A New Approach for Indoor Customer Tracking Based on a Single Wi-Fi Connection

Yuntian Brian Bai¹, Suqin Wu¹, Yongli Ren², Kevin Ong³, Guenther Retscher³, Allison Kealy⁴, Martin Tomko⁵, Mark Sanderson², Hongren Wu⁶, Kefei Zhang¹

¹ SPACE Centre, Sch. of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia
² Sch. of Computer Science and Information Technology, RMIT University, Melbourne, Australia
³ 3 Dept. of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria
⁴ 4 Dept. of Infrastructure Engineering, University of Melbourne, Melbourne, Australia
⁵ 5 Department of Geography, University of Zurich, Zurich, Switzerland
⁶ Sch. of Electrical and Computer Engineering, RMIT University, Melbourne, Australia

Abstract—Wi-Fi and smartphone based location tracking technologies have been widely applied in location based services (LBS). This research is a part of an Australian Research Council (ARC) project required by a large global shopping mall company located in Australia. It aims to develop an effective customer tracking approach for acquiring shopping behavior of customers and providing them better services. Currently, the log data provided by the company only recorded one Wi-Fi connection at a time for one smartphone user. This makes most of the conventional tracking and positioning methods inapplicable. In order to solve this problem, a new approach was developed, which takes advantage of both of the cell of origin (CoO) method and the fingerprinting method. Its implementation mainly includes two phases: 1) the cell determination phase, including partitioning of the floors of the shopping mall using a Voronoi diagram and determining a unique cell for each wireless access point (AP), adjusting the cells manually according to the real floor layout, obstructions as well as an on-site observing and calibrating process. An AP-shop/room matching table and other relevant information are stored in a database; and 2) the user location tracking phase, including determination of the cell that the customer locates in according to the detected AP linked with the customer’s mobile phone. A back-tracing method is also used to improve the tracking performance. This new approach was tested in a large shopping-mall-like space covering an area of 1000 square meters, containing 8 simulated shops and 11 APs. The results showed a 96% correct detection rate whilst the 4% erroneous detections occurred when a user was near the cell boundaries. This new approach is also easy to implement and integrate with other location tracking systems.

Keywords—LBS; Wi-Fi; smartphone; indoor positioning; single wireless access point connection

I. INTRODUCTION

Wi-Fi is a technology that allows an electronic device to exchange data or connect to the Internet wirelessly using microwave signals in the 2.4 GHz and 5.0 GHz bands. It has been widely used for communications and entertainments for more than ten years and has facilitated the growth of modern social media services. Wi-Fi provides services in private homes, businesses, as well as in public spaces. Organizations wishing to provide advanced services or for advertising purposes, such as in airports, hotels and shopping malls, often provide free use of Wi-Fi access to attract customers. In this case, Wi-Fi based passive tracking systems are used for tracking customers and collecting their information (e.g., location and shopping behavior information). In fact, research in this area is of highest significance as characterized by the advent of the Internet of Things [1]. Passive tracking technologies such as radio frequency (RF) based and vision-based tracking technologies have the advantage of user device independence. Vision-based technology uses cameras to track movement of people and is truly passive, as target users do not need to carry any equipment [2]. However, it requires a line-of-sight between the cameras and users, and still poses significant challenges due to complex background environments and multiple dynamic targets. RF-based technology is not limited by the line-of-sight condition as radio signals can penetrate walls and diffract around objects. Users can be passively identified by information that is unique to the user, such as the MAC address of the user’s Wi-Fi enabled smartphone.

In recent years, Wi-Fi has become a default feature of all smartphones. Consequently, society has become increasingly dependent on their phones as the main source of telecommunication, entertainment, and location-based services (LBS). Various sensors, such as Global Positioning System (GPS) receivers, accelerometers, gyroscopes, digital compasses, cameras, Wi-Fi, and Bluetooth, have been embedded in smartphones for these purposes. Driven by the rapidly increasing demand from the smartphone industry, wireless local area network (WLAN) technologies experienced significant development. According to Donovan’s report, Wi-Fi market revenues are forecasted to reach US $12 billion by 2017, a 57% increase over 2012 revenues [3]. For example, around 80 million passengers pass through a typical large hub airport each year and there are over 250 such hubs in the world [4]. Other records show that annual visitor traffic for a large shopping mall can reach 28–75 million [5, 6]. The efficient management and provision of personalized services for people through an airport or a shopping offers the potential to increase customers’ satisfaction and consequently generate more revenues.

