LocAuth: A fine-grained indoor location-based authentication system using wireless networks characteristics

Mohsen A. Alawami a, Hyoungshick Kim a,b,*

a Sungkyunkwan University, South Korea
b Data61 CSIRO, Australia

ABSTRACT

Location-based information has become an attractive attribute for use in many services including localization, tracking, positioning, and authentication. An additional layer of security can be obtained by verifying the identity of users who wish to access confidential resources only within restricted, small, indoor trusted zones. The objective of this paper is to construct highly secure indoor areas primarily by detecting only legitimate users within their work cubicles. In this paper, we present a fine-grained location-based authentication system (LocAuth) which ensures the physical presence of the user within his/her small trusted zone (2 m² area). To do this, LocAuth exploits the ambient wireless network characteristics (e.g., BSSID, SSID, and RSSI) of nearby Wi-Fi and Bluetooth devices observed from each trusted zone. We propose a novel technique called Top-Ranked Network Nodes (TRNNs) to accurately overcome the fluctuations in wireless signals and enhance the ability to distinguish targeted trusted zone from neighboring areas. In addition, we developed an application to implement LocAuth on Android-based smartphones and tested it in a real indoor environment. The tested area is composed of seven adjacent and closely spaced work cubicles located in our lab. We evaluated LocAuth in two ways: through RSSI-based nearest neighbors (RSSI-based NN) and through supervised machine learning algorithm (Support Vector Machines). The results of the experiment show the effectiveness of LocAuth by achieving a high classification accuracy (above 98%). This demonstrates its feasibility in terms of both accuracy as well as fine-grained classification.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

With the growth of wireless communication technologies and the development of smartphones, the demand for location-based information is rapidly increasing for many services. Nowadays, determining user location has been intensively studied and several technologies have been proposed for this purpose. Global positioning system (GPS) technology is the most widely adopted technique for procuring user location information and authentication. GPS uses line-of-sight radio signals transmitted from three or more visible satellites to accurately measure user’s outdoor location (Gao and Groves, 2018; Okamoto and Chen, 2015; Oshin et al., 2012). In contrast, for the inside of buildings with no line-of-sight, other solutions are usually employed such as wireless network technologies. Many recently proposed works exhibit user location estimation inside buildings, rooms, or offices based on the fingerprints of signal strengths (Afzalan and Jazizadeh, 2019; Chen et al., 2018; Jiang et al., 2012; Youssef and Agrawala, 2005).

Many techniques in literature have been proposed by exploiting the fact that all phones have the Wi-Fi capability to estimate phone location using Wi-Fi fingerprints characteristics seen in an environment that is well-covered by Wi-Fi signals (Khalajmehrabadi et al., 2017; Palacios et al., 2019; Wang et al., 2015; Wu et al., 2012). Nevertheless, the deployment of these techniques were done in indoor environments with open layouts that have large area sizes (i.e., spaced and unconstrained locations), and public areas (e.g., common rooms, corridors, and halls in homes, universities, and malls) environments. Other techniques have exploited the fact that all smartphones are supported with cellular technology and have utilized the location information extracted from the signature of the received signal strength (RSS) from cell towers to provide a ubiquitous location determination service (Rizk, 2018; Rizk et al., 2019; Tian et al., 2015). Such proposals rely on getting location information for a user based on the density of cell towers observed in the area where the user is located. However, according to the cellular standard, current high-
ends smartphones can receive signals from up to seven surrounding cell towers but in practice it provides access information (i.e., RSSI, Cell ID, Type of network) of only the single associated cell tower. This limitation drops the accuracy of traditional techniques in determining user location.

Many other techniques have also been proposed for indoor-outdoor location detection based on observing the user's movement styles and the signatures’ fluctuations of the surrounding readings (i.e., cellular, Wi-Fi, Light, and GPS) while he/she transiting from the inside of the building to outside. Ali et al. (2018); Capurso et al. (2016); Zhang et al. (2019). Such proposals focus only on providing binary classification results to infer whether a user is indoors or outdoors without computing the current location’s coordinates. Although all the aforementioned literature techniques are considered the de facto approaches in the location determination domain, it eventually concentrates only on addressing location-estimation problems of a user for purpose of indoor-outdoor localization, tracking, and positioning services.

Currently, wireless network data with short-communication range have aroused extensive attention in the area of location determination services, as mass indoor rooms, offices and companies have been covered with developed and deployed networks access nodes (e.g., Wi-Fi and Bluetooth), as well as current smartphones that have been equipped with sensors to easily connect these network nodes once they are within the coverage range. Thus, the objective of this study is to exploit user’s location information to provide an important layer of security in smart and sensitive indoor environments such that it can be used for authentication purposes in situations where legitimate users must physically reside within restricted trusted zones and specifically request to gain access to a system.

The most common three types of credentials that can be used in the authentication process are: known shared secrets (e.g., keys, passwords, PIN, etc.), physical characteristics (e.g., fingerprint, retina, etc.), and habitual behaviors. Each type has its own advantages and disadvantages. For example, known shared secrets are easy to use, convenient, and do not require high maintenance cost for the systems. However, this type of authentication is vulnerable to cyber attacks, especially for low-security level passwords, thus users should choose complex passwords to preserve the security level. As a result, users are required to invest significant effort to remember complex passwords. On the other hand, users who authenticate by physical characteristics are not required to remember complex passwords, but sensors used for identifying and correctly classifying physical characteristics are often expensive and inaccurate. Habitual behaviors seem to provide a solution to the disadvantages of the previous two types; however if these traits or behaviors are copied, the user is unable to replace or change the credentials, thus allowing an adversary to attack the credentials directly.

Often, to implement a highly secure indoor environment, access to confidential resources might be restricted to a very close, inside one room, and small trusted zones that is inside government agencies, research labs, healthcare industries, or military secure rooms. Therefore, only authenticated users who are authorized would be allowed to access confidential resources. In this paper, we focus on demonstrating the feasibility of adopting wireless network nodes (i.e., Wi-Fi and Bluetooth) and smartphones to implement a fine-grained location-based authentication process under a small area (≤2 m²), adjacent (having distance with approximately less than 2 m) and restricted trusted zones located inside the same room in indoor environments. To the best of our knowledge, this is the first work that aims to identify and verify a user’s presence inside his/her trusted zone as well as precisely distinguish the targeted location’s wireless fingerprints from fingerprints of the other trusted zones that are very close, adjacent, and inside the same room. This supports safety and security issues for the important systems such that only a legitimate person who is distinguishable at his/her area can get access to the information on the systems by authenticating the location information with the system login password, for example.

Therefore, this work studies the feasibility of determining the trusted zone to which the user’s smartphone is restricted. Through this restriction, the system may limit the sensitive data of the system to be accessed only based on location authentication. We verify the feasibility by collecting various wireless characteristics of the network nodes (e.g., Wi-Fi and Bluetooth) seen at each trusted zone using the Android application that we developed. Finally, we apply our LocAuth system on these collected dataset for authentication purposes and to verify that the selected user belongs to his/her trusted zone accurately.

The contributions of this paper are summarized as follows:

1. We present a realistic system (LocAuth) to authenticate users within fine-grained indoor trusted zones which are very close, adjacent, small area sizes (around two meters), and located inside the same room.
2. We propose a novel approach, called LocAuth TRNNs-based technique, that implements the LocAuth method based on the characteristics of the Top-Ranked Network Nodes (TRNNs) and demonstrate an accurate trusted zone detection.
3. We develop a smartphone application to collect dataset consisting of the ambient wireless network characteristics.
4. We apply this work in seven neighboring work cubicles in our research lab and compare our classification accuracy with two different detection methods.

Moreover, LocAuth is an infrastructure-less system which depends only on the existing pre-installed indoor network nodes (Wi-Fi and BT) and the user’s smartphone.

The three expected significant advantages that come from distinguishing the trusted zone to which the authorized user’s smartphone belongs are as follows. First, the authenticated user at his/her trusted zone can get access directly to sensitive and confidential data on the system without the need to remember a complex password or shared secret keys. Second, the authenticated user’s smartphone can remain unlocked while in his/her trusted zone (e.g., on the desk or work cubicle). Third, attacks in sensitive locations (e.g., at airports, military zones, and company offices) are significantly restricted because they would require the adversary to be physically located within these places.

The rest of this paper is organized as follows. In Section 2, we briefly explain the LocAuth system design, and in Section 3, we provide a detailed analysis of the LocAuth TRNNs-based technique. We evaluate the performance of our work in Section 4. In Section 5, we discuss the purpose and benefits of LocAuth. In Section 6, a description of related work is given. Finally, in Section 7, we conclude the paper and suggest future work.

2. LocAuth system design

For ubiquitous and realistic applicability, LocAuth works with no prior knowledge of the coordinates of the surrounding wireless network nodes and requires no need for additional external hardware. We refer to wireless Wi-Fi access points and Bluetooth devices as network nodes (NN) for the remainder of this paper. LocAuth uses only the characteristics information (e.g., BSSID, SSID, and RSSI) of all visible wireless network nodes observed indoors. The motivation behind the use of radio signals for location-based authentication is to take advantage of the building structure such that the characteristics of the wireless network can be observed uniquely at different indoor rooms. However, it is difficult
to distinguish two trusted zones whenever they are in the same room and each with the area less than 2 m (e.g., adjacent users’ seats at a research lab or company). Fig. 1 shows the screenshot of the LocAuth Android application installed on the user’s smartphone to collect the wireless characteristic fingerprints observed at each trusted zone. Fingerprints for both Wi-Fi APs and Bluetooth devices are recorded at each trusted zone such that each scan consists of various wireless characteristic (e.g., BSSID, SSID, and RSSI), the number of all visible nodes, and the average RSSI values. Additionally, the app shows a per-second increment counter, the number of scans where each scan is recorded every 5 seconds, start/stop scan buttons, and “Edit Text” to enter the label of the tested zone. All these collected data are saved in text files for further processing.

2.1. Radio signals analysis

According to the nature of wireless radio signals, the signal strength will decrease as the propagation distances increase and as they are impeded by building walls, doors, and windows. Therefore, it seems possible to distinguish between rooms using the radio signals spread from the same network nodes which exhibit different characteristics and different ranges of their values. This is a challenge in the case of adjacent trusted zones inside the same room which exhibit network characteristics with very similar ranges of their values. Aiming at this problem, we carried out experimental analyses and collected Wi-Fi characteristics of 120 scan iterations at four adjacent trusted zones, each 2 m area in size in our research lab room. Fig. 2 shows the average RSSI values of Wi-Fi APs collected, which are very close to each other and range from -70 dBm to -80 dBm. Fig. 3 exhibits the total number of Wi-Fi APs observed at four adjacent trusted zones; the Wi-Fi APs are random and range from 20 to 50 APs. From these experimental observations, we find that the wireless characteristics of all visible network nodes are not useful to differentiate between adjacent trusted zones in the same room. Therefore, in this paper, we consider the concept of Top Ranked Network Nodes (TRNNs) instead of all visible nodes observed at each trusted zone.

2.2. Overview of LocAuth

Here, we briefly describe how LocAuth system works starting from collecting the relevant trusted zone wireless fingerprints data, passing to Top Ranked Network Nodes (TRNNs) analysis, and reaching to the final location determination for the authentication. Actually, before using LocAuth system, one pre-step is necessary for determining the number of trusted zones that are very close, adjacent, and inside the same room (e.g., in company offices or research labs) to know what locations should be covered by LocAuth during authenticating the user. After that, LocAuth starts to trigger the first stage, called LocAuth learning, as an offline process to collect wireless fingerprints from surrounding network nodes observed at each of the predetermined trusted zones and hence ends by storing the relevant processed information into the database. Since, the network nodes, as well as the trusted zones, have fixed locations, this stage often occurs once at the beginning to learn about the environment. The details of the learning stage are explained in Section 2.3.

During the execution time (i.e., the online process) when a user wants to get authenticated to access his system, LocAuth triggers the second stage, called LocAuth TRNNs-based technique, includes four consecutive steps which perform the core process to determine top-ranked network nodes and extracting its fingerprints’ features. This technique, explained in detail in Section 3, aims to
use the user’s personal smartphone within the targeted trusted zone for the purpose of authentication. Technically, it works for only the Top networks that have unique attributes and features among all-the-others network nodes as well as focuses only on the KNN trusted zones from the targeted location to accurately conduct location-based determination process. Finally, the outcomes of the second stage, (i.e., The selected TRNNs and its extracted features that observed at the KNN trusted zones), pass to the third stage, called detection methods, for conducting comparison and classification processes with the data stored in the database using two different detectors to infer the final location by distinguishing the targeted trusted zone from neighbors.

To be precise, Fig. 4 shows an overview of the LocAuth system which accommodates the whole processes into three main stages executed in series as follows:

i. **LocAuth Learning**: The first stage is responsible for collecting the fingerprints for long per-interval periods at each trusted zone. During this stage, we execute three steps in cascade which are (1) building up the input data matrix from all network nodes observed at each trusted zone, (2) sampling period determination and filtering of unnecessary network nodes, and (3) constructing the wireless characteristics data of only the registered network nodes. Finally, the fingerprints of all learned trusted zones inside the room are stored in the database.

ii. **LocAuth TRNNs-based technique**: Provides the following steps, (1) online scan processing, (2) selecting only the K-nearest neighbors (KNN) trusted zones from the target one among all learned trusted zones stored in the database based on MAC addresses, (3) inferring the Top Ranked Network Nodes (TRNNs) fingerprints of the selected KNN trusted zones, and (4) extracting RSSI-based features from the inferred TRNNs.

iii. **Detection methods**: Responsible to detect the final trusted zone that authenticates the user. This will be done using two different detection methods which are nearest neighbors (NN) and Support Vector Machines (SVMs).

