - 1)$ as the *matched path* with regard to $\mathbb{C}^{m_1}$. By retrieving the warping path with the matched path, we can get the correspondences of each locations on the compound trajectory segment to those on the matched path. With such correspondences, a GSM power vector measure on the compound trajectory segment can be assigned to the match path. For one-to-one or one-to-many (e.g., l'_3–l_2, l'_3–l_3 and l'_3–l_4 in Fig. 8) alignments, a measure can be simply copied. In contrast, for many-to-one alignments (e.g., l'_1–l_1 and l'_2–l_1), measures are first averaged before assigned. By this means, for each location on each path in the digital map, the edge server can collect a set of power vectors measures of individual vehicles.

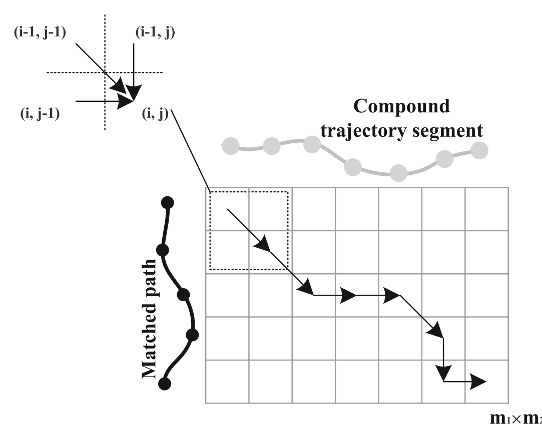

Fig. 7 The process of warping path generation in DTW

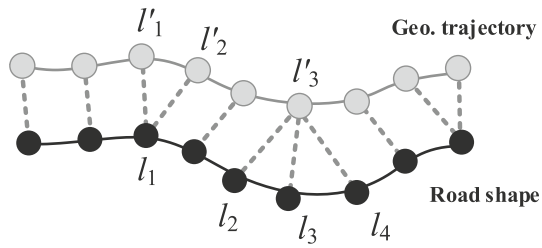

Fig. 8 Aligning an individual compound trajectory to the matched path on a digital map using DTW

4.3.2 Outlier measure removal and consistent fingerprints aggregation

For each channel at a certain position on the digital map, outlier measures can be removed by computing the distribution of differences between a RSSI reading and all other RSSI readings in the dataset. For each RSSI reading, we compute the mean difference from it to all its neighbors. By assuming that the resulted distribution is Gaussian with a mean and a standard deviation, all readings whose mean differences are outside an interval defined by the global differences mean and standard deviation can be considered as outliers and trimmed from the dataset (Otsason et al. 2005). After all outlier measures are removed, a reliable and accurate GSM power spectrogram map can be finally constructed by aggregating all consistent power measures into a mean digital map. However, the whole data set containing all consistent power measures should be kept and maintained in an edge server for further usage of removing the abnormal measures and aggregating consistent fingerprints.

5 Prototype implementation

We first study the impact of radio posture on the quality of collected GSM power spectrogram and conduct static experiments in our lab. In specific, as depicted in Fig. 9, we put five Motorola C118 phones with different postures on a plane and put another two phones with 0.5 m higher and 0.5 m lower, respectively. Each phone continuously scans 128 channels for one hour. Figure 10 shows the cumulative distribution of correlation coefficient values between GSM spectrogram measurements from one of the horizontal phone and all other radios at the same scanning cycle. It can be seen that over 90% of the correlation coefficient values are larger than 0.6, which indicates a satisfactory similarity between radios with different placements. Therefore, we draw the conclusion that the localization performance would hardly be affected by radio placement.

