Transparent Location Fingerprinting for Wireless Services

Mauro Brunato Csaba Kiss Kalló
Università di Trento
Dipartimento di Informatica e Telecomunicazioni
via Sommarive 14, I-38050 Pantè di Povo (TN) — ITALY
Email: brunato|kkcsaba@science.unitn.it

***Abstract*— Detecting the user location is crucial in a wireless environment, not only for the choice of first-hop communication partners, but also for many auxiliary purposes: Quality of Service (availability of information in the right place for reduced congestion/delay, establishment of the optimal path), energy consumption, automated insertion of location-dependent info into a web query issued by a user (for example a tourist asking informations about a monument or a restaurant, a fireman approaching a disaster area).**

The technique we propose in our investigation tries to meet two main goals: transparency to the network and independence from the environment. A user entering an environment (for instance a wireless-networked building) shall be able to use his own portable equipment to build a personal map of the environment without the system even noticing it.

Preliminary tests allow us to detect position on a map with an average uncertainty of two meters when using information gathered from three IEEE802.11 access points in an indoor environment composed of many rooms on a 625m^2 area. Performance is expected to improve when more access points will be exploited in the test area. Implementation of the same techniques on Bluetooth are also being studied.

***Index Terms*— Ad Hoc Routing, Interconnection Ad Hoc - wired, QoS, Middleware, Location Management**

I. Introduction

Location detection and management is rapidly becoming a crucial issue in wireless environments [7], [1], [2], [8]. The advantages of a network node (meaning both a router and a terminal host) knowing its own position, and sharing this information with others, are becoming more and more evident as routing algorithms are becoming smarter and mobile-specific applications are being introduced at the user level [12].

For instance, the ability to build the network topology based on real-world node dislocation can help building more robust routing algorithms, reducing dependence from unwanted behavior of radio wave propagation: if we only use radio strength to build the routing scheme, two distant nodes may become prime neighbors at the expense of nearby nodes, because of self-interference and multipath fading effects; this situation, however, can lead to unstable topologies, since small movements are likely to substantially decrease the signal level of distant nodes.

QoS-enabled middleware can also benefit from user location information from many viewpoints: routing schemes can be calibrated in order to obtain the desired delay, the user's movements can be tracked in order to put relevant information as near as possible to his location in order to reduce the wireless link congestion; it is also possible to model the user's future behavior in order to reduce the expected network load by distributing information along his possible path and by prefetching data (which will be likely requested by the user in a future time) under good radio link conditions if substantial degrade is foreseen along the modeled user path, resulting in faster perceived service and equipment battery savings.

Finally, end applications can take advantage from location information by partially automating user queries. Consider a tourist asking for information about the monument in front of him. If the application (browser) is aware of the user's location, a lot of typing by the tourist can be avoided.

This paper is organized as follows. In Section II we introduce the context of our work, previous results in the field of location discovery. In Section III we describe the hardware and software equipment we are using for experiments. In Section IV we show some results we obtained in our tests. Section V discusses briefly our current work, extending the results reported in this paper. Finally, some conclusions and indications for future work are outlined in Section VI.

This research is partially supported by the Province of Trento (Italy), in the framework of project WILMA[1].

II. Context

The technique we propose in our investigation tries to meet two main goals.

- The first is transparency to the network: a node should be able to run the location algorithm without requiring any algorithm on the other nodes, and without the rest of the network even noticing it (the information will be spread according to the user's privacy policy).
- The second goal is independence from the environment: no prior knowledge of the environment should be required. A user entering an environment (for instance a wireless-networked building) must be able to use his own portable equipment to build a personal map of the environment.

These goals cannot be met by a standard positioning system. In fact, while satellite positioning systems such as USA's GPS, former Soviet Union's GLONASS and the planned EU's

[1]WILMA is an acronym for *Wireless Internet and Location Management Architecture*; more information can be gathered at the project's web site: `http://www.wilmaproject.org/`.

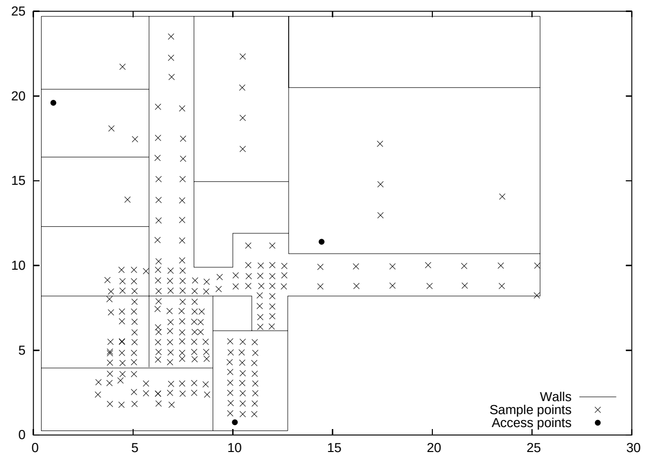

Fig. 1. The experimental environment.

