ISBN 978-3-200-02328-4

INDOOR PERSON LOCALIZATION SYSTEM THROUGH RSSI BLUETOOTH FINGERPRINTING

Héctor José Pérez Iglesias, Valentín Barral, Carlos J. Escudero

Department of Electronics and Systems, University of A Coruña, A Coruña, Spain
{hperez, valentin.barral, escudero}@udc.es

ABSTRACT

The growth of wireless and mobile communications technologies offers new possibilities for context driven information systems. Specifically, nowadays, mobile phones are equipped with several radio-frequency technologies, like Global System for Mobile Communications (GSM), WiFi or Bluetooth. In this way, the idea of using them to create a location system arises. Typical location algorithms can extract relevant information about Radio Frequency (RF) signals and estimate the position of the device. In this paper, we present a study of Bluetooth signal as source of information in one of these location systems. We analyze its capabilities and we create a set of algorithms to transform Bluetooth data in order to improve the location process. The system consists of N nodes connected to a cable network disposed in a scenario divided in cells, where each of them is composed of a thin client PC with a Bluetooth device and a directional antenna. It works by inquiring the Received Signal Strength Indicators (RSSI) values of all the visible Bluetooth devices. Employing a fingerprinting technique, it determines the most probable grid cell for each device.

1. INTRODUCTION

Location systems are widely used. Particularly in outdoor scenarios, there are a lot of applications making use of positioning data to bring services to the user. In these scenarios, Global Positioning System (GPS) uses to be the positioning data provider. However, indoor case is considerably different. The GPS signal is not be able to go through buildings walls, so is needed another solution in this situation. Signals like WiFi, Bluetooth, Radio-Frequency Identification (RFID) or Ultra Wide Band (UWB) are normally used as the base repository of information, whereas different location algorithms make use of them to obtain a estimation of the physical position. One of these typical algorithms is fingerprinting. This algorithm is based on the comparison of the signal detected by a device and a set of signal fingerprints created in a previous phase.

These kind of algorithms are very powerful, because they can be adapted to the signal nature and make itself more or less accurate by reducing or increasing the separation across the fingerprints. There are several implementations of this type of algorithm in the literature, like [1, 2]. As may be seen, fingerprinting algorithms can be used with any kind of signal, but its physical properties have a critical influence in the performance of the algorithm. There are some studies about this influence, like [3] with IEEE 802.11b. One of the technologies supporting fingerprinting is Bluetooth. Nowadays, the use of devices with Bluetooth capabilities is widespread, especially in the mobile phone world. Most of present-day mobile phones have Bluetooth support, and the idea of making a good positioning system using only this signal is very interesting. However, the nature of the Bluetooth signal (not designed for positioning purposes) makes implementations of algorithms of this kind not achieve very high accuracy. Problems arise due to the variability of the signal power, that makes the took measurements in a particular position not be stable enough to be successful compared to the stored fingerprints for this point. Thus, with Bluetooth, the fingerprint cells normally must be placed significantly distant in order to minimize the number of mistakes made. This affects the final accuracy of the whole system, that is directly proportional with the distance between these fingerprint positions.

This paper shows a new pre-processing algorithm to extract the relevant information and create a quality set of Bluetooth fingerprints. Additionally, an example of fingerprinting algorithms is shown, using a set of samples captured in a real scenario to build the fingerprints and test the different solutions. Finally, a way to measure the quality of different location algorithms is developed, in order to compare the different solutions.

This paper is structured as follows. Section 2 describes how the system works, explaining the calibration and the localization algorithm. Section 3 describes a test scenario, and subsequently, we present the performance of the system under this scenario in Section 4. Finally, Section 5 shows the conclusions of this work.

2. SYSTEM WORKING

There are two distinct phases in the operation of the system: Calibration of the system and Localization of the devices.

