

Positioning Support in Pervasive Environments

Vasile-Marian Scuturici¹, Dejene Ejigu¹

Abstract— In order to implement reactive and proactive functionalities in a pervasive environment, contextual data must be processed. One of the most important features of the context is the position of the users and the devices. In this paper, we describe a method to determine the position of a WiFi enabled device. The prediction is based on the signal strength of the available access points. The prediction model is built from a database containing the signal strength measured in some known locations. The result is the name of the room/office where the device is localised. We also present a usage scenario, in which the user/device position is used to start proactive actions in our pervasive service environment called PerSE.

I. INTRODUCTION

The development of pervasive computing environments enforces some new ways of human-machine interaction.

In a pervasive service environment, the system middleware should take care of capturing the user's intentions, solving ambiguousness, and executing the corresponding action. Knowledge about location is a key element to determine user's intention in such an environment. Different methods are used to detect the position of a device in the space as an (x, y) vector. But the (x, y) position is difficult to integrate with the user perception of the neighbourhood inside a building. The neighbourhood relation between the users is expressed not by the physical proximity but by the perception of the presence in the same room. Our vision of a pervasive environment therefore uses the room/office positioning information instead of the precise physical position.

Local Positioning System (LPS) as contrasted to Global Positioning System (GPS) are areas that attract attention of researches that involve indoor location awareness support in pervasive computing environment. Our work here is to use an existing WiFi infrastructure to detect location of a device. We modeled the positioning process using two phases: the learning phase and the prediction phase. During the learning phase, data about the real situation is collected, classified, and interpreted into knowledge. A person holding a PDA (Personal Data Assistant) moves around the rooms and other spaces in the building to capture the signal strength. All recorded data are associated with literal location names like room labels. The collected data is then used to build a prediction model for the location name based on a decision tree method. Prediction phase then uses this model for location prediction based on the real-time data values.

Pervasive environments can use this contextual information in order to offer reactive or proactive functionalities. Following the detection of location of a mobile user, a pervasive system must support proactive actions based on the user's intension that is learned from previous experiences. Data mining techniques are used for acquisition of knowledge about the pattern of user intension from this history data. To demonstrate the link between our positioning system and a proactive pervasive support for determining user intension, we also present a usage scenario in which the user position is used to start proactive actions in our pervasive platform called PerSE [15].

A test-bed for our work is located in a WiFi enabled three storey building. Signal strength of radio frequencies between the WiFi enabled devices and the WiFi access points are used to localise these devices in an indoor situation. The WiFi access points in the building are not setup for the purpose of location detection but as a regular network infrastructure setting for WLAN connection. Our effort here is, therefore, to include localization as an added value service to the existing network infrastructure.

The rest of the paper is organized as follows: related positioning methods are presented in section 2. Section 3 presents our vision of a pervasive environment and the need for room/office based positioning services to handle proximity in the sense of user perception. Section 4 presents our positioning algorithm, which uses data mining techniques to predict the room where the WiFi device is situated. Thereafter we present the results of our experiments in section 5. A scenario of how we use results from our work is given in section 6. Section 7 concludes the paper.

II. RELATED WORKS

Among localization and distance measuring methods are Global Positioning Systems (GPS), Radio Frequency (RF) delay measurement, association to nearest Access Point (AP), and received RF signal strength. GPS is widely used in cellular handsets to provide location information and location-based services. It is based on signals from a network of satellites placed in a circular orbit around the earth [7]. The use of GPS is limited to outdoor systems. RF delay measurement or round trip flight time measurement [2] used by local positioning systems is a method in which the location of the client is derived from the known location of the APs using triangulation [13] or other methods. It is based on the fact that RF time requirement is proportional to the distance traveled. This method assumes known locations of APs in the RF space which may not always be true under pervasive environments.

¹ LIRIS-UMR 5205 CNRS, INSA de Lyon, France

Nearest associated AP is a method that assumes a client associated to an AP is closer to it than to any other AP and therefore must be located within a radius (cell) of that AP. In an office environment, the cell size in the WiFi space is too wide to be useful.