Due to the complexity of most indoor environments, the integration of multiple positioning technologies and/or
algorithms could lead to a better solution for indoor location tracking. For example, RFID-based tracking systems have been successfully demonstrated in recent years. Smartphones as RFID readers have been implemented for both the iPhone and Android phones [7, 8].

Although the provision of personalized services to customers can improve business performance, it remains a challenge. Information that is directly relevant to a customer’s interests relies on reliable knowledge on their location and the environment context for this location. This level of information is currently not accessible in a robust and reliable manner, primarily due to the difficulties associated with location determination in indoor environments where GPS does not work. The widespread use of smart phones and the potential commercial and social benefits achievable from the delivery of a robust indoor positioning capability has driven the trend over the last decade towards the development of a broad range of indoor positioning technologies and techniques. Combined with the availability of nontraditional technologies and signals that can be used for positioning, as well as dramatic reductions in the cost and size of non-GNSS based augmentation sensors, the solutions available for robust positioning in difficult GNSS environments are evolving rapidly. In this paper, we focus on Wi-Fi based positioning, given they require no additional or specialized infrastructure for positioning.

There are four types of measurement models commonly used for Wi-Fi based indoor positioning [9, 10]: time of arrival (ToA), time difference of arrival (TDoA), angle of arrival (AoA) and received signal strength (RSS) based models. The RSS-based model is the most popular one. The three most commonly used location estimation methods for indoor positioning and location tracking are Cell of origin (CoO), trilateration and fingerprinting. CoO is less popular than trilateration and fingerprinting, and its accuracy depends on the means of implementation, where different implementation methods may lead to significantly different location accuracies. Trilateration usually requires at least three access points (AP) connections. Fingerprinting consists of two phases: an off-line training phase and a localization phase. Deterministic and probabilistic techniques are often used for its location estimation [11]. During the off-line training phase, a range of known points, the so-called reference points (RP), covering the area of interest are selected and the received signal strength indicator (RSSI) values are collected at each RP. These observation datasets are stored in a fingerprinting database (DB). In the localization phase, a set of RSSI observations are collected from the surrounding APs and the observations are then compared with those stored in the fingerprinting DB for identification of the best matched dataset in the DB. The location of the RP associated with the best matched dataset is taken as the user’s location estimate.

Like most shopping malls or infrastructure deployments, the Wi-Fi system used for this project was designed predominantly for wireless communications and entertainments rather than for tracking and location determination. For example, the Wi-Fi system used here recorded only one Wi-Fi hotspot (i.e., one AP) connection for a customer at a time in a log file. Our primary interest is to track customers’ shopping behavior based on the log data. However, most of the conventional localization models or methods are not applicable to the single connection scenario, a new approach therefore had to be developed as is presented in the following sections.

II. METHODOLOGY OF NEW APPROACH

A. Analysis of Collected Data

This research is part of the TRacking Indoor Information BEnavior (TRIBE) — a project funded by the Australian Research Council (ARC). A CISCO™ Wi-Fi system was installed in our industrial partner’s shopping mall located in Melbourne and a vast amount of log data from and opt in free Wi-Fi service (> 1 million records for a 6-month period) was provided. The log data includes a RSSI value, a smartphone ID of the user being tracked, an AP ID associated to the user’s smartphone with the strongest RSSI value received, the duration of the user connection to the AP and sequential connection and disconnection time. Further statistical analysis of the data showed that about half of the durations of AP connection are equal to or less than 5 minutes and 80.9% of the durations are less than 29 minutes as shown in Fig. 1.

![Fig. 1. Statistic of the duration of user Wi-Fi connections](image)

Due to the single Wi-Fi connection recorded, the only estimation method that can be used for this tracking is CoO. An AP covered cell, which is usually a polygon, is formed using the rule that the strength of the signals from this AP is stronger than those from the other surrounding APs. Fig. 2a demonstrates how the cells are formed for AP5 and AP6 in a free space. According to the design requirement of the CISCO™ Wi-Fi system, there was a 10-20% overlap between the coverage areas of two adjacent APs as shown in Fig. 2b.

![Fig. 2. (a) Formation of a cell for each AP; (b) Overlap between two adjacent cells](image)
Other features of the Wi-Fi system in relation to location tracking are listed below [12]:

- The default value for RSSI cutoff is -75 dBm.
- The default relative RSSI discard time is 3 minutes.
- The default absolute RSSI Discard Time is 60 minutes, which means that RSSI samples older than 60 minutes are not considered in location tracking.

