Quest Journals
Journal of Electronics and Communication Engineering Research
ISSN:2321-5941 Volume1 ~ Issue 3 (2013) pp: 23-28
www.questjournals.org

Research Paper

Secured and Robust Wi-Fi Shotgun Reads Based Mobility Mapping In Indoor User Tracking

[1]Chitra Sudhakaran, [2]Mr. Franklin George Jobin M.E(PhD),

[1]*M.E Applied Electronics, Anna University, Chennai.*
[2]*Assistant professor(ECE Dept.), Anna University, Chennai*

Received 20 November, 2013; **A**ccepted 05 December, 2013 © The author(s) 2013. Published with open access at **www.questjournal.org**

ABSTRACT : *The user mobility tracking plays an important role in identifying human activities and providing future user centered location based services (LBS's).Estimating the geographical position of mobile device such as smart phone in an indoor environment is not easy without the use of specific infrastructures. But in this paper, we present a secured and robust adaptive mobility map construction scheme for large scale which does not require any offline fingerprinting efforts.*

KEYWORDS: *Mobility tracking, WI-Fi, shotgun reads, spectral clustering, Mobility map*

I. INTRODUCTION

The location information promises to provide attractive services in ubiquitous computing environments. In an outdoor environment, Global Positioning Systems (GPS) [1] provide precise location of mobile devices with worldwide coverage.GPS, however has a shadow problem and is not available in indoor environment. In indoor environments, radio technologies such as Wi-Fi and cellular signals are used as observations for location estimates and provide a wide coverage especially in urban environments.

The widespread availability of wireless networks (Wi-Fi) has created an increased interest in harnessing them for other purposes, such as localizing mobile devices. While outdoor positioning has been well received by the public, its indoor counterpart has been mostly limited to private use due to its higher costs and complexity for setting up the proper environment. In this paper, we use local Wi-Fi network to localize a mobile user in an indoor environment. Wi-Fi (or 802.11 networking) works on the basic principle that data packets are sent using radio waves. These radio waves can be received by any compatible receiver placed in a pc, mobile phone, tablet pc or any other circuit. Through Wi-Fi, one may be able to track objects or people in real time, while adapting to changes in both the environment, and the Wi-Fi network, in a reliable manner. There has long been interest in the ability to determine the physical location of a device given only Wi-Fi signal strength.

The data distribution may vary based on changes in temperature and humidity, as well as the position of moving obstacles, such as people walking throughout the building. This uncertainty makes it difficult to generate accurate estimates of signal strength measurements. The Received Signal Strength (RSS) values measured by most radio transceivers can be used to estimate the distance between nodes and implement range-based localization schemes. Received signal strength indicator (RSSI) is a measurement of the power present in a received radio signal. RSSI is an indication of the power level being received by the antenna. Therefore, the higher the RSSI number (or less negative in some devices), the stronger the signal. The Received Signal Strength (RSS) values measured by most radio transceivers can be used to estimate the distance between nodes and implement range-based localization schemes. These schemes are popular because no additional hardware is required on the nodes to localize.

The Transmitter of the signals is known as an Access point. In computer networking, a wireless access point (AP) is a device that allows wireless devices to connect to a wired network using Wi-Fi, or related standards. The AP usually connects to a router (via a wired network) if it's a standalone device, or is part of a router itself.

One successful approach for indoor user tracking is a Wi-Fi based fingerprint [2],[3].But it involve some cumbersome work of fingerprint landmark calibration. Previous works in mobility tracking most often rely on the integrations of assisted GPS (A-GPS), Cell-ID, Bluetooth and Wi-Fi with the use of fingerprints based technologies for indoor [4]-[6].In contrast to our previous tracking systems which are based on GPS measurements and geometric clustering, the adaptive mobility mapping is designed to adaptively construct a mobility map of environment using randomly selected and unlabelled sequences of Wi-Fi received signal strength (RSS).

The RSS sample vectors are recorded by many individuals moving around the environment as they conduct their daily activities. Each signal typically covers only a small part of the coverage area and the idea is to piece their sequences together by treating them like shotgun reads. Shotgun read is a term used in DNA sequencing [7].