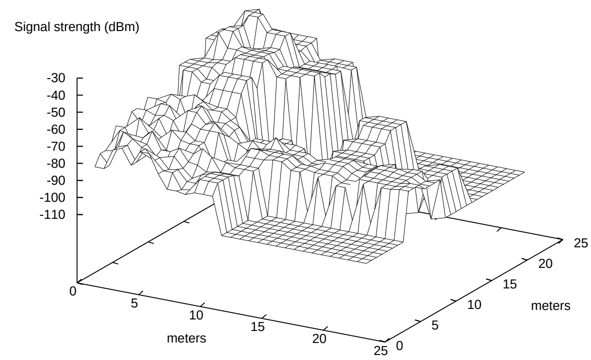

Fig. 2. Radio signal strength for AP1 of Figure 1.

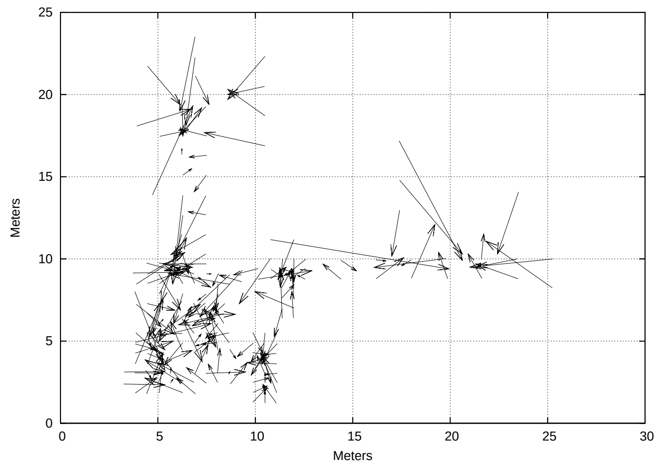

Fig. 3. Displacement error (194 pairs, leave-one-out estimates, $k = 6$).

GALILEO offer a rather good position estimate together with other interesting services, they cannot be operated indoors or in a town with tall buildings.

Other common systems suitable for indoors localization require an appropriate infrastructure, such as infrared or radio beacons.

To achieve our proposed goals, we assume the existence of non-mobile nodes (which are likely to exist even in an ad-hoc network in the form of access points to the wired network).

We use signal strength information to build a location fingerprint map of the environment. When enough information has been collected, it can be used to derive the unknown location based on signal strengths of the various transmitters.

III. Equipment and experimental settings

The IEEE802.11b wireless LAN technology (also known as WiFi) was selected for the initial part of the project due to many reasons: widespread use, fairly low cost, and above all the fact that signal strength measurements must be reported by the card as part of standard compliance.

Three IEEE802.11b Lucent Technologies Avaya AP-II access points have been placed as shown in Figure 1, connected to external antennas, while a laptop equipped with a Lucent Technologies ORiNOCO Silver PC card was used to build a radio map of the environment; the map consists of a sequence of pairs (ss_i, p_i) where ss_i is a triplet of radio signal strengths and p_i is the corresponding physical coordinate in the map.

Figure 2 shows the signal strength received from access point AP1 (the black dot at coordinates (1m, 19.6m) in Figure 1) along the map; the -102dBm level (the lower flat portions of the graph) is used to represent areas not covered by measures.

IV. Results

After collecting several example pairs as described above, in our case 194 samples, the algorithm chosen for determining the unknown position, given a triplet ss of radio strength levels expressed in dBm units, was the k-nearest-neighbors technique. Given a positive integer number k, the algorithm works as follows:

1) Find among the known signal strength ss_i the k that are nearest to the given ss triplet; let $i_1, i_2, \dots, i_k$ be their indices.
2) Calculate the estimated position by the following average, weighted with the inverse of the distance between signal strengths:

$$p = \frac{\displaystyle\sum_{j=1}^{k} \frac{1}{d(ss_{i_j}, ss) + \varepsilon} \cdot p_{i_j}}{\displaystyle\sum_{j=1}^{k} \frac{1}{d(ss_{i_j}, ss) + \varepsilon}},$$

where $d(ss_i, ss)$ is the Euclidean distance between the two triplets, and ε is a small real constant ($\varepsilon = .01$ in our tests) used to avoid division by zero.

