WBroximity: Mobile Participatory Sensing for WLAN- and Bluetooth-based Positioning

Farid Zaid, Diego Costantini, Parag S. Mogre, Andreas Reinhardt, Johannes Schmitt, Ralf Steinmetz
Multimedia Communications Lab, Technische Universität Darmstadt
Rundeturmstr. 10, 64283 Darmstadt, Germany
{fzaid, dcosta, pmogre, areinhardt, jschmitt, rst}@kom.tu-darmstadt.de

Abstract—Recently, there has been much interest in positioning based on the widespread WLAN technology, notably observed in the increasing number of hotspots and mobile devices equipped with WLAN interface. One technique to use WLAN for positioning is location fingerprinting, where WLAN networks in preselected sample locations are collected and used as fingerprints for those locations. However, to collect such fingerprints, existing services typically need to employ many skilled wardrivers who scan networks in the streets. This approach turns out to be very costly, especially when a large scale system coverage with acceptable positioning accuracy is required. Therefore, we propose WBroximity as a novel solution for the aforementioned concerns. With WBroximity, not only WLAN but also Bluetooth fingerprints are collected, therefore benefitting from the short range of Bluetooth to enable more precise positioning. Furthermore, such hybrid fingerprints are collected by using the paradigm of participatory sensing, thus cutting the extra costs needed to employ special personnel for this task, and allowing the system coverage to expand to wherever participants reach. In this paper, we present the technical details of realizing WBroximity as a location provider and its usage for collecting real fingerprint datasets. We evaluate the achieved accuracy in light of combining WLAN and Bluetooth, and the inherent aspects of participatory sensing, like number of participants and quality of participation. We give also an initial design and evaluation of a countermeasure to mitigate the effects of malicious participation.

Index Terms—Participatory sensing, mobile devices, indoor positioning, WLAN and Bluetooth.

I. INTRODUCTION

Nowadays, mobile devices that benefit from the Global Positioning System (GPS) capabilities are proliferating. Only in 2009, about 150 millions GPS-empowered units were shipped [1]. However, the fact that GPS performs poorly in indoor settings due to signal blockage and multipath effects disables these devices from benefiting from GPS in a range of indoor location-based services (e.g., navigation and object tracking). One promising technique for indoor positioning is to utilize the Wireless Local Area Network (WLAN) signals, because they show location-dependent characteristics. A WLAN-based technique has also economic advantages. WLAN installations are ubiquitous in indoor scenarios (office environments, as well as apartment blocks). Hence, no additional installation costs are needed. Additionally, the number of mobile devices with WLAN interface is ever increasing, allowing such devices to seamlessly benefit from a WLAN-based positioning system.

Performance evaluations of WLAN positioning [2] shown that a location fingerprinting scheme is most relevant for indoor environments because such scheme does not require a line-of-sight between the transmitter and the receiver as normally required by other schemes like Time of Arrival (TOA) [3]. Commonly, location fingerprinting models the positioning problem as a data classification problem. During the offline phase a radio map is built in the target environment by collecting the Received Signal Strength (RSS) measurements of the nearby access points at specific sampling locations. Location models are then built given the radio map using a supervised machine learning algorithm, e.g., a Decision Tree or Naïve Bayes [4]. In the online phase, RSS measurements are used to calculate the estimated location coordinates in real-time based on the models built in the offline phase.

II. PROBLEM STATEMENT

Unsurprisingly, collecting fingerprints turns out to be the most costly phase, especially when we are aiming at wide system coverage while maintaining acceptable levels of accuracy. For example, the commercial positioning service Skyhook [5] employs many drivers to scan WLAN hotspots by driving through streets in cities and towns. Scanning should be even done on a periodic basis to cater for the fact that WLAN hotspots can change over time, e.g., when some access points are shut down, relocated or replaced. Besides the incurred high costs, this approach has limitations when accuracy of indoor positioning is concerned. Firstly, as scanning is done from vehicles in the streets, real scanning of fingerprints inside buildings remains unapproachable, which in general affects the accuracy obtained indoors. Secondly, because WLAN signals are known to travel relatively long distances (up to 40m indoors [6]), precise localization requires gathering a large number of fingerprints and probably high-dimensional fingerprints (i.e. many networks per each fingerprint) which may not be affordable at all places.

In this paper, we propose WBroximity (short for WLAN and Bluetooth Proximity sensing) as a novel solution to address the aforementioned concerns.