The output in DNA sequencing is a set of linear sequences of genomes. The output in the present system is a directed and weighted graph .This graph is called the mobility map. It abstracts the environment under coverage into a finite set of unlabelled location point(LP's).After the map is constructed,labelling of location can be done. Although the user can label the place on the map, simple automatic methods can be used to construct semantically meaningful place names from the observed data.

The LBS market has grown significantly in the past decade and is expected to reach a size of $12.7 billion in 2014.In addition to where the user is currently located, it is useful for the system to have knowledge about where the user is expected to go to, what activity the user will do next etc.The location information of a user promises to provide attractive services in ubiquitous computing environments. Over the years, diverse user tracking systems have been developed. In an outdoor environment, Global Positioning Systems (GPS) provide precise locations of mobile devices with world-wide coverage.GPS, however has a shadow problem and is not available in indoor environment. In indoor environments, radio technologies such as Wi-Fi and cellular signals are used as observations for location estimates and provide a wide coverage, especially in urban areas. One successful approach for indoor user tracking is a Wi-Fi fingerprint based method. The technique builds a radio map by measuring the Wi-Fi signals at each reachable calibration point, a priory and tracks a mobile device based on run-time observation of the Wi-Fi signals.

II. SYSTEM DESCRIPTION

The aim of the work is to sporadically collect sequences of Wi-Fi RSS sample vector to create a mobility map. The map is G= (V_c, E_φ) where C€V_c represents a LP.A LP is formed by clustering together of RSS sample vectors with large similarities. The weight of each φ € E_φ represents the transition matrix between two neighboring LP's.During mobility tracking, the location of a user can be determined by filtering the real time RSS trajectory of user into this mobility map.

The RF Signal Tracker is an engineering application for doing impromptu hand-held drive-tests with our Android phone. We can monitor the RF and WiFi signal strength for the device as well as the serving cell locations and hotspots, describe a cell site's zone of coverage, identify changes in technology and handover points, and save and playback that data. While many of the phone stats in the app can be displayed on the phone already (go to Settings -> About -> Status to see them). The advantage of the app is you can then map, record, and analyze, and share that data in a meaningful way.

Wi-Fi signal measuring tool is shown below:

Fig1: WiEye

a. The process for mobility map:

The process for mobility map construction consists of 3 steps:-

A) Shotgun reads collection

B) Spectral clustering on shotgun reads

C) Estimation of first order Markhov transition probabilities.

*Corresponding Author: Chitra Sudhakaran
M.E Applied Electronics, Anna University, Chennai

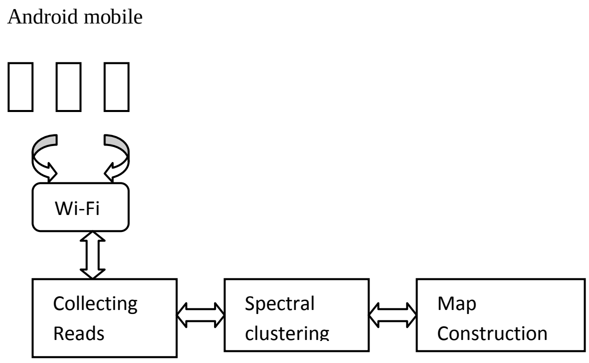

Figure2: Map construction scheme

2.1.1 Shotgun reads collection:

In this scheme, a subset of user population is equipped with an Android phone which will start recording RSS measurements at a regular time interval of 1 second by default, when the user is on the move. At this stage, starting and ending of recording is manually triggered. The phone is equipped by a v1.4 pedometer to detect movement.

Each measurement is a sample vector that contains the measured RSS's from a set of Wi-Fi AP's.Any sequence i.e.) too long is fragmented into shorter sequences of less than 600 samples to limit the computation time. Every resulting sequence is a shotgun read. Once a sufficient number of shotgun reads have been collected, then the clustering and mapping procedures are done.

Assume N raw RSS shotgun reads have been collected. The lth shotgun read is represented as $R^l = \{\mu_1^l \ldots \mu^l_{N^l}\}$ where μ_i^l is the collected RSS sample vector and N^l is the number of sample vectors in the read.
Assume that the total number of Wi-Fi AP's in the entire coverage is M. Each raw RSS sample is treated as M dimensional vector. Because M is large, direct clustering of raw RSS sample is difficult. Since RSS measurements are subjected to random variations, clustering based on RSS values alone leads to unstable results.

