WiFi Received Signal Strength (RSS) Based Automated Attendance System for Educational Institutions

Sidratul Muntaha Khan
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1905009@cse.buet.ac.bd

Mehreen Tabassum Maliha
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1905078@ugrad.cse.buet.ac.bd

Md Shahedul Haque
Department of Computer Science
Virginia Polytechnic Institute and State University
Arlington, USA
mdshahedul@vt.edu

Ashikur Rahman
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
ashikur@cse.buet.ac.bd

Abstract

Smartphones are becoming part of people's day-to-day activities now-a-days. Globally, almost 90% of cellular phones are smartphones. "Personnel tracking" is a viable usage of smartphones. Automated attendance system, an application of personnel tracking, is efficient and essential for enhancing productivity and streamlining operations in modern workplaces and educational institutions. While traditional methods of attendance tracking are prone to inaccuracies and inefficiencies, smartphone-based systems offer a smoother approach. In this paper, we present a smartphone-based attendance system that leverages Wi-Fi signal strength for indoor localization. Instead of requiring precise positioning, we propose a system with zone-based localization approach. We divide the whole area into distinct smaller zones and determine the users' location within these zones. Through real-world deployment, we demonstrate that our novel approach reduces the complexity of exact positioning while still achieving high accuracy in identifying a user's presence within specific areas, which are often enclosed by boundaries. Comparative evaluations with implementing an algorithm show the superiority of our proposed method in terms of accuracy and practicality, making it suitable for deployment in large-scale organizational settings.

CCS Concepts

• **Human-centered computing** → **Smartphones**; • **Networks** → *Location based services.*

Keywords

Attendance System, Wi-Fi, RSS, Localization, Smartphone App

NSysS '24, December 19–21, 2024, Khulna, Bangladesh

ACM ISBN 979-8-4007-1158-9/24/12
https://doi.org/10.1145/3704522.3704523

ACM Reference Format:
Sidratul Muntaha Khan, Mehreen Tabassum Maliha, Md Shahedul Haque, and Ashikur Rahman. 2024. WiFi Received Signal Strength (RSS) Based Automated Attendance System for Educational Institutions. In *11th International Conference on Networking, Systems, and Security (NSysS '24), December 19–21, 2024, Khulna, Bangladesh.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3704522.3704523

1 Introduction

Accurate attendance tracking plays a vital role in various settings, particularly in educational institutions and workplaces. In schools and universities, attendance often directly impacts students' academic performance and compliance with institutional policies. Traditional methods for taking attendance, such as manual roll calls or signing in an attendance register, are not only time-consuming but also non-scalable. As organizations grow larger or classes become more populated, these manual methods become increasingly time-consuming and complex, leading to lost time and inaccuracy in record-keeping. Also vulnerability exists as there are ways to exploit the manual attendance system by providing proxy attendance or back-date signing. To overcome these aforementioned challenges, an automated attendance system is a viable alternative that provides an effective solution.

Researchers have proposed different approaches for building automated attendance system. Some of them propose using RFID([17]), Bluetooth ([5]), or Wi-Fi ([2]) as the communication technology; some of them argue for Biometric features such as fingerprints [1], human speeches [7], or even facial images [22]. Existing automated systems, such as RFID-based solutions, bluetooth-based approaches, and biometric scanners, require expensive infrastructure to be installed in the workplace or academic institutes. More specifically, RFID-based attendance systems face several challenges due to high implementation costs, the risk of tag loss or damage, limited detection range, and potential security issues like tag cloning. On the other hand, Bluetooth-based attendance system might not work correctly for the larger classrooms as its range is very limited and not even anywhere near the range of WiFi signals. An average Bluetooth signal range is approximately 30 feet, whereas an average 2.4GHz Wi-Fi router can cover a 150 feet area in all directions in an indoor environment. Last but not least, Biometric based approaches suffer from computation overhead since they need more time to check attendance than other available alternatives. Also, these attendance systems might be hacked or hijacked, raising concerns related to privacy and data security.

To overcome the challenges faced by the above-mentioned systems, a Wi-Fi based system utilizes existing infrastructure, such as Wi-Fi networks, minimizing the need for additional hardware and providing a seamless and automated approach to attendance management as demonstrated in [9], [12], [2], and many more. Eventually, Wi-Fi-based systems have become cost-effective solution across a range of applications, including indoor positioning, monitoring, and communication. One of the most significant advantages of these systems is their ability to leverage existing Wi-Fi infrastructure, significantly reducing the need for additional hardware and associated maintenance costs [11]. Wi-Fi networks help to offer wide coverage, providing consistent services across extensive areas in a bounded region such as a home [8]. Another noteworthy benefit is their utility for location-based services. Wi-Fi signals enable effective *indoor positioning* through techniques such as signal fingerprinting, allowing for scalable and accurate tracking solutions in environments where GPS is unreliable [4]. Moreover, employees and students are already equipped with smartphones and/or laptops or personal digital assistance (PDA), which already has a Wi-Fi interface to connect to the automated attendance system. A recent study shows globally, almost 90% of cellular phones are smartphones [15].

Despite the advantages of existing Wi-Fi-based systems, several limitations remain at work. One of the challenges is accurately distinguishing users' presence within a confined boundary (e.g. within a classroom, or a office space) as discussed in the survey [18]. Variations in Wi-Fi signal strengths are affected by environmental factors like room layout, physical obstacles etc., which lead to error-prone results in determining users' exact locations. As a result, many false positives (user is outside the boundary, but classified as inside of it) or false negatives (user is inside the boundary, but classified as outside of it) might be triggered in the attendance records. Moreover, Wi-Fi signals often vary with time due to changes in network load conditions, interference generated from nearby devices, or due to human movements, which further complicates precise localization.