• WBroximity combines the good of WLAN and Bluetooth technologies for location fingerprinting. Similarly to WLAN, Bluetooth is already integrated in modern mobile devices, and moreover in many consumer gadgets like game consoles, printers, and speakers [7]. However, in contrast to WLAN, Bluetooth provides a short-range wireless coverage (up to 10m for Bluetooth Class 3),...
which makes it relevant for fingerprinting locations indoors (up to the room level).

- **WBromixy uses participatory sensing** to collect location fingerprints. In participatory sensing [8], a user of a mobile device can be viewed as a mobile sensor node that gathers data and shares it among all users. For our specific purposes (i.e., positioning), users participate with the gathered fingerprints plus user-generated **fingerprint labels**. A fingerprint label, is a small piece of information indicating where the user was when a fingerprint was collected. The motivation for using a participatory sensing paradigm here is threefold:

  1. Cutting the costs. With participatory sensing we eliminate the extra costs that are otherwise needed to employ drivers to scan fingerprints. Normal users will do this job on a voluntary basis.
  2. Increasing the coverage. Because of the always-with always-on nature of mobile devices, fingerprints scanning can reach wherever the users go.
  3. Increasing the accuracy. The fact that the same location can be sampled by many users independently provides useful redundancy and a way to control the quality of data gathered by each user. Moreover, with participatory sensing, it is also possible to scan fingerprints directly inside buildings, as contrary to scanning them in the street only.

Besides proposing WBromixy as a solution for WLAN- and Bluetooth-based positioning, we propose the following contributions:

- We realize WBromixy as a location provider for Android mobile phones [9], allowing location-based applications to seamlessly access locations determined by WBromixy in the same manner GPS locations are accessed.
- We collect real world datasets and apply to them different machine learning algorithms to find the one providing the best positioning accuracy.
- We assess the inherent effects of participatory sensing on achieved accuracy, including both the impact of number of participants and quality of user participation.
- We give an initial design and evaluation of a countermeasure to mitigate the effects of bad user participation.

As an additional contribution, we made the WBromixy dataset publicly available online\(^1\), together with a video showing WBromixy in action at our laboratory.

The rest of this paper is organized as follows: Section III describes the technical details for realizing WBromixy. Section IV presents our evaluation methodologies and discusses the results. Section V highlights the major literature related to our approach. Finally, Section VI concludes the paper and gives an outlook on future work.

\(^1\)http://www.kom.tu-darmstadt.de/~fzaid/wbromixy.html

![Figure 1: WBromixy as a location provider for Android](image-url)

III. SYSTEM DESCRIPTION

One design goal when we realized WBromixy was to allow seamless binding to it as a location provider. Using Android as a development environment, custom location providers are easily plugged in by extending the `LocationProviderImpl` class. As shown in Fig. 1, analogous to the `GPS_PROVIDER` and `NETWORK_PROVIDER` already shipped in the Android SDK, the `WBROXIMITY_PROVIDER` offers the methods needed to enable, disable and query the status of the location provider. The specific functionality of the WBromixy provider is offered through two methods:

- The `getLocation()` method, which provides the latest location fix by scanning the WLAN and Bluetooth fingerprints and sending them for evaluation by the WBromixy server. The return value of the method is a `Location` object as shown in Fig. 1. This object holds the GPS longitude and latitude coordinates of the location fix, location accuracy, as well as a key-value map (accessed through the `getExtras()` method), which stores additional information specific to the location provider. Therefore, this map is a good place to store the fingerprint label. Because a WLAN or Bluetooth scan can take several seconds to finish, the `getLocation()` returns the result of evaluating the last scanned fingerprint. Therefore, the key-value map stores as well a timestamp to indicate the freshness of the returned value.
- The `setLocation()` method, which is specific to the `WBROXIMITY_PROVIDER` (not existing in the base location provider or other providers). This method stresses the fact that WBromixy is based on participatory sensing, where the users can provide location information too. The `Location` object passed to this method will typically contain the user-generated fingerprint label plus a GPS location that is either automatically or manually determined, as we will explain next.

To demonstrate all these functionalities, we implemented a basic frontend as shown in Fig. 2. The frontend features both textual and map views to display the current location of the user and to allow her to create new fingerprint labels. The server part of WBromixy is realized as a module in our
ContextFramework.KOM [10], which provides, among others, an evaluation service that seamlessly interfaces to the different machine learning algorithms provided by the Weka tool [4].

A. Collecting and labelling the Fingerprints

Collecting and labelling the fingerprints form the part of the system where users contribute to improve the overall positioning accuracy. Most WLAN and Bluetooth network cards can read at least the following information about networks in range:

- The Basic Service Set Identifier (BSSID), representing the name of the WLAN/Bluetooth network.
- The RSS, indicating the strength of the received signal.
- The Link Quality Indicator (LQI), indicating the quality of the received signal.