2.1. Calibration of the System

We assume, for any scenario, that we have deployed N directional antennas in the perimeter, trying to establish the same distance between each pair of contiguous antennas. Also, it is desirable that their position in height is the same and the direction of each antenna is pointing the center of the scenario. Thus, after having divided the stage into a grid $P \times Q$ of equal square cells, we will assign a number to each one.

Measurements

The system sampler are the RSSI values observed on the system antennas coming from the mobile devices. We summarize this measurement in a vector $[timestamp\ antenna\ cell\ MAC\ RSSI]$ where *timestamp* is the instant that sample was acquired, *antenna* and *cell* are identification numbers and *MAC* is the Media Access Control (MAC) address of the observed device.

Thus, we can proceed to realize the measures for each cell. That consists in getting a phone with Bluetooth, enabled and visible, and putting it on the cell during a minimum predetermined time T storing all measures. Note that in this process we will have the same MAC in all the measurements, because we are using only one mobile for the calibration.

Position Vectors Algorithm

At this point, there are a lot of samples (vectors of 5 dimensions), in general with different timestamps. Considering only one device and one cell, for reasons of simplicity, the aim is to obtain new N-vectors formed by $[RSSI_1\ RSSI_2\ \cdots\ RSSI_N]$.

For only one device and each cell k, we employ the Algorithm 1, where $S_{M\times 3}$ is a matrix composed by M rows contains all the measurements sorted by ascending timestamp with the following form $[timestamp\ antenna\ RSSI]$.

```
C_k ← {} /* Cluster for cell k */
for i ← N to M
  v ← [ ] /* dimension N */
  a' ← S_i2 /* antenna */
  v_a ← S_i3 /* RSSI */
  A ← {1, 2, ..., N} /* set of antennas */
  A ← A − {a'}
  for each A as a
    j ← i − 1
    continue ← true
    while continue do
      if (S_i1 − S_j1) > WINDOW then
        continue ← false
      else
        if a = S_j2 then
          v_a ← S_j3 /* RSSI */
          continue ← false
        else
          j ← j − 1
          continue ← j ≠ 0
        end if
      end if
    end while
    if is_null(v_a) then
      break
    end
  end for
  if every_component_is_not_null(v) then
    v ← v/||v||_2 /* normalization */
    C_k ← C_k ∪ {v}
  end if
end for
```

Algorithm 1: Position Vectors Algorithm

Given a device and a cell, for every S_i we search the most recent previous samples for each antenna which have a timestamp whose difference with the timestamp of S_i would be lower than the parameter WINDOW. If there are samples for every antenna we will obtain a vector with N RSSIs and we will add this vector to the cluster for the given cell k, i.e. C_k. Thus, we will have $R = P \cdot Q$ (number of cells in the grid) clusters, each one with different number of vector of dimension N.

Note that in Algorithm 1, the same measurement for a specific antenna can be employed to compose different vectors, and note that the resultant vectors are always normalized. This is important because every mobile device can obtain different signal power, also, the signal can suffer impairments depending on its exact position, i.e., the signal will be different depending where the user carry the device, for example, in his pocket or in his hand. The idea is to obtain the underlying relation between all antennas on each cell.

Statistical Information Retrieval

At this point we compute the histogram of RSSIs for every pair $(Cell,\ Antenna)$ obtained in the previous stage. We have to decide the number of intervals, with the same length, in order to determine the resolution of the histogram. Thus, we will obtain the mean and the variance, comparing betweens cells and antennas. Also, it is possible to apply statistical hypothesis test with different distributions, if we find a positive will make easier to obtain the localization of a device. However, the algorithms proposed in this paper use directly the histogram values as source of information for take their decisions.

Considering $L \equiv$ number of intervals, all this data is stored in $\mathbf{H}^i_{N\times L}$ matrices (with $i = 1\cdots\mathbf{R}$), where the rows are the antenna and columns are the interval. Every value in this matrices means the frequency of occurrence in an RSSI interval.