2.3. LocAuth learning stage

This stage describes in detail the first steps of user authentication based on location fingerprints.

**Building input data matrix**: First, the user who wants authenticate in his/her trusted zone uses the android application shown in Fig. 1 to collect the fingerprints (e.g., BSSID, SSID, and RSSI) of the ambient wireless network nodes. Through this learning stage, the user collects fingerprints for repeated time intervals \( T_i \) for each of \( L \) scans in the series as shown in Fig. 5. Each scan records the MAC addresses and RSSI readings of all network nodes observed at a specific trusted zone and constructs the input data matrix of dimension \( (m \times 4L) \) as shown in Fig. 6a. In our experiments, we assume the time sampling interval \( T \) is 10 minutes and the sampling rate is \( h \) (1 scan per 5 sec). So, the total number of scans per one time interval is:

\[
L = T_i \times h. \tag{1}
\]

**Filtering process**: Due to the indoor environment interference, the collected input data are always fluctuating and not reliable information for fingerprinting. Thus, in order to gain consistency within the data, we always first collect the fingerprints and then filter the unnecessary data. This filtering aims to ignore/remove some interference network nodes from the input data matrix. For example, some temporary network nodes or some other faraway devices are detected only occasionally at some current trusted zones. Therefore, in each trusted zone, we use MAC addresses of the network nodes in the input matrix and count number of detection times out of the total \( L \) scans. Then, we filter out a network node (AP/BT) that has a lower detection number than the duplication threshold \( D_{th} \) and ignore their fingerprints according to (2). The others are kept as registered network nodes.

**Constructing data matrix of registered network nodes**: Based on the sampling and filtering step, we extract \( R \) number of network nodes that considered as registered nodes observed at each trusted zone.

\[
AP/BT = \begin{cases} 
\leq D_{th} & \text{Ignored.} \\
\geq D_{th} & \text{Registered.} 
\end{cases} \tag{2}
\]

Then, we construct the registered data matrix with dimension \( p \times R \) where the columns represent the RSSI values of each registered network node for \( p \) times, where \( D_{th} \leq p \leq L \) as shown in Fig. 6b. After that, we extract four important parame-
ters for each column: the corresponding MAC address, average signal strength (Avg), correlation factor (C) and duplication number (D). We explain these parameters in more detail in Section 3.2. Finally, by completing the learning stage, all data fingerprints that contain MAC addresses, RSSI values and the parameters (Avg, C, D) of the registered network nodes of all trusted zones are stored in the database for further analysis using LocAuth Top Ranked Network Nodes (TRNNs) technique. All notations used in this paper are shown in Table 1.

3. LocAuth TRNNs-based technique

In this section, we describe the online LocAuth TRNNs technique which relies only on some important network nodes. These nodes are called Top Ranked Network Nodes (TRNNs) and have specific attributes to be selected from the registered nodes. Authentication using the online LocAuth technique starts with two steps. Firstly, as shown in Fig. 4, the online per-scan fingerprint is collected using the user’s personalized smartphone at his/her trusted zone. Then, pre-processing sampling and filtering are applied to enhance the quality of the fingerprint and to record only the registered network nodes as explained in the Section 2.3.

3.1. MAC based KNN trusted zones

After the learning stage, users who wishes to authenticate use a personal smartphone to measure the online per-scan fingerprint at his/her trusted zone. This fingerprint is then compared with the fingerprints of all learned trusted zones stored in the database in order to only select the best matching K trusted zones among all learned locations. The fingerprints in the database are tagged with corresponding positions of the trusted zones locations. We exploit the MAC addresses of the stored fingerprints to extract the corresponding K-Nearest Neighbors (KNN) trusted zones that have the highest similarity to the targeted one by means of Euclidean distance. We assume the measured online fingerprints have MAC addresses \( f = (MAC_1, MAC_2, MAC_3, \ldots, MAC_n) \), then compare with the MAC addresses of the fingerprints of all learned trusted zones according to (3) and (4) below:

\[
KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2},
\]

and,

\[
Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases}
\]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.

3.2. Analysis of top ranked network nodes (TRNNs)

After constructing the registered data matrix of the KNN trusted zones, we still have too many irrelevant network nodes that can decrease the accuracy of authentication. Therefore, we propose a new concept called Top Ranked Network Nodes (TRNNs) such that we focus on selecting the fingerprints of only the top-ranked wireless nodes of each KNN trusted zone from the database. The top-ranked network nodes are one or more network nodes which are

Table 1

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRNNs</td>
<td>Top Ranked Network Nodes.</td>
</tr>
<tr>
<td>TZ</td>
<td>Trusted zone.</td>
</tr>
<tr>
<td>BSSID</td>
<td>Basic service set identifiers.</td>
</tr>
<tr>
<td>SSID</td>
<td>Service Set Identifier.</td>
</tr>
<tr>
<td>RSSI</td>
<td>Received Signal Strength Indicator.</td>
</tr>
<tr>
<td>AP</td>
<td>Wi-Fi access point.</td>
</tr>
<tr>
<td>BT</td>
<td>Bluetooth node.</td>
</tr>
<tr>
<td>( T_i )</td>
<td>The ith time interval.</td>
</tr>
<tr>
<td>( h )</td>
<td>Sampling rate (1 scan per 5 sec).</td>
</tr>
<tr>
<td>L</td>
<td>The number of scans per time interval.</td>
</tr>
</tbody>
</table>

\( m \) The number of observed network nodes per scan.

(3) and (4) below:

\[ KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2}, \]

and,

\[ Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases} \]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.

3.2. Analysis of top ranked network nodes (TRNNs)

After constructing the registered data matrix of the KNN trusted zones, we still have too many irrelevant network nodes that can decrease the accuracy of authentication. Therefore, we propose a new concept called Top Ranked Network Nodes (TRNNs) such that we focus on selecting the fingerprints of only the top-ranked wireless nodes of each KNN trusted zone from the database. The top-ranked network nodes are one or more network nodes which are

\[ KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2}, \]

and,

\[ Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases} \]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.

\[ KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2}, \]

and,

\[ Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases} \]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.

\[ KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2}, \]

and,

\[ Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases} \]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.

\[ KNN(f, TZ_q) = \sqrt{\sum_{j=1}^{n} (Comp(MAC(f, TZ_q)))^2}, \]

and,

\[ Comp(MAC(f, TZ_q)) = \begin{cases} 
1 & MAC(f) \in TZ_q, \\
0 & \text{Otherwise}.
\end{cases} \]

where \( q = (1, 2, 3, \ldots) \) is number of learned trusted zones stored in the database. Therefore, the objective of this step is to filter out fingerprints of unnecessary trusted zones and turn the focus of the LocAuth on fingerprints of the K-Nearest Neighbors to the targeted trusted zone. This is useful in the case of indoor environments that have big offices with a large number of adjacent cubicle works. Also, this step makes computations confined to more efficient measurements with less overhead.
preferable, nearer and more confidential nodes observed at each of the KNN trusted zones and contribute most to the predicted output.

Three benefits from performing TRNNs analysis:

i. Removes noisy nodes: selecting highly reliable nodes provides less radio signal fluctuations.
ii. Enhance accuracy: less deceptive data results in an improvement of model quality.
iii. Small computation time: less unnecessary data means that the algorithm processes will be faster and more accurate.

Therefore, the aim is to make our system more accurate with the lowest misclassification and benefit from better selection of the best network nodes (TRNNs) seen at each trusted zone. However, we must also determine which parameters to base on the top-ranked network nodes (TRNNs). To do so, we select the TRNNs that perform well for the following three parameters: strongest RSSI signal values (Avg), highest correlation factors (C), and highest duplication numbers (D).

**TRNNs parameters estimation:** According to the matrix shown in Fig. 6b which includes the fingerprints of the registered network nodes at specific trusted zones, each column in this matrix corresponds to the fingerprints of one network node. For each column, we estimate the three proposed parameters (Avg, C, D) in order to select the TRNNs observed at each trusted zone as follows:

i. Average signal strength (Avg): The network node which provides strong RSSI readings appears as a good quality reference point. We estimate the average value of all RSSI readings recorded for each column. From here, the network nodes that exhibit the highest average RSSI values are properly classified as important nodes observed at the certain trusted zone.
ii. Correlation factors (C): Through this parameter, we measure the correlation relation between the p RSSI readings that belong to the same column and provided by the same network node. Then, the columns that exhibit the lowest correlation factors means the corresponding network nodes provide inconsistencies readings and then not preferred as reference nodes because of high readings fluctuations.
iii. Duplication number (D): We estimate the number of times such that one network node can be detected out of L scans in each time interval T_p. From this, we focus on those non-noisy nodes which are observed almost with every scan at each trusted zone and provide the highest duplication number (D).

**TRNNs selection and fingerprints extraction:** Based on the parameters values (Avg, C, D) estimated from the fingerprints of each column, the next step is to select the top-ranked network nodes (TRNNs) of KNN trusted zones. Firstly, we apply the weighted Eq. (5) and compute the S_k values at the kth trusted zone, k = (1, 2, 3, …, K), as follows:

\[ S_{k} = \alpha_{1} \times \text{Avg}_{k} + \alpha_{2} \times C_{k} + \alpha_{3} \times D_{k}, \]

s.t.
\[ \alpha_{1} + \alpha_{2} + \alpha_{3} = 1, \]
\[ 0 < \alpha_{1}, \alpha_{2}, \alpha_{3} \leq 0.5, \]

where \( \alpha_{1}, \alpha_{2}, \) and \( \alpha_{3} \) are designated weights factors denoting the importance value of the three parameters. \( S_{k} \) denotes the estimated value based on these three weighted parameters of the \( r \)th column’s fingerprints, \( r = (1, 2, 3, \ldots, R) \), at the \( k \)th trusted zone. This process will be repeated for all KNN trusted zones to construct S matrices that have entries of \( S_{k} \) values. After that, we sort the S matrix in descending order and select the \( \mathcal{V} \) window size network nodes that provide the highest \( S_{k} \) values as the top-ranked network nodes (TRNNs). Finally, all fingerprints (i.e., MAC addresses, the three parameters values, and RSSI values) of the selected \( \mathcal{V} \) TRNNs are extracted and saved together with the tags of trusted zones into the database for further processing. In rest of this paper, LocAuth relies only on the fingerprints of these \( \mathcal{V} \) TRNNs of each KNN trusted zone in order to authenticate the user at a specific trusted zone.

**RSSI-based features extraction:** This step is used to build an SVM model by extracting important features based on the RSSI values of the TRNNs. Therefore, for each top-ranked network node, we first infer its corresponding RSSI column from the registered data matrix and then extract 16 different features for SVM-based detection (as in Section 4.3).
all the data saved in text files labeled with the trusted zone name and date/time. We collect dataset of \( L = 120 \) scans for each time interval \( (T_i = 10 \) minutes) during the learning stage.

4.2. First detection method: RSSI-based NN

We use the nearest neighbor concept based on RSSI values as a detection technique through estimating distances between online fingerprints and those fingerprints of learned trusted zones stored in the database. Through each online scan, we use the MAC addresses to infer the \( W \) TRNNs of the online fingerprints which are same to those \( W \) TRNNs of each selected KNN trusted zone. Then, we extract the corresponding absolute RSSI values of these \( W \) TRNNs for both online fingerprint and learned KNN trusted zones. Fig. 8 shows the process of comparison using the Euclidean distance in signal space according to (6).

Assuming the extracted RSSI values of \( W \) TRNNs selected vector of the online fingerprint are:

\[
U = \{\text{RSSI}_1, \text{RSSI}_2, \ldots, \text{RSSI}_W\}
\]

and the RSSI values for the same \( W \) TRNNs vector of the \( i \)th trusted zone in the database are:

\[
TZ_i = \{\text{RSSI}_{1Z_i}, \text{RSSI}_{2Z_i}, \ldots, \text{RSSI}_{WZ_i}\}, \quad i = 1, 2, 3, \ldots, k.
\]

The Euclidean distances can be expressed in the following formula:

\[
d_{\text{RSSI}}(TZ_i, U) = \sqrt{\sum_{j=1}^{W} (\text{RSSI}_{ij} - \text{RSSI}_{jZ_i})^2}. \tag{6}
\]

Following this, the distances between the online fingerprints and each KNN trusted zone stored in the database are estimated by comparing RSSI-values of the TRNNs. Then we select the minimum distance \( d^* \) and the corresponding trusted zone \( TZ^* \) that represents the final trusted zone into which the user authenticates. During the evaluation of LocAuth using RSSI-based nearest neighbor (NN), we performed 20 online iteration of scans to collect the dataset in each of the 7 trusted zones separately at different periods of time and randomly spaced days with varying two main factors.