Wi-Fi RSSI (Received Signal Strength Indicator)-based attendance systems such as [2], [20], etc. have gained attraction and trust because their initial setup cost is less and they promise privacy, and data security. These systems gained significant attention in recent years as it can be built upon existing wireless networking infrastructure with minimal additional costs. However, they lack precise validation in class-room boundary regions, where vulnerabilities may permit cheating in attendance checks. To this extent, we propose an interesting idea of Wi-Fi fingerprinting based indoor localization [13], focusing on precisely verifying presence within specific boundaries like classrooms or offices. Our work focus on eliminating privacy risks, proxy attendance, and minimizing disruptions during the class lecture. In addition, employing these algorithms has the potential to enhance attendance accuracy and mitigate loopholes in boundary areas.

The main contributions of this paper are as follows:

(a) We combine Wi-Fi fingerprint based indoor localization, focusing on accurately determining presence within specific boundaries.
(b) Instead of determining the precise location of the user, we focus on confirming if the user is inside the designated area or not.
(c) We develop a testbed and use it to conduct extensive real-life experiments. The results demonstrate the efficacy and accuracy of our system. We observe that our system is robust in determining the presence of users in the concerned areas.

The rest of this article is organized as follows: In Section 2, we present related work. In Section 3, we provide a detailed overview of the proposed system and formulate the problem. In Section 4, we focus on the system methodology and design. We present the experiment and result in Section 5. In Section 6, we discuss future work and draw our conclusions.

2 Related Works

There has been ample studies related to automated attendance systems with a variety of techniques. In this section, we briefly introduce state-of-the-art attendance systems based on RFID, bluetooth, biometric attributes, and WiFi.

RFID Based: The attendance systems proposed in [17] and [6] introduce RFID based attendance system. In this system, readers automatically record and track attendance by scanning RFID tags. However, RFID-based attendance systems incur high implementation costs, pose the risk of tag loss or damage, suffer from limited detection range, and raise potential security issues like tag cloning.

Bluetooth Based: Bluetooth devices installed around the classroom for automated attendance techniques is proposed by [16]. Similar to RFID, bluetooth-based systems also suffer high initial setup cost and low detection range.

Bio-metric Based: Large scale organization implements bio-metric attendance system for example, fingerprint based system mentioned in [1], [23], [19], [14], etc. Speech based attendance system is also explored in [7]. Even some companies employ facial detection based attendance system as in [22] or [10]. Compared to RFID and bluetooth-based solutions, biometric systems need more additional cost for setup and require extensive maintenance. Furthermore, they have computational overhead while registering attendance. Also, there are possibilities to exploit these attendance system.

Wi-Fi Based: Anand et al. propose an attendance technique with the help of Wi-Fi in [2]. Here, a WiFi access point manages an access point. When a person is close enough to the manager, the attendees will receive the signal of that access point, and thus his/her attendance will be confirmed. But this does not work for a classroom, as the teacher's access point can also be visible outside the classroom, though it may be weaker. So, a student can get attendance from outside the room boundary. Another work by [20] focuses on attendance monitoring with the help of WiFi access points and their corresponding signal strengths. Here, every location has a calculated coordinates and a list of APs with their signal strengths. When a test data arrives with an AP list with RSS values, they calculate the top 3 nearest points according to the RSS values with the help of Euclidean distance. Then, the selected location coordinates are averaged, which are test point's coordinates. Then, the coordinate is checked, whether it is inside or outside the classroom boundary region.

Another optimized idea for classroom based attendance comes from indoor localization techniques. For example, Anzum et al. ([3]) utilizes the idea of WiFi access points (APs) and their corresponding received signal strength (RSS) values to automatically track people or objects indoors. Similarly, Schepers et al. propose privacy preserving positioning with WiFi [21].

3 Problem Formulation and System Overview

In this section, we present a formal description of the problem that we aim to solve in this paper. In addition, we provide an overview of the proposed system architecture.

3.1 Problem Formulation

Consider an area of interest which is a classroom and the strip areas right by the perimeters. The presence of students inside the room needs to be detected. Ignoring the height of the classroom, it can be viewed as a 2-dimensional grid.

Formally, we are given a room A and its nearby perimeter areas that is virtually divided into a set of $m \times n$ grid cells (each cell is called a zone) consisting of m rows and n columns. Thus the zones $G = \{g_{11}, g_{12}, \ldots, g_{mn}\}$ cover both inside and nearby outside area of the room. Moreover the entire area has wi-Fi coverage with a set of q stationary access points $AP = \{AP_1, AP_2, \ldots, AP_q\}$.

3.1.1 Fingerprinting: Wi-Fi fingerprinting involves collecting and assigning to each grid point $g_{ij} \in G$ a set of q different access points along with their received signal strength (RSS) values. The number of access points detected can vary at each scan and at different grid points. If an access point is not present at a grid location during a scan, we assign that access point a minimum RSS value. Thus, every grid point g_{ij} has a Wi-Fi signal vector that looks like:

$$V_{\text{ref}_{ij}} = \{AP_1 : RSS_1^{ij}, AP_2 : RSS_2^{ij}, AP_3 : RSS_3^{ij}, \ldots, AP_q : RSS_q^{ij}\}$$

Thus, the entire area simply can be represented by a set of reference vectors, denoted by R (covering all the grid points):

$$R = \{V_{\text{ref}_{11}}, V_{\text{ref}_{12}}, \ldots, V_{\text{ref}_{mn}}\}$$

In other words, each vector of a grid point acts as a virtual q-dimensional coordinate of that grid point. Therefore, the room, which is surrounded by boundaries, is divided into suitable grid points both inside and outside, with each grid point having q-dimensional virtual coordinates.