In general, the greater the RSSI value received, the better the signal quality was. Table I shows the signal quality for various RSSI ranges. However, it should be noted that the RSSI values always vary between different phones, even the phones are of the same model and in the same environment. A simple test indicates this difference, in which three smartphones with the same distance to the same AP were used to collect RSSI values. The test results are displayed in Fig. 3.

### Table I. Relationship between RSSI range and Signal Quality

<table>
<thead>
<tr>
<th>RSSI range</th>
<th>Signal quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better than -40 dB</td>
<td>Exceptional</td>
</tr>
<tr>
<td>-40 dB to -55 dB</td>
<td>Very Good</td>
</tr>
<tr>
<td>-55 dB to -70 dB</td>
<td>Good</td>
</tr>
<tr>
<td>-70 dB to -80 dB</td>
<td>Marginal</td>
</tr>
<tr>
<td>-80 dB and beyond</td>
<td>Intermittent to No Operation</td>
</tr>
</tbody>
</table>

Fig. 3. Comparison of 33 RSSI values received from three smartphones at a 5-second sampling rate

### B. Location Estimation Methods

Given the constraint imposed for the Wi-Fi system used in this project of being tied to a single AP, trilateration and fingerprinting methods are not applicable. As previously mentioned, the only method that can be considered for this project is CoO. In fact, CoO is the easiest and simplest mechanism among all the location estimation methods. It does not require the implementation of complicated algorithms and thus its tracking speed is very fast. The word “cell” in CoO is associated with an AP in the Wi-Fi system involved. On the server side, the system tracks a mobile device connected to an AP and the location of the device can be considered within the cell associated with the AP. Almost all cell-based WLANs and other cellular-based RF systems can be easily and cost-effectively adapted to provide CoO location estimation capability.

However, CoO has the drawback of coarse granularity. For various reasons, mobile devices can be associated to cells that are not in close physical proximity. This coarse granularity can be especially challenging when attempting to resolve the actual location of a mobile device in a multi-story structure where there is considerable floor-to-floor cell overlap.

To improve location granularity over the CoO method, the highest signal strength technique can be applied when receiving cells provide RSSI for mobile devices. In this approach, the localization of the mobile device is performed based on the cell that detects the mobile device with the highest signal strength, which is also the only RSSI value recorded for each session of the Wi-Fi connection. The dual-phase idea of the fingerprinting method is also used as part of this approach in order to improve the accuracy of the cell boundaries. The combination of CoO and fingerprinting methods formed the theoretical base of this approach as discussed in the sections below.

### C. Floor Partition and Cell Determination

#### 1. Generation of Initial Cells

The core of the CoO method is the determination of the cells, i.e., dividing the area of interest into many small cells/polygons and making the cells optimum. Other researchers have studied various locational optimization problems where an optimal location of people and/or facilities is found in a continuum on a plane that can be solved by dividing the plane into a number of regions through the Voronoi diagram [13].

Conceptually, the generalization of the ordinary Voronoi diagram is straightforward. Let S be a space (e.g. the Euclidean space, denoted by \( \mathbb{R}^n \), as in the ordinary Voronoi diagram); \( s_i \in S, i = 1, 2, ..., n \), satisfying \( s_i \cap s_j = \emptyset, i \neq j \), called a generator (e.g., a point, a line, an area etc.), \( p \) be an arbitrary point in S; and \( d(p, s_i) \) be the distance between \( p \) and \( s_i \) (e.g., the Euclidean distance, the Manhattan distance etc.), then, a set \( V_i \) of points is defined by

\[
V_i = \{ p | d(p, s_i) \leq d(p, s_j); \quad p \in S \setminus \bigcup_{j=1}^{n} s_j, i \neq j, i = 1, \ldots, n \} \tag{1}
\]

If the resulting set \( \mathcal{V} = \{ V_1, ..., V_n \} \) is a tessellation of \( S \setminus \bigcup_{j=1}^{n} s_j \), then \( \mathcal{V} \) is called the generalized Voronoi diagram [9].