We implement a prototype system using two cars, i.e., one for map construction and the other for localization, as

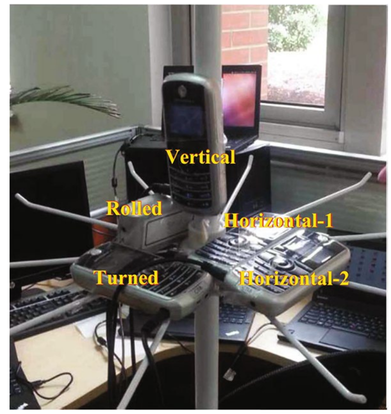

Fig. 9 GSM radios set in different orientations

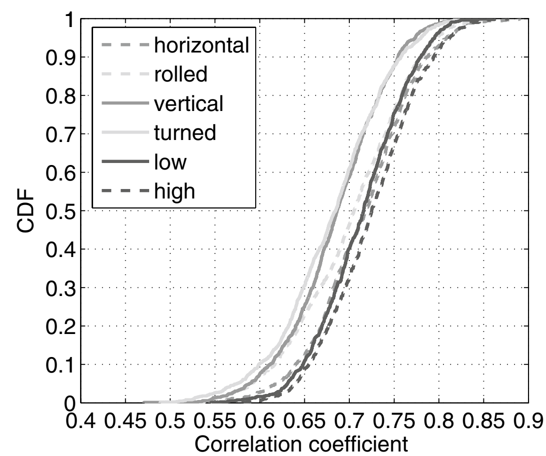

Fig. 10 Similarity between different GSM radios with different placements

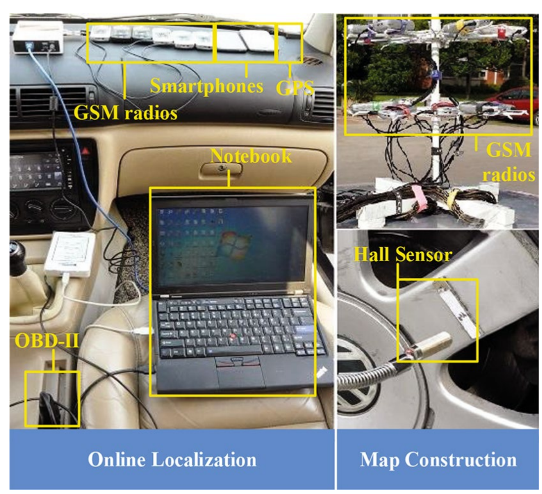

Fig. 11 Our prototype implementation of UPS

shown in Fig. 11. For online localization, the vehicle integrates a set of GSM radios (e.g., Motorola C118 cellphones are used), an inertial motion sensor to estimate the heading direction (e.g., we get motion sensor readings from two Android smartphones, one HTC S720t and one Samsung Galaxy S4), an OBD-II interface to read the instant speed of the car, a GPS module for rough location information, and an onboard computing unit (e.g., a laptop). To examine the impact of different GSM radio deployments to the performance of localization, we mount three sets of one, two, and four C118 phones, respectively, on the dash panel, and mount one set of one C118 phone on the roof of the vehicle. Each set of these phones scan the GSM band in parallel. For vehicles with no OBD interface available, the SenSpeed (Han et al. 2014) scheme based on motion sensors can be used to estimate speed information. Both methods can achieve good accuracy as about 90% errors for using OBD are about 0.8 km/h and 3.2 km/h for using vibration data. Using OBD and vibration data have an average error of 0.4 km/h and 2 km/h, respectively. For map construction, the map-construction vehicle integrates two smartphones, an OBD-II interface and a GPS receiver. In addition, to boost the procedure of map construction, we mount more GSM radios (i.e., 16 C118 phones) on the roof of the vehicle and a Hall sensor to detect the revolution of wheels in order to get the real travelled distance.

6 Performance evaluation

6.1 Methodology

With our prototype system, we war drive a typical urban area in our city within an area of about six square kilometers and a route about 30 km long (as illustrated in Fig. 12) for 16 days from May 4th to May 19th. With 16 radios, the 125-channel GSM power spectrogram can be sampled at a rate of 10 Hz. By deliberately driving at a speed lower than 35 km/h, we can obtain a power vector, which refers to the vector of RSSI values over all 125 channels measured at one location, at every 1 m. In addition, the trace of all sensors is collected. We encountered both heavy and light traffic when the trace was collected. In order to mitigate the accumulative errors in travel distance derived from the Hall sensors, we enforce exactly the same 23 stop locations on the route of each day.