3.1.2 Attendance Checking: After fingerprinting, when the real-time attendance system is in operation, the main objective is to solve a binary classification problem that determines whether a new smart device (SD) is inside the polygonal boundary region or not. This, in turn, indicates if the student is inside the designated classroom or not. Given a new WiFi signal vector (V_a) observed by a user's SD, $V_a = \{AP_1 : RSS_1, AP_2 : RSS_2, AP_3 : RSS_3, \ldots, AP_q : RSS_q\}$ the system determines the most matching k grid points $\{1 \leq k \leq m\}$ according to the matching algorithm. This algorithm compares V_a with all the vectors in the previously mentioned set of reference vectors, namely R, and selects the k-closest vectors:

$$R_{\text{selected}} = \{V_{\text{ref}_1}, V_{\text{ref}_2}, \ldots, V_{\text{ref}_k}\}$$

The selected vectors correspond to k points. If the majority of the selected points are inside the polygonal boundary, the new SD is classified as being inside the boundary. If the majority of the selected points are outside, the SD is classified as being outside the boundary. Thus, using the matching technique for virtual coordinate vectors, the binary classification problem for determining attendance is solved.

3.2 System Overview

We utilize the traditional fingerprinting approach to implement the attendance system. Similar to other fingerprinting approaches, the zone-based localization consists of two phases: (1) an offline profiling phase, and (2) an online localization phase.

(a) **Profiling Phase:** During the profiling phase, a central fingerprint database is maintained to store vectors of Wi-Fi access points (APs), also known as SSIDs (Service Set Identifiers), along with their Received Signal Strength (RSS). A smart device (SD) is taken to all selected grid cells (zones) throughout the area of interest, and a certain number of signal strengths along with the access point names are captured by placing the SD at the centroid of every grid for a period of time. These captured signal vectors, along with the identification number of the grid points, are sent to the central database. Instead of capturing any real (x, y) coordinates, the vector captured at each centroid of the grids is represented as the vector coordinate of that point. These vector coordinates are captured both inside and outside the room.

(b) **Localization Phase:** Localization is performed in real-time. In this phase, a new SD collects real-time Received Signal Strength (RSS) values with corresponding SSIDs and sends them to the server. Then, the server compares the vector comprising the RSS values with the available fingerprint database using a pattern matching algorithm. The result of the algorithm determines whether the new SD is inside or outside the boundary of the zone. Thus, the final problem converges into a binary classification problem.

The next subsection formally describes these two phases in details.

4 Proposed Methodology

Our methodology for smartphone based attendance system with the help of Wi-Fi signal strength could be divided into two major parts.

4.1 Data Collection

4.1.1 Radio-map Creation. The first phase of developing the automated attendance collection system involves constructing a radio map of the classrooms. This process transforms the classroom environment into a virtual grid-based region, enabling precise data collection of Wi-Fi signal strength across predefined points. The signal strengths are measured at the center of each grid point, both inside and outside the classroom, to see how the Wi-Fi signal changes across different areas.

4.1.2 Fingerprint Data Generation. A custom mobile application is used to facilitate data collection. This application collects the list of following data at each grid point:

- **Wi-Fi Access Point (AP) Names:** An Access Point (AP) is either a hardware device or a software application that allows wireless devices, such as smartphones, laptops, and tablets, to connect to a network via Wi-Fi. Examples of APs can be routers, hotspots, repeaters, etc. In our context, we focus on standalone APs, specifically routers and repeaters. Although a wireless device can connect to only one AP at a time, it can still detect signals generating from multiple APs within its range. These signals include the names (SSIDs) of the detected APs.
- **Received Signal Strength (RSS):** RSS represents the strength of the Wi-Fi signal from each AP detected at a specific grid point. When a wireless device detects the signal of an AP, it also receives a quantitative value indicating the strength of the signal. This strength is measured in decibels relative to 1 milliwatt (dBm), where dBm is a logarithmic unit. The formula to convert power P (in milliwatts) to dBm is:

$$\text{dBm} = 10 \times \log_{10}(P)$$

RSS values are typically negative because the received signal is weaker than the reference value of 1 mW. For instance, a value of -30 dBm indicates a stronger signal than -80 dBm. Generally, RSS values can range from 0 dBm (ideally the strongest signal) to around -100 dBm (very weak signal). The interpretation of RSS values is as follows:

- Excellent signal: -30 dBm to -50 dBm.
- Good signal: -50 dBm to -60 dBm.
- Fair signal: -60 dBm to -70 dBm.
- Weak signal: -70 dBm to -75 or -80 dBm (may cause slow speeds or dropped connections).
- Very weak signal: below -80 dBm (signal might not be detectable by a smart device).

Several factors can influence RSS values. The further a device is from the access point, the weaker the signal becomes. Physical obstructions like walls, floors, and large objects can also attenuate the signal strength.

4.1.3 Handling External Factors. To ensure accuracy and consistency, data is collected multiple times at each grid point. Averaging these repeated measurements helps mitigate fluctuations caused by environmental factors, human interference, device variability and noise from transient devices. The data collection process covers multiple classrooms and laboratories within the university, utilizing different smartphones at various times of the day. This diversity in device type and time ensures that the collected data captures the dynamic nature of Wi-Fi signal fluctuations, thereby improving the robustness of the dataset. To implement the data collection, we divided a room and its nearby perimeter areas into a $m \times n$ grid points virtually. We define the center of each of these grid points inside and outside as **reference points** such as $r_i c_j$ where i is the row number and j is the column number. In each reference point, we have run the Android application and collected necessary data, which later is processed as fingerprint data for each reference point. Given that existing Wi-Fi routers are generally located in the corners of rooms, their current placement is suboptimal for this study. To enhance the quality of the dataset, additional access points are installed in central locations within the testbed classrooms. Once collected, the data is automatically transmitted to a cloud-based database for secure storage and further analysis.

4.1.4 Test Data Generation. In the testing phase, we conduct the same data collection procedure inside and outside at various random points. Each of these data collection points is called a **test point**.