2.2. Localization of Devices

Principally, we can consider two kinds of algorithms to make the localization [4] of the devices: memoryless and sequential. The main difference is that sequential employs the previous estimated position in order to achieve the new position. In this paper we focus in a memoryless localization.

Obtaining several measures for a device during a certain time, we will obtain a set of position vectors employing the Algorithm 1. Note that they are normalized. After that, we locate for every vector the corresponding interval in the histogram for every cell and antenna, obtaining N values for each cell. This values represent the probability of obtaining these RSSI measurements for a specific antenna on a particular cell. Thus, at this point we have to find the most probably cell according with these probabilities. With this aim, in order to obtain a norm, we have decide to employ one of the following criteria:

1. **Addition** Make the addition of all probabilities of antennas for every cell.
2. **Square Addition** Make the addition of squares of all probabilities of antennas for every cell.
3. **Multiplication** Make the multiplication of all probabilities of antennas for every cell.

We summarize this in the Algorithm 2. Employing the criterion 1, where R is the number of cells, $\mathbf{D}_{W\times N}$ is the matrix with W position vectors of N antennas, and q_i is the value of the chosen cell for the position vector i.

```
for i ← 1 to W
  for j ← 1 to R
    P_ij ← 0
    for k ← 1 to N
    I ← find_interval_in(D_ik, H^j_k.)
    P_ij ← criterion_addition(P_ij, H^j_kI)
    end for
  end for
  q_i ← index_of_max(P_i.)
end for
```

Algorithm 2: Localization Algorithm

2.3. Improvement of Localization Algorithm

We can improve the algorithm described in Section 2.2 considering groups of G position vectors being adjacent in time. Thus, we have decided to employ the jointly information of the set following the next two criteria:

- Count
- Sum

For Count the idea is simple, we employ the Algorithm 2 for every position vector in the set obtaining a cell for each one. After, we choose the cell that have more occurrences in the set.

In other hand, employing the Sum criterion we use the Algorithm 2 for every position vector in the set obtaining a value for each cell depending on the chosen criterion to compute the norm. After that, we will make the addition of all values for each cell and we will choose the cell with highest value for this set.

3. TEST SCENARIO

We have develop some test in an area with dimensions of 6.6×7 meters. There are four cells centered at the points $(3.3, 0.5)$, $(3.3, 2.5)$, $(3.3, 4.5)$ and $(3.3, 6.5)$. There are also four nodes disposed at the corners, with their respective antennas which are orientated to the center of the scenario. Each node is composed by the following hardware:

- eBox-4300: VIA Eden Processor 500 MHz, 512 MB RAM
- Cambridge Silicon Radio, Ltd Bluetooth Dongle
- SMC SMCHMANT-6-EU Directional Home Antenna

Thus, we have the following parameters:

- $N = 4, P = 4, Q = 1, (R = P \cdot Q = 4)$

The scenario is a laboratory in the Facultad de Informática od Universidad da Coruã, Spain, which is equipped with a set of WiFi access points and other Bluetooth devices that can produce impairments in the signal captured by the antennas [5]. Also, the size of the room (too small), furniture and people can modify signal properties. All these issues make the scenario a good approximation to the usual real-world problems that happen in indoor environments. The calibration process was made by using a single mobile phone, that remained static during several minutes while antennas were detecting it. The fingerprint vector was created using this data, and later a new set of measurements was captured with the same device to test the fingerprinting location algorithm. In both cases, measurements were taken for 5 minutes on each cell.

4. RESULTS

Considering the scenario described in Section 3, we show in Table 1 the hit probability for the different criteria to ponder the probability of every antenna and the two methods to obtain the best cell with all the information of the group. These results are computed employing 60 position vectors to make the decision.

We can see, in this specific case, that the Count column achieves better results than Sum using Addition or Multiplication methods, while the Squares Addition method is not able to obtain a good result.

The Multiplication method normally gets better results when RSSI values are different enough in distinct cells. Using this method, the lower probability values in wrong cells "propagate" very easily, so those cells will not be chosen as frequently as the correct one. In contrast, the sum of probabilities across all the antennas does not penalize the wrong choices so much, because a low probability in a particular antenna can be balanced with a high probability in another one. But if RSSI values are very similar for each antenna in distinct cells, the Multiplication method will achieve a worse results. In this case the product of high values of probability will make wrong cells be chosen more frequently than if these values were only added.

	Count	Sum
1. Addition	93,48%	88,14%
2. Squares Addition	44,22%	57,39%
3. Multiplication	95,79%	52,14%

Table 1. Probability of success for Group Size of 60 vectors

Figure 1 shows the Probability of success versus group size (the number of adjacent position vectors used), for the three criteria employing the Count Method to use the whole group information. Figure 2 shows the equivalent for the Sum Method. Note that these curves have been obtained averaging the hit probability between the different cells.

In a similar way, Figures 3 and 4 show the Probability of success versus capture time (in seconds). This means that we employ all the previous vectors obtained from current instant to the current instant minus the chosen capture time. This is showed for the three same criteria employing the Count and Sum Methods. Again, curves represent average probability across all the cells.

The Count and Sum methods work differently according to the nature of the data. Thus, in a scenario where the measures are stable enough, and the Probability of success in a wrong cell has no very high values, the Sum method will achieve better results. This occurs because this criterion can detect the probability "trend", and even if in a specific observation the correct cell is not selected, its probability (that could be high) is reflected in the final decision.

However, if measurements are not stable, and success probability on a wrong cell fluctuates from low to very high values, Count method will be the best. This is because with the Count method the error is contained in a single observation, and even if its value is very high, it only affects to this specific measurement and not to the rest of them.

In the test shown in this paper, we can see how the Count criterion achieves better results than the Sum one. As we have said before, this happens because the Bluetooth signal is not very stable, so it can get high probability values sporadically in wrong cells that ruin the Sum criterion.

We can also see how with the Count criterion, the Product method achieves better results in less time than the Addition method. This happens because normally the probability values in wrong cells are very low, but sometimes they are very high for a specific one. Using the Count criterion these abnormal values are isolated into a unique iteration, and they do not affect the rest of measurements. With the Sum criterion the Addition method is better than the Product one, because in this case the higher abnormal values achieved by the Product method ruin the rest of the data, whereas using the Addition method the abnormal values have lower values and their impact in the final result is also smaller.

After many tests, in the Algorithm 1 a `WINDOW` parameter equal to 4 seconds was chosen. This parameter is related to the data capture frequency of the antennas, and is also related to the device capabilities. In same way, number of intervals taken into account to calculate probabilities was set to 60. This parameter is related to the number