We construct an aggregated GSM map using the trace of the first 14 days collected by the group of 16 radios according to the scheme described in Sect. 4.3 and use the trace of the last 2 days collected by groups of one/two/four

Fig. 12 The 2 km × 3 km urban experiment area

radios for online vehicle localization test. We use group of fewer radios to simulate the practical situation where cars are usually driven at a higher speed and missing channels are more common.

We evaluate the localization performance of UPS and compare with GPS using the following metric.

Localization error refers to the absolute distance difference between the estimated location and the ground truth. The ground truth for UPS is calculated as the distance since last stop, which can be derived from the wheel revolution information. We measure the circumference of the wheel on every experiment. The ground truth of GPS is the aggregated (average) GPS reports in the first 2 weeks on the same spots.

In the following trace-driven simulations, we first investigate the effect of system parameter configuration to the system performance, and then use the optimal parameter configuration to compare the performance of UPS with that of GPS.

In this section, we first study the effect of the sliding window length and radio configuration to the localization performance of UPS. We vary the length of the sliding window from 25 to 300 m with an interval of 25 m and vary the number and placement of GSM radios. For each setting, we randomly select 1000 locations for test from the trace collected on May 18th and 19th. For each location, we use the most-recent compound trajectory segment of 200 m to conduct the maximum similarity search within a disk range with a radius of 50 m centered at the GPS coordinates of that location on the aggregated GSM map.

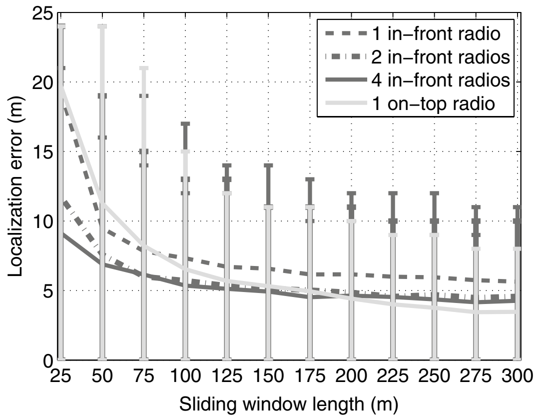

Fig. 13 Localization errors vs. the sliding window length under different radio configurations

6.2 Effect of sliding window and radio configuration

Figure 13 plots the average and 90% confidence interval of localization errors as a function of the sliding window length under different radio configurations. It can be seen that, as the window length increases, the localization errors decrease gradually and plateau when the window length is larger than 200 m. The reason is that the temporal stability goes better when long GSM power spectrogram segments are used for comparison. It can also been seen that adding more GSM radios can help improve the localization performance of UPS. From the figure, it can also be seen that the radio placement also accounts. For example, with the only one GSM radio placed on the roof of the vehicle, UPS can achieve even better localization performance than using a set of four radios placed in the vehicle when the window length is sufficient large.

6.3 Impact of urban environment

To study the impact of urban environment to the localization performance, we divide the trace collected on May 18th and 19th according to the four different scenarios, i.e., on two-lane roads (normal urban roads), on four- and eight-lane roads (major urban roads) and on roads that are under elevated roads (major urban roads with semi-open condition). For each scenario, we use the radio configuration with four radios placed in front of the vehicle and randomly select 1000 locations for test. For each location, we use the most-recent compound trajectory segment of 200 m to conduct online localization algorithm. In addition, we also use independent compound trajectory segments to localize three and five LRPs. With multiple LRPs, two aggregation schemes, i.e., normal average and selective average (the maximum and minimum estimates are discarded before the rest estimates are averaged), are adopted to retrieve the final localization results.

Figure 14 plots the average and the 90% confidence interval of localization errors under different urban scenarios with various aggregation schemes. It can be seen that, in all scenarios, UPS can achieve extraordinary localization performance. For example, the average localization error is less than 5 m. This implies that UPS is resilient to complex urban environments. It is also surprising to see that UPS can achieve best localization performance when it is on under elevated roads. The main reason is that the success of UPS relies on the wide availability and rich temporal–spatial features of GSM signals and therefore has less to do with the road type. In addition, using multiple LRPs and the selective aggregation scheme can effectively reduce the localization errors. Thus, we choose to use the selective average of five LRPs to estimate vehicle locations and compare UPS with GPS with this setting in the next experiment.