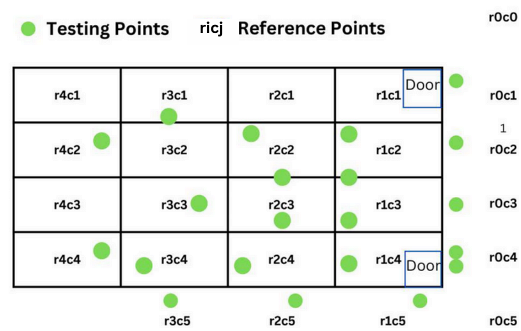

Figure 1: A Grid based Classroom. Here, $r_i c_j$ are reference points at $i-th$ row and $j-th$ column and green points are test points

From Figure 1, we can see a full radio-map for a classroom. Hypothetical grids are generated both inside and outside of the room according to available space. At the center of each grid, there is a **reference point** where fingerprint data is taken via our Android application. Also, green dots are the random **test points** where test data is taken via the same Android app.

4.2 Data Analysis

During the data collection process, it is expected that at certain grid points, not all wireless access points (APs) can be detected due to limited transmission range of access points and/or environmental obstacles such as walls or furniture. These incomplete data result in missing values (NaNs) for undetected APs. To handle these missing values, we have adopted a combination of predefined substitution and imputation techniques:

4.2.1 Data Pre-Processing. Initially, where an AP is undetected, the missing RSS values are set to a minimum possible value. The value is chosen as if it represents a signal level sufficiently weak to indicate the AP is likely outside the range but ensures that a numerical data is present for subsequent analysis.

In order to enhance dataset manageability and improve the overall accuracy of the model, SSIDs that are deemed less impactful, such as those corresponding to personal hotspots or transient networks are excluded from the dataset. This process ensures that only permanent access points contributing to the classroom environment are retained for analysis.

To maintain the integrity of the dataset, suitable outlier detection and removal techniques are employed. One such technique is Interquartile Ranging (IQR), which calculates the range between the 25th (Q1) and 75th (Q3) percentiles. In this method, outliers are points that lie below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$, where $IQR = Q3 - Q1$. Outliers, which could result from interference, momentary signal fluctuations, or device issues, are carefully identified and excluded from further analysis so that they do not mislead the original values. Now, for each reference point, we have a vector consisting of one set of pairs as a component of it. Each pair consists of an AP name and its signal strength, which looks like the following list in Table 1:

$\mathbf{AP_1}$	$\mathbf{AP_2}$	$\mathbf{AP_3}$	...	$\mathbf{AP_q}$
RSS_1	RSS_2	RSS_3	...	RSS_q

Table 1: RSS Values from Different Access Points (APs)

Then, for a room, we take the most frequently available access points that are present in most of the reference points. These selected access points with their averaged signal strength are the fingerprint data for each reference point in the room. For a room with m reference points and n selected access points, our fingerprint data looks like the following Table 2 :

	AP				
	$\mathbf{AP_1}$	$\mathbf{AP_2}$	$\mathbf{AP_3}$	...	$\mathbf{AP_q}$
$V_{\mathrm{ref}_{11}}$	RSS_1^{11}	RSS_2^{11}	RSS_3^{11}	...	RSS_q^{11}
$V_{\mathrm{ref}_{12}}$	RSS_1^{12}	RSS_2^{12}	RSS_3^{12}	...	RSS_q^{12}
...	...	...	...	...	...
$V_{\mathrm{ref}_{mn}}$	RSS_1^{mn}	RSS_2^{mn}	RSS_3^{mn}	...	RSS_q^{mn}

Table 2: Reference Points and RSS Values from Different Access Points (APs)

4.2.2 Implementing the Algorithm. Next, we provide an algorithm for attendance checking system which checks if a test point lies inside or outside the classroom boundary region. The algorithm is a binary classification algorithm. To determine the location classification of the test point, we use the Wi-Fi RSS-Based Location Classification algorithm (see Algorithm 1). This algorithm calculates distances, selects nearest neighbors, and performs majority voting to classify the test point as inside or outside the room. In our analysis, we have introduced several distance metrics to calculate the similarity between the test point and the reference points. These metrics include Euclidean distance, Manhattan distance, Cosine similarity, and Hamming distance, which are defined as follows:

- **Euclidean Distance** The Euclidean distance between two points $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ in an n-dimensional space is defined as the straight-line distance between them, and is given by:

$$d_{\mathrm{Euclidean}}(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{1}$$

- **Manhattan Distance** The Manhattan distance, also known as the L1 distance or taxicab distance, is the sum of the absolute differences of their corresponding coordinates, and is defined as:

$$d_{\mathrm{Manhattan}}(p, q) = \sum_{i=1}^{n} |p_i - q_i| \tag{2}$$

- **Cosine Similarity** The Cosine similarity measures the cosine of the angle between two vectors in an inner product

Algorithm 1 Wi-Fi RSS-Based Location Classification

Input:

R: Set of reference points $\{V_{\mathrm{ref}_{11}}, V_{\mathrm{ref}_{12}}, \dots, V_{\mathrm{ref}_{mn}}\}$ where each vector is with q number of (AP, RSS) pairs and the corresponding labels (Inside or Outside)

T_v: Test Vector $T_v = \{AP_1 : RSS_1, \quad AP_2 : RSS_2, \quad AP_3 : RSS_3, \dots, AP_q : RSS_q\}$ with (AP, RSS) pairs

k: Number of nearest reference points (odd integer)

Output: Classification of T as **inside** or **outside**

Initialize $D \leftarrow$ empty list

for each reference point $V_{\mathrm{ref}_{ij}}$ in R **do**
  $d \leftarrow$ calculate_distance$(T_v, V_{\mathrm{ref}_{ij}})$
  D.append$(d, \mathrm{label}(V_{\mathrm{ref}_{ij}}))$
end for