2.1.2 Spectral Clustering on shotgun reads:

In recent years, spectral clustering has become one of the most popular modern clustering algorithms. It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm. Clustering is one of the most widely used techniques for exploratory data analysis, with applications ranging from statistics, computer science, biology to social sciences or psychology. In virtually every scientific field dealing with empirical data, people attempted to get a first impression on their data by trying to identify groups of 'similar behavior' in their data.

Given a set of data points $x_1, \ldots x_n$ and some notion of similarity $s_{ij} \geq 0$ between all pairs of data points x_i and x_j, the intuitive goal of clustering is to divide the data points into several groups such that the points in the same group are similar and points in different groups are dissimilar to each other, if we do not have more information than similarities between data points. If we do not have more information than similarities between data points, a nice way of representing the data is in form of the similarity graph G = (V, E). Each vertex v_i in this graph represents a data point x_i. Two vertices are connected if the similarity s_{ij} between the corresponding data points x_i and x_j is positive or larger than a certain threshold, and the edge is weighted by s_{ij} . The problem of clustering can now be reformulated using the similarity graph: we want to find a partition of the graph such that the edges between different groups have very low weights (which means that points in different clusters are dissimilar from each other) and the edges within a group have high weights (which means that points within the same cluster are similar to each other).
Two steps are involved in spectral clustering-
-Use of low dimensional space to map each RSS sample vectors
-Use of K-means clustering in low dimensional space to congregate similar vectors into clusters.

Step 1: Objective Function Minimization:

For each shotgun read, we construct a weighted graph $\psi^l = (V^l, E^l)$.The weight of an edge represents the similarity between two nearby vertices. The simplest way to represent the RSS data in a low dimensional space is to map the weighted graph ψ^l onto a line while ensuring that any vertices which are nearby will have corresponding mapping points.

M.E Applied Electronics, Anna University, Chennai

Within R^l similarity between the i-th and j-th sample vectors can be calculated by $S^l_{ij}=e^{-[\alpha_R F_R(\mu_i^l,\mu_j^l)+\alpha_T F_T(T_i^l,T_j^l)]}$ where α_R and α_T are tunable weighting coefficients for the RSS and for timestamps. $F_R(\mu_i^l, \mu_j^l)$ and $F_T(T_i^l,T_j^l)$ are the features of RSS and of timestamp respectively.They are calculated as:

$$F_R(\mu_i^l,\mu_j^l) = \frac{\left\| \mu_i^{\,l}, \mu_j^{\,l} \right\|}{\max\limits_{s,t} \left\| \mu_i^{\,l}, \mu_j^{\,l} \right\|} \tag{1}$$

$$F_T(T_i^l,T_j^l) = \frac{\left| T_i^{\,l}, T_j^{\,l} \right|}{(N^l - 1)\delta} \tag{2}$$

The features are symmetric and hence so are the similarities: $S_{ij}^{\,l} = S_{ji}^{\,l}$ ($S_{ij}^{\,l} \geq 0$).

$\sum^{N^\wedge l} (r_i^{\,l} - r_j^{\,l})^2 S_{ij}^{\,l}$ We minimize the objective function $F(r^l) = \sum^{N^l} (r_i^{\,l} - r_j^{\,l})^2 S_{ij}^{\,l}$ by choosing the optimal set $\check{r}^l = (\check{r}_1^{\,l}, \ldots \check{r}^l_N{}^l)^T$ such that:

$$\check{r}^l = \text{argmin}\{F(r^l)\} = \underset{i,j=1}{\arg\min} \left\{ \sum^{N^l} (r_i^{\,l} - r_j^{\,l})^2 S_{ij}^{\,l} \right\} \tag{3}$$

Step 2:

K-means clustering is a method of vector quantization originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

The most common algorithm uses an iterative refinement technique. Due to its ubiquity it is often called the k-means algorithm; it is also referred to as Lloyd's algorithm, particularly in the computer science community. The algorithm is often presented as assigning objects to the nearest cluster by distance. This is slightly inaccurate: the algorithm aims at minimizing the WCSS objective, and thus assigns by "least sum of squares". Using a different distance function other than (squared) Euclidean distance may stop the algorithm from converging. It is correct that the smallest Euclidean distance yields the smallest squared Euclidean distance and thus also yields the smallest sum of squares. Various modifications of k-means such as spherical k-means and k-medoids have been proposed to allow using other distance measures.