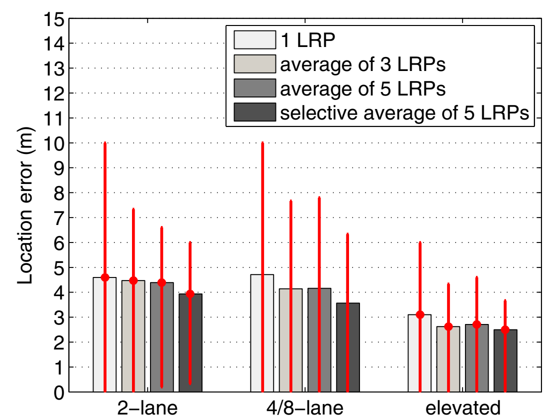

Fig. 14 Localization errors vs. various urban scenarios using variant aggregation schemes

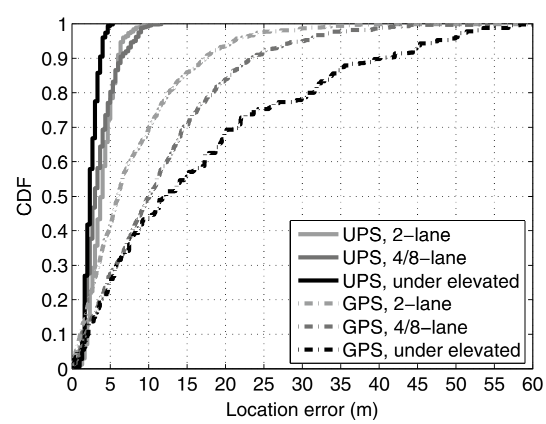

Fig. 15 CDFs of Localization errors of UPS and GPS under different urban scenarios

Table 1 Performance comparison between different localization schemes

		UPS			GPS			Fingerprinting
		2-lane	4/8-lane	Under elevated	2-lane	4/8-lane	Under elevated	Chen et al. (2006)
Localization error(m)	50%	4	3	2	6	10	12	94
	90%	6	6.5	3.5	17.5	23.5	40	291

6.4 Performance comparison

In this experiment, we compare UPS with GPS under urban settings. We divide the test trace and use the same radio configuration as described in the above experiment. For each scenario, we randomly select 1000 locations for test. For each location, the selective average over five LRPs is adopted.

Figure 15 plots the CDFs of localization errors and Table 1 shows the comparison between UPS, GPS and a system called POLS (Chen et al. 2006). It can be seen in Fig. 15 that UPS is more robust and stable in all kinds of urban scenarios compared with GPS. For example, when tested on 2-lane, 4/8-lane and under-elevated roads, UPS achieves low localization errors of 6, 6.5 and 3.5 m with a precision of 90%, respectively, whereas GPS achieves much higher localization errors of 17.5, 23.5 and 40 m, respectively. As a result, UPS can outperform GPS by five times on average. When compared with POLS, UPS has an absolute advantage although POLS needs a fine-grained fingerprinting by war driving. This is mainly because POLS only utilizes the RSS information of channels allocated in the current and neighboring cells and the information is limited to a single position.

6.5 Effect of traffic load

To verify that the pre-built map can be generally employed in dynamic urban environments with different traffic load, we divide the collected data into two groups according to whether the traffic is heavy or light while collecting. Then we match the trajectories with the map built under the similar, opposite and average traffic conditions. For each scenario, we use the radio configuration with two radios placed in front of the vehicle and randomly select 1,000 locations for test. For each location, we use the selective average of five LRPs algorithm with a window size of 200 m.

In Fig. 16, we show how localization error contributes when matched to maps built in different urban scenarios under various traffic conditions. As the contributions are only slightly vary from each other, we can simply neglect the impact of traffic on map construction. As a result, we can aggregate all collected data into one map and match to it regardless of the current traffic conditions.