D.sort_by_distance()
$k\_nearest \leftarrow D[0 : k]$

$inside\_count \leftarrow 0$
$outside\_count \leftarrow 0$
for each point in $k\_nearest$ **do**
  if point.label == inside **then**
    $inside\_count \leftarrow inside\_count + 1$
  else
    $outside\_count \leftarrow outside\_count + 1$
  end if
end for

if inside_count > outside_count **then**
  Classify T as **inside**
else
  Classify T as **outside**
end if

space. It ranges from −1 to 1, where 1 indicates that the vectors are identical. The formula for Cosine similarity is:

$$\mathrm{Cosine}(p, q) = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}} \tag{3}$$

- **Hamming Distance** The Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. It is given by:

$$d_{\mathrm{Hamming}}(p, q) = \sum_{i=1}^{n} \delta(p_i, q_i) \tag{4}$$

where

$$\delta(p_i, q_i) = \begin{cases} 0, & \text{if } p_i = q_i \\ 1, & \text{if } p_i \neq q_i \end{cases}$$

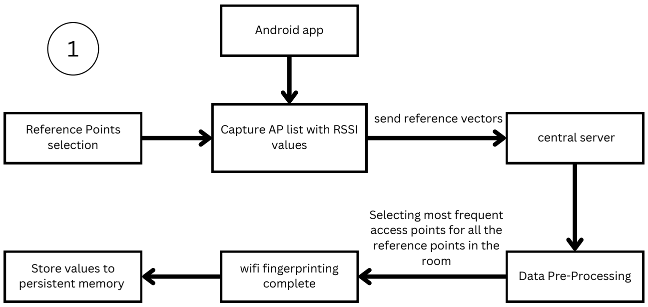

Figure 2: Fingerprinting Phase

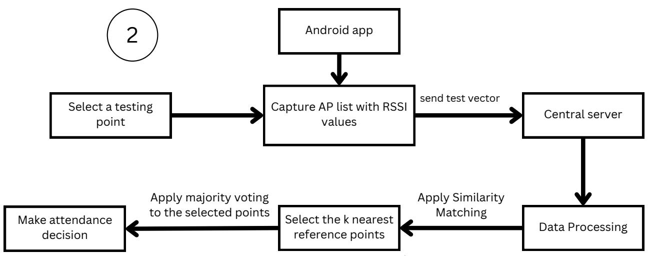

Figure 3: Testing Phase

By varying the distance metrics and number of most similar reference points, we have evaluated the accuracy of the test points.

In our study, we also compare the average signal strengths measured by different devices, keeping in mind that students will be using a variety of device models in the classroom. To reflect this, we present comparisons of the RSS values for selected access points at each reference point in the room across various devices. Additionally, since it is possible that some access points may fail or their signals may become unavailable during the attendance period, we test how well our system performs by simulating the removal of certain access points and measuring the resulting accuracy.

Here is a brief summary of our implementation illustrated in Figure 2 and Figure 3. In the first diagram, the process begins by selecting reference points and capturing RSS values through the Android app. These reference vectors are sent to a central server for data preprocessing, where the most frequent access points in a room are identified after outlier elimination and averaging the RSS values of an AP on a reference point. Once the Wi-Fi fingerprinting is complete, the data is stored in a persistent memory. The second diagram continues the process by selecting a test point and capturing a new list of RSS values. These are sent to the central server, where with the help of various distance metrics, similarity matching is applied to find the nearest reference points. A majority voting system is then used to decide the test point's proximity to reference points, allowing an attendance decision to be made, which in general solves a binary classification problem.

4.2.3 Real Life Implementation. While implementing this attendance system in real life, both teachers and students use an app to interact with Wi-Fi signals to record data.

As we can see from the Figure 4, First, the teacher logs into the app, selects the relevant room, course, and attendance period, and scans the available Wi-Fi access points (APs) within the classroom. This list of APs is sent to the server, which contains pre-stored fingerprint data. The server processes the teacher's AP list and compares it with the stored fingerprint list to create an intersection of available access points at that time, thereby identifying the Wi-Fi signals present in the classroom.

Simultaneously, students also log into the app, select their course, and scan the available Wi-Fi signals using the app. Their scanned data is sent to the server, which compares it against the reference points based on the proposed algorithm. Thus, the server marks the students' attendance. The result is then immediately sent to the students, indicating whether their attendance is complete. If a student's attendance is incomplete, they are asked to try again, provided the class time has not ended. At the end of the session, the teacher receives a complete attendance list for the class.

5 EXPERIMENT AND RESULTS

In this section, we present the detailed experiment and corresponding results obtained from a real testbed.

5.1 Experimental setup and technical details

As a testbed, we utilize the WLAN infrastructure of the ECE Building in Bangladesh University of Engineering and Technology (BUET). We have chosen two general classrooms of this building, which are room no 203 and 204. Let us mark these rooms as room A and room B, consecutively. For both rooms, we create a grid based radio-map inside and outside to get reference points. Data collection in these two rooms is performed independently.

There are plenty of Wi-Fi access points perceived in both of the rooms. But, to increase accuracy, we place an extra Wi-Fi source in the center of the room while conducting data collection in that room. We have divided the room A by a 4×4 grid inside (hence, $m = 4$ and $n = 4$ in the algorithm). Middle point of each grid is a reference point. Also, there are 9 reference points outside the room. Similarly, in room B, there are also 4×4 grid inside where 16 reference points are present. And there are 10 reference points outside. So, for room A, we get 25 reference points (16 inside + 9 outside). On the other hand, we get 26 reference points (16 inside + 10 outside) for room B.