Received signal strength is another important value to predict location of mobile clients. This method depends on the principle that the loss of signal strength between transmitter and receiver is related to the distance traveled. The RADAR project [5] uses the received signal strength indication (RSSI) measured at multiple APs to triangulate coordinate of mobile clients. In [4] and [6], signal strength measured both at the APs and at the mobile clients (MCs) is used to locate the MCs. They use the TMI (triangulation, mapping, interpolation) and the TIX (triangulation, interpolation, and extrapolation) algorithms respectively. They also assume known locations of APs in the signal space and are therefore dependant on the information about the WiFi space infrastructure. The RightSPOT [3] system uses the RSSI information measured at the MCs to locate MCs. It uses specialized SPOT (Smart Personal Objects Technology) devices to measure the RSSI from FM radio stations in and around a city to detect location of MCs. Bulusu et al [1] have used a low-cost approach based on commercial off-the-shelf radios but these devices do not provide software-accessible signal strength reading.

Wang et al [17] have indicated in their experiment that indoor positioning is feasible using signal strengths of WLAN transmissions. Their technique, however assumes the knowledge of position of the access point infrastructure. They have built relationship between distance from the access point and the signal strength by the triangulation method.

Positioning is becoming an important research concern of a context-aware service in a pervasive world. The aim of our work here is therefore to use the existing WiFi infrastructure and non-specialized client devices like laptops, PDAs and smart phones to detect their own location so as to get pervasive (ubiquitous) access to location dependent services. Our prediction method works principally in an indoor environment and can also be extended to WiFi enabled outdoor spaces. It does not also require prior knowledge of the infrastructure like location and number of APs.

III. POSITIONING NEEDS IN A PERVASIVE ENVIRONMENT

PerSE (Pervasive Services Environment) platform is a middleware environment that supports the interaction of independent services which collaborate to perform an intended action. The services are located on distributed processing units which compose the pervasive environment.

In order to provide the user with the means to express his intentions, we have defined a language called PsaQL (Pervasive Service Action Query Language). PsaQL plays a similar role in the PerSE-enabled pervasive environment as SQL does for relational database management systems. The PsaQL language and the process of building a pertinent service graph from a user intention expressed in PsaQL are detailed in

[15].

For example, a user can express the intention to display one presentation from his notebook to the video projector being in the room:

```
USE sunrise.ppt ON BASE notebook WITH
SERVICE projector
```

In this case the user already has the knowledge that his presentation is stored in a file named “sunrise.ppt” on his computer accessible as “notebook”, and that the video projector is piloted by a service named “projector”. The environment must select the best service graph with respect to the constraints of the whole system (i.e. data transmission speed, reliability ...). A history database is also used to improve the result by taking into account the satisfactory actions executed in similar contexts.

In the presented PsaQL query, the user assumes that the video projector is situated in the same visually and physically accessible space. The use of a video projector situated in another room to display his presentation will not be a satisfactory solution. To ensure the neighbourhood constraint, the query can be expressed as:

```
USE sunrise.ppt ON BASE notebook WITH
SERVICE projector IN NEIGHBOURHOOD
```

We consider two devices as neighbours if they are positioned in the same room/office. The same presentation can be shared with the neighbours using:

```
USE sunrise.ppt ON BASE ALL IN
NEIGHBOURHOOD
```

Generally, we consider three degrees of accessibility in our pervasive environment: the same device (*local*), the same visually accessible space (*neighbourhood*), and the global pervasive environment (*world*). Checking if one device is in the local or in the world space is simple to verify, but checking if the processing unit is in the neighbourhood (in the same visually and physically accessible space) is a difficult task. Compared to the physical distance between two entities, the proposed neighbourhood localisation brings some semantic information concerning the space containing the entities.

IV. A MODEL FOR PREDICTING LOCATION IN PERVASIVE ENVIRONMENT

Like in any prediction process, modelling localization involves two basic activities: learning and prediction. During the learning phase, data about the real situation are collected, classified, and interpreted into knowledge. The prediction phase then uses this knowledge for location prediction based on the real-time data values. We plan our work in such a way that during the learning phase, a person holding a PDA moves around the rooms and other spaces in the building to capture the signal strength. All recorded data are associated with literal location names like room numbers. We have developed a WiFi-Spotter program that tracks received signal strength and

then keep the record on a file for further analysis. We have compared signal strength captured by our WiFi-spotter with that from the shareware programs like MiniStumbler [8] and found them similar.