The first factor is the weight values where we put two different settings: Case\(_1 = \{\alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2\} \) and Case\(_2 = \{\alpha_1 = 0.333, \alpha_2 = 0.333, \alpha_3 = 0.333\} \). The second is TRNNs window size value \( (W) \) where we set three different values \( (W=5, W=10, W=15) \). After conducting the experiments and detecting the trusted zones, the results are listed using the confusion matrix to show the effectiveness of this detection method. The confusion matrix is constructed such that each row indicates the true location of the trusted zone and the corresponding column indicate the output identity predicted by LocAuth. Each matrix element represents how many times (out of 20 times) each truly trusted zone in the row gets correctly classified in the column. Fig. 9 shows the confusion matrix for 7 trusted zones detected when the Case\(_1 \) of the weight values are used with TRNNs window sizes \( (W=15, W=10, \text{and } W=5) \) respectively. Similarly, Fig. 10 shows the confusion matrix for 7 trusted zones detected when the Case\(_2 \) of the weight values are used with TRNNs window sizes \( (W=15, W=10, \text{and } W=5) \) respectively.

**Performance Metrics:** We used the following metrics in the evaluation of LocAuth: (1) True Positive Rate (TPR) = TP/(TP+FN), TP rate of trusted zone \( X \) is the probability of detection which measures the proportion of online scan iterations that are correctly identified as trusted zone \( X \). (2) False Negative Rate (FNR) = 1-TPR. FN rate of detection of trusted zone \( X \) is the fraction of online scans of trusted zone \( X \) which are incorrectly detected as other trusted zones. As expected, we observe that a trusted zone is often misclassified as a neighbor trusted zone and this misclassification has a fraction varies according to the above two mentioned factors. This evaluation of the probability of detection for each of 7 trusted zone under specific settings for the two main factors is shown in Figs. 11 and 12 receptively.

Since the correctly identifying positives is the most important thing to do with the dataset, we should choose factors with the higher overall sensitivity. Thus, we considered the average TPR values of all the trusted zones. As shown in Table 2, we observe that the best overall sensitivity is 87.143% with the minimum number of required TRNNs at \( W = 5 \). and this is achieved when Case\(_1 \) of the weight values are used. However, the worst overall sensitivity

![Diagram](image-url)
is 80% with the maximum number of required TRNNs at \( W = 15 \), and this is achieved when Case2 of the weight values are used. Therefore, based on these experimental results, the first detection technique provides good authentication performance with smaller computational cost by setting weight values to Case1 and TRNNs window size to \( W = 5 \).

### 4.3. Second detection method: SVM

The previous classification method which is RSSI-nearest neighbor detection technique gives modest performance results (i.e., less than 90%). Here, we use supervised machine learning algorithms as a second detection technique. To achieve classification task, Support Vector Machines are the most popular method to perform maximum margin between two classes. Recently, SVMs have gained much popularity for localization and authentication on smartphones because of the effective in high dimensional spaces and memory efficient.

Fig. 13 illustrates the five steps of SVM implementation in our work. We conducted five iterations of collecting online dataset at different days to authenticate users in their trusted zones, and these five steps are repeated with each iteration. In the first step,
we sensed the online fingerprints at each trusted zone and then inferred the corresponding fingerprints of the k-Nearest Neighbors (KNN) trusted zones from the database by comparing the MAC addresses.

Following this, we filter out the unnecessary network nodes and constructed the registered data matrix shown in Fig. 6b. In the second step, LocAuth estimates (Avg, C, D) parameters observed at each KNN trusted zone to select best the window size (\(\mathcal{V}\)) TRNNs. By the end of this step, the TRNNs matrix 1 of dimension (\(\mathcal{V} \times K\)) is constructed where \(K\) denotes the number of k-Nearest Neighbors trusted zones and \(\mathcal{W}\) denotes the TRNNs window size seen at each KNN trusted zone. In the third step, LocAuth filters out the RSSI columns from the registered data matrix that correspond to the selected (\(\mathcal{V}\)) TRNNs and then extracts (Z) RSSI-based features of each column. As a result, the RSSI-based features of each TRNN at each KNN trusted zone are represented as one row in the features matrix.

Table 3 summarizes the list of features used. The matrix 2 of dimension (\(\mathcal{V} \times K\) × \(\mathcal{Z}\)) is constructed and contains the features values of all \(\mathcal{W}\) selected TRNNs used along with the tagged KNN trusted zones. Additionally, the RSSI-based features of the online scans are extracted. However, all above steps are exactly repeated at the other trusted zones such that the data of all matrices are aggregated in one matrix called (Alldata) as shown in Algorithm 1. In the fourth step, we input Alldata matrix to train the SVM classifier on the predictor variables (features values) through one-to-all binary classification. The trained SVM model is built and has the ability to predict new unseen dataset. For the final step, this identified classification model (SVM model) returns a classification accuracy value which indicates the correct rate for authentication purposes. Note that these five steps will be repeated again for all new online iterations and the final vector of classification accuracy is returned to show the efficiency of LocAuth to authenticate the user in his/her trusted zone.

From the results of the 7 tested trusted zones shown in Tables 4 and 5, we see that LocAuth works better with SVMs and provides high output detection accuracy that will be used for legitimate user authentication. To get an indication about the effect of each case as well as the values of the window size TRNNs, we average the classification accuracy of each trusted zone for all iterations. Since these iterations are executed in the same location at different periods of time, this averaged result visualizes how much the detection process of the tested trusted zone is guaranteed in both stable and unstable wireless signal environments. Also, it shows the geometry relationship between the trusted zone location with the entire environment structure complexity.

The average classification results are plotted in Figs. 14 and 15 receptively, which show the detection accuracy that each trusted zone can provide when the user uses the LocAuth for multiple iterations to authenticate his/her location at any period of time.

---

**Table 3** Features list used for each TRNN.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Extracted RSSI-based features</th>
</tr>
</thead>
</table>
| RSSI measurements for Wi-Fi APs and BT devices | - Mean, Median, Standard deviation (STD)  
- Maximum (max): maximum RSSI reading  
- Minimum (min): minimum RSSI reading  
- Range value: (Maximum - Minimum)  
- Mode value: most frequently occurring value  
- RMS: root mean square value  
- Sum of absolute deviation from mean  
- First, second and third quartiles (Q1, Q2, Q3)  
- CV percentage (%) ratio of the STD to the mean  
- 3rd moment (skewness); 4th moment (kurtosis)  
- Inter Quartile Range (ICR): (Q3 - Q1) |

---

**Algorithm 1:** LocAuth using machine learning (SVM).

**Input:** Online fingerprints of the trusted zones

**Output:** Classification accuracy

1. Itr: Number of iterations.
2. N: Number of existing TZs in the tested environment.
3. Data\(_1\): accumulative offline features data matrix
4. Data\(_2\): accumulative online features data matrix

5. for \(i \leftarrow 1\) to Itr do

6. for \(j \leftarrow 1\) to N do

7. Take online scan at TZ\(_j\)
8. Infer the KNN TZs from the databases
9. Estimate (Avg, C, D) parameters of each KNN
10. Select (\(\mathcal{V}\)) TRNNs for the KNN TZs
11. Build matrix\(_{1}\) of dimension (\(\mathcal{V} \times K\))
12. Compute (Z) RSSI-based features
13. Build matrix\(_{2}\) of dimension [(\(\mathcal{V} \times K\) × \(\mathcal{Z}\)]
14. Data\(_1\) ← [Data\(_1\) + matrix\(_{1}\)]
15. Compute (Z) online RSSI-based features of TZ\(_j\)
16. Data\(_2\) ← [Data\(_2\) + online features of TZ\(_j\)]

17. end
18. All data ← [Data\(_1\) + Data\(_2\)]
19. Label Alldata with “1” and “0”
20. Apply SVM classifier
21. end
22. Get classification accuracy vector

---

**Fig. 14.** Average classification accuracy using SVM machine learning technique for 7 trusted zones with TRNNs window sizes (\(\mathcal{W}\)=5, \(\mathcal{W}\)=10, \(\mathcal{W}\)=15) when (\(\alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2\)).
Table 4
List of classification accuracy results using SVM classifier with Case 1: (α1 = 0.5, α2 = 0.3, α3 = 0.2).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>T2</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99.44</td>
<td>98.15</td>
<td>99.24</td>
<td>97.62</td>
<td>98.07</td>
<td>99.09</td>
<td>98.11</td>
</tr>
<tr>
<td>2</td>
<td>99.25</td>
<td>97.40</td>
<td>99.32</td>
<td>98.37</td>
<td>97.77</td>
<td>99.13</td>
<td>98.32</td>
</tr>
<tr>
<td>3</td>
<td>99.40</td>
<td>96.98</td>
<td>99.62</td>
<td>98.75</td>
<td>97.35</td>
<td>99.93</td>
<td>99.43</td>
</tr>
<tr>
<td>4</td>
<td>99.29</td>
<td>97.20</td>
<td>98.90</td>
<td>98.49</td>
<td>96.86</td>
<td>96.24</td>
<td>98.22</td>
</tr>
<tr>
<td>5</td>
<td>99.54</td>
<td>97.05</td>
<td>99.11</td>
<td>97.22</td>
<td>97.96</td>
<td>99.05</td>
<td>96.11</td>
</tr>
</tbody>
</table>

(a) TRNNs window size (W=5).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>T2</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.55</td>
<td>98.88</td>
<td>98.65</td>
<td>97.83</td>
<td>96.16</td>
<td>98.98</td>
<td>96.80</td>
</tr>
<tr>
<td>2</td>
<td>98.74</td>
<td>99.37</td>
<td>97.68</td>
<td>98.69</td>
<td>95.81</td>
<td>99.11</td>
<td>97.64</td>
</tr>
<tr>
<td>3</td>
<td>97.12</td>
<td>97.18</td>
<td>99.36</td>
<td>99.36</td>
<td>98.11</td>
<td>98.86</td>
<td>98.83</td>
</tr>
<tr>
<td>4</td>
<td>96.51</td>
<td>96.96</td>
<td>98.71</td>
<td>98.28</td>
<td>97.45</td>
<td>99.35</td>
<td>96.70</td>
</tr>
<tr>
<td>5</td>
<td>98.48</td>
<td>98.20</td>
<td>97.30</td>
<td>98.44</td>
<td>96.08</td>
<td>98.94</td>
<td>97.84</td>
</tr>
</tbody>
</table>

(c) TRNNs window size (W=15).

Table 5
List of classification accuracy results using SVM classifier with Case 2: (α1 = 0.333, α2 = 0.333, α3 = 0.333).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>T2</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99.44</td>
<td>98.15</td>
<td>99.24</td>
<td>97.62</td>
<td>98.07</td>
<td>99.09</td>
<td>98.11</td>
</tr>
<tr>
<td>2</td>
<td>99.25</td>
<td>97.40</td>
<td>99.32</td>
<td>98.37</td>
<td>97.77</td>
<td>99.13</td>
<td>98.32</td>
</tr>
<tr>
<td>3</td>
<td>99.40</td>
<td>96.98</td>
<td>99.62</td>
<td>98.75</td>
<td>97.35</td>
<td>99.93</td>
<td>99.43</td>
</tr>
<tr>
<td>4</td>
<td>99.29</td>
<td>97.20</td>
<td>98.90</td>
<td>98.49</td>
<td>96.86</td>
<td>96.24</td>
<td>98.22</td>
</tr>
<tr>
<td>5</td>
<td>99.54</td>
<td>97.05</td>
<td>99.11</td>
<td>97.22</td>
<td>97.96</td>
<td>99.05</td>
<td>96.11</td>
</tr>
</tbody>
</table>

(a) TRNNs window size (W=5).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>T2</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.11</td>
<td>99.43</td>
<td>99.24</td>
<td>97.44</td>
<td>97.85</td>
<td>99.32</td>
<td>97.51</td>
</tr>
<tr>
<td>2</td>
<td>97.54</td>
<td>98.62</td>
<td>98.11</td>
<td>98.52</td>
<td>96.83</td>
<td>96.22</td>
<td>95.14</td>
</tr>
<tr>
<td>3</td>
<td>98.30</td>
<td>98.00</td>
<td>99.54</td>
<td>94.33</td>
<td>97.81</td>
<td>98.73</td>
<td>98.14</td>
</tr>
<tr>
<td>4</td>
<td>95.00</td>
<td>98.29</td>
<td>92.45</td>
<td>98.48</td>
<td>97.62</td>
<td>94.27</td>
<td>96.64</td>
</tr>
<tr>
<td>5</td>
<td>97.03</td>
<td>98.44</td>
<td>94.33</td>
<td>96.22</td>
<td>97.28</td>
<td>98.90</td>
<td>98.11</td>
</tr>
</tbody>
</table>

(c) TRNNs window size (W=15).

![Fig. 15.](image) Average classification accuracy using SVM machine learning technique for 7 trusted zones with TRNNs window sizes (W=5, W=10, W=15) when (α1 = 0.333, α2 = 0.333, α3 = 0.333).

Finally, the overall accuracy of detection using the machine learning (SVM) technique is computed based on averaging the accuracy results of all the trusted zones that are considered within a specific case and window size value. As shown in Table 6, we observe that when Case 1 of the weight values is used, the overall detection accuracy is the best (98.366%) with the minimum computations cost and number of required TRNNs (W=5). However, all other results are still above (97%) which give us an indication that the SVMs can provide better classification on all cases compared to the first detection technique. Therefore, based on these experimental results, we can consider LocAuth with the machine learning (SVM) detection technique a promising approach that provides better authentication performance for very close trusted zones.

5. Discussion

Nowadays, users who are inside buildings (e.g., rooms, corridors) can easily detect the wireless network node characteristics such as MAC address (BSSID), nodes names (SSID) and signal strength (RSSI) using their smartphones. Here, we emphasize that the work in this paper concentrates on determining the presence of a user inside the targeted trusted zone based on the readings of surrounding wireless network nodes to provide authentication service. Compared with previous authentication work, our work is seamless since it does not rely on any cartographic proofs (i.e., complex authentication protocols) from a central authentication authority. In addition, LocAuth is different from the state-of-art location determination work (see Section 6) since it is implemented in a fine-grained manner with specific kind of locations layout in smart indoor environments.

More specifically, we provide a mechanism to verify the identity of users who wish to access confidential systems in smart indoor environments. Often, to implement this, access to the confidential systems might be restricted to a very close, adjacent, inside one room, and small trusted zones, as shown in Fig. 7, that is inside

Table 6
Overall detection accuracy using SVM machine learning method.

<table>
<thead>
<tr>
<th>Weights values</th>
<th>Overall detection accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>W = 5</td>
<td>Case 1: 98.366% Case 2: 97.384%</td>
</tr>
<tr>
<td>W = 10</td>
<td>Case 1: 97.911% Case 2: 97.522%</td>
</tr>
<tr>
<td>W = 15</td>
<td>Case 1: 97.917% Case 2: 97.695%</td>
</tr>
</tbody>
</table>
government agencies, research labs, healthcare industries, or military secure rooms. According to the nature of radio signals, this is a challenge in analysing the wireless characteristics while considering this kind of fine-grained locations layout, because the radio signals exhibit the same range of readings for the adjacent trusted zones (as explained in Section 2.1). Therefore, LocAuth addresses this challenge and acts as a second useful factor of authentication based on user location (i.e., in addition to the system login passwords) whenever a user wants to login to his/her system. In particular, LocAuth distinguishes wireless location-based characteristics of the targeted trusted zone from the other neighbor locations to ensure that the legitimate user using his personal smartphone is present in the targeted trusted zone.

As a result, we describe a possible scenario about how we can exploit the location information from the LocAuth as an additional authentication factor within a trusted zone. In the beginning, LocAuth Android application should be installed on the user's smartphone and the learning stage is already conducted such that the wireless fingerprints of all adjacent trusted zones that exist inside the targeted room are collected. After finishing the learning stage, we build a database that will be used later for the online process. Then, when a user wants to authenticate himself, LocAuth triggers stage 2 and stage 3, and finally infers the current location of the legitimate user to conclude whether he/she is inside the targeted trusted zone or not. If so, an Android application on the user's smartphone generates a QR code to prove that a user is personally present in his trusted zone. This generated QR code should be sent to a server to be validated whether it is generated from the legitimate smartphone as well. Finally, user needs to point his smartphone to the webcam on the system to authenticate the generated QR code as a second login method, in addition to the system login passwords.

We believe that the location-based authentication is more secure since it confirms that the user is currently and personally logging to the system from the authorized location. As a result, an attacker can not mimic the legitimate user within the targeted trusted zone even in the situation of a smartphone stolen or lost. This is because of two reasons: (1) We assume that the user's smartphone is often protected, such that an anomaly first needs to unlock the smartphone in order to run LocAuth application and generate QR code based on the location information. (2) The anomaly needs to be personally present in the authorized zone and using the stolen smartphone, which is practically difficult especially in sensitive rooms where entry is restricted.

6. Related work

Location-based authentication methods belong to both location determination technologies as well as authentication security mechanisms that aim to use smartphone’s sensor data to precisely identify user’s location information for verifying a user’s presence inside his/her trusted zone. This kind of location-based authentication mostly conducted in smart and secure indoor environments that have sensitive information resources and systems. In this section, we focus on two groups of related work for location-based authentication: indoor location determination techniques and authentication-based techniques.

6.1. Indoor location determination techniques

Nowadays, location determination techniques have become the most common technology for addressing localization, positioning, tracking and navigation problems either in outdoor, indoor, or indoor-outdoor areas for various applications in industrial, networking, E-Healthcare, and public safety and security (Alinsavath et al., 2019; Choi and Kim, 2019; Kulshrestha et al., 2019; Laoudias et al., 2018). In this study, we focus only on indoor environments to address user’s location detection based on characteristic of wireless networks for purpose of authentication.

Besides its inherent usage in wireless communications for data transmission, radio signals are fundamental in the area of user location information determination and can be classified into two categories of approaches (Nurminen et al., 2017; Zafari et al., 2019). The first category uses a signal propagation model and computes the distances between transmitters (network nodes) and receivers (targets) based on signal strengths. This category, called range-based, uses the proximity and triangulation techniques that rely on calculating Angle-of-Arrival (AOA), Time of Arrival (TOA), and Time Difference of Arrival (TDOA) features of the wireless signal strength propagation to estimate the location of a user (Chen and Wang, 2019; Kotaru et al., 2015; Qi et al., 2019; Xu et al., 2016; Yang and Shao, 2015; Zhang et al., 2019). However, these techniques require pre-information about network topology and pre-area planning such that the coordinates of the transmitters (i.e., network nodes) should be known accurately beforehand as well as such techniques are not suitable for indoor location estimation applications, especially as distance increases, because of high sensitivity to multi-path calculations. The other category, called range-free techniques, is more related to our work and uses location fingerprints composed of two phases. The first phase consists of the offline build of the locations’ fingerprint database, and the second phase is the localization itself. The most widely used algorithms in location fingerprints for the location estimation based on Wi-Fi and Bluetooth measurements are probabilistic methods, k-Nearest Neighbors, Support Vector Machines (SVMs), and deep learning using (Chen et al., 2019; Martínez del Horno et al., 2019; V. et al., 2018; Xia et al., 2019; Zhuang et al., 2015). Nevertheless, the deployment of these techniques was done in indoor environments with open layouts that have large area sizes (i.e., locations are unconstrained, spaced, and in different rooms), and public areas (i.e., non-sensitive areas such as corridors, inside malls, and halls) environments. Consequently, location estimation performance of traditional techniques drop significantly since the analysis of Wi-Fi signals will be more challenge for the location determination demand in some real situations where the locations are very close, small, adjacent, and in the same room.

6.2. Authentication-based techniques

Generally, in the recent years, several approaches use data collected from smartphone’s sensors to identify the specific user for authentication purpose. A significant research proposes that VLC is an emerging technology to transfer important data (e.g., passwords, PIN, or shared secret) for location-based detection and authentication purposes. In particular, data can be transmitted through a wireless communication medium via either the radio spectrum or the visual lighting spectrum. Visible light communication (VLC) uses existing Light Emitting Diodes (LEDs) for both illumination and data communications services. The work in Bakar and Dahnal (2017) proposed implementing location-based authentication using visible light communication to collect location information of the user within a small area (e.g., the room). They used the existing LED light-bulbs infrastructure with specially-designed control circuits (transmitters) to transmit the shared secret keys to the smartphones (receivers). They only explained their work theoretically, and assumed only one trusted zone per room in order to infer user’s location information for authentication. Also, the work in Suduwella et al. (2017) proposed a protocol to restrict Internet connectivity to be confined to small restricted areas. They targeted the location-based Internet access protocol which depends on location-based information so that any user who authenticates using VLC also gains internet access even in highly restricted areas.
However, VLC technology has many challenges summarized in Mukherjee (2017), and unrealistic for practical deployment because VLC is still quite theoretical. Moreover, VLC requires additional external hardware such as VLC transmitters, which are control circuits that must be installed within each LED light bulb to transmit the data, as well as at the receiver side, the photo-diodes circuits should also be mounted to the smartphones for analysis processes such as decoding the received data.

Besides VLC, other works have been emerged to use location information as an additional factor of authentication for various applications to enhance security and privacy issues and stop adversaries when sufficient evidence shows that a legitimate user is not in-present at the required location. The researchers in Agadakos et al. (2016) proposed a method called Icicus that provides an additional factor of authentication based on the Internet of Things (IoT) to ensure that a user can not be in two places at once. Icicus leveraged the increasing number of smart things that users carry to locate a user in a specific location, where the physical presence is required, and enhance location estimation to sense the user more robustly than smartphones based solutions in smart environments. The work in Mohamed and Cheffena (2018) presented an RSSI-based gait authentication process using wireless body area networks (WBANs). They collected radio RSSI signals from on-body wearable devices (e.g., smartwatches), transmitted the information to smartphones mounted on the waist of the same user, extracted three channel features and finally used them as inputs with four different classification techniques for user gait authentication purposes.

Furthermore, the current wireless technology that uses radio waves (e.g., Wi-Fi and Bluetooth) is commonly used to provide indoor wireless network connections, localization, and positioning. The author in Kobayashi and Yamaguchi (2015) presented a user’s behavior authentication method by exploiting only Wi-Fi BSSID addresses that the user’s smartphone observes them on a daily basis during conducting his/her normal daily patterns. They examined the possibility of the Wi-Fi BSSID to provide authentication factor through collecting 30 days of data that is related to the geographical location histories of the most visited outdoor places in their living area (e.g., taking the train every morning) as a template data. Then, tested the data of last 24 hours to characterize the user’s behavior for authentication purpose. In addition, the same author in Kobayashi and Yamaguchi (2016) optimized the authentication process to be only on one-hour data instead of the last 24 hours’ data to avoid spoofing possibility and save the user’s information when a smartphone device is stolen. The authors in Chen et al. (2016) proposed a condition-based location authentication method that contains four parties: Prover, witness, verifier, and server. The mobile device of a user (i.e. the prover) can be authenticated in a location by the aide of the information from the devices of neighboring users as well as the verified messages from the verifier and server. This method has two weaknesses: 1) It has been applied onto a wide outdoor area of geographical map location; 2) the user and neighbors may cheat the verifier and the server.

The authors in Ahn and Cho (2019) theoretically presented the method of exploiting four types of data to provide location-based information that authenticates a specific user who wants to login on a secret financial-based web browser (e.g., bank or government agency). Instead of using only ID/passwords and public certificates, they proposed a structure of the location-based login process using GPS location information from a smartphone, IP address and time from a PC, and weather data that provides secure data with location information. To gain access to the targeted web browser, the above four location information types should match and then the user can login using ID/Password. However, while this method is convenient and safe, they have assumed that the user already has access to the PC and the user interactions on the devices or systems are required to collect the location-based information. This may provide an opportunity for an attacker to steal a victim’s sensitive data. The work in Zeng et al. (2016) investigates the possibility that smart homes and places can identify user’s identity without carrying/wearing smart devices (i.e., with device-free). They proposed a Wi-Fi based person identification framework called WiWho that uses channel state information (CSI) captured by Wi-Fi endpoints to detect user’s steps and walking gaits. Finally, WiWho achieved an average accuracy of 80% to 92% for distinguishing the specific user from a small group of users (2 – 6) in a device-free manner.

The work in Wang et al. (2017) applied Identity-based Signature (IBS) into Internet of Things (IoT) to satisfy identity-based authentication to support security issues when use ubiquitous connectivity at dynamic trusted zones under the framework of 5G oriented networks (ION). However, this method involves some sensitive identification information (e.g., phone number, IP, email address) as public keys for identity verification. The authors in Yuan et al. (2019) presented a method to authenticate a location of a targeted user device (UE) by the use of information collected from the peer UEs within the coverage area of an Wi-Fi AP. Since the strength of the confidence level of the authentication depends on the high capacity of UEs, this will effect negatively the quality of the wireless network (QoS). To overcome this, they proposed an access control algorithm to balance location authentication accuracy as well as network channels consumption. However, the tested coverage area in this work is sufficiently wide which reach to tens of meters as well as involves existence of large number of peer UEs for verification, this limits the applicability for authentication scenarios within small areas of trusted zones.

The work in Mainali et al. (2019) presented a contextual authentication method, called ConSec, to enhance the privacy of device data based on user location measurements. ConSec creates a locality-sensitive hashing (LSH) representation from the general location data that collected from user’s device such as GPS coordinates, Wi-Fi ESSID, and barometric altitude. Then, they used the LSH models to learn machine learning algorithms for authenticating users based on their contextual behaviors. The authors in Rexha et al. (2018) improved securing data access on mobile devices by exploiting three components of attributes which are user face biometrics using built-in camera, gesture recognition, and frequent visited locations’ information (using Wi-Fi networks) to support the trustworthiness of user detection and enhance the confidence level of the biometrics-based authentication.

7. Conclusion and future work

This paper presented (LocAuth) a fine-grained location-based authentication system used to authenticate the user in his/her trusted zone by utilizing wireless networks characteristics observed from the surrounding indoor environments. The system is implemented within very small and adjacent workplaces located in the same room such that each place owner can use LocAuth system using his/her smartphone to perform the authenticating process. Correctly authenticated users are only the authorized users who have gain to access the confidential systems. Top-Ranked Network Nodes (TRNNs) technique is proposed to improve classification accuracy for authentication purpose. We conducted real experiments in our lab and used two different evaluation methods. Our results demonstrate that performing the LocAuth technique, which depends on wireless network characteristics as input features and machine learning as detection methods, can provide high classification accuracy and show the effectiveness of the proposed authentication approach. Since LocAuth is applied only in smart places where wireless network nodes are available, we suggest fu-
ture work to exploit other input features such as the lights and sounds of the surrounding environment to enhance system applicability in more places.

Acknowledgments

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) [No. 2017-0-00545]

References