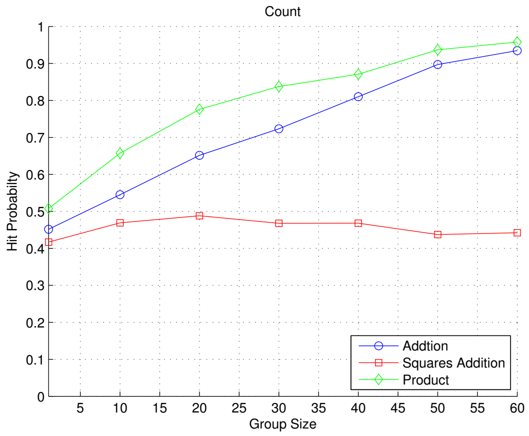

Fig. 1. Count: Probability of success vs Location Group Size

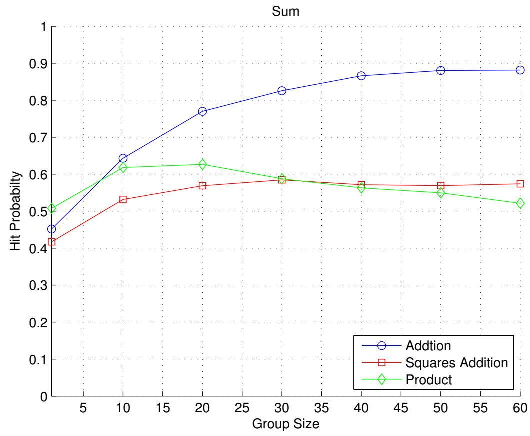

Fig. 2. Sum: Probability of success vs Location Group Size

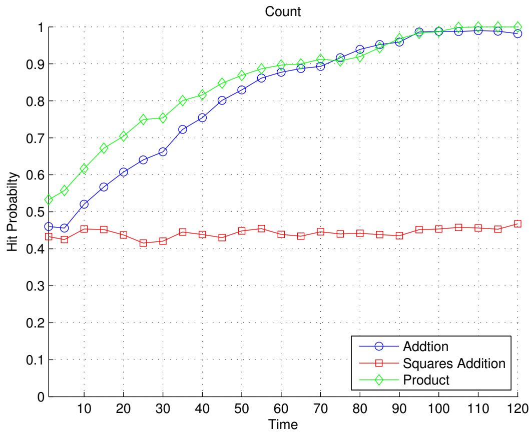

Fig. 3. Count: Probability of success vs Time

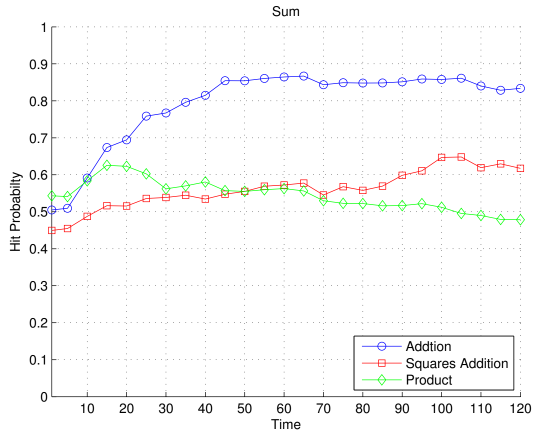

Fig. 4. Sum: Probability of success vs Time

of measurements captured. A high number of measurements implies that a high number of intervals can be used.

5. CONCLUSIONS

In the present paper we have described an algorithm to process Bluetooth data acquired by several antennas. The algorithm builds a set of measurements, grouping them according to closer timestamps, so that they can be used efficiently by a location system. Furthermore, we conducted a study about the Bluetooth signal and its capabilities as a source of information in a fingerprinting location algorithm. We also showed a set of methods to compare the quality of these fingerprinting location algorithms, and we proposed an algorithm of this kind.

Future lines of work include the improvement of the fingerprinting location algorithm and test in different scenarios. Another important future line of work is related to taking measurements of moving devices, and not only static, also, taking into account the previous chosen cells to decide which is the best option for current cell. This can be a more realistic scenario, keeping in mind a possible implementation in a real system. Another important improvement will be the construction of multidimensional histograms, considering in each interval the RSSI of the different antennas.

6. ACKNOWLEDGEMENT

The authors acknowledge the support of Ministerio de Industria, Turismo y Comercio with FEDER funds of the European Union (project ref. IPT-020000-2010-35).

7. REFERENCES

[1] P. Prasithsangaree, P. Krishnamurthy, and P. Chrysanthis, "On indoor position location with wireless lans," in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*. IEEE, 2002, vol. 2, pp. 720–724.

[2] W.M. Yeung and J.K. Ng, "An enhanced wireless lan positioning algorithm based on the fingerprint approach," in *TENCON 2006. 2006 IEEE Region 10 Conference*. IEEE, 2006, pp. 1–4.

[3] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*. IEEE, 2004, pp. 14–23.

[4] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[5] N. Golmie, RE Van Dyck, and A. Soltanian, "Interference of bluetooth and ieee 802.11: simulation modeling and performance evaluation," in *Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2001, pp. 11–18.