A hybrid BLE and Wi-Fi localization system for the creation of study groups in smart libraries

Kiril Antevski, Alessandro E. C. Redondi, Razvan Pitic
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Italy
Email: {name.surname}@polimi.it

***Abstract*—Campus libraries in modern universities provide students with group study areas where they can work and study collaboratively. In this paper, we propose a complete solution for the creation of study groups in future smart libraries featuring (i) a smartphone application to create study groups, (ii) a hybrid Bluetooth Low Energy (BLE) and Wi-Fi indoor positioning system to localize study groups and (iii) a server-based infrastructure based on MQTT and Node-RED to advertise study groups to other students. We describe in details all components of the architecture and perform an experimental evaluation of the indoor positioning system in a realistic scenario.**

***Index Terms*—Smart Library, Indoor Localization, BLE, Wi-Fi**

I. Introduction

Libraries are one of the most convenient resources that modern universities have to offer, and they are undoubtedly one of students' most favourite spaces. They provide students with an optimal place to study, either alone in absolute silence or in groups, as well as to borrow study material such as printed books and magazines, electronic resources (full-text articles, journals), DVDs, etc... Most libraries also have free Wi-Fi internet, as well as other capabilities such as inexpensive access to printers, laptops and tablets, digital calculators and so on. Finally, university libraries are generally open until late at night and most of them encompass a coffee shop to provide students with their caffeine fix. For all these reasons, campus libraries are generally very appreciated and they are often very crowded, at each hour of the day.

Among all facilities provided by campus libraries, group study areas where students can work and study collaboratively are of particular interest: although individual study allows one to concentrate and minimizes distraction, group study offers the opportunity to reinforce ideas and share information, finally building and broadening the understanding of the subject under study. Indeed, 20 years of educational research has consistently demonstrated that cooperative classroom groups result in greater learning than competitive or individualistically-structured learning environments [1], [2].

With the advent of the Internet of Things paradigm, university campuses are evolving to the new concept of *smart spaces*, and libraries are no exception. According to such a vision, *smart libraries* will leverage networks of intelligent sensors and actuators connected to the internet, able to provide a vast range of high-level services to both users and administrators. Examples include smart lighting [3], localization of books on bookshelves [4], quality of air monitoring [5], automatic ventilation [6] and many others.

In this paper, we propose a complete system for the management of study groups in smart libraries. The system includes a smartphone application that allows a student to create a study group anywhere in the library and to advertise it through a server-based infrastructure to other students. Besides general information such as the topic under study and the number of students involved in the group, the advertisement includes the position of the study group inside the library, so that students willing to join can locate it easily. To this end, we develop a hybrid Bluetooth Low Energy (BLE) and Wi-Fi indoor positioning system whose main feature is the ability to operate with both wireless technologies, following the user preferred configuration. When only Bluetooth is enabled on the user's smartphone device, the positioning system leverages the presence of BLE beacons deployed in the library to determine the user location. When only Wi-Fi is enabled, the system uses a fingerprinting approach and relies on a database of received signal strength indicators (RSSI) fingerprints to estimate the user's position. When both BLE and Wi-Fi are enabled, the system estimates the user's location fusing the information coming from the two technologies, with clear improvements on the localization performance. Moreover, in this latter configuration, the availability of Bluetooth positioning allows to automatically update the Wi-Fi RSSI fingerprint database, minimizing the cumbersome burden generally associated with off-line Wi-Fi fingerprints collection. The system is designed leveraging software tools widely accepted in the IoT world, such as the Message Queueing Telemetry Transport (MQTT) protocol for information exchange between users and IBM Node-RED for the back-end.

The rest of the paper is organized as follows: Section II reviews related works in the area of smart spaces and indoor localization systems in general; Section III describes the hybrid approach taken for implementing the localization system. Section IV describes each component of the proposed systems, while Section V reports on its performance evaluation. Finally, Section VI comments on future research directions and concludes the paper.

II. Related Work

In the last few years, the advent of the IoT paradigm and the availability of low cost sensors and actuators coupled with wireless communication have stimulated an enormous amount of works about the so-called *smart spaces*. Smart spaces include all those physical locations which are enriched with sensing and reaction capabilities with the final goal to offer high-level services to users. This general definition is valid at different levels of spatial granularity, starting from the broad concept of *smart cities*, narrowing the discussion to *smart buildings* and even more to *smart office* and *smart home* environments. The types of services offered by such smart spaces is twofold: on the one hand, they are expected to enhance the quality of life of their occupants with applications such as air quality monitoring, noise monitoring, automation of electronic devices (lights, VAC systems, etc.). On the other hand, smart spaces are expected to be a great asset for their administrators, providing services to increase their efficiency and decrease their management cost. Examples of such services include energy consumption monitoring and optimization, waste management and many others [7].