Using this algorithm, leave-one-out error estimates were performed by removing one couple from the training set and using all other couple in the previous algorithm in order to get an estimation of its position based on the signal strength triplet. This procedure was repeated for every point; displacements of the estimated from the true position are shown as arrows in Fig-

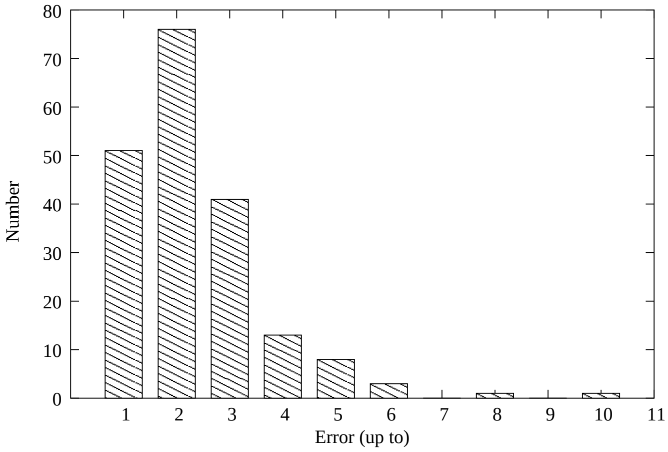

Fig. 4. Experimental error distribution (194 pairs, leave-one-out estimates, $k = 6$).

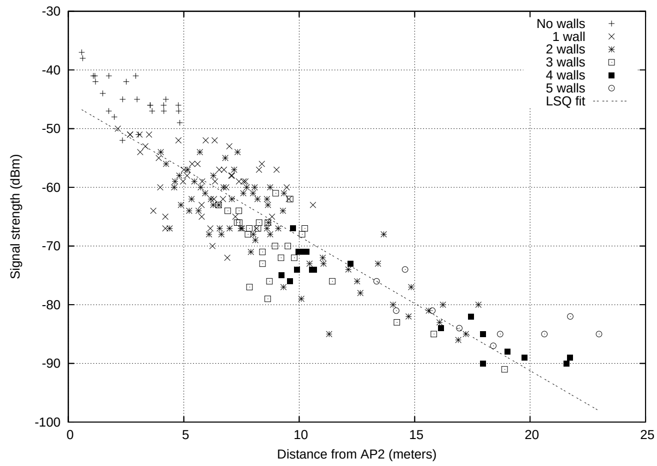

Fig. 5. Scatterplot of signal strength against distance for AP2; the number of wall crossings from the AP to each test point is reported.

ure 3 for $k = 6$ (weighted average of 6 nearest neighbors in the radio signal space).

Distribution of the error is shown in Figure 4; every histogram bar represents the number of couples for which the leave-one-out position estimate resulted in a given error class (up to one meter for the first, from one to two meters the second, and so on). The average positioning error is about 1.78 meters, even though occasional errors up to 10 meters show up. The parameter value $k = 6$ was chosen because it returned the lowest average error; however, all values from $k = 2$ to $k = 25$ return an average error below 2 meters.

V. Ongoing work

A. Different techniques and problem evolutions

The technique we proposed is substantially training by examples; the nearest-neighbors technique has been used because the structure of the radio space is reasonably smooth (apart from wall crossings, as we can see in Figure 2). Other training techniques are being developed and studied by our group: in particular, neural network models and support vector techniques are good candidates; their positioning error is comparable with the nearest-neighbors technique, and while the training algorithm takes a rather long time, the complexity of position estimation is lower. Another technique that can take advantage from this kind of measurements employs the Bayes theorem to derive a conditioned probability distribution for placement.

More precision can probably be attained when the past history can be considered, by tracking user movements and computing mobile average. To perform these tests, a PDA was equipped with the same PC card and a graphical program that allows the user to insert his current position while detecting signal strengths.

1) Neural networks [3]: Learning by example is the natural scope of neural networks. In our context the multi-layer feed-forward perceptron model has been applied with 3 input neurons (one for each access point), two outputs (the x and y coordinates) and a hidden layer with 4, 8 or 16 neurons. The best results reported an error of around two meters.

2) Probabilistic models: Probabilistic methods based on Bayesian theory require the knowledge of the signal propagation model in the form of a probability distribution. There are two possible approaches to building a reliable model. With the first approach [9] a suitable radio propagation model is selected, then experimental observations are used to infer its parameters. This method is particularly suitable for open environments, where distance is the main cause of signal fading and a fairly simple model can be used. The second approach [5] is based on repeated observations of the received signal strength for each sampled point; once enough data have been collected, empirical distributions of individual signal strengths at different locations can be computed. In this case, no analytical model of signal propagation is built, and complex environments can be mapped, where walls and multipath fading are not negligible. The main drawback of this approach is the large number of experimental observations needed to calculate reliable distributions of signal strengths at every sample point. Once the signal propagation model has been built, the Bayes theory of conditioned probability can be used to infer a position probability distribution, given the signal strength distribution detected at one point. This distribution can be used to calculate a representative point (the average of the distribution or the maximum). Preliminary tests using the same 194-measurements set report an average error of above 3 meters. The large error can be justified by the inadequate radio model we were forced to use. In fact, while the training set is large enough to estimate a few parameters in an analytical radio model, it is too small to calculate individual signal strength distributions for every sample point, so the first of the two mentioned approaches had to be used. The plot of signal strength against distance in Figure 5 shows that signal strength (reported in dBm) decreases in a linear fashion with distance. The number of walls crossed by the straight line from the access point to the test point is not influent, as we infer by observing that all plotted points seem to adjust along the same straight line. Linear fit tests confirm that adding the number of crossed walls in the model does not improve the dependence.