Fig. 4. A Voronoi diagram example for one level of the shopping mall (roughly 220*380 m² and 52 APs: the red spots)
Many application software packages such as Easy Contour™ and QGIS™ provide a function for automatically generating Voronoi diagrams. In this research, ArcMap™ was used to generate a Voronoi diagram based on the AP distribution for each level of the shopping mall involved in this research. Each cell (or polygon) of the diagram contains a unique AP. These cells can be called the initial cells. At this stage, the environmental affecting factors are not considered for the cell formation, the diagram will be adjusted later according to the real indoor environment. Fig. 4 shows the Voronoi diagram for one level of the shopping mall.

2. Cell Adjustment

The initial cells can be directly used for free space environments but can hardly be used for a complex indoor environment. A manual adjustment and an on-site calibration process are designed to determine the proximity of the cell boundaries. Several factors need to be taken into account during the manual adjustment, including but not limited to:

- The physical layout of the floor of the shopping mall, e.g., the architectural structure, location of goods and furniture.
- Real distribution of APs.
- The dynamic environment such as day-to-day mobility and relocation of goods and facilities, varying crowd densities at different times, etc.
- Building materials and appliances that may cause signal attenuation [14, 15] (see Table II).

### TABLE II. ATTENUATION WHEN SIGNALS ARE PROPAGATED THROUGH A WALL OR MULTIPLE WALLS [14]

<table>
<thead>
<tr>
<th>Attenuation properties of common building materials and appliances</th>
<th>2.4 GHz attenuation &amp; penetration loss (dBm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interior Office Door</td>
<td>4</td>
</tr>
<tr>
<td>Steel Fire/Exit Door 1.75&quot;</td>
<td>13</td>
</tr>
<tr>
<td>Brick 3.5&quot;</td>
<td>6</td>
</tr>
<tr>
<td>Concrete Wall 18&quot;</td>
<td>18</td>
</tr>
<tr>
<td>Glass Divider 0.5&quot;</td>
<td>12</td>
</tr>
<tr>
<td>Interior Hollow Wall 6&quot;</td>
<td>9</td>
</tr>
<tr>
<td>Interior Solid Wall 5&quot;</td>
<td>14</td>
</tr>
<tr>
<td>Common Floor</td>
<td>15</td>
</tr>
<tr>
<td>Refrigerator</td>
<td>19</td>
</tr>
</tbody>
</table>

Among these affecting factors, the physical layout and associated use of patterns is the most important factor being considered, e.g., it is unlikely to detect a user at some places/spaces where people are impossible to be. These “dead-spots” should be removed from the locations of interest. The walls, floors and other facilities affect the Wi-Fi signal propagation characteristics with transmission not conforming to an ideal isotropic sphere. A real Wi-Fi signal map captured from the commercial Ekahau Site Survey™ system demonstrates this characteristic in Fig. 5.

A new group of cells (see Fig. 6) is obtained after a manual adjustment based on the surrounding physical environment, but an on-site calibration process is still needed to make all the cells perfectly match the tracking requirement.

3. On-site Calibration Process

To evaluate the approach developed in this research and to avoid potential disturbances to the daily business operation of the shopping mall, a shopping-mall-like simulation test bed was set up in the Design Hub building of RMIT University for further testing and implementation of the new approach as shown in Fig. 7.

The next step of this new approach is the on-site calibration process for observing and improving the proximity of the cell boundaries. This process is critical and a large amount of work may be required. Probabilistic decision making was used in determining the cell boundaries, i.e., if less than 70% of the signals are received in a boundary-close area of a cell from its associated AP (from a 10-minute test at a 5-second sampling rate), the boundary will be reformed. For example, in Fig. 8a, the thick glass door is within the cell-A1007, the connection probability from AP1008 and AP1007 are 72% and 28% as tested within the cell of AP1007 in the area close to the door, so the cell was adjusted from Fig. 8a to Fig. 8b. Once the AP-cell pair is confirmed, the relationship between the AP and specific shops can be easily defined. Finally, an AP-shop matching table is generated and stored in the database for use in real-time tracking.

![Fig. 5. Map of Wi-Fi signal propagation (tested in the Indoor Positioning Lab, RMIT University)](image-url)

![Fig. 6. The Voronoi diagram after manually adjusted based on the surrounding physical environment](image-url)
D. Real-time Tracking

Once steps 1 and 2 have been completed, the real-time tracking process is straightforward. The AP with the strongest RSSI value received by a smartphone user is detected by the system and the location of the user is then considered to be located somewhere in the shops within the AP-associated cell.

A DB was developed for the test and implementation, which includes 5 main tables named User, UserGroup, AP, Shop and ShopGroup (see Fig. 9).