Focusing to indoor spaces, the availability of the knowledge of each user's position is key to implement several of the aforementioned services: for this reason, a huge amount of work has been recently done to implement reliable indoor positioning systems. Due to the impossibility of using the global positioning system (GPS) indoor, researchers have tried to use many other signals for developing positioning techniques, including a vast range of radio signals (Wi-Fi, Zigbee, RFID, Bluetooth, etc.), sounds and ultrasounds, magnetic fields, images etc. [8]

Among all these, Wi-Fi has clearly attracted a lot of attention due to its widespread usage and availability, both in terms of network infrastructure (WLANs and access points) and user equipments (Wi-Fi enabled mobile devices). Clearly, Wi-Fi based IPSs constitute the most cost-effective way to implement location-aware services in smart environments, even without the need of installing additional infrastructures.

There exist many ways in which Wi-Fi signals can be used for positioning. A class of techniques relies on trilateration or triangulation approaches, by first converting Wi-Fi measurements to distances and then applying geometry based algorithms to determine the position of a user. The measurements used for such approaches include time-of-arrival (TOA) and time-difference-of-arrival (TDOA), angle of arrival (AOA) and received signal strength (RSS). The latter signal in particular has been subject of many studies: indeed, RSS measurements can be performed on off-the-shelf mobile devices without any extra sensing unit, do not require complex synchronization algorithms compared to time-domain methods (TOA, TDOA) and work reasonably well in non line of sight environments [9]. Also, RSS measurements are used to enable a second class of location techniques known as *fingerprinting*, which are based on the collection of the signal patterns characterizing each physical position. Fingerprinting methods are generally composed of two phases: in an off-line phase, the space under consideration is sampled at many known locations. For each location, the vector of received signal strength from all the in-range Wi-Fi access points (a.k.a. the fingerprint) is stored in a database. In the online phase, a user measures the RSS from all the detected access points and transmits this vector to the database: here, the closest match in the RSS space is found, e.g., using a similarity metric according to a k-nearest neighbour approach, and the target position is estimated using the location of the most similar fingerprints. Several other techniques can be applied on the fingerprinting database to estimate the location of the user, including machine learning tools such as neural networks and support vector machines. Regarding the position accuracy, many works have revealed that Wi-Fi fingerprints allows to reach positioning accuracies in the range of 2-5 meters, depending on the number of access points available in the area and on the fingerprinting technique chosen. A detailed survey of Wi-Fi fingerprint-based indoor positioning techniques can be found in [10].

Motivated by the introduction of Bluetooth Low Energy (BLE), starting from 2010 there has been increasing interest for indoor positioning systems based on Bluetooth signals in addition to classical Wi-Fi solutions. In such systems, battery-operated, low cost BLE *beacons* play the role of Wi-Fi access points. Being cheap and operated with long-lasting batteries, BLE beacons allow for quick and flexible IPS deployments, at a smaller operation cost for the user than Wi-Fi-based systems. However, due to Bluetooth reduced communication range (tens of meters in indoor scenario) which would require many beacons to be deployed to cover even small areas, BLE-based systems generally use proximity analysis for positioning rather then fingerprinting. That is, they assign to the user the location of the strongest BLE beacon received by its device. Therefore, Bluetooth-based positioning systems are often used in hybrid configurations paired to Wi-Fi, rather than alone. In [11], the authors propose a hybrid localization system based on combined RSS fingerprints and using machine learning methods. In particular, the fingerprints are vectors of Wi-Fi and BLE RSSI and the classification method used is based on boosting and weak-learners classifiers. The authors show that their hybrid approach slightly improves the performance of a BLE-only and Wi-Fi only system by 3% and 2%, respectively. The work in [12] proposes a hybrid system where Wi-Fi is used as the main infrastructure for fingerprints, while Bluetooth serves to partition the indoor space and the fingerprint database. As a result, Wi-Fi positioning is improved in a divide-and-conquer manner. To partition the entire Wi-Fi map optimally, the authors propose a deployment algorithm for Bluetooth beacons taking into account the number of available beacons and the size of the Wi-Fi database. The authors also propose different system architectures, depending on where the position estimation is performed (on the device or on the server) and evaluate them with respect to client-side energy consumption, delay and coexistence of Bluetooth and Wi-Fi. Other examples of hybrid Wi-Fi and Bluetooth systems can be found in the research papers [13] and [14]. All these systems report improvements in the accuracy of the position estimation compared to Wi-Fi only solutions, generally lower than 3 meters on average. Recently, hybrid systems have appeared also in commercially available solutions for indoor localization, such as [15], [16] and [17]. Broadly speaking, most of the existing works on hybrid localization tends to concatenate the Wi-Fi and Bluetooth RSS in a single vector to be used as fingerprint. Differently, we propose a novel approach in which the information of the two localization systems are fused together using an a-priori information on their accuracy.

III. Indoor Localization System

Knowing the position of a student in the library with high accuracy is key to the success of the proposed system. To this end we propose three approaches, detailed in the following subsections: (i) Bluetooth-only localization, (ii) Wi-Fi fingerprints localization and (iii) hybrid Bluetooth and Wi-Fi localization. As a general notation, let $\mathbf{x}$ be the position of a student in the library. Differently from other works, we assume that the set of possible positions in which a student may be located is a finite discrete set $\mathcal{X}$, that is $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^2$. The set $\mathcal{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_M}\}$ contains M known locations in the library, such as single desks, tables or study rooms. The goal of the localization system is to compute $\hat{\mathbf{x}} \in \mathcal{X}$, i.e., the estimated location of a student in the library.

A. Bluetooth-only localization

In this scenario, we leverage the presence of N BLE beacon devices deployed in the library. Given their small size, such beacons can be deployed practically everywhere: a natural deployment strategy is to place them so that the set $\mathcal{X}$ is covered as much as possible: when $N \geq M$ the solution is trivial. When $N < M$, one may choose to deploy them in the most important spots (e.g., only in study rooms) at the cost of reduced performance. Once the BLE beacons are deployed, they periodically broadcast beacon messages which are received by any BLE-enabled mobile device in their communication range. Here we rely on a proximity algorithm, and the estimated location $\hat{\mathbf{x}}$ of a student carrying such a mobile device is considered to be collocated with the BLE beacon from which it receives the strongest signal. Although such a method may appear simplistic, Section V will show that it performs remarkably well when the number of available beacons is enough to cover all possible locations. Clearly, this behaviour is due to the particular application and assumptions under consideration and cannot be easily generalized to other scenarios.

B. Wi-Fi fingerprint localization

When the mobile device carried by a student is not Bluetooth-enabled, we leverage the presence of any existing Wi-Fi infrastructure to perform localization. Given the widespread availability of Wi-Fi networks (both public and private), a mobile device in any location receives a great number of beacon frames from L different access points, denoted with $l = 1, \ldots, L$. Each beacon frame carries information on the Wi-Fi network served by the l-th access point, such as the network Basic Service Set Identifier (BSSID), the capability of the network, the supported rates, etc.. Depending on the relative position of the l-th access point compared to the mobile device, its beacons will be received with a signal strength s_l, usually expressed in dBm. The m-th location of the set $\mathcal{X}$ can be therefore described with one or more signal strength vectors, or fingerprints, $\mathbf{s}_{m,i} = \{s_1, \ldots, s_L\}^{\mathrm{T}}$, where i is used to denote different signal strength measurements taken at the same location. Such measurements are collected for each location in an offline phase and stored in a database, together with the location at which they were collected. In the online phase, a user measures a signal strength vector $\tilde{\mathbf{s}}$, known as *query*, and transmits it to the database. Here, several strategies can be adopted to estimate the location $\hat{\mathbf{x}}$ of the user:

1) k-Nearest Neighbour (k-NN): The Euclidean distance between the query $\tilde{\mathbf{s}}$ and each entry $\mathbf{s}_{m,i}$ in the database is computed. The k vectors with shortest distances are retained and the estimated location $\hat{\mathbf{x}}$ is assigned to the most common location among the k nearest neighbours.

2) k-Means clustering: Although widely used, the k-NN approach suffers from two main drawbacks: first, different fingerprints from the same location can be influenced by large signal noise, which may lead to localization errors. Second, k-NN requires pairwise matching with all fingerprints, which can be computationally intensive for large databases. To overcome both problems, k-means clustering can be used: first, the fingerprint database is partitioned in k different clusters where fingerprints are characterized by the nearest mean. Each cluster has a cluster-head fingerprint, which serves as a "prototype" for the cluster itself. Generally, the number of clusters to be created is a free parameter and must be tuned according to the specific application. In our scenario, we set $k = M$ and we create cluster-heads by averaging together different observations belonging to the same location, that is:

$$\mathbf{s}_m = \frac{1}{N_m} \sum_i \mathbf{s}_{m,i} \tag{1}$$

where N_m is the number of observation corresponding to the m-th location. This operation filters the noise in the observations and reduces the size of the database. After this step, 1-NN is performed on the reduced database composed by cluster-heads only, and the estimated location $\hat{\mathbf{x}}$ is returned.

3) PCA-based fingerprinting: Each fingerprint is composed of L signal strength measurements from different access points. However, only a subset of size $P < L$ of such access points may provide useful information. As an example, if one of the access points is far from every location in $\mathcal{X}$, the corresponding received signal strength will be very low (or even not detected) in every fingerprint. Similarly, two collocated access points will generate correlated entries in the fingerprints. To overcome such problems, principal component analysis (PCA) can be applied to the fingerprint database. PCA is a very popular technique used in machine learning for dimensionality reduction: it finds an orthogonal transformation that converts the original observations in a set of linearly uncorrelated *principal components*. Such principal components are also sorted in decreasing order of their variance, naturally identifying a "ranking" among them. Operatively, we learn the PCA transformation $\mathbf{T}$ from the fingerprint database after the offline phase and communicate it to the mobile devices. In the online phase, a mobile device transforms the query $\tilde{\mathbf{s}}$ in the PCA domain by multiplication with $\mathbf{T}$, that is:

$$\tilde{\mathbf{w}} = \mathbf{T}\tilde{\mathbf{s}} \tag{2}$$

Then, only the first $P < L$ elements of $\tilde{\mathbf{w}}$ (the first P principal components) are retained and transmitted to the database, where either k-NN matching or k-Means clustering can be applied on the PCA-transformed and reduced database.

C. Hybrid Bluetooth and Wi-Fi localization

When both Bluetooth and Wi-Fi are enabled on a student's smartphone, it is possible to set up a hybrid localization framework that fuses together the information coming from the two systems. In particular, we observe that the error probability density function $P(E)$ of the BLE or Wi-Fi systems is not uniform with respect to space, but estimation errors of the two systems always occur in particular locations. The reason of this non-uniformity is clearly due to the particular propagation characteristics of the environment under consideration, paired with the specific deployment of the BLE beacons and Wi-Fi access points. Since the position of beacons and access points is generally fixed and does not change after the first deployment, one may estimate the spatial distribution of the location errors $P(E \mid \hat{\mathbf{x}})$ in an offline phase, and use it a-posteriori to correct the estimation.

In details, the process is as follows: first, two users' locations $\hat{\mathbf{x}}_{\text{BLE}}$ and $\hat{\mathbf{x}}_{\text{Wi-Fi}}$ are estimated independently using Bluetooth and Wi-Fi, respectively. Then, a new location estimation $\hat{\mathbf{x}}_{\text{H}}$ is produced according to the following:

$$\hat{\mathbf{x}}_{\text{H}} = \begin{cases} \hat{\mathbf{x}}_{\text{BLE}} & \text{if } P(E \mid \hat{\mathbf{x}}_{\text{BLE}}) < P(E \mid \hat{\mathbf{x}}_{\text{Wi-Fi}}) \\ \hat{\mathbf{x}}_{\text{Wi-Fi}} & \text{otherwise.} \end{cases} \tag{3}$$

As an example, Figure 1 shows the estimated conditional probabilities $P(E \mid \hat{\mathbf{x}}_{\text{BLE}})$ and $P(E \mid \hat{\mathbf{x}}_{\text{Wi-Fi}})$ for twenty locations of the set $\mathcal{X}$. Assume that the BLE and Wi-Fi localization systems compute as estimated positions $\hat{\mathbf{x}}_{\text{BLE}} = 8$ and $\hat{\mathbf{x}}_{\text{Wi-Fi}} = 14$. Since $P(E \mid \hat{\mathbf{x}}_{\text{Wi-Fi}} = 14) < P(E \mid \hat{\mathbf{x}}_{\text{BLE}} = 8)$, the newly estimated location according to the hybrid system is $\hat{\mathbf{x}}_{\text{H}} = 14$.

D. Online fingerprints collection

One of the most time-consuming steps of the fingerprint based approach is the offline phase in which the fingerprint database is constructed. This requires to put a device on each of the locations in $\mathcal{X}$ and to collect many fingerprint vectors to build the database. Moreover, the process should be repeated periodically to avoid that possible changes in the monitored area (e.g., movement or addition of furniture) impact on the signal propagation environment making the already-collected fingerprints obsolete.

Hybrid localization can be used to alleviate the cumbersome task of updating the fingerprints database. In particular, when

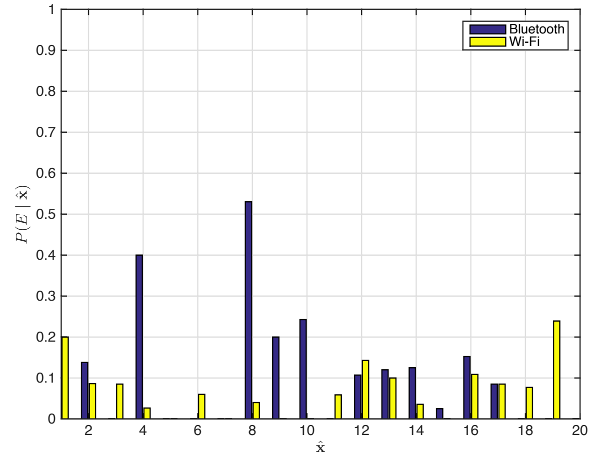

Fig. 1. The conditional error probability $P(E \mid \hat{\mathbf{x}})$ estimated for Bluetooth and Wi-Fi (6-NN with PCA) for 20 locations of the set $\mathcal{X}$

hybrid localization is enabled, the query vector transmitted by a user is added to the database and labeled with the estimated location returned by the Bluetooth-only system. To avoid that a wrong location label is appended to the fingerprint, the system may explicitly ask the mobile device owner to confirm the estimated location.

Note that the same explicit notification may be used to update the conditional probability density functions $P(E \mid \hat{\mathbf{x}}_{\text{BLE}})$ and $P(E \mid \hat{\mathbf{x}}_{\text{Wi-Fi}})$ online.

IV. Proposed System

In this section we provide a detailed description of the proposed system, which is illustrated in Figure IV. As one can see, the system is composed by four main building blocks: (i) the hybrid BLE and Wi-Fi network infrastructure, described in Section IV-A; (ii) the mobile client application (Section IV-B); (iii) the back-end server application and (iv) the front-end graphical interface, described in Section IV-C.

A. Hybrid Network Infrastructure

As aforementioned, the proposed system uses BLE beacons and Wi-Fi access points for localizing students in a library.

1) BLE infrastructure: Several battery-operated Bluetooth beacons are deployed in the area of interest, one per location in the set $\mathcal{X}$. Each beacon periodically broadcasts advertising packets carrying the transmitter identifier, with a transmission interval of one second. Recalling that the BLE localization system works by assigning to a user the position of the beacon with strongest received signal, the transmission power of the beacons is set to the lowest value, which corresponds to a transmission range of about 2 meters. By doing this we obtain a twofold positive effect on the performance of the system: (i) we minimize the probability of localization errors due to shadowing/multi-path fading and (ii) we prolong the battery lifetime of the BLE beacons.

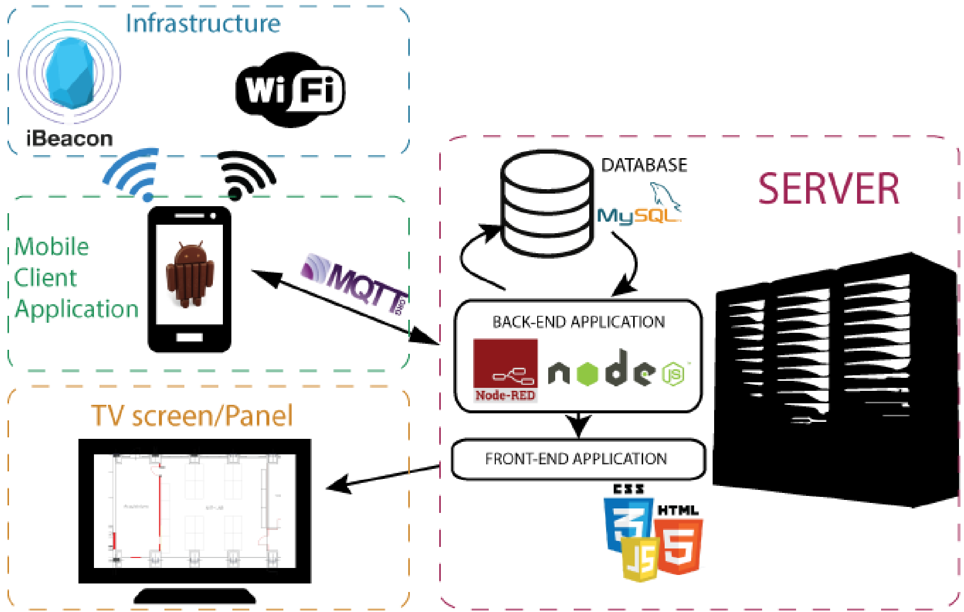

Fig. 2. Architecture of the proposed system

2) Wi-Fi infrastructure: For what concerns the Wi-Fi infrastructure, we leverage the network of access points already installed in the proximity of the library to implement the fingerprint-based localization system. As explained in Section III-B, we create fingerprints by measuring RSS from beacons emitted by L different access points. However, the number of access points covering the library may be greater than L: operatively, we chose the L access points with strongest average observed RSS and we store their BSSID in a whitelist which is disseminated to each client device. In the online localization phase, if a client device does not receive beacons from one of the L access points, a value of $s_{\min} = -80$ dBm is inserted in the query vector before transmission to the database. At the same time, all beacons received from access points not present in the whitelist are discarded.

B. Mobile Client Application

Each student in the library uses its smartphone to access the localization service. To this end, we develop an Android application which is responsible for measuring the signals coming from BLE beacons and Wi-Fi access points for localization purposes and to offer the student smart library services. At start-up, the application checks the network interfaces enabled on the device and asks the user to activate at least the Wi-Fi interface, if not already active. If the user also activates the BLE interface, all localization techniques presented in Section III can be performed. After that, the application connects to the public back-end server via Wi-Fi or 3G/LTE protocol and downloads the Wi-Fi access point whitelist and other useful application data such as the PCA transformation matrix $\mathbf{T}$, and is ready to transmit to the server the Wi-FI or BLE measurements for localization. Data transfer between the client and the server is done through the Message Queuing Telemetry Transport (MQTT) protocol, which can be considered the de-facto standard for low-power IoT and Machine-to-Machine (M2M) applications due to its simplicity, efficiency and scalability. MQTT works according to the publish/subscribe pattern: each device can publish messages with a particular topic, and only subscribers of that topic will receive the messages. All messages pass through an MQTT broker (the back-end server in our case), which filters messages based on their topics. To publish a message, a single TCP/IP connection from a client to the broker is required, without the need of knowing the set of subscribers. In our application the client device publishes packets with two topics: *measurements* and *whitelist*. The former packet contains the Wi-Fi and BLE measurements made by the mobile device and a *study subject* field which contains the textual description of the subject under study at the student's desk. The latter MQTT packet is used by the client device to retrieve from the server the whitelist of access points to be included in the *measurements* packet.

C. Back-end and Front-end Server Application

The back-end server is the core of the system and it is responsible of (i) running the localization engine and (ii) managing the smart library service. The back-end system is entirely implemented using IBM Node-RED, a visual tool built on top of the Node.js server-side framework [18] that has recently become very popular in the development of IoT applications, due to its flexibility in creating quick software prototypes. Node-RED is an event-processing engine that already includes building blocks for MQTT and MySQL connections, as well as the possibility of writing ad-hoc Javascript functions to process the incoming/outgoing data, thus greatly simplifying the development process. As shown in Figure IV, the back-end server acts as a MQTT broker for mobile clients: upon reception of a MQTT *measurements* message from a client, the server runs the localization engine to estimate the position of the client. If Wi-Fi localization is performed, the back-end server performs customized queries on the MySQL database of fingerprints to estimate the user's location. Once the location is computed, the server uses the *study subject* field of the *measurements* message to update the front-end, a simple HTML/CSS web page which features a map of the library with rooms and desks highlighted with occupancy information and topic under study. The front-end page can be displayed on a big screen at the entrance of the library or directly from students on their mobile phone, thus providing quick information on the topics under study in the library and their location, the location of free desks, etc.

V. Experiments

To evaluate the performance of the system in a realistic indoor scenario, we have implemented a proof-of-concept in a 100 m^2 indoor space of our university. The space has 23 desks positioned as shown in Figure 5: each desk is 1.2 m long and 0.5 m wide, and is used by one single student/researcher. For what concerns the Bluetooth infrastructure, we deployed one BLE beacon per desk, setting its transmission power to the lowest value of -70 dBm. As for the Wi-Fi infrastructure, the space is covered by 18 different access points deployed in its proximity, whose BSSID are stored in a whitelist which is disseminated to each client. The mobile device used for the experiments is a Samsung Galaxy Ace Style LTE G357, with BLE communication running Android 4.4.4 (KitKat). On the device we installed the mobile client application described in Section IV-B.

A. Scenario and accuracy metrics

The experiments are performed as follows: for each one of the 23 desks, the mobile application is used to measure the received signal strength from all whitelisted access points and BLE beacon devices. To mimic a realistic scenario the measurements are performed during working hours, when desks in the indoor space are occupied by users. For each desk, 40 measurements are performed for a total of 920 measurements: each measurement lasts one second and two consecutive measurements are spaced by 100ms. Multiple received signal strength measurements from the same Wi-Fi access point or BLE beacon device are averaged during the same one-second measurement interval. Finally, each measurement is also labeled with its ground truth position $\mathbf{x}_{\text{GT}}$. We divide the complete set of 920 measurements in two sets: 70% of the data is used for training purposes and 30% for testing. The test set is used to compute an *accuracy* metric defined as:

$$A = \frac{1}{N}\sum_{i=1}^{N} e_i, \tag{4}$$

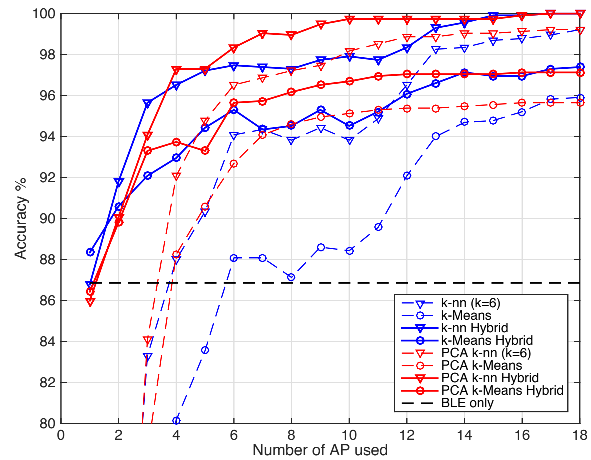

Fig. 3. Accuracy of different localization methods on the test set

where e_i is defined as:

$$e_i = \begin{cases} 1 & \text{if } \hat{\mathbf{x}} = \mathbf{x}_{\text{GT}} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

and N is the number of queries in the test set. The training data is used for building the fingerprint database and estimate the conditional probabilities $P(E \mid \hat{\mathbf{x}}_{\text{BLE}})$ and $P(E \mid \hat{\mathbf{x}}_{\text{Wi-Fi}})$. To estimate these probabilities, the training set is further divided in two sets (again in 70:30 proportion) and the smaller set is used as a new test set for computing the conditional probabilities as:

$$P(E \mid \hat{\mathbf{x}}) = 1 - A(\hat{\mathbf{x}}), \tag{6}$$

where $A(\hat{\mathbf{x}})$ denotes the accuracy measure computed only for those locations estimated as $\hat{\mathbf{x}}$ by the localization system.

B. Performance comparison

Figure 3 shows the performance of the different localization methods as computed on the test set, where we varied the number of access points or principal components to be used for fingerprints, i.e., the fingerprint vector dimension. Such dimension impacts on the storage requirements of the fingerprint database and the time needed for estimating a location using k-NN based approaches, and it is therefore good practice to keep it low. For methods in the original RSS domain (shown as blue curves), we sorted the access points in descending order of their average signal strength, while for PCA-based methods (shown in red) the principal components are naturally sorted. For all methods based on k-NN matching, we only show the curve corresponding to the best performance, which in our case is obtained for k=6.

From the inspection of Figure 3, several considerations can be made:

1) we observe that the performance of the bluetooth-only system are surprisingly good, with a correct position returned almost 87% of the times. Considering the size of

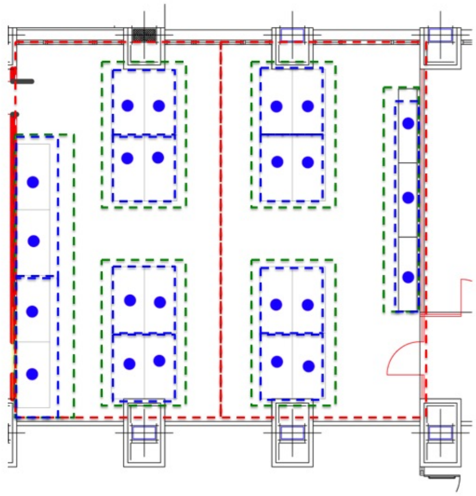

Fig. 4. The indoor area used for the experiments: 23 desks are present (each one with a BLE beacon), represented as blue dots. The three additional scenarios composed of 11, 6 and 2 locations are represented as well, in blue, green and red dashed lines, respectively.

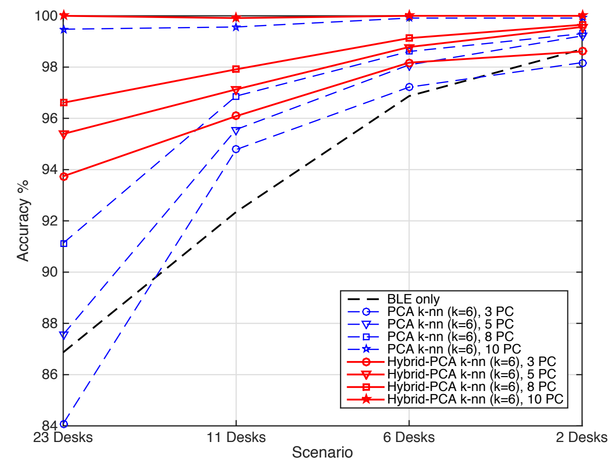

Fig. 5. Accuracy of different localization methods on the test set

the desks, this means that the average localization error of the BLE-only system is below 1.2 meters 87% of the times.

2) the accuracy of Wi-Fi methods generally increase as the size of the fingerprints increase. For methods in the RSS domain, the accuracy tends to saturate after 13-14 used access points, while for PCA-based methods 8-10 principal components are enough to obtain the maximum performance.
3) PCA-based methods outperform methods in the original RSS domain in terms of maximum accuracy for small size of the fingerprints. The gap is as high as 8% for k-Means approaches, while it is limited to 4% for approaches based on k-NN. This means that the PCA transformation learnt from the available data is able to identify the most important components and to filter out noisy and correlated ones.
4) k-nearest neighbours matching generally outperforms k-Means based matching in terms of localization accuracy. The gap can be as high as 6% for small size of the fingerprints, while it is limited to 4% for high-dimensional fingerprints. The drawback of k-NN methods compared to k-Means is their high complexity: this issue should be taken into account when using large fingerprint databases and when computational resources on the central server are scarce.
5) Finally, hybrid methods always show the best performance, demonstrating that fusing together the information coming from the two independent systems allows for almost perfect location estimation. In particular, the hybrid method obtaining the best performance is the one fusing BLE with PCA fingerprints and k-NN matching, which allows to obtain almost perfect location estimation (99.74%) with as few as 10 principal components.

In addition to testing the scenario composed by 23 desks, we also analyze how the localization accuracy varies when grouping together different nearby desks: this allows to obtain some insights on the performance of the system when the spatial granularity of the locations in $\mathcal{X}$ varies. In particular we tested three additional scenarios in which the number of locations in $\mathcal{X}$ is varied in the range $\{23,11,6,2\}$ as illustrated in Figure 4. Figure 5 reports the performance of the BLE-only system, the 6-NN PCA Wi-Fi fingerprint system and the 6-NN PCA hybrid system on these new scenarios. For the two latter methods, we computed the performance when using 3,5,8 or 10 principal components in the fingerprints. As expected, the performance of all systems increase as the number of locations in $\mathcal{X}$ is decreased. In particular, the hybrid approach always allows to obtain an accuracy equal or greater than 94% for all tested scenarios even using as low as 3 principal components in the fingerprint. When using 10 principal components, perfect recognition is obtained for all scenarios.