3) Support vector machines: The Support Vector algorithm is based on the statistical learning theory developed over the last three decades by Vapnik, Chervonensis and others [11]. See, for example, [6] for details. The algorithm can be used for classification (i.e., mapping samples on a two-valued set, usually ± 1), scoring (mapping on small integers) and regression. Various implementations can be found on the Internet; in particular we used the packages `SVMlight` developed by T. Joachims [4] and `mySVM` by S. Rüping. In this case, current leave-one-out error estimates are about 2 meters.

B. Bluetooth scatternets

Beside WiFi, we are also working on localization issues with Bluetooth. In particular, localization of Bluetooth devices can help optimize interconnection topologies from the point of view of communication speed and energy consumption.

Interconnected piconets are called *scatternets*, and their aim is to allow more than eight active Bluetooth devices in the same network while augmenting their range by bridging. However, scatternet formation and operation algorithms are not part of the Bluetooth specifications [10] yet. In the frame of our work we try to develop new methods for optimizing communications in scatternets taking advantage of localization information that we can gather from the mobile devices.

The signal strength measurement problem with Bluetooth is not as straightforward as in the case of IEEE 802.11b. The latest version of the Bluetooth Specification does not require the device manufacturers to provide a means for software developers for the exact measurement of the signal strength, as in the case of WiFi. A Bluetooth device only needs to be able to tell whether the signal strength is acceptable, too strong or too weak. This granularity is not enough for developing a positioning system similar to the one presented in this work. Since the localization problem is very important in context-aware computing, a standard way for measuring the signal strength between Bluetooth radios would be extremly useful.

Another open issue when extending our work to Bluetooth is the series of interworking problems experienced with systems from different producers. These problems originate from the different implementations of the higher layer protocols.

VI. Conclusions

We discussed experiments to determine the user's position in a wireless networked environment without the need of additional infrastructures or of particular network configuration.

Preliminary tests allow us to detect position on a map with an average uncertainty of two meters when using information gathered from three IEEE802.11b access points in an indoor environment composed of many rooms on a 625m^2 area. Performance is expected to improve when more access points will be exploited in the test area. Implementation of the same techniques on Bluetooth, aimed at providing localization-based services as well as topology formation algorithms, are also being studied.

References

[1] P. Bahl, V. N. Padmanabhan, and A. Balachandran. A software system for locating mobile users: Design, evaluation, and lessons. Technical report, Microsoft Research, MSR-TR-2000-12, April 2000.

[2] Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *IEEE INFOCOM 2000*, pages 775–784, March 2000.

[3] Roberto Battiti, Thang Le Nhat, and Alessandro Villani. Location-aware computing: a neural network model for determining location in wireless lans. Technical Report DIT-5, Università di Trento, Dipartimento di Informatica e Telecomunicazioni, 2002.

[4] T. Joachims. Making large-scale SVM learning practical. In B. Schvlkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11. MIT-Press, 1999.

[5] Andrew M. Ladd, Kostas E. Bekris, Guillaume Marceau, Algis Rudys, Lydia E. Kavraki, and Dan S. Wallach. Robotics-based location sensing using wireless ethernet. Technical Report TR02-393, Department of Computer Science, Rice University, 2002.

[6] Edgar Osuna, Robert Freund, and Federico Girosi. Support vector machines: Training and applications. Technical Report AIM-1602, MIT Artificial Intelligence Laboratory and Center for Biological and Computational Learning, 1997.

[7] K. Pahlavan, P. Krsihnamurty, and J. Beneat. Wideband radio channel modeling for indoor geolocation application. *IEEE Communications Magazine*, 36(4):60–65, Apr 1998.

[8] K. Pahlavan, Xinrong Li, and Juha-Pekka Makela. Indoor geolocation science and technology. *IEEE Communications Magazine*, 40(2):112–118, Apr 2002.

[9] Teemu Roos, Petri Myllymäki, and Henry Tirri. A statistical modeling approach to location estimation. *IEEE Transactions on Mobile Computing*, 1(1), January 2002.

[10] Bluetooth SIG. Bluetooth core specification, version 1.1, February 2001. http://www.bluetooth.com/.

[11] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[12] R. Want and B. Schilit. Expanding the horizons of location-aware computing. *IEEE Computer*, 34(8):31–34, August 2001.