To collect data for a room, in each test point, we run our custom build app, providing the reference point position and timer. Then, the app automatically collects Wi-Fi AP list with their corresponding signal strength. We collect signals 20 times at each reference point with a 20 seconds gap in between. These data are our reference point data.

5.1.1 Outlier Detection. For each reference point in a room, from the 20 set of data, we first remove outliers by applying Interquartile range outlier elimination techniques for each access point data and then average the remaining data.

5.1.2 Fingerprint Data Generation for a Room. For each reference point, we have one set of pairs. Each pair consists of an AP name and its average signal strength.

Then, for each room, we determine the most frequent access points, which are present in most of the reference points inside and outside of that room. Even if some reference points do not capture a Wi-Fi access point due to their weak signal strength there, we do not ignore the access point for that reference point. Instead, we

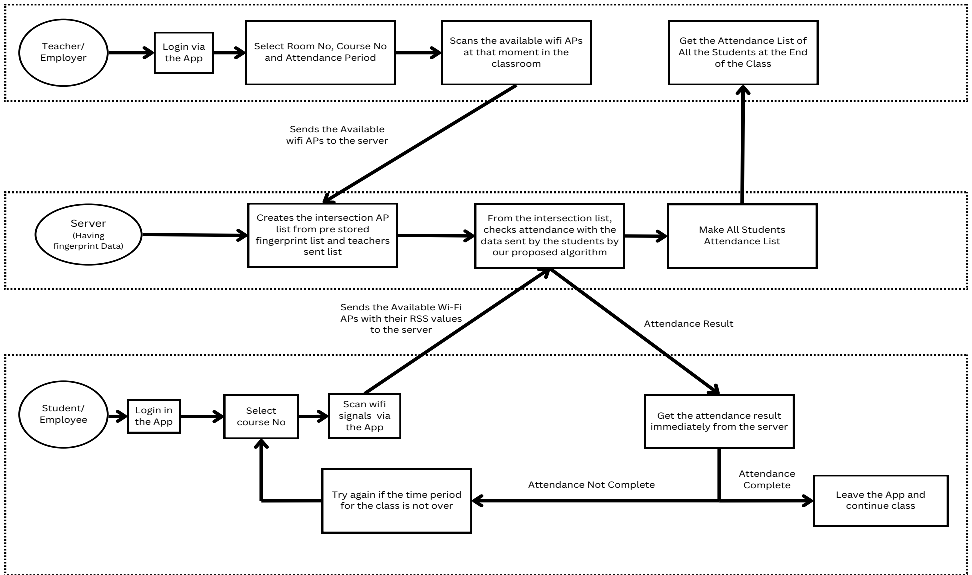

Figure 4: Automated attendance checking procedure in real life

assign them a minimum signal strength value of $-90dB$ for all the missing access points.

So, for a room, we have finally determined some fixed access points that are present in all the reference points. In this way, our reference dataset generation as well as fingerprinting is complete for a room.

5.2 Experimental Results

Once the reference dataset is ready, we utilize test samples to determine the accuracy of our algorithm. The test sample contains a list of Wi-Fi APs with RSS values. For a particular test point, the signal is collected multiple times. Then, for every access point, we exclude the outlier from its RSS values and then take the average values like the way we averaged RSS values for the reference points. Thus, for a test point, we also have a list of AP with their averaged RSS values.

Now, we run the algorithm to the test point against the reference points. Here, we take three different values of k eg. $3, 5, 7$ where k is the number of best matching reference points. We utilize the four different distance metrics, namely euclidean distance, manhattan distance, cosine similarity and hamming distance to calculate the distance between test point vector and every reference point vector. The algorithm here determines the k nearest reference points with respect to the test point. Next, we perform the majority voting to the k selected reference points to determine whether the test point is inside or outside the room.

From the results presented in Figure 5 and Figure 6, we can observe how different values of k and various distance metrics influence the accuracy of test points in both Room A and Room B. In Room A, for $k = 3$, the accuracy achieved with the Euclidean and Manhattan distance metrics is 87% and 92%, respectively, while cosine similarity also reaches 87%, and Hamming distance trails behind at 61%. With $k = 5$, Euclidean and Manhattan accuracies remain consistent at 87% and 92%, though cosine similarity drops to 77% and Hamming distance stays at 61%. For $k = 7$, the accuracy declines further for all metrics except Hamming (61%), with Euclidean at 82%, Manhattan at 84%, and cosine similarity at 69%.

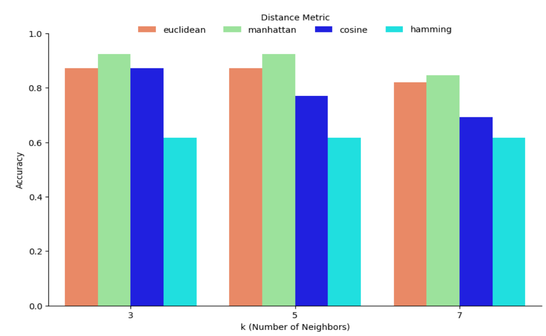

Figure 5: Accuracy of test data of Room A

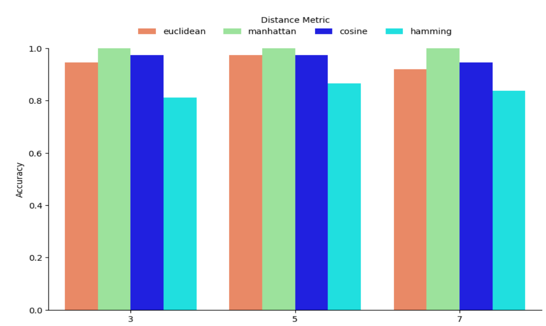

Figure 6: Accuracy of test data of Room B

Similarly, in Room B, the performance for $k = 3$ shows that Euclidean yields 94.6%, Manhattan gets 100%, cosine similarity achieves 97.3%, and Hamming distance provides 81%. With $k = 5$, accuracy for Euclidean and cosine metric remain steady at 97.3%, while Manhattan gets perfect accuracy (100%), and Hamming improves slightly to 86.5%. However, at $k = 7$, accuracy begins to decline across most metrics, with Euclidean at 91.9%, cosine at 94.6%, and Hamming at 83.8%, though Manhattan remains consistently perfect with 100% accuracy.

Overall, both $k = 3$ and $k = 5$ yield strong results for the Euclidean and Manhattan distance metrics, but Manhattan distance proves superior across the board, particularly for Room B where it achieves 100%accuracy at both values of k over around 40 test points. In Room A, Manhattan also performs well with 92% accuracy, misclassifying 3 out of about 40 test points. The poorer performance observed at $k = 7$ can likely be attributed to boundary conditions, where fewer reference points are available outside the room, limiting the number of suitable neighbors for the test points situated at outside. This suggests that higher k values, such as 7 or above, may not be optimal for this attendance system. Additionally, other distance metrics like cosine similarity and Hamming distance do not perform well in either room.

5.2.1 Missing Access Point. On a particular day, one or two access point(s) might be malfunctioning. We simulate this scenario and find how robust our system is. For this purpose, we deliberately remove one AP at a time and evaluate the overall accuracy using the remaining access points.

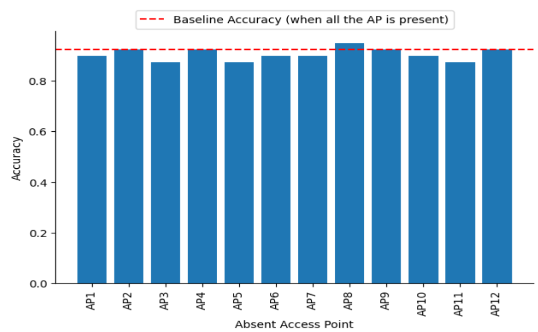

Figure 7: Accuracy of test data of Room A by removing one AP at a time

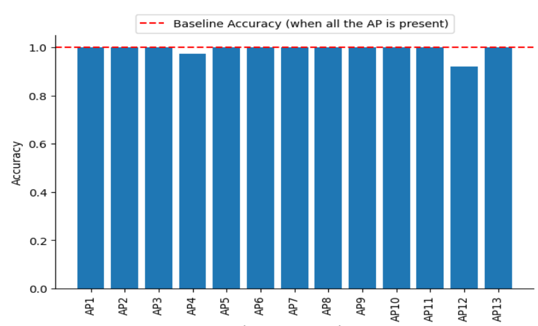

Figure 8: Accuracy of test data of Room B by removing one AP at a time

As shown in Figure 7, the performance remains relatively stable across most APs, with only slight fluctuations in accuracy. While the absence of some APs slightly decreases accuracy, the absence of one particular AP surprisingly results in the highest accuracy, even surpassing the baseline. Similarly, Figure 8 shows that removing a single AP has little impact on accuracy. Although accuracy decreases slightly when two APs are removed, for most APs, the accuracy remains unaltered. These findings suggest that during the attendance phase, if an AP fails, the built attendance system would still continue to operate with reasonable accuracy.

5.2.2 Device Heterogeneity. To observe the impact of device heterogeneity on RSS values, we collect the signal strength data using two different Android smartphones: a) Device Model RMX3363, and b) Device Model SM-M127G. Signals are captured at various points using both devices, and some of the results are displayed in the following figures.

As seen in Figure 9, while the average signal values from the two devices are not identical, they exhibit similar patterns. The relative signal strength across all access points remains almost consistent between the two devices. For example, if a signal is weak at a specific point, it is weak on both the devices, and the same argument holds true for stronger signals. These findings suggest that our proposed method is robust enough to handle heterogeneity of user devices effectively.

6 Conclusion and Future Work

We present a Wi-Fi RSS-based automated attendance system for accurate and efficient attendance tracking within a designated area while addressing the shortcomings of previous works. Our approach leverages the unique characteristics of Wi-Fi signal strength to identify the presence of individuals within the concerned area, providing a reliable alternative to traditional attendance methods. The results demonstrate that this proposed algorithm effectively balances accuracy and computational efficiency, making it an efficient solution for real-time attendance monitoring. We are planning to enhance our system to distinguish attendance given from areas above or below the specified room in the future. Currently, our application is developed exclusively for Android devices, but we aim to expand support for iOS users as well as for laptops and tablets. Also, combining crowd-sensing techniques in the decision making process regarding the presence of users in the designated area can enhance the overall efficiency of the attendance system.

References

[1] CO et al. Akinduyite. 2013. Fingerprint-based attendance management system. *Journal of Computer Sciences and Applications* 1, 5 (2013).
[2] S et al. Anand. [n. d.]. Attendance monitoring in classroom using smartphone & Wi-Fi fingerprinting. In *2016 IEEE Eighth Int. Conf. on Technol. for Educ. (T4E)*.
[3] Nafisa Anzum, Syeda Farzia Afroze, and Ashikur Rahman. 2018. Zone-based indoor localization using neural networks: A view from a real testbed. In *2018 IEEE Int. Conf. on Communications (ICC)*. IEEE.
[4] MM Atia, A Noureldin, J Georgy, and M Korenberg. 2011. Bayesian Filtering Based WiFi/INS Integrated Navigation Solution for GPS-Denied Environments. *Navigation* 58, 2 (2011).
[5] Vishal et al. Bhalla. 2013. Bluetooth based attendance management system. *Int. Journal of Innovations in Engineering and Technol. (IJIET)* 3, 1 (2013).
[6] O.G. Chiagozie and O.G. Nwaji. 2012. Radio frequency identification (RFID) based attendance system with automatic door unit. *Acad. Res. Int.* 2, 2 (2012).
[7] Subhadeep et al. Dey. 2014. Speech biometric based attendance system. In *2014 twentieth national conf. on communications (NCC)*. IEEE.
[8] A. ElShafee and K.A. Hamed. 2012. Design and implementation of a WIFI based home automation system. *Int. J. Comput. Inf. Eng.* 6, 8 (2012).
[9] B Ganesh. 2019. WIFI BASED ATTENDANCE SYSTEM. (2019).