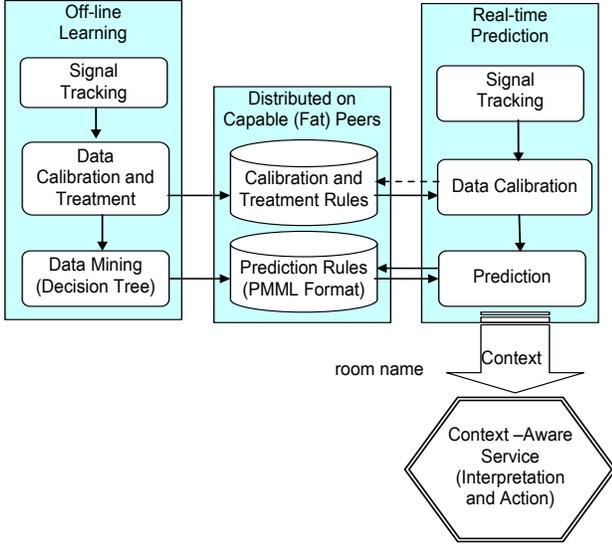


Figure 1: Architecture of our learning and prediction model

Such recorded data are further calibrated for variations that may happen due to the type of the tracking devices or other environmental influences. We have then planned a further step of processing for data and pattern classification using a data mining tool and a decision tree model. The result from this process is our working model that can later be used for real-time location detection. The model is represented in the Predictive Model Markup Language (PMML) format. Our positioning module (WiFi-Spotter) uses this model to predict locations on real-time bases. Figure 1 shows the architecture of our learning and prediction process. It also indicates the link of this prediction process to a context-aware service for triggering reactive or proactive actions.

To provide positioning support, as shown on the architecture of the model, we have indicated the learning phase (off-line) and the prediction phase (real-time). During the learning phase, signal strength at selected locations of the rooms in the building including meeting halls, offices, common rooms, printing rooms and corridors are collected and classified for pattern identification and learning. The prediction or real-time phase uses patterns identified in the learning phase to detect location for the real-time values of signal strength.

A. Learning Phase

The learning phase involves mobile clients that scan a list of radio frequency signal strength values from all n known access points at each tracking location. For each tracked point k we have a vector with the signal strength values and a label corresponding to the room/office where the point is situated:

$$(ap_k^1, ap_k^2, \dots, ap_k^n, room_k)$$

The signal strength values are situated between -40 and -90. The number of APs that can be seen varies with the location. Corresponding to the data collected from the tracked points we build an attribute-value table. The attributes of the table corresponds to the access points (identified by the MAC address) and to the room/office label. We replace the missing values (corresponding to an AP not seen in a location) with the value -100. An example of a portion of a table is presented in Figure 2.

Room	00:06:5A:40:0D:C6	00:06:5A:40:0D:D7	00:06:5A:10:0D:C6	00:06:5A:10:0D:D7
501.317	-60	-60	-60	-57
501.317	-60	-60	-60	-57
501.317	-68	-63	-59	-65
501.319	-60	-62	-64	-100
501.319	-57	-57	-60	-100
501.319	-57	-66	-57	-100

Figure 2 A portion of an attribute-value table built on the calibration points.

We use a decision tree method for data classification and learning. Decision trees are popular knowledge representation, classification and learning tools as they are easy to use and interpret [9]. In decision trees, learned patterns are represented as a tree where nodes in the tree embody decisions based on the values of attributes and the leaves of the tree provide predictions. A new situation can be classified simply by tracing a path from the root of the tree to a leaf, with the path taken being determined by the input attribute values. These input values in our case are real-time array of signal strength values received from the APs.

Among the number of algorithms and programs that implement decision tree, we used MCubiX [10]. We found MCubiX to be well structured tool whose output can be collected, viewed and stored in different forms. Demonstration of classification of a region into sub regions (into sub spaces for more than two dimensions) by the learning algorithm using a simple example of two APs and three rooms under a simple decision tree with depth two is given in Figure 3.

B. Prediction phase

Prediction phase involves detection of the room in which a mobile client is located. The two important input parameters for the prediction module, therefore, are decision rules obtained from the learning phase and the real-time signal strength values collected at a specific location. The model is stored in Predictive Model Markup Language (PMML) format and also contains the probability values of each prediction. PMML [14] is an XML-based language which provides a way for applications to define data mining models and to share models between PMML compliant applications.