K-means uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters:

The first phase uses batch updates, where each iteration consists of reassigning points to their nearest cluster centroid, all at once, followed by recalculation of cluster centroids. This phase occasionally does not converge to solution that is a local minimum, that is, a partition of the data where moving any single point to a different cluster increases the total sum of distances. This is more likely for small data sets. The batch phase is fast, but potentially only approximates a solution as a starting point for the second phase.

The second phase uses online updates, where points are individually reassigned if doing so will reduce the sum of distances, and cluster centroids are recomputed after each reassignment. Each iteration during the second phase consists of one pass though all the points. The second phase will converge to a local minimum, although there may be other local minima with lower total sum of distances. The problem of finding the global minimum can only be solved in general by an exhaustive (or clever, or lucky) choice of starting points.

Basic mean shift clustering algorithms maintain a set of data points the same size as the input data set. Initially, this set is copied from the input set. Then this set is iteratively replaced by the mean of those points in the set that are within a given distance of that point. By contrast, k-means restricts this updated set to k points usually much less than the number of points in the input data set, and replaces each point in this set by the mean of all points in the input set that are closer to that point than any other (e.g. within the Voronoi partition of each updating point). A mean shift algorithm that is similar then to k-means, called likelihood mean shift, replaces the set of points undergoing replacement by the mean of all points in the input set that are within a given distance of

M.E Applied Electronics, Anna University, Chennai

the changing set. One of the advantages of mean shift over k-means is that there is no need to choose the number of clusters, because mean shift is likely to find only a few clusters if indeed only a small number exist. However, mean shift can be much slower than k-means, and still requires selection of a bandwidth parameter. Mean shift has soft variants much as k-means does.

We run the K-means clustering on each read $\acute{R}^{l} = \{C_1^{l}, \ldots C_l^{\varphi l}\}$ where C_r^{l} and φ^{l} are the r-th RSS cluster and the number of clusters in R^{l}.Since the clustering for each read is based on the similarity of both RSS and relative timestamp values, the clustering outcome is more deterministic and stable than using RSS values alone as random variations in RSS measurements are filtered out.

2.1.3 First order Markhov transition probabilities:

A Markov chain is a stochastic process with the Markov property on a finite or countable state space. The term "Markov chain" refers to the sequence (or chain) of states such a process moves through. Usually a Markov chain is defined for a discrete set of times (i.e., a discrete-time Markov chain) although some authors use the same terminology to refer to a continuous-time Markov chain. The changes of state of the system are called transitions, and the probabilities associated with various state changes are called transition probabilities. The process is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state (or initial distribution) across the state space. By convention, we assume all possible states and transitions have been included in the definition of the process, so there is always a next state, and the process does not terminate.

A discrete-time random process involves a system which is in a certain state at each step, with the state changing randomly between steps. The steps are often thought of as moments in time, but they can equally well refer to physical distance or any other discrete measurement. Formally, the steps are the integers or natural numbers, and the random process is a mapping of these to states. The Markov property states that the conditional probability distribution for the system at the next step (and in fact at all future steps) depends only on the current state of the system, and not additionally on the state of the system at previous steps.

Since the system changes randomly, it is generally impossible to predict with certainty the state of a Markov chain at a given point in the future. However, the statistical properties of the system's future can be predicted.

We order the LP's formed by reads in chronological order $\{L_1 \ldots L_N^{L}\}$ based on mean relative timestamp of the LP's, where N^{L} is the number of LP's.The transition probabilities among LP's are calculated by the first order Markhov model. The transition probability matrix is created such that

$$\varphi_{uv} = \frac{\textbf{Number of transitions from } \mathbf{L_u} \textbf{ to } \mathbf{L_v}}{\textbf{Number of transitions from } \mathbf{L_u}}$$

where φ_{uv} denotes the transition probability from L_u to L_v.

2.2 Naming policy:

Labeling the place helps the user recognize the current place. Although the user can manually assign a name to the place after the map is built, our system supports automatic extraction of semantically meaningful name of the place. Initially, and we look for the strongest access points from the current location. The access point, which emits a strong radio signal, is simply assumed as being installed in the current location. We can assume unlabelled places with the name of the nearest access point where the access point has strong signal strength over a threshold -56dBm in our experiment.

With this method, we extracted meaningful words from our data set:''restaurant",''school gate" and so on. Many access points were operating in their factory configuration, hence sophisticated filters are required to prevent assigning meaningless names.

III. RESULTS

We conducted experiments in our college campus. The RSS data is recorded by the Android phone at a scanning rate of 1 sample/second. The similarity between the shotgun reads is shown by a cluster gram as shown in figure 4 and the corresponding connecting graph of vectors is shown in figure 5.Then the clustering of sample vectors is done using K-means clustering algorithm as shown in figure 6.

*Corresponding Author: Chitra Sudhakaran
M.E Applied Electronics, Anna University, Chennai

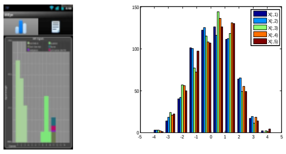

Figure 3: Shotgun read collection

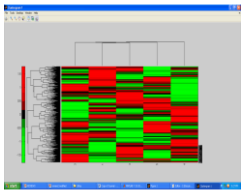

Figure 4: Cluster gram-Similarity between Vectors

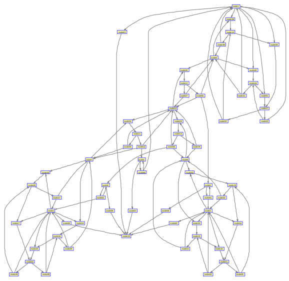

Figure 5: Graph construction

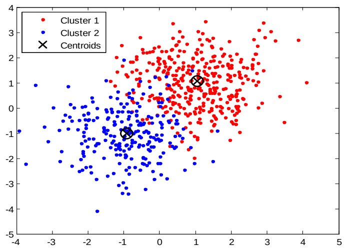

Figure 6: Clustering between sample vectors

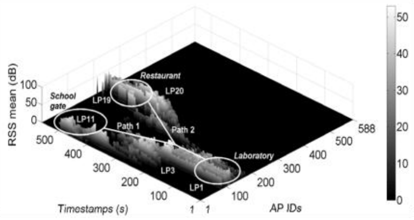

Figure 7: The construction of adaptive mobility map.

IV. CONCLUSION

We described an approach to use ubiquitously and randomly recorded Wi-Fi RSS data to construct an adaptive mobility map for mobility tracking. With the use of unlabelled Wi-Fi shotgun reads, our approach can avoid the cumbersome work of collecting labeled data for fingerprinting. The mobility map created is a graph consisting of LP's, which are formed automatically through the clustering process. The mobility map will provide guidance to people tracking during the online phase.

REFERENCES

[1]. D.Ashbrook and T.Starner,"Using GPS to learn significant locations and predict movement across multiple users", Personnel and Ubiquitous Computing, vol.7, no.5, pp.275-286, Oct.2003.

[2]. M.Zhou, P.Krishnamurthy, Y.B.Xu and L.Ma,"Physical distance vs. signal distance: an analysis towards better location fingerprinting," in Proc.2011 IEEE HPCC, pp.977-982.

[3]. A.K.S.Wong, K.T.Woo, A.T.L.Lee, X.M.Xiao and V.W.H.Luk."An AGPS-based elderly tracking system", in Proc.2009 IEEE ICuFN, pp.100-105.

[4]. M.Anisetti, C.A.Ardagna, V.Bellandi, E.Damianiu and S.Reale,"Map based location and tracking in multipath outdoor mobile networks,"IEEE Trans.Wireless Commun., vol.10, no.3, pp.814-824.

[5]. M.Y.Lee and D.S Han,"Voronnoi tessellation based interpolation method for Wi-Fi radio map construction," IEEE Commun.Lett., vol.16, no.3, pp.404-407.

[6]. X.M.Xiao, A.K.S.Wong, K.T.Woo and R.S.Cheng,"Energy efficient elderly tracking algorithm", in Proc.2011 IEEE ICC, pp.1-5.

[7]. S.Anderson,"Shotgun DNA sequencing using cloned DNase I-generated fragments," Nucleic acids research,vol.9,no.13,pp.3015-3027,July 1981.

*Corresponding Author: Chitra Sudhakaran
M.E Applied Electronics, Anna University, Chennai