The current implementation of WBroximity allows fingerprints to be labelled with a symbolic location, i.e., a string representing the user’s location (e.g., office, kitchen, administration building, etc.). As shown in Fig. 2a, a symbolic label is created by manually entering the name of the location the user thinks she is at. However, to display on a map a location determined by WBroximity, GPS data (longitude and latitude), automatically obtained by the on-board GPS receiver or inserted manually by the user, can be appended to a symbolic label. For the demo frontend, we allow the user to manually pinpoint her current position on a map view as shown in Fig. 2b. The user can zoom in the map to identify a point more accurately. For our experiments, we allowed a zoom level higher than normally supported by the map view. This is done by displaying an indoor view as shown in Fig. 2c. This view is geo-referenced, meaning that the relative coordinates of points inside the view can be easily mapped to absolute GPS coordinates.

B. Building the Model and Classification

On the server, labelled fingerprints are used as input to a classifier to build the location model. As we are working with symbolic labels, we are applying classifiers that support discrete class attributes. Therefore, the positioning accuracy of the system equates to the classification accuracy of the used classifier, measured as the ratio of the correctly classified fingerprints. As a response to an unlabelled fingerprint, WBroximity will then return the label with highest classification accuracy (see Fig. 2d).

Because the location model is initially empty, the performance of the system is expected to be quite poor at the beginning. To deal with this slow-start problem, we bootstrap the model with a set of valid fingerprints. Such fingerprints can be obtained for many locations in the world using an online hotspot directory [11][12], which lists information about the geographic locations of the hotspots and nearby important points of interest.

IV. EVALUATION

For our evaluation, we used the WBroximity frontend to collect real world fingerprints in locations like our laboratory and the university campus and over a period of two weeks and different times of the day. As some of the WLAN networks are built up by several access points, we considered each access point as a single hotspot and used its MAC address as a BSSID. For example, throughout our building a total of 5 WLAN networks can be detected and are formed by 26 access points. Having collected the dataset, we applied to them data analysis techniques to study the applicability of different machine learning algorithms to the positioning task. We evaluated the performance trends in light of fingerprint characteristics and the impact of user participation.

A. Setup

1) Classifier Selection: The initial step in our analysis was to choose a classifier to recognize fingerprint patterns, and then to use such classifier as a baseline for the rest of the evaluation. For this purpose, we selected subsets of the dataset and applied to them these classifiers: J48, NB Tree, REP Tree, Naïve Bayes, Logistic, JRip, PART and Ridor2. As shown in Fig. 3, with an accuracy of at least 92%, Naive

2For a detailed description of the tested classifiers, the reader is referred to the documentation of the Weka tool [4].
Bluetooth networks are only visible at one location: thus, the system classifies correctly more often. The reason is that fingerprint per user per time unit. Obviously, with Bluetooth Fig. 4b shows the effect of injecting one static Bluetooth injected our real data with synthesized Bluetooth fingerprints. effect of integrating Bluetooth over time. For this purpose, we more often than stationary ones. Therefore, we examined the because mobile Bluetooth neighbourhoods are encountered play a role in distinguishing the fingerprints. wide area where networks are disjoint, then the BSSID can location has a characteristic LQI. However, if we consider a different networks disappears, normally it does not affect the other networks. Based on this analysis, we adopted NB for the rest of our evaluation.

2) Selecting Fingerprint Attributes: The kind of attributes (BSSID, LQI) included in the collected fingerprints relates directly to the storage and processing requirements of fingerprints. Therefore, besides selecting the best classifier, it is useful to decide at an early stage which attributes are critical for the positioning task and which are unnecessary, if any. To assess this aspect, we tried omitting different attributes and attribute combinations, and measured the obtained accuracy (again for NB). Fig. 4a stresses that LQI is essential for positioning. This is due to the fact that in locations which are close to each other, most often the same networks will be seen, so the BSSID alone does not provide much information. Interestingly, the BSSID can even be completely left out without affecting the accuracy, assuming that each sampled location has a characteristic LQI. However, if we consider a wide area where networks are disjoint, then the BSSID can play a role in distinguishing the fingerprints.

B. Integrating Bluetooth Information

Bluetooth fingerprints can exhibit temporal dependency because mobile Bluetooth neighbourhoods are encountered more often than stationary ones. Therefore, we examined the effect of integrating Bluetooth over time. For this purpose, we injected our real data with synthesized Bluetooth fingerprints. Fig. 4b shows the effect of injecting one static Bluetooth fingerprint per user per time unit. Obviously, with Bluetooth the system classifies correctly more often. The reason is that Bluetooth networks are only visible at one location: thus, detecting one of these networks is equivalent to detecting an individual location, like a room. While we made this evaluation using static Bluetooth networks, mobile networks can be included as well in the fingerprints without causing the overall positioning accuracy to degrade, as they will not be contributing any additional information.

C. Effects of Participatory Sensing

Our approach is affected by two inherent factors of participatory sensing: the number of participants and the quality of user participation. A participating user is a user who is contributing location fingerprints to the system, and not a user who is simply using the system for localization.

1) Effect of the Number of Participants: We measured the accuracy over time when different number of users are participating. As shown in Fig. 4c, after the same amount of time units, as expected, the system depicts higher accuracy for higher number of users. However, no matter what the number of users is, the system tends to learn exponentially, and converges towards a maximum accuracy (about 90%) after a while. This is actually an interesting observation, because it suggests that for a desired level of accuracy, we need to collect a sufficient number of fingerprints. For example, to reach the 90% accuracy at a specific location, it is enough to collect about 200 fingerprints. Although it looks irrelevant at a first glance if this amount of fingerprints comes from the same user or different users, it is beneficial in a participatory sensing scenario to have these fingerprints coming from different users to increase the trustworthiness of the gathered data.

2) Effect of the Quality of User Participation: As users have typically no access to tamper with the collected fingerprints, the only remaining factor affecting the quality of user participation is the user-generated label. In practice, different users may provide different labels for the same location, or even the same user may at different times provide different labels for the same location. In the worst case, malicious users may intentionally mislabel the fingerprints. To evaluate this aspect, we mimicked bogus participation by reassigning wrong labels to different number of fingerprints in our dataset. Fig. 4d shows that accuracy gradually degrades as the ratio of mislabelled fingerprints increases. However, the system still depicts an accuracy over 90% after introducing 20% wrong labels.

3) Mitigating the Effect of Mislabelled Fingerprints: As a countermeasure against wrong labels, we adapted WBroximity such that it restricts the range of symbolic labels a user can assign to fingerprints. As shown in Fig. 2d, WBroximity returns a list with the four most probable locations. The user can then select one of these labels only. Fig. 4d shows the accuracy when the countermeasure is being applied. A 90% accuracy is still achievable even when 80% of the fingerprints are mislabelled. However, this approach has the downside that it limits the freedom of honest participants to correct the system. Therefore, we initially provided some users with administrative privileges to extend the set of labels, if required.
Figure 4: Evaluation results applied to NB as a baseline classifier

D. WBroximity Location Fix: An Example

In this example we show how the getLocation of the WBROXIMITY_PROVIDER transforms symbolic labels into GPS geographic locations and how it computes the location accuracy. As we mentioned in Section III-A, whenever available, GPS coordinates are appended to the symbolic labels. We take the test case shown in Fig. 5a, where we pinpointed our actual location (marked by the green spot) on an accurate indoor map of our lab building. The map is also accurately geo-referenced so that GPS coordinates can be assigned to the locations inside the building. At this test location, 10 WLAN hotspots were detected, 3 of them are known to be operated within our building. With 220 fingerprints used to build the location model, the gray spots in the figure represent the locations of the 4 most likely labels \( \{l_1, l_2, l_3, l_4\} \) and the probability of each location. Therefore, the worst location estimate has simply the GPS coordinates of the smallest spot (i.e., with 9%), the best location estimate has the coordinates of the largest spot (i.e., with 28%), and the coordinates of the average estimate (marked by the red spot) is computed according the following formula:

\[
x_{avg} = \frac{1}{4} \sum_{i=1}^{4} (p_i \times x_i)
\]

where \( p_i \) is the probability for location \( l_i \), and \( x_i \) is the coordinate (longitude or latitude) of location \( l_i \). It is to be noticed here that the probabilities do not necessarily sum to 1.0, because only the 4 most likely labels are returned. We assume that labels with very low probability do not have big influence on the average estimate. The overall accuracy is then represented by the minimum circle enclosing the 4 most likely locations, indicated by the dotted red line (with radius \( \approx 17m \)). This accuracy is acceptable given the relatively low number of reference fingerprints used to cover the whole area of our lab. However, even with this low number, the possibility to scan fingerprints indoors will commonly give higher accuracy compared with approaches that depend on only in-street fingerprint scanning. For example, Fig. 5b
depicts the accuracy achieved by Skyhook WPS [5] at the same test location. We used the Skyhook SDK available for Android, this way we made sure that we used the same WLAN network card to collect the same fingerprints that we collected in the case of the WBroximity test. However, it is unclear from the SDK documentation which fingerprint attributes (i.e., RSS, LQI, etc.) are being included in the fingerprints. In this test, Skyhook reported a location with an accuracy of about 147m, which, although much better than the nominal ±750m accuracy officially claimed, hardly covers the actual location. This observation suggests that it can be beneficial to use our approach to extend the reach and accuracy of existing services depending solely on in-street network scanning.

V. RELATED WORK

Participatory sensing is a field that has been recently receiving serious attention. For example, NoiseTube [13] is a participative approach to measure noise pollution by turning mobile phones into noise sensors and automatically sharing the geo-localized measurements with the community. In the area of location-based services, CellSpotting [14] counts on users’ participation of detected GSM and UMTS cells, their geographic locations, and nearby touristic information. Closely related to our work is the Jiwire Wi-Fi Finder [15], offering a mobile application with a feature that allows users to submit newly discovered hotspots. However, in contrast to our approach, the aim of [15] is WLAN-based advertising and finding free and fee-based hotspots, therefore, accurate positioning is not a goal here.

In general, since participatory sensing is heavily based on voluntary contribution by users, designing incentives for participation remains a hot research topic. For example, some work proposed to take advantage of gaming to obtain useful labelled images [16]. [17] studied economic models and proposed an auction-based mechanism for commercial participatory sensing. A give-and-take scheme where users had a balance between requested and answered queries is proposed by [18], while [19] proposed an application where users can get information about cheapest prices for one product as long as they submit one themselves.

In the area of localization, fingerprinting is not the only technique but there are several other works, especially on triangulation/trilateration. We will present the most significant here. GPS systems cover already outdoors positioning within a few meters error rate [20], but different localization methods are needed where GPS signal is not available, for example indoors or in dense high rise urban environments (called “urban canyons”). Alternative methods, such as ultrasound [21] or infrared [22], require the deployment of specific devices, while others can rely on existing infrastructures, as GSM, WLAN and Bluetooth.

In [23], GSM-based trilateration techniques alone in a city environment yielded an accuracy of 100 – 200m, improved to 15 – 20m in conjunction with WLAN beacons. Such work relied on a Database filled with the absolute positions of beacons provided by institutions or war-drivers (like the already mentioned Skyhook). Although this is a feasible approach, there is no guarantee or incentive to provide such data. In [24] similar results using fingerprinting were achieved without the help of WLAN beacons, but the GSM stations had to be known in advance, and their RSS and locations collected and mapped.

Another fingerprinting technique used the 6-strongest GSM cells achieving 44m accuracy at best [25], while [26] reached around 5m by using the fingerprints from all the detected GSM cells in range. Unfortunately, they achieved this result with up to 29 additional cells, hardly available everywhere, thus deteriorating such accuracy.

Bluetooth fingerprinting has been used in [27] for indoor localization, but although good room level results have been achieved, such system alone can provide location information only for a limited area and requires an extensive deployment of fixed Bluetooth devices due to their limited range.

WLAN solutions are proved to reach up to meter level accuracy, especially through fingerprinting [28][29], but normally these approaches require initial (and periodical) calibration or several input information from expert users.

VI. CONCLUSIONS AND OUTLOOK

In this paper, we have presented a hybrid approach towards an accurate indoor positioning system based on WLAN and Bluetooth fingerprints. Participating users can contribute by providing labelled fingerprints taken using their mobile phones, and thus help in the creation of a full model capturing the real environment. Once the model is initialized, localization can be performed by classifying the current fingerprint received from a mobile device using a Naïve Bayes classifier. The system is resilient even when a subset of fingerprints have accidentally or maliciously been mislabelled, and only encounters a small drop in the classification accuracy. WBroximity was implemented as a location provider for Android mobile devices, and it was evaluated by collecting and analyzing real fingerprint datasets.

The next steps for improving WBroximity include GPS-based labelling, which provides a consistent way for labelling locations and thus a higher classification accuracy. Additionally, we are considering to apply a filtering of Bluetooth networks to find which Bluetooth devices have stationary locations.Finally, to motivate users to participate, we will be focusing on both optimizing the fingerprint scanning in terms of battery consumption, and designing incentive schemes for more and good participation.

ACKNOWLEDGMENTS

Parts of this research have been supported by the German Research Foundation (DFG) within the Research Training Group 1362 “Cooperative, adaptive and responsive monitoring in mixed mode environments”. Special thanks go as well to our students Adrian Loch and Pablo Lopez for implementing the WBroximity application and running the experiments at the TU Darmstadt.
Figure 5: (a) WBroximity worst, average and best position fixes (b) Performance of WBroximity vs. performance of Skyhook

REFERENCES