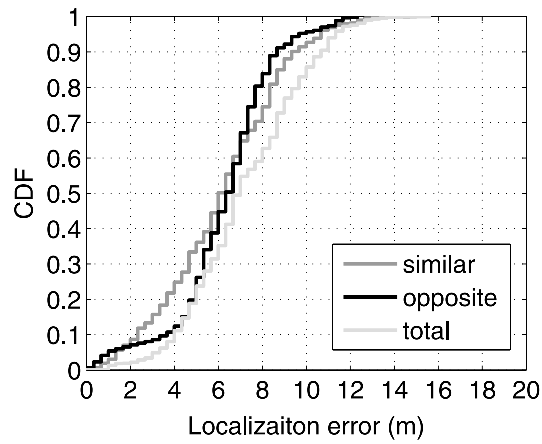

Fig. 16 CDFs of localization errors of UPS with different traffic loads

7 Discussion

In this section, we discuss some design issues of UPS encountered in practice.

Computational complexity To conduct the online vehicle localization, the most expensive step is to identify the optimal location in the map starting from which a GSM power spectrogram segment picked from the map has the maximum correlation coefficient value. Therefore, the algorithm complexity is bounded by the distance needed for search along a candidate road (can be identified with GPS). Given a search distance of k meters and a sliding window of n channels wide and m meters long, the computational cost for localization is $O(kmn)$. In our implementation, we consider a search distance of 50 m (larger than the maximum GPS errors) and use a sliding window of 41 channels wide and 200 m, respectively. We implemented UPS on a laptop with an Intel i7-2640M processor and measured the average processing time of our algorithm which is about 0.2 ms.

Map updating In UPS, a vehicle needs a pre-downloaded GSM map for online localization. The reliability of the map not only significantly affects the final localization accuracy but also decide whether using GSM power spectrogram for localization is feasible. For example, if the GSM map was transient and tended to vary along time, this would require edge servers to keep updating the map. In addition, this would also require a vehicle to frequently download the most up-to-date map before conducting online localization, which makes the scheme hard to use. With the temporal stability analysis on GSM power spectrogram, we have the observation that GSM power spectrogram has good temporal stability over a period of 2 weeks (restricted by the limited duration of our trace) and would not unlimitedly and monotonously become more and more unstable along the time (as shown in Fig. 2, the coefficient values calculated with an interval of 7 days is even smaller than that calculated with an interval of 15 days). In addition, in UPS, a GSM power vector measurement on the map is achieved by aggregation a large number RSSI readings collected from individual vehicles, which can greatly help achieve a reliable map by reducing the impact of noisy and random measures. In practice, an edge server can keep collecting individual measures and only calculate the aggregated results after a long period of time or a sufficient number of individual measures are collected. If the difference between the newly resolved results and the current aggregated measures is larger than a threshold, the map will be updated. This situation may happen when some GSM base stations have been rearranged or a new building has been constructed, etc. During the beginning stage of these changes, there would be a buffer period for the system to fit the changed environment, when the localization error may be temporarily large. However, the system will automatically adopt to the new environment by updating GSM maps.

8 Conclusion and future work

In the paper, we have observed that GSM power spectrogram has ideal temporal–spatial characteristics for fingerprinting. With this observation, we have proposed an online vehicle localization scheme, UPS. UPS can achieve stable localization performance under variant urban environments and at the same time needs a minimum hardware deployment. In addition, UPS can cost-effectively obtain a fine-granularity and reliable GSM power spectrogram map at a metropolitan scale. We have implemented a prototype system of UPS and verified the efficacy of UPS design through extensive field experiments. The experimental results show a promising performance of UPS that UPS can achieve a high localization accuracy of 5.3 m with a 90% precision.

In future, we will further improve UPS in the following two directions. First, we will further evaluate the effectiveness of UPS at a broader area such as in different cities and countries. Second, we will continue to investigate other frequency band such as FM and TV for accurate outdoor localization. Last, we will consider to use GSM power spectrogram for accurate indoor localization.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user location and tracking system. Proc. IEEE INFOCOM **2**, 775–784 (2000)

Biswas, P., Aghajan, H., Ye, Y.: Integration of angle of arrival information for multimodal sensor network localization using semidefinite programming. In: Proceedings of 39th Asilomar Conference on Signals, Systems and Computers (2005)

Chan, Y.T., Tsui, W.Y., So, H.C., Ching, P.C.: Time-of-arrival based Localization under NLOS Conditions. IEEE Trans. Veh. Technol. **55**(1), 17–24 (2006)

Chandrasekaran, G., Vu, T., Varshavsky, A., Gruteser, M., Martin, R.P., Yang, J., Chen, Y.: Tracking vehicular speed variations by warping mobile phone signal strengths. In: Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 213–221 (2011)

Chen, M.Y., Sohn, T., Chmelev, D., Haehnel, D., Hightower, J., Hughes, J., LaMarca, A., Potter, F., Smith, I., Varshavsky, A.: Practical metropolitan-scale positioning for GSM phones. In: Ubiquitous Computing (UbiComp), pp. 225–242. Springer, Berlin (2006)

Elnahrawy, E., Francisco, J.A., Martin, R.P.: Bayesian localization in wireless networks using angle of arrival. In: Proceedings of the 3rd ACM International Conference on Embedded Networked Sensor Systems, pp. 272–273 (2005)

Elnahrawy, E., Austen-Francisco, J., Martin, R.P.: Adding Angle of arrival modality to basic RSS location management techniques. In: Proceedings of the 2nd IEEE International Symposium on Wireless Pervasive Computing (ISWPC) (2007)

Google Maps for Mobile (2018). http://www.google.com/mobile/maps/. Accessed 26 Nov 2018

Gwon, Y., Jain, R.: Error characteristics and calibration-free techniques for wireless LAN-based location estimation. In: Proceedings of the 2nd ACM International workshop on Mobility Management and Wireless Access Protocols, pp. 2–9 (2004)

Han, H., Yu, J., Zhu, H., Chen, Y., Yang, J., Zhu, Y., Xue, G., Li, M.: SenSpeed: sensing driving conditions to estimate vehicle speed in urban environments. In: Proceedings of IEEE INFOCOM (2014)

Ibrahim, M., Youssef, M.: Cellsense: an accurate energy-efficient GSM positioning system. IEEE Trans. Veh. Technol. **61**(1), 286–296 (2012)

Kjærgaard, M.B., Bhattacharya, S., Blunck, H., Nurmi, P.: Energy-efficient Trajectory Tracking for Mobile Devices. In: Proceedings of the 9th ACM International Conference on Mobile Systems, Applications, and Services, pp. 307–320 (2011)

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., et al.: Place lab: device positioning using radio beacons in the wild. In: Pervasive Computing, pp. 116–133. Springer, Berlin (2005)

Li, M., Lu, Y.: Angle-of-arrival estimation for localization and communication in wireless networks. In: Proceedings of 16th European Signal Processing Conference (2008)

Lin, K., Kansal, A., Lymberopoulos, D., Zhao, F.: Energy-accuracy trade-off for continuous mobile device location. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services, pp. 285–298 (2010)

Osmocombb project. http://bb.osmocom.org/trac/. Accessed 26 Nov 2018

Otsason, V., Varshavsky, A., LaMarca, A., De Lara, E.: Accurate GSM Indoor Localization. In: Ubiquitous Computing (UbiComp), pp. 141–158. Springer, Berlin (2005)
Paek, J., Kim, J., Govindan, R.: Energy-efficient rate-adaptive GPS-based positioning for smartphones. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services, pp. 299–314 (2010)
Paek, J., Kim, K.H., Singh, J.P., Govindan, R.: Energy-efficient positioning for smartphones using cell-ID sequence matching. In: Proceedings of the 9th ACM International Conference on Mobile Systems, Applications, and Services, pp. 293–306 (2011)
Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket location-support system. In: Proceedings of the 6th ACM Annual International Conference on Mobile Computing and Networking, pp. 32–43 (2000)
Ramos, H.S., Zhang, T., Liu, J., Priyantha, N.B., Kansal, A.: Leap: a low energy assisted GPS for trajectory-based services. In: Proceedings of the 13th ACM International Conference on Ubiquitous Computing, pp. 335–344 (2011)
Skyhook. http://www.skyhookwireless.com/. Accessed 26 Nov 2018
Usman, A.U., Okereke, O.U., Omizegba, E.E.: Instantaneous GSM signal strength variation with weather and environmental factors. Am. J. Eng. Res. (AJER) **4**(3), 104–115 (2015)
Varshavsky, A., Chen, M.Y., de Lara, E., Froehlich, J., Haehnel, D., Hightower, J., LaMarca, A., Potter, F., Sohn, T., Tang, K., et al.: Are GSM phones the solution for localization? In: Proceedings of the 7th IEEE Workshop on Mobile Computing Systems and Applications, pp. 34–42 (2005)
Xiao, J., Wu, K., Yi, Y., Wang, L., Ni, L.M.: Pilot: passive device-free indoor localization using channel state information. In: Proceedings of the 33rd IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 236–245 (2013)
Youssef, M., Agrawala, A.: The Horus WLAN location determination system. In: Proceedings of the 3rd ACM International Conference on Mobile Systems, Applications, and Services, pp. 205–218 (2005)
Youssef, M., Youssef, A., Rieger, C., Shankar, U., Agrawala, A.: Pinpoint: an Asynchronous Time-based Location Determination System. In: Proceedings of the 4th ACM International Conference on Mobile Systems, Applications and Services, pp. 165–176 (2006)

Hongzi Zhu received the B.S. and M.S degrees from Jilin University, in 2001 and 2004, respectively, and the PhD degree in computer science from Shanghai Jiao Tong University, in 2009. He was a post-doctoral fellow in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, and the Department of Electrical and Computer Engineering, University of Waterloo, in 2009 and 2010, respectively. He is now an associate professor in the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include Internet of Things, vehicular networks, and mobile computing. He received the Best Paper Award from IEEE Globecom 2016 and Best Paper Award Runner-up from IEEE SECON 2018. He is an associate editor of IEEE Transactions on Vehicular Technology and Peer-to-Peer Networking and Applications and was a leading guest editor of IEEE Network. He is a member of CCF and CCF Technical Committee on Internet of Things. For more information, please visit http://lion.sjtu.edu.cn.

Fan Wu is a professor in the Department of Computer Science and Engineering, Shanghai Jiao Tong University. He received his B.S. in Computer Science from Nanjing University in 2004, and Ph.D. in Computer Science and Engineering from the State University of New York at Buffalo in 2009. He has visited the University of Illinois at Urbana-Champaign (UIUC) as a Post Doc Research Associate. His research interests include wireless networking and mobile computing, algorithmic game theory and its applications, and privacy preservation. He has published more than 170 peer-reviewed papers in technical journals and conference proceedings. He is a recipient of the first class prize for Natural Science Award of China Ministry of Education, NSFC Excellent Young Scholars Program, ACM China Rising Star Award, CCF-Tencent "Rhinoceros bird" Outstanding Award, CCF-Intel Young Faculty Researcher Program Award, Pujiang Scholar, and Tang Scholar. He has served as the chair of CCF YOCSEF Shanghai, on the editorial board of Elsevier Computer Communications, and as the member of technical program committees of more than 60 academic conferences. For more information, please visit http://www.cs.sjtu.edu.cn/~fwu/.

Siyuan Cao is currently a fourth year Ph.D. student in the Department of Computer Science at Purdue University, working in SIMBA Lab with Professor He Wang. She received her B.S. degree in 2015 from Shanghai Jiao Tong University. During Summer 2016 and Summer 2017, she worked as an intern at Microsoft Research Lab - Asia and Microsoft Research AI, respectively. Her research interests include mobile computing and mobile sensing.

Shan Chang received the Ph.D. degree in computer software and theory from the Xi´an Jiaotong University in 2013. From 2009 to 2010, she was a visiting scholar in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. She was also a visiting scholar with BBCR Research Lab, University of Waterloo, from 2010 to 2011. She is now an associate professor with the Department of Computer Science and Technology, Donghua University, Shanghai. Her research interests include security and privacy in mobile networks and sensor networks. She received the Best Paper Award from IEEE Globecom 2016 and Best Paper Award Runner-up from IEEE SECON 2018. She is a member of CCF.

Li Lu received his B.E. and M.S degrees in 2000 and 2003, respectively, all in Automation Control. He received PhD in the Key Lab of Information Security, Chinese Academy of Science, in 2007. He is a Professor with School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research interests include Battery-free systems and RFID technology, wireless network and network security. He is a member of CCF.