C. Online fingerprint collection

As explained before, the system requires an offline phase in which the Wi-Fi fingerprint database is created. Such a cumbersome step should be repeated from time to time to ensure that the information in the database is up-to-date compared with the localization queries coming from the mobile devices. As an example, in Figure 6 we compare the performance of the best performing hybrid localization system (PCA 6-NN) obtained in two cases in which the information in the fingerprint database is up-to-date or 2 weeks old. During these two weeks, the indoor space has been subject to some changes in terms of furniture and people usually occupying it. These changes impact on the accuracy of Wi-Fi localization, with a clear decrease in the accuracy, as high as 70%. Instead of re-populating the database with updated groundtruth fingerprints, we use a new database obtained with the fingerprints coming from the hybrid localization system. Although these fingerprints are affected by the error of the bluetooth-only system,

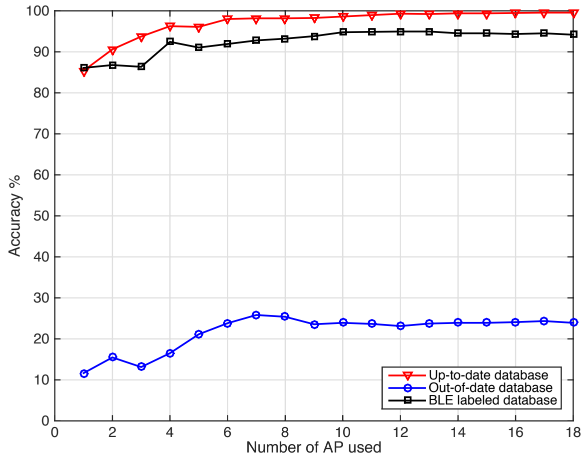

Fig. 6. Accuracy of hybrid localization when using different Wi-Fi fingerprints database: up-to-date, out-of-date and BLE labeled

Figure 6 shows that using such an approach allows to avoid the cumbersome procedure of repopulating the database, still achieving acceptable accuracy performance.

VI. Conclusions

We have presented a system for the creation of study groups in a smart library. The core of the system is an hybrid BLE and Wi-Fi indoor localization system able to work in different configuration, namely BLE-only, Wi-FI-only or fusing the information coming from the two technologies. The complete system has been implemented and evaluated in a realistic scenario, demonstrating promising performance for the task at hand. Future research directions will investigate the client-side and server-side performance (e.g., energy consumption and computational overhead of the application running on the mobile device, end-to-end delay, etc..) as well as a large-scale deployment in a real library.

References

[1] R. K. Sawyer and S. Berson, "Study group discourse: How external representations affect collaborative conversation," *Linguistics and Education*, vol. 15, no. 4, pp. 387 – 412, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0898589805000288

[2] D. W. Johnson and R. T. Johnson, "Instructional goal structure: Cooperative, competitive, or individualistic," *Review of Educational Research*, vol. 44, no. 2, pp. 213–240, 1974.

[3] S. Bhardwaj, T. Ozcelebi, R. Verhoeven, and J. Lukkien, "Smart indoor solid state lighting based on a novel illumination model and implementation," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1612–1621, November 2011.

[4] A. d'Alessandro, A. Buffi, P. Nepa, and G. Isola, "A localization technique for smart bookshelves based on uhf-rfid systems," in *Antennas and Propagation Society International Symposium (APSURSI), 2013 IEEE*, July 2013, pp. 1112–1113.

[5] C. Peng, K. Qian, and C. Wang, "Design and application of a voc-monitoring system based on a zigbee wireless sensor network," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2255–2268, April 2015.

[6] Z. Wang and L. Wang, "Intelligent control of ventilation system for energy-efficient buildings with co2 predictive model," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 686–693, June 2013.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[8] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 13–32, First 2009.

[9] A. Redondi, M. Chirico, L. Borsani, M. Cesana, and M. Tagliasacchi, "An integrated system based on wireless sensor networks for patient monitoring, localization and tracking," *Ad Hoc Netw.*, vol. 11, no. 1, pp. 39–53, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.adhoc.2012.04.006

[10] S. He and S. H. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 466–490, Firstquarter 2016.

[11] M. Cooper, J. Biehl, G. Filby, and S. Kratz, "Loco: boosting for indoor location classification combining wi-fi and ble," *Personal and Ubiquitous Computing*, vol. 20, no. 1, pp. 83–96, 2016. [Online]. Available: http://dx.doi.org/10.1007/s00779-015-0899-z

[12] A. Baniukevic, C. S. Jensen, and H. Lu, "Hybrid indoor positioning with wi-fi and bluetooth: Architecture and performance," in *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on*, vol. 1, June 2013, pp. 207–216.

[13] J. Zhu, K. Zeng, K. H. Kim, and P. Mohapatra, "Improving crowd-sourced wi-fi localization systems using bluetooth beacons," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012 9th Annual IEEE Communications Society Conference on*, June 2012, pp. 290–298.

[14] P. Mirowski, T. K. Ho, S. Yi, and M. MacDonald, "Signalslam: Simultaneous localization and mapping with mixed wifi, bluetooth, lte and magnetic signals," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, Oct 2013, pp. 1–10.

[15] Insiteo, http://www.insiteo.com.

[16] Wifarer, http://www.wifarer.com.

[17] Indoo.rs, http://indoo.rs.

[18] L. M. Surhone, M. T. Tennoe, and S. F. Henssonow, *Node.Js*. Mauritius: Betascript Publishing, 2010.