![Fig. 7. Testing environment for the TRIIBE project including 11 APs (red spots) and 30 main testing points (green spots)](image)

![Fig. 8. An example of cell adjustment based on on-site testing](image)

- **AP_NEIGHBOURHOOD**: maximum 3 levels of weighted neighbor APs are stored in this attributes (e.g., 1AP1010-1AP1008-2AP1011-2AP1007 for AP1009, the first digit represents the floor number of the AP).
- **CELL_CONN_HIST**: a cluster of records of a user’s location history (e.g., AP1011-AP1010-AP1009).
- **CELL_PROBABILITY**: probability of 4-29% of the neighbor AP detections within this cell (e.g., 28%AP1007-11%AP1009 for AP1008). Detections of less than 4% are omitted and if the detections are more than 29% then the boundary is required to be reformed.

An example describes the principle of the back-tracing method, as shown in Fig. 10, assuming the target user moves from position A approaching to B, the user may receive the strongest signal either from AP1003 or AP1009 based on fluctuations in the Wi-Fi signals. In this case, the system can check the CELL_CONN_HIST to see if the last cell associated to the user is AP1010 or AP1002. Once the system finds that the user is from AP1010 and also with the assistance of the AP_NEIGHBOURHOOD of AP1010, the cell then associated with AP1009 is determined as the user’s current location.

![Fig. 9. Part of the tables and attributes in the DB](image)

![Fig. 10. Example of the back-tracing method](image)

III. EXPERIMENTAL TEST

A. Cell Determination

The experimental test was conducted in the test bed (see Fig.7) established in the Design Hub Building of RMIT.
University. The test bed was a large shopping-mall-like space covering a floor area of 1000 m² and containing 8 simulated rooms/shops, 11 APs (red spots) and 30 test points (green spots). The initial Voronoi diagram (see Fig. 11a) was created first by the ArcMAPTM software and was then manually adjusted based on the floor layout (see Fig. 11b). After an on-site calibration test was conducted, a final unique AP-cell pair (see Fig. 11c) was identified for each AP. All the strongest RSSI values collected within a cell were considered from its associated AP.

![Initial Voronoi diagram](image1)

![Manually adjusted Voronoi diagram](image2)

![Final AP-cell pair after the on-site test and adjustment](image3)

Fig. 11. Determination of AP-cell pairs

According to the number of rooms in each cell, an AP-room/shop matching table (Fig. 12b) was defined. The matching table and other information such as the coordinates of APs and cells were then stored in the database for use in the location tracking phase. E.g., if a user is detected by AP1007, then the user can be considered in ShopS00005.

During the real-time tests, the relative RSSI discard time was reset to 10 seconds, which meant that a connection duration would be omitted if it was less than 10 seconds. The AP-room/shop matching table makes the location tracking simple and straightforward. As shown in Fig. 12b, a user’s location (i.e., room or shop) can be determined by the AP that transmits the strongest RSSI value to the user’s smartphone. The new approach was finally tested through the 30 TPs and 100 samples at each TP were collected. The matched and mismatched user detections are displayed in Fig. 13.

![Matching detections](image4)

**B. Location Tracking Result**

Part of the test result from the 8 shops is displayed in Fig. 13. There are 100 samples collected from each shop, the correct detecting rate is 96% and the other 4% erroneous detections all occurred when the user was near the boundaries of the cells/rooms (e.g., somewhere around TP23 and TP 30).

![Testing result](image5)

![Cell matching and location tracking](image6)

**Fig. 13. Testing result of the new approach**
IV. CONCLUSION AND FUTURE WORK

This research provides a potential solution based on Wi-Fi positioning techniques to indoor customer tracking for a large shopping mall. A new approach was developed and tested using a simulated shopping-mall-like indoor test bed. The new approach used a combination of the traditional Wi-Fi CoO and fingerprinting methods to address limitations of the individual approaches. Two adjustment processes were performed as part of this technique in order to comprehensively account for many of the affecting factors that limit robust cell determination. A back-tracing method was also used to improve the overall tracking performance. The test results obtained using 30 points with known locations is at the level of 96% correct detection rate.

One limitation of this approach is that in the current deployment, not every AP is associated with a unique room/shop and vice versa. In addition, the new approach has not been tested in a multi-story environment. Further work will be focus on these issues.

ACKNOWLEDGMENTS

This work was supported in part by the Australian Research Council (LP120200413).

REFERENCES