Real-time signal strength values collected by the mobile

client from the APs are given in the form of an array. For each of the m rooms involved, our prediction module calculates the probability that these values are observed in the room and then selects the room with the maximum probability. Given $P(R_k)$ as the probability that the given array of values are observed from room R_k , an expression for room R is given by:

$$R = R_k : P(R_k) = \underset{j=1}{\overset{m}{\text{Max}}} P(R_j)$$

V. EXPERIMENTAL RESULTS

We have used data from selected locations in 33 rooms of

different size, from two floors of a three storey building. We have collected some 13,500 records of arrays of signal strength values in all the 33 rooms, corresponding to four hours learning and calibration phase. The collected data contains information about 100 distinct access points. Only 20 of these access points are situated in the building. The topology of the two floors constituting the test-bed is depicted in Figure 4. As the rooms in our experimental test-bed are located side by side, knowledge about their width helps us to investigate the resolution of our detection method.

The average width of the small rooms is 3m for offices and 7m for amphitheataters with an overall average of 5m interval

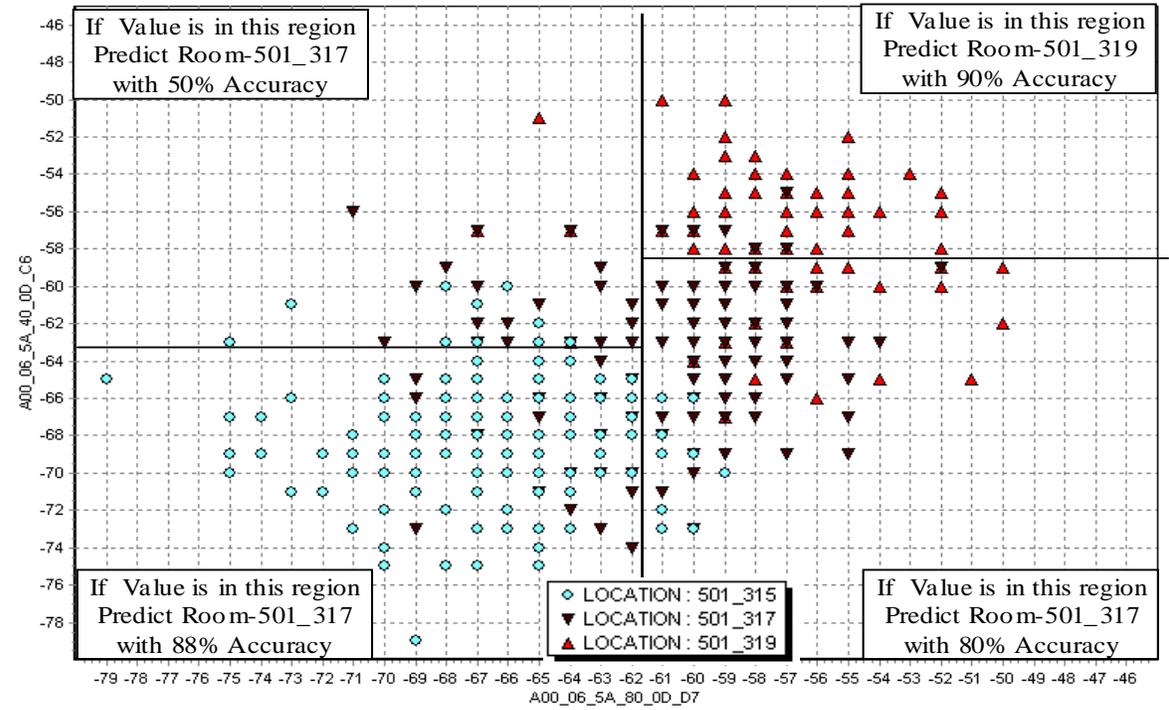


Figure 3: Example of the effect of classification of a region into sub regions by the learning algorithm using a simple two APs and three rooms.

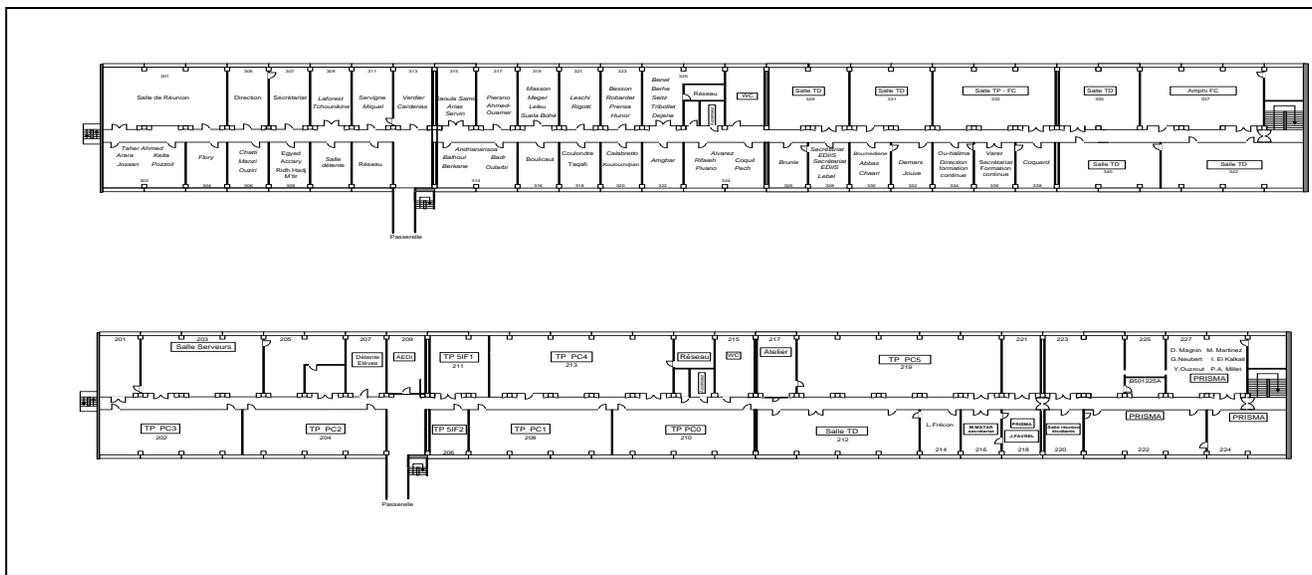


Figure 4: The topology of the two floors used as test-bed in our experiments.

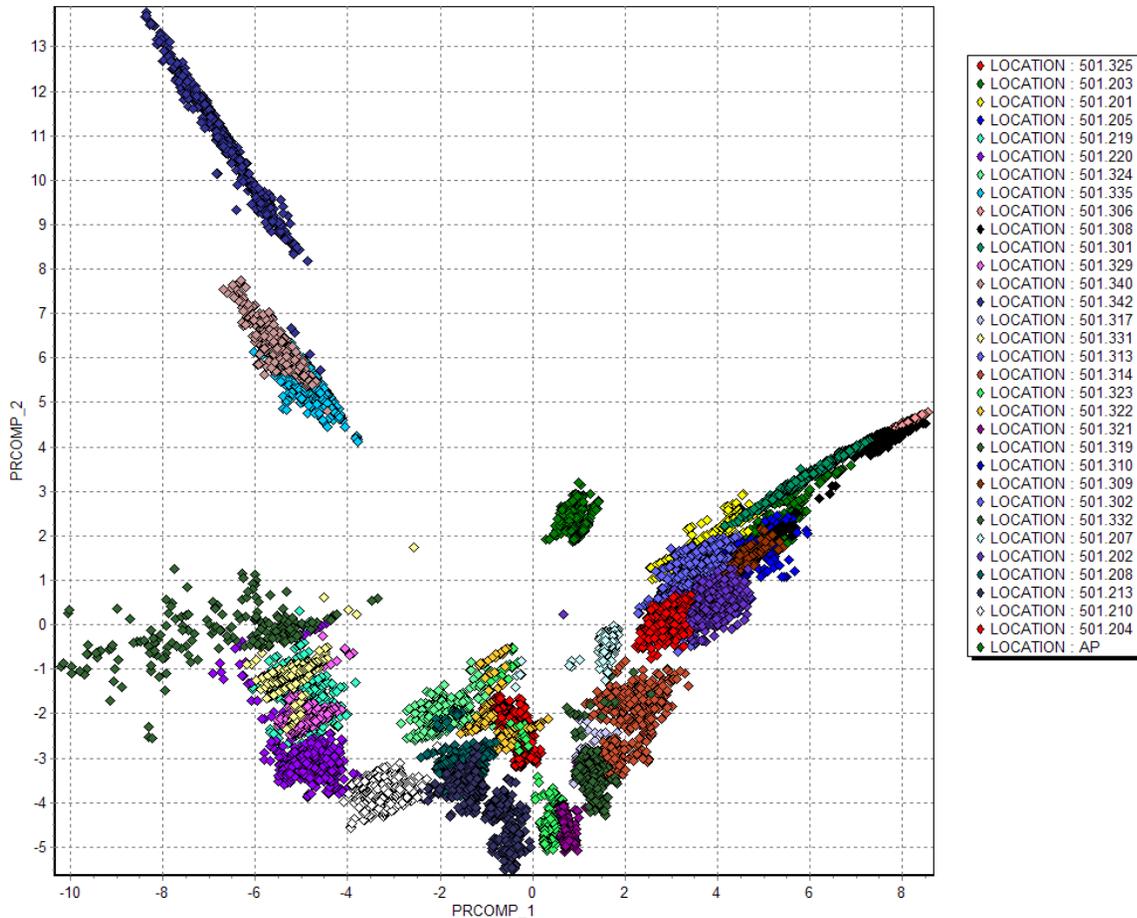


Figure 5: Two dimensional space representation of the collected data using PCA algorithm.

between the corresponding points of references in the neighboring rooms. This shows that our method is detecting separate locations at an interval of an average of 5m and in most of the cases it involves detecting small rooms with intervals as small as 3m.

To see the structure of the collected data in a two dimensional space, we have used a Principal Component Analysis (PCA) algorithm [16] to project the 100 dimension space (corresponding to all visible APs) in a 2 dimensions space. The data seems to be separable (Figure 5).

Using the collected data as a learning dataset we build a decision tree model. The decision tree has been tested using a cross validation technique. The results are very encouraging: the error rate is situated below 5%, corresponding to a 95% hit rate.

The number of rules generated by the decision tree is about 200 (Figure 6). The size of the PMML file containing the decision tree model is about 320 KB.

A. Influence of the type of tracking device

The strength of the installed WiFi card is a major source of variation that contributes to detection error. We can not use one generic model for all types of devices that are using different types of cards. There are two possible approaches to

solve this problem. The first and the straight forward solution is to create separate models for every card and make the models available for use during prediction. The drawback here is that the more cards we involve in our model the more storage is required in an already storage hungry environment. This limits the scalability of the model.

```
IF 00_06_5A_E0_0D_FA < -74,00 and 00_06_5A_80_0D_C9 >=-87,00
THEN LOCATION in [501.342] with accuracy 1,0000

IF 00_06_5A_E0_0D_FA < -74,00 and 00_06_5A_80_0D_C9 < -87,00
and 00_06_5A_60_0D_C6 < -68,50 and 00_06_5A_20_0D_DB < -77,50
and 00_06_5A_80_0C_BD >=-73,50 and 00_06_5A_C0_0D_D7 >=-92,50
and 00_06_5A_E0_0D_D7 < -85,00 THEN LOCATION in [501.329] with
accuracy 0,9877

IF 00_06_5A_E0_0D_FA < -74,00 and 00_06_5A_80_0D_C9 < -87,00
and 00_06_5A_60_0D_C6 < -68,50 and 00_06_5A_20_0D_DB < -77,50
and 00_06_5A_80_0C_BD >=-73,50 and 00_06_5A_C0_0D_D7 >=-92,50
and 00_06_5A_E0_0D_D7 >=-85,00 THEN LOCATION in [501.210] with
accuracy 1,0000
```

Figure 6 Example of rules generated by the decision tree model

The second approach is data normalization using trend analysis. Our trend analysis experiment with three different products (Dell-Axim3[®], Dell-Axim50[®] and HP-Hx470[®])

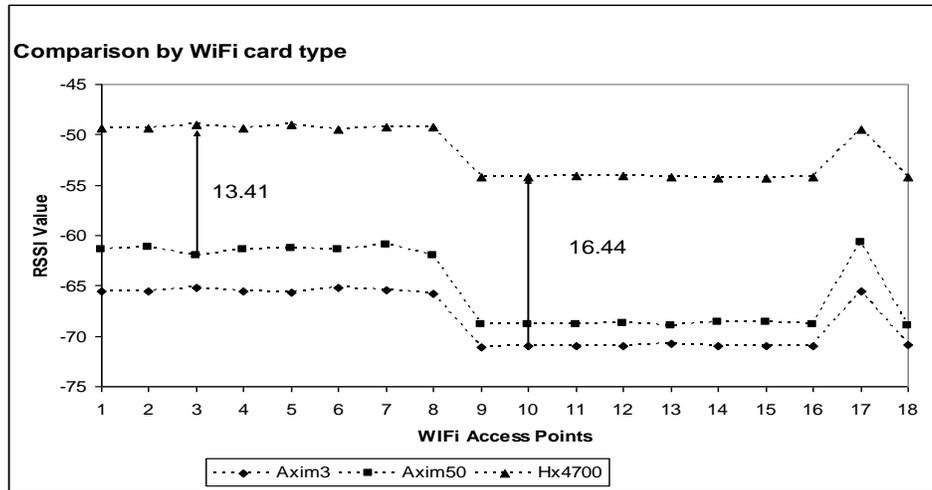


Figure 7 Average values of signals measured by three different devices at the same location. There are 18 access points available at this location.

Base ¹	BaseName	Location	Running Action	TriggerAction	...
Localhost	PDA_David	Common_Room	USE mybaby.mpg WITH SERVICE multimedia_player ON BASE LOCALHOST	SHARE SERVICE multimedia_player ON BASE LOCALHOST WITH SERVICE multimedia_player ON BASE ALL IN NEIGHBORHOOD	
Localhost	PDA_David	Common_Room	USE vacances.mpg WITH SERVICE multimedia_player ON BASE LOCALHOST	SHARE SERVICE multimedia_player ON BASE LOCALHOST WITH SERVICE multimedia_player ON BASE ALL IN NEIGHBORHOOD	
...	

shows a significant variation in the received signal. Figure 7 shows the existence of an important constant trend between the lines. These constant values calculated as an overall average variation between the lines can be used to normalize signal values during data calibration and treatment phase.

VI. SCENARIO FOR USING LOCATION AS A CONTEXT IN A PERVASIVE SYSTEM

Consider a scenario where David, one of the users of our system, is given a proactive multimedia entertainment service on his PDA while he is in the common room for the tea break. He is watching a video recorded on the last holyday he had with his baby. The common room is also used by some friends of David. They too are also equipped with PDAs. David wants to share the seen of his video with his friends. In this case he will use the middleware PerSE to express his intention (using PsaQL) as:

```
SHARE SERVICE multimedia_player ON BASE
LOCALHOST WITH SERVICE multimedia_player
ON BASE ALL IN NEIGHBORHOOD
```

This is one example of a reactive functionality of a pervasive environment where positioning using NEIGHBOURHOOD as the presence in the same room is useful. David want to share the image of his multimedia player running on his PDA with the other PDAs situated in the same room. The PerSE middleware will propose this action to each PDA located in the common room (maybe not all the people present in the room are interested in the seen of David's baby). In this process, detection of other PDAs in the common room is one of the important applications of our proposed method.

A. Proactive functionalities

Consider that the following set of facts is known by the middleware as a history on David's PDA:

Based on a history like this, a predictive model to start proactive actions can be built using decision trees. The

¹ The term *base* means device or processing unit in the PerSE middleware.

predicted attribute is the action to be triggered (TriggerAction) based on an existing context. This context will contain the base where the context has been stored (localhost), the name of this base ("PDA_David"), the location ("CommonRoom") and the actually running action (the intention expressed by David using PsaQL to display a video on his PDA).

The resulted decision tree will then contain the following rule:

```
IF BASE = LOCALHOST AND BASE_NAME =
"PDA_DAVID" AND LOCATION = "CommonRoom"
AND RunningAction = "USE * WITH SERVICE
multimedia_player ON BASE LOCALHOST"
THEN TriggerAction = "SHARE SERVICE
multimedia_player ON BASE LOCALHOST WITH
SERVICE multimedia_player ON BASE ALL IN
NEIGHBORHOOD"
```

This scenario demonstrates the situation where location prediction is used in combination with the context history to trigger a proactive action that represents David's intension. The primary use of our localisation service in this example is to detect that PDA_DAVID is in the Common_Room. Detecting who else is present in the room can also be considered as point where the service is called.

VII. CONCLUSIONS AND FUTURE WORK

With the growth of enabling technologies in pervasive computing and the increasing users' interest, knowing the position of involved devices and users is of crucial importance. Different existing methods are used to detect the position of a device in the space as an (x, y) vector. But often, the (x, y) position does not express the user perception of the neighbourhood inside a building. The neighbourhood relation between the users may be represented not by the physical proximity, but by the perception of the presence in the same room. We have presented here our vision of a pervasive environment using the room/office positioning information instead of the precise (x, y) physical position.

The signal strength values from WiFi access points collected by mobile clients (PDAs in this particular case) have been used to formulate the prediction model. Collected signal strength data are classified and translated to the prediction model using a decision tree algorithm. We have tested the model using our interactive PDA based interface for detection. The obtained results are encouraging and the prediction hit rate is found to be more than 95%. As indicated on the scenario, the results of this work are of paramount importance for localization of users and resources in pervasive environment to trigger proactive actions.

The PerSE middleware is still a prototype, without a positioning support. It is able to execute user intentions expressed in PsaQL, using the available services. The positioning module presented in this paper has been tested as an independent module. We will integrate this work in the PerSE middleware.

REFERENCES

1. Bulusu, N., Heidemann J., Estrin D., GPS-less Low-Cost Outdoor Localization for Very Small Devices, In Proceedings of the IEEE Personal Communications Magazine, Volume 7, No. 5, pp. 28--34, October 2000.
2. Werb J., Lanzl C.: Designing a positioning system for finding things and people indoors, IEEE Spectrum, Volume 35, Issue 9, pages 71 – 78, September 1998.
3. Krumm, J., Cermak G., Horvitz E.: RightSPOT: A Novel Sense of Location for a Smart Personal Object, Adaptive Systems and Interaction, Microsoft Research, Redmond, Washington 98052 (<http://research.microsoft.com>, last checked 23/12/2005).
4. Smailagic, A., Siewiorek, D. P., Anhalt J., Kogan D., Wang Y.: Location Sensing and Privacy in a Context Aware Computing Environment, IEEE Wireless Communications, 9(5), pages 10-17, October 2002.
5. Bahl, P., Padmanabhan, V. N.: RADAR: An In-Building RF-based User Location and Tracking System. In Proceedings of IEEE INFOCOM 2000, Tel-Aviv, Israel, March 2000.
6. Gwon, Y., Jain, R.: Error characteristics and calibration-free techniques for wireless LAN-based location estimation, in Proc. of ACM MobiWac, October 2004.
7. Global Positioning System (GPS), <http://samadhi.jpl.nasa.gov/msl/Programs/gps.html>, last checked 23/12/2005.
8. MiniStumbler v0.4.0 Release Notes, <http://www.netstumbler.com>, last checked 23/12/2005).
9. Quinlan, J.R.: Induction of Decision Trees, Machine Learning, Volume 1, Issue 1, pages 81-106, 2003
10. MCubiX: <http://www.diagnos.ca/>, last checked 2/1/2006
11. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1, pages 108–121, 1997
12. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York, pages 415–438, 1997
13. Global Positioning Satellites, <http://hyperphysics.phy-astr.gsu.edu/hbase/gps.html>, last checked 23/12/2005.
14. Predictive Model Markup Language (PMML): Data Mining Group, <http://www.dmg.org/>, last checked 23/12/2005.
15. Bihler, P., Brunie L., Scuturici, V.M.: Modeling User Intention in Pervasive Service Environments, Lecture Notes in Computer Science, Volume 3824, Pages 977 – 986, Novembre 2005
16. Pearson, K.: On Lines and Planes of Closest Fit to Systems of Points in Space, Philosophical Magazine 2(6), pages 559-572, 1901.
17. Wang, Y., Jia, X., Lee, H.K., An Indoors Wireless Positioning System Based on Wireless Local Area Network Infrastructure, the 6th International Symposium on Satellite Navigation (SatNav 2003), Melbourne, Australia, 22–25 July 2003.