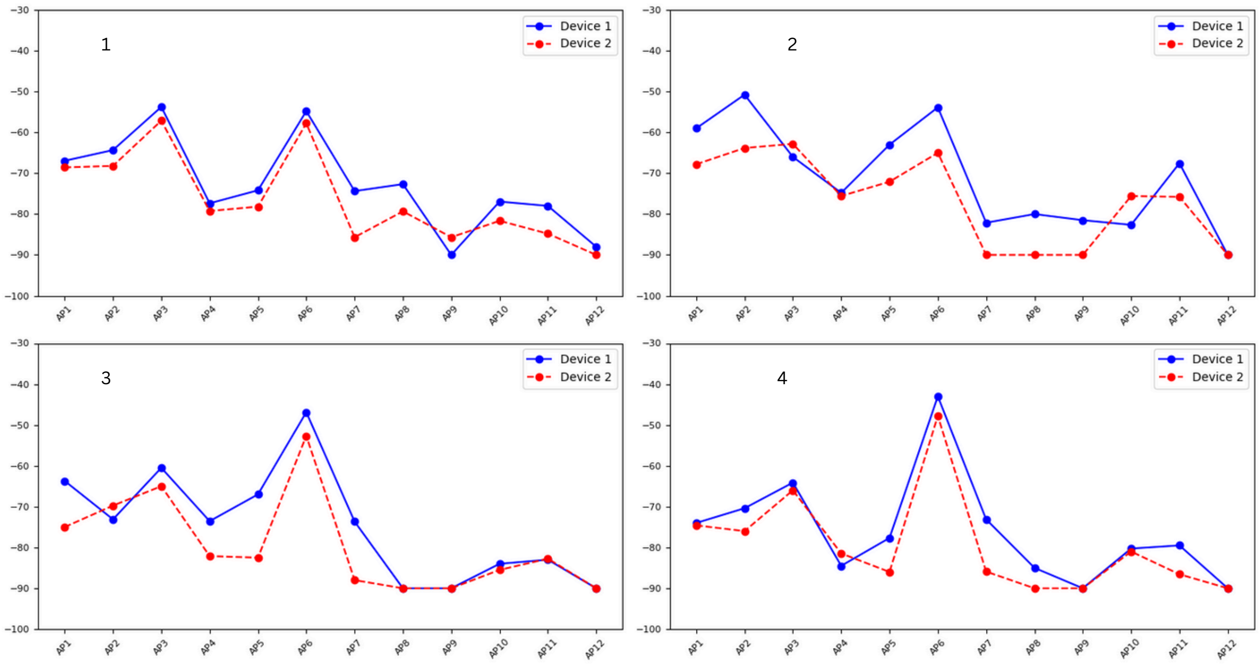

Figure 9: Signal Strength Variations for different APs at 4 different points.

[10] Clyde et al. Gomes. 2020. Class attendance management system using facial recognition. In *ITM Web of Conf.*, Vol. 32. EDP Sciences.
[11] V. Gunasekaran and F.C. Harmantzis. 2008. Towards a Wi-Fi ecosystem: Technol. integration and emerging service models. *Telecommun. Policy* 32, 3-4 (2008).
[12] Mahadi et al. Hasan. 2019. Bssid based monitoring class attendance system using wifi. In *2019 Third Int. conf. on I-SMAC*. IEEE.
[13] S-H Gary He, Suining Chan. 2015. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Commun. Surveys & Tuts.* 18, 1 (2015).
[14] Seng Chun Hoo and Haidi Ibrahim. 2019. Biometric-Based Attendance Tracking System for Education Sectors: A Literature Survey on Hardware Requirements. *Journal of Sensors* 2019, 1 (2019).
[15] Josh Howarth. 2024. How Many People Own Smartphones? (2024-2029). https://explodingtopics.com/blog/smartphone-stats. Accessed: 2024-09-24.
[16] Myungchul et al. Kwak. 2016. A Smartphone-based Tool for Checking Attendance of Students in Classroom Automatically.(2016). (2016).
[17] TS Lim, SC Sim, and MM Mansor. 2009. RFID based attendance system. In *2009 IEEE Symp. on Industrial Electronics & Applications*, Vol. 2. IEEE.
[18] Fen et al. Liu. 2020. Survey on WiFi-based indoor positioning techniques. *IET communications* 14, 9 (2020).
[19] O et al. Oloyede Muhtahir. 2013. Fingerprint biometric authentication for enhancing staff attendance system. *system* 5, 3 (2013).
[20] Nithin et al. Ramakrishnan. 2023. Wi-Fi Based Smart Attendance Monitoring System. In *2023 7th Int. Conf. on Computation System and Information Technol. for Sustainable Solutions (CSITSS)*. IEEE.
[21] Domien Schepers and Aanjhan Ranganathan. 2022. Privacy-preserving positioning in wi-fi fine timing measurement. *Proceedings on Privacy Enhancing Technologies* (2022).
[22] Tata Sutabri et al. Tata Sutabri. 2019. Automatic attendance system for university student using face recognition based on deep learning. *Int. Journal of Machine Learning and Computing* 9, 5 (2019).
[23] Hitesh Walia, Neelu Jain, et al. 2016. Fingerprint based attendance systems-a review. *Int. Research Journal of Engineering and Technol.* 3, 5 (2016).