LocID: A Secure and Usable Location-Based Smartphone Unlocking Scheme Using Wi-Fi Signals and Light Intensity

Mohsen Ali Alawami, Aishwarya Ram Vinay, and Hyoungshick Kim

Abstract—A user’s location information can be used to identify the user. For example, in Android, we can keep our smartphone unlocked when it is located near a place that was previously registered as a trusted place. However, existing location-based user authentication solutions failed to support fine-grained indoor location registration. In this article, we present a highly accurate identification scheme named location-based identification (LocID) using the smartphone’s physical fingerprints that are sensed at the user’s secure area. Our key idea is to use Wi-Fi signals and the light intensity of smartphones’ fingerprints to accurately identify their locations within adjacent indoor areas. To generate stable and reliable location fingerprints, we develop a novel mathematical location-wise signal signature (LSS) algorithm by minimizing the bias highly affected by various environmental factors. To show the feasibility of LocID, we evaluated the performance using machine learning classifiers with a real-world data set containing Wi-Fi and light fingerprints from 68 locations in a building for five weeks. The experimental results show that LocID achieved an F1-score of 98.3%, false-negative rate (FNR) of 2.9%, and false-positive rate (FPR) of 1.7%, respectively, when LocID was retrained daily. Moreover, LocID achieved an F1-score of 94.4%, FNR of 2.7%, and FPR of 3.7%, respectively, when LocID was trained using the data set during the first week and tested with the remaining four weeks data set. Also, we tested LocID using only one-scan of location fingerprints and achieved 90.6% and 90.8% AUC scores for support vector machine and K-nearest neighbor classifiers, respectively.

Index Terms—Indoor identification, light intensity, location information, smartphone unlocking, Wi-Fi signals.

I. INTRODUCTION

LOCATION information collected by a user’s smartphone can be used as an additional identification factor to enhance the security and usability of user authentication on smartphones [1]. In practice, users on a daily basis visit and use specific locations and areas (e.g., offices, homes, and research labs) as their trusted places and frequently perform services on their smartphones. A risk-based authentication assessment has been conducted for users to study the validity of allowing them to perform frequent services without authentication mechanisms while they are well within the trusted areas [2]. For example, users can perform services or leave their smartphones unlocked at secure locations to reduce the burden of explicit unlocking of phones [3], [4].

Since 2014 Android v5.0, Google provided several authentication options under the menu called Smart Lock [5], [6]. A Smart Lock has an option called “Trusted Places,” which lets users set specific physical locations/areas as trusted places where their smartphones will automatically stay unlocked. That is, whenever a user is close to a trusted place, her phone will skip over the standard lock screen and let her access the phone without any explicit authentication mechanism. However, when they are anywhere else, the standard security methods on the phones will show up and apply. Since Smart Lock’s trusted places feature [7] relies mainly on GPS, phones remain unlocked within a radius of around 80 m to detect users’ trusted locations, and users cannot customize (increase or decrease) trusted location sizes. Therefore, this GPS-based user authentication method does not authenticate trusted locations at close and adjacent indoor areas, limiting its universal applicability and adaptation. Recently, Cho et al. [8] conducted two user studies that aimed to investigate users’ requirements for location-based authentication schemes. Their study results revealed that users prefer registering fine-grained indoor locations as trusted places and adjusting location sizes, which Smart Lock does not support. Based on such observations, we aim to design a location-based authentication scheme to address those requirements.

Several indoor location determination technologies using data from surrounding radio signals of wireless networks and lights have been proposed [9]–[12]. Although these techniques have mainly focused on finding characteristics of fingerprints of a specific location to estimate its coordinates, they have at least one of the following limitations that make them not applicable to our design goals. First, most of the approaches often depend on collecting radio signals and analyzing characteristics, such as Angle of Arrival (AoA) and Time of Flight (ToF) that require specific hardware (e.g., antenna arrays or benchmarks) to find the precise location of a user/device [13], [14]. Second, other approaches depend on collecting the physical channel state information (CSI) from unique Wi-Fi network interface cards available only on particular laptops (not supported on smartphones and tablets) to calculate the user’s location coordinates [15]–[18]. Third, visible light
communications (VLCs)-based approaches [19], [20] require bulbs installed with specially designed electronic circuits to perform data modulation/demodulation between transmitters and receivers.

We were inspired by these studies using Wi-Fi logs (e.g., SSID, BSSID, and RSSI) and light intensity (Lux values) to identify geographical locations. This work specifically aims to develop a highly accurate location-based smartphone unlocking scheme named LocID supporting fine-grained indoor location areas (≤ 2 m²) using Wi-Fi RSSI and light intensity features. Here, we summarize the key challenges of successfully developing LocID schemes. First, raw Wi-Fi data are unstable over time, even in the same place. Consequently, the instability of RSSI measurements can significantly reduce the performance of LocID schemes using the RSSI data. Therefore, we need to address the instability of the RSSI data. Our techniques filter out less visible network nodes (noisy nodes that appear occasionally) and leverage the measurements of the top-ranked network nodes (stable and reliable) to reduce the instability of Wi-Fi RSSI data. Second, the error rates of existing LocID schemes using Wi-Fi RSSI data are too high to satisfy users’ expectations. To improve the location detection accuracy, we consider both Wi-Fi (SSID, BSSID, and RSSI) and light intensity values as features. We note that light intensity readings are available everywhere and easily collected using typical smartphones and could be used as a complementary feature with Wi-Fi RSSI features. Finally, we carefully optimized hyperparameters to improve the detection accuracy of LocID.

We highlight this article’s contributions as follows.

1) We propose a highly accurate fine-grained LocID system named LocID using Wi-Fi RSSI and light features for secure and usable smartphone unlocking within small and adjacent indoor areas. LocID relies only on measurements transmitted from surrounding commercial off-the-shelf (COTS) Wi-Fi access points and light sources installed in indoor environments and received on smartphones without any additional specialized hardware. To specifically explore the effect of randomness in radio signals observed from Wi-Fi APs, we develop a new and robust mathematical algorithm called location-wise signal signature (LSS) that enables stable and reliable Wi-Fi signal signatures.

2) To show the feasibility of LocID, we developed a data collection application and collected real-world Wi-Fi and light data from 68 different locations distributed among four close and adjacent Area of Interest (AoI) for a period of five weeks on the same floor inside our university’s building. In addition, we conducted extensive experiments to extract feature vectors of Wi-Fi and light fingerprints and evaluated LocID performance under both area-wise and period-wise settings as well as different evaluation impacts.

3) We computed three evaluation metrics: a) false-negative rate (FNR) (the proportion of trusted locations inside trustworthy areas that are wrongly classified as untrusted locations); b) false-positive rate (FPR) (the proportion of untrusted locations at untrustworthy areas that are wrongly identified as trusted locations); and c) F1-score [a weighted average metric emphasizing on the model’s performance regarding false positive (FP) and false negative (FN)]. Our results show the effectiveness of the proposed technique achieving an overall F1-score of 98.3%, FNR of 2.9%, and FPR of 1.7% for the daily evaluation approach, and F1-score of 94.4%, FNR of 2.7%, and FPR of 3.7% for a long period (over weeks) evaluation approach. Further, we demonstrate the usability of LocID with only a one-scan (one-time online scan of Wi-Fi and light fingerprints) test set of location data, and our reported results show the overall AUC [90.6% support vector machine (SVM) and 90.8% K-nearest neighbor (KNN)] scores when models are trained over the first week but tested for the remaining weeks of the data set.

The remainder of this article is structured as follows. In Section II, we review related works. We present the system design overview of LocID in Section III. Section IV demonstrates the methodology details of LocID, including an analysis of the LSS algorithm. We proceed with detailed experimental results from our evaluation approaches in Section V. We finally provide our conclusion in Section VI.

II. RELATED WORK

A. Indoor Localization Techniques

In the past two decades, many approaches, such as OIL [21], Hours [22], and RADAR [23], have been proposed for the purpose of indoor localization in which radio signal measurements can be used to find location information. Further, several works, such as Zee [24], LiFS [25], and Unloc [26], use crowdsourcing applications to locate users in a fingerprint space. However, those solutions require using benchmarks to generate radio maps and heavy signal processing techniques that are difficult to employ for real-world scenarios on smartphones with limited computing resources. In contrast, we present a lightweight and efficient indoor LocID solution that does not require any factor affecting the practicality, such as radio maps, computing resources, or specialized devices.

To obtain precise location information with a low localization error, several approaches (e.g., [27]–[30]) have tried to use a complex physical layer and extract CSI, which is available only on some specialized Wi-Fi network interface cards (e.g., AR9380 and Intel 5300) equipped on devices such as laptops and cannot be obtained on smartphones. Other proposals require the use of special hardware, e.g., antenna array [31], [32] or depend on complicated methods such as ToF information with MIMO AP to localize users within a decimeter level [33].

Note that, although location-based approaches inspire our work, we do not need to estimate the precise user’s location coordinates since it is not the goal of this article. Instead, our work introduces the concept of location identification based on classifying fingerprints of a group of locations rather than only one location as indoor localization systems. In other words, LocID aims to develop models that can correctly identify the
trustworthy area in which all its locations have the same fingerprints and meanwhile differ from all locations’ fingerprints of untrustworthy neighbor areas. Classifying adjacent and close areas that have a group of locations is more complex than a single location classification because it requires finding common and distinctive characteristics among a group of locations at the same AoI.

B. Location-Based Identification and Authentication Approaches

The concept of using location information in indoor environments was also used for user identification/authentication on the devices. Denning and MacDoran [34] were the first who introduced the idea of using geographic location information of people when they log into computers and transact business electronically. Banking services [3], [4] also have applied the idea of using location data as a second-factor authentication solution on smartphones to detect fraudulent transactions and improve the security of sale transactions. Li et al. [35] presented a system called iLock that automatically locks the smartphone once a significant physical separation from its owner is detected to prevent data theft on a lost/stolen mobile device. iLock relies on acoustic signals and adopts frequency modulated carrier wave (FMCW) with at least one speaker and one microphone. However, iLock was only tested in two ideal and separated areas: 1) a library and 2) a lab. Chen et al. [36] presented a system called Chaperone that relies on acoustic sensing to capture the user’s departure patterns (motion and distance of the owner from the device) using the device’s microphone and speaker to lock it and prevent the loss in real time. However, acoustic sensing is severely affected by environmental factors, such as movement of nearby people, background noise, the presence of obstacles that highly affect the distance estimations and result in false detection.

Agadakos et al. [37] proposed the idea of determining the user’s location using the Internet of Things (IoT)—by collecting reports of location information and proximity between the user’s phone and paired IoT devices owned by the user to act as an additional factor of authentication. Alawami and Kim [38] investigated the possibility of classifying users’ seats inside the same room (research lab) for the authentication purpose on their work cubicles using radio signals fingerprints. Fridman et al. [39] demonstrated that the physical location information could be unique to identify users for improving active and continuous authentication on mobile devices—using GPS (when outdoors) and Wi-Fi (when indoors). Cho et al. [8] focused on investigating users’ requirements and perceptions and conducted two user studies about using location-based authentication for smartphone unlocking in indoor environments. We designed LocID in the light of their findings and implemented an accurate and fine-grained unlocking smartphone system in practice.

Furthermore, other studies have relied on location information for network security and access control purposes. Xiong and Jamieson [40] proposed a system called SecureAngle to improve the security of wireless networks by exploiting location information using the AoA method with aiding the Rice WARP boards that provide an array of Wi-Fi antennas. Yu et al. [41], in order to catch a spoofer located a few meters from the legitimate user, measured the similarity of subsequent network behaviors when a legitimate device is connected to Wi-Fi AP and continuously monitors whether there is a significant change in the values of physical characteristics (e.g., RSSI and RTT) or not. The work in [42] proposed a system, called SWISH, that exploited location information to address the security challenges when a mobile user connects to a foreign network instead of his faraway home network by authenticating three parties involved in the sharing process. Yuan et al. [43] presented an authentication system such that a user who wants to be authenticated seeks the physical location information from the UE of peers in the same network. The work in [44] exploited context information based on location changes to assess the state changes of patients suffering from disorders. They rely on the availability of Wi-Fi hot-spots as the data source for the approach by analyzing the long-term stay locations visited by patients.

Also, visible lights have been exploited for developing location-based Internet access protocols when users are physically present under light bulbs installed in the ceilings [45]–[48]. They used the existing LED light-bulbs infrastructure equipped with control circuits (transmitters) to transmit the shared secret keys to the smartphones (receivers). However, in practice, VLC-based works have disadvantages that limit the employment of the protocols as follows. First, every VLC-enabled bulb should be installed with specially designed and expensive electronic circuits to perform the modulation/demodulation processes and share secret keys with the receivers. Second, although photodiodes receivers can be mounted on smartphones, they suffer from transmitters’ alignment and synchronization issues. The work in [47] proposed a light-based user identification system using intensity readings collected from built-in light sensors on smartphones and distinguishing users’ behaviors when walking through specific locations and paths in indoor environments. In addition, students’ location information are used to enhance the academic performance of the education in the schools and universities. The system in [49] presented an attendance-based application using location information when connecting to the Wi-Fi AP and the facial recognition property to mark the attendance when the student is inside the classroom. Similarly, smart classroom work in [50] developed an Android application installed on students’ smartphones and remotely controlled by authorities (e.g., university administration) to automatically block all less essential functions and allow to use only the permitted list of applications while they are inside lecture rooms.

III. SYSTEM DESIGN

In this section, we introduce an overview of the LocID system and its architecture.

A. Overview of LocID System

Here, we give an explanation of the terms used in this article, the scenario of how LocID operates, and challenges we
faced during the design. At the beginning, we describe terms of trusted/untrusted locations and trustworthy/untrustworthy A0Is used during the implementation of the work. The term trusted/untrusted locations refers to a place of the small area having size ($\leq 2$ $m^2$) while the term A0Is refers to the entire room-size, such as research lab rooms, classrooms, offices, or lounges that should contain many adjacent locations. The concept of trustworthy A0I means an indoor area in which all its locations are considered as trusted locations where the smartphone should be unlocked. In contrast, all locations in adjacent untrustworthy areas, such as corridors that connect these trustworthy areas, are considered as untrusted locations where the smartphone should be locked. Therefore, the users can unlock their smartphones as long as they provide valid location fingerprints inside the trustworthy A0I, and the user is then authorized to unlock his/her smartphone. In contrast, when the locations’ fingerprints belonging to an untrustworthy A0I are identified, a smartphone is potentially locked.

To develop a practical location-based device identification scheme, we faced some challenges as follows: first, the stability of fingerprint readings (RSSI signal values and light intensity) are highly affected by the layout of the place, walls, and other obstacles that leads to different reflections and multipath over each A0I and time. Second, the collected fingerprints are also influenced by device diversity issues when tested over various Android-based smartphones due to the manufacturing differences of the built-in sensors. These substantial challenges are the reasons to explain why we do not directly rely on raw collected fingerprints. Instead, we developed our robust location-signal-signature (LSS) algorithm so that the LocID system is neither subordinated nor affected by these challenges and can accurately identify the area to which the smartphone belongs.

B. Architecture of LocID

The LocID system architecture is shown in Fig. 1 where the LocID process using Wi-Fi and light fingerprints in an indoor environment consists of five major components as follows.

**Data Collection:** Users have to first install our developed application on their smartphones, where LocID collects raw Wi-Fi and light signal readings from the surrounding environment. Then, LocID sends the collected fingerprints to the trusted third-party server (the server is under our control) and applies the location signal signature (LSS) algorithm for extracting feature sets. After that, LocID implements the classification process based on the pretrained location-identity models for location identification purposes. Based on the classification result, LocID identifies the location’s identity as “Trusted” as long as the location fingerprints are correctly classified within the trustworthy A0I, and the user is then authorized to unlock his/her smartphone. In contrast, when the locations’ fingerprints belonging to an untrustworthy A0I are identified, a smartphone is potentially locked.

**Detection:** A0I-wise Score

Period-wise Score

***Learning***

Location-Identity Models

Location-Identity Learning

Fig. 1. Overview of LocID.
ALAWAMI et al.: LocID: A SECURE AND USABLE LOCATION-BASED SMARTPHONE UNLOCKING SCHEME

Fig. 2. LocID Android application.

location within each AoI: 1) Wi-Fi characteristics of surrounding wireless network nodes (i.e., BSSID, SSID, and RSSI) and 2) light readings (intensity values in Lux). Additionally, the app shows four “Edit Text” labels to enter the IDs used in our experiments: the ID of the interested area, the ID of the location inside the area, the day of collection, and the location’s label (Trusted or Untrusted). Eventually, the app shows a per-second increment counter and start/stop buttons for controlling the data collection process. The data collection process has been conducted at a total of 68 different locations distributed randomly in four adjacent AoI areas inside our university’s building as follows: research laboratory (ID: AoI_1, contains 20 locations), lounge room (ID: AoI_2, contains 14 locations), corridor (ID: AoI_3, contains 22 locations), and another lounge room (ID: AoI_4, contains 12 locations). More details about the layout of the tested environment are given in Section V-A. All these recorded fingerprints of locations are automatically saved in (.csv) files and sent to our server to implement location-based identification using LocID.

**Location-Identity Association:** In this step, for each AoI in $\Psi = \{AoI_1, AoI_2, AoI_3, AoI_4\}$, we predetermined $M$ locations $\Phi = \{1, 2, 3, \ldots, m\}$ as group training points for the users who belong to the area and want to use smartphone unlock property. We collected a quantity of $N = \{1, 2, 3, \ldots, n\}$ scans of wireless fingerprints at each location during every collection day while the smartphones were put on the tables or seats of users. Then, we preprocessed the collected data to remove the irrelevant network nodes (less visible APs) that are rarely seen from the surrounding environment. After that, we vectorized the location RSSI readings with the corresponding network node MAC address at each location. Finally, we augment light readings to the vectorized Wi-Fi data and then associate the required labels (i.e., Location id, AoI id, and trusted/untrusted) for identification.

**LSS Algorithm and Feature Extraction:** Moving to this step, we developed a novel LSS algorithm based on the wireless Wi-Fi and light data seen at the location in a specific AoI. We rely only on network nodes’ good stability and reliability attributes for Wi-Fi data: the most stable nodes have minor fluctuations in their RSSI readings collected over scans, and the most reliable nodes have strong RSSI values on average. After that, we use the created signal signature vectors of both Wi-Fi and light fingerprints to extract feature vectors of each location in an AoI for learning models.

**Location-Identity Learning:** Here, we feed the combination feature sets of Wi-Fi and light fingerprints extracted from the previous step to the machine learning techniques to train and create the location-identity models. In this work, we used two popular types of supervised machine learning algorithms: SVMs and KNN with a binary classification approach to evaluate LocID performance.

**Detection Process:** During this step, we tested new unseen fingerprints data of locations by considering two types of estimated scores: AoI-wise and period-wise under different evaluation scenarios. Then, we aggregated the final score to evaluate identification performance and usability.

**IV. METHODOLOGY**

In this section, we describe in detail the methodology of LocID, including data preparation, location-identity association, and LSS algorithm. Fig. 3 depicts the data preparation of the LocID system. Specifically, we selected four adjacent AoI areas inside our building, three of them (AoI_1, AoI_2, and AoI_4) are considered trustworthy areas, and all its inside locations 20, 12, and 14 are listed as trusted locations, respectively. In contrast, the remaining area (AoI_3) with its inside locations (22) is considered untrustworthy. A list of locations’ IDs is created for both trustworthy and untrustworthy areas. Data was collected with the location-wise collection process.
in which smartphones have been put in every single location to collect \( N \) scans during three time periods in a day (Morning, Afternoon, and Evening). All the collected data are then saved into (.csv) files that are named with labels: AoI_ID, location_ID, period (Morning, Afternoon, or Evening), and collection day. Each file contains data of \( N \) consecutive scans and each scan represents a list of readings in three columns (Wi-Fi MAC, Wi-Fi RSSI, and Light).

### A. Removing Irrelevant Data From Fingerprints

As known, Wi-Fi readings are random measurements that have several fluctuations and are highly affected by the surrounding environments’ obstacles, such as walls, distances between transmitters and receivers, and area geometry. Smartphones, during the sensing process, can hear up to hundreds of network nodes for every scan. However, when exploring the data collected at every location during experiments, we found that low visibility APs (faraway and temporary) do not appear in all \( N \) scans; instead, it occasionally appears out of a set amount of scans for the exact location. Therefore, to collect a consistent data set, we defined a threshold \( \alpha \) to validate network data of high visible APs in which any node that can be detected for more than \( \alpha \) times out of \( N \) scans will be considered as a visible network node. If not, the network node is considered irrelevant and its fingerprints will be ignored from the data set.

### B. MAC-RSSI Vectorization

In this step, we aim to vectorize filtered data of the visible network nodes in which each node’s MAC address is listed with all its RSSI readings of scans at every location. To do this, we implement the MAC-RSSI process as shown in Fig. 4 where LocID explores data of every network node and extracts the RSSI values of all scans associated with the targeted MAC address. In detail, suppose we have \( N \) scans collected at a certain location and each scan contains data (MAC and RSSI) of hundreds of network nodes, then LocID transforms location data into \( d \) data vectors that represent \( d \) network MACs (\( MAC_1, MAC_2, MAC_3, \ldots, MAC_d \)) with their \( N \) RSSI measurements (\( RSSI_1, RSSI_2, RSSI_3, \ldots, RSSI_N \)). Note that this step is done after removing irrelevant nodes from each scan in which we guarantee that all \( d \) vectorized nodes are the visible ones for identification and have RSSI measurement counts more than the threshold \( \alpha \). Moreover, some of the visible network nodes were missed at times in all of the scans (red crosses indicate this in Fig. 4), whereas some other visible nodes are not. For example, network nodes of \( MAC_1 \) and \( MAC_d \) have been detected for all \( N \) scans; however, nodes of \( MAC_2 \) and \( MAC_3 \) have been missed during only the first and second scan, respectively. Finally, we augmented the collected light readings and the Wi-Fi measurements of each location associated with the labels.

### C. Location-Wise Signal Signature Technique

Until now, we prepared the data (Wi-Fi and Light) vectors for each location collected at a specific period in a certain AoI. However, by exploring these Wi-Fi fingerprints, we found that the raw readings of sensed network nodes are not stable and have unreliable RSSI values when scanning several times at a specific location. This is because of many factors, such as rooms’ geometry, obstacles between nodes and smartphones, and network properties. In other words, we focused on addressing two main issues of Wi-Fi RSSI measurements, which are instability and unreliability due to the following reasons. First, for instability, a network node provides fluctuations in RSSI values for \( N \) consecutive scans even when the smartphone is located in the same location. These fluctuations vary from one network node to another and from one location to another. Second, for unreliability, we found that the RSSI values of a network node collected at the same location vary over scans from \(-30\) dBm (Amazing value means the AP can only be a few meters away from the location) to \(-90\) dBm (unusable value means the AP is noisy or faraway)

\[
CSSD = \sum_{j=1}^{n-1} (RSSI_{j+1} - RSSI_j) \quad (1)
\]

\[
RSSI = \frac{\sum_{j=1}^{n} RSSI_j}{n} \quad (2)
\]

\[
F = \left\{ \sum_{\forall x \in \Delta} f(x), \text{for all } x \in \psi_{MAC} \right\} \quad (3)
\]

where

\[
f(x) = \omega \times idx_{RSSI}(x) + (1 - \omega) \times idx_{CSSD}(x). \quad (4)
\]

To visualize the instability and unreliability of radio signals, we plot the signal signatures of randomly selected 20 network nodes observed at three different locations (Location ID = 6, ID = 10, and ID = 20) that are located in three adjacent areas (AoI_1, AoI_2, and AoI_3), respectively, as shown in Fig. 5. For each location, we plot three characteristics of

---

**Fig. 4.** MAC-RSSI vectorization process.
the network nodes that are collected over a sequence of \(N\) scans: 1) the raw RSSI readings; 2) the cumulative sum of signal difference (CSSD) using (1); and 3) the average RSSI values using (2). First, given the original values of RSSI readings of each network node for a sequence of consecutive \(N\) scans, we computed the CSSD values to measure the number of fluctuations that represent the stability of signal signature at a specific location. CSSD values range from 0 to 15 for the three locations, in which, the more minor the CSSD values, the more stable network nodes and vice versa. Then, we computed the average RSSI values (ranges from \(-90\) to \(-37\)); the larger the RSSI averages (near \(-30\) dBm), the more reliable network nodes and vice versa.

The plots show that some network nodes exhibit high average signal strength (more reliable) than others, and similarly, some nodes exhibit low CSSD values (more stable) than others, and these observations vary from one location to another. Therefore for a better identification process, this analysis motivates us to focus only on those network nodes that provide the most stable and reliable measurements of Wi-Fi signal signatures at each location for all \(N\) scans. These selected network nodes, called top-ranked nodes (\(\beta\)), provide the most important and consistent fingerprints that we rely on to infer stable and reliable signal signatures of locations.

**Optimization of Used Hyperparameters:** Here, we demonstrate how we decided to set the appropriate values of the basic parameters (\(\alpha\) and \(\beta\)) used during the experiments. First, the objective of determining the \(\alpha\) value is to ignore data from the less visible network nodes that rarely appear at each location and keep using fingerprints of only network nodes that are frequently visible more than the \(\alpha\). Specifically, in our experiments, we set \(N = 12\) scans when collecting fingerprints in each period (Morning, Afternoon, and Evening) at each location. Therefore, to decide the appropriate value of \(\alpha\), we plot the histogram distribution of all locations’ data collected over the course of one day. Particularly, Fig. 6 visualizes the number of network nodes (y-axis) that can be seen for more than the \(\alpha\) threshold at locations of three trustworthy areas (AoI_1, AoI_2, and AoI_4) and an untrustworthy area (AoI_3); given that smartphones can hear up to hundreds of network nodes.

The \(\gamma\)-axis in Fig. 6(a) and (b) shows how many trustworthy and untrustworthy locations belong to the same bins (range of network nodes).

As the goal is to choose the \(\alpha\) value that gives the least overlapping between locations of trustworthy and untrustworthy areas, we can see (inside dashed rectangles) that when \(\alpha = 5\), it provides less overlapping than when \(\alpha = 7\). Note that increasing the overlap in locations’ fingerprints will negatively impact the classification performance. Thus, empirically we considered the data from visible network nodes that are consistently detected for more than \(\alpha = 5\) and filtered out the others as irrelevant nodes. Second, the objective of determining the \(\beta\) value is to select the amount of top-ranked visible network nodes that have high stability and reliability among hundreds of nodes. Although this step is applied on visible network nodes (i.e., nodes are detected more than \(\alpha\) times), we noticed that not all of them perform well with stability and reliability properties. Therefore, to provide effective performance for our system, we infer data from only top \(\beta\) visible network nodes, which show the highest values of both stability and reliability among all nodes. To determine the appropriate value of \(\beta\), we trained models using KNN and SVM classifiers across various possible values of \(\beta = 10, 20, 30, 40, 50, 60\) and then calculated the performance metrics (F1, FNR, and FPR) over one-day of fingerprints size. Fig. 7 shows the impact
of \( \beta \) parameter on the identification performance between locations’ data of all trustworthy and untrustworthy areas using two classifiers. The figure shows clearly that the value \( \beta = 50 \) gives the best results with F1 up to 99%, FNR = 3.7%, and zero FPR for both classifiers. Therefore, in our experiments, we considered \( \beta = 50 \) as an appropriate value for evaluation and believe that it could be suitable in ubiquitous scenarios where locations in smart buildings can be covered with this value of network nodes.

Description of the LSS Algorithm: We explain the details of LSS Algorithm 1 as follows. As input, for each location, we used the outcome data of location-identity association stage which are: vector of all visible network nodes \( \psi_{MAC} = (MAC_1, MAC_2, MAC_3, \ldots, MAC_d) \), the RSSI matrix \( D_{RSSI} \) in which each column in the matrix represents the RSSI readings of corresponding node over all \( N \) scans (RSSI1, RSSI2, RSSI3, \ldots, RSSI_n), the light vector \( \phi_L \) which represents intensity measurements of \( N \) scans, the location id, and the AoI id.

In the beginning, lines 1–4 show the acronyms used in the algorithm. During the whole process, we iterate over each tested AoI area in the building, each location inside the AoI, and every network node, respectively. In lines 8–12, we handle the two properties: 1) stability by computing the CSSD using (1) and 2) reliability by averaging the RSSI values using (2). In lines 14–17, we create CSSD and RSSI vectors for all network nodes seen at \( m \)th location, sort them in ascending and descending order, respectively, and then tag the MAC addresses according to the sorted vectors. Continuing, in lines 18–21, considering that there are up to hundreds of network nodes seen at each location, this step aims to infer the highest indexed network nodes in both CSSD and RSSI vectors to take advantage of stability and reliability. First, we apply (3) to compute the importance value \( f(x) \) of each network node \( x \) (how much the network node is important in order to extract only the high-ranked important nodes) by accumulating the weighted indexes \( \Delta \) for all \( \omega \) values belonging to \( \Delta \) (containing values ranging between 0 and 1 to balance the effect of the two properties). Then, we sort the nodes in ascending order based on their accumulated index values and finally extract the top \( \beta \) network nodes with their RSSI measurements. Following, lines 22–24 augment the light measurements \( \phi_L \) to the top \( \beta \) APs data to construct beneficial signal signatures, \( \langle \vec{F}, \vec{RSSI}, \phi_L \rangle \), of the \( m \)th location. After that, we use these signature vectors with the list of statistical features shown in Table I for extracting feature sets of both Wi-Fi and light fingerprints (E). Similarly, we repeat this process to iterate all \( m \) locations inside each AoI area to get the final data matrix, \( (m \times E) \), as outputs to train the models.

V. EVALUATION

A. Experimental Setup and Data Set

As the primary aim of our work is to focus on the smart buildings that have several AoIs covered by wireless network service (e.g., universities, companies, airports, malls, and smart homes), we choose the third floor of our university’s building as a testing environment. In this work, we evaluate LocID on a large-scale and real-world data set collected at adjacent and close trustworthy and untrustworthy indoor areas, each of which has many locations, to perform realistic location-based smartphone unlocking. The top view layout of the dedicated floor is shown in Fig. 8, which contains four adjacent rooms (research labs and lounges) as well as corridors. We named the tested areas as follows: one research laboratory (AoI_1) consists of 20 locations, the first lounge
(1) Wi-Fi RSSI measurements.
- Mean, Standard deviation (std).
- Median, Variance values (Var).
- Autocorrelation coefficients: (Avg ACF, Var ACF).
- Average of Ten maximum values.
- Average of Ten minimum values.
- Range value: (Maximum - Minimum).
- Mode value: A number that appears most often.
- Root mean square value (r.m.s).
- Q1, Q2, Q3: First, Second, and Third quartiles.
- CV percentage (%): Ratio of the std to the mean.

(2) Light intensity readings.
- Mean, Standard deviation (std).
- Median, Variance values (Var).
- Autocorrelation coefficients: (Avg ACF, Var ACF).
- Average of Ten maximum values.
- Average of Ten minimum values.
- Range value: (Maximum - Minimum).
- Mode value: A number that appears most often.
- Root mean square value (r.m.s).
- Q1, Q2, Q3: First, Second, and Third quartiles.
- CV percentage (%): Ratio of the std to the mean.

that, some smartphones are equipped with RGB light sensors, our work focuses on getting ambient brightness as light intensity measurements in Lux regardless of the sensor type. We implemented $N = 12$ scans at each location with a sampling rate of one scan every $5$ s to receive network broadcasts and $1$ s to receive light intensity readings. The details of our real-world collected data set are shown in Table II. To qualify the experiments against changing measurements over time, we collected our data set during three time periods (Morning, Afternoon, and Evening) every day and for several days. For example, we put the mobile phone at the user’s location (e.g., on his table, sofa, chair, or on hand) with the screen facing up orientation and collected scans of surrounding Wi-Fi and light fingerprints for a specific time interval (e.g., $1$ min). This process was started from location 1 to location 68 in a cyclical movement and repeated for all 68 locations during the three periods a day (morning, afternoon, and evening) using two phones.

In detail, locations’ fingerprints were collected at three periods per day during the following chosen times: [Morning: (9:30–12:00), Afternoon: (14:00–18:00), Evening: (20:30–23:00)]. Moreover, we conducted extensive experiments at all 68 locations for collecting data set for ten days over five weeks. Specifically, the data set collected for ten days ($D_i$) is distributed among five consecutive weeks ($W_j$) as follows: \{(D_1, W_1), (D_2, D_3, D_4, W_2), (D_5, D_6, W_3), (D_7, D_8,
False Negative: untrustworthy located outside the trustworthy areas are correctly identified as users located inside trustworthy areas. True Positive (TP): LocID can identify users located inside the trustworthy areas that are wrongly classified as untrusted locations and hence unlocks the smartphone at insecure locations. The formula for the F1 score is: F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, which is a weighted average metric emphasizing on the model’s performance regarding FP and FN.

Our evaluation approach to assess the performance of LocID was based on two main settings: 1) Area-wise profile and 2) Period-wise profile. Specifically, we focus on investigating the effectiveness of the system through estimating performance scores of every two neighboring trustworthy and untrustworthy areas during three time periods (Morning, Afternoon, and Evening). Then, we aggregate the results by averaging all scores of the two profiles for various impacts as follows. First, because all indoor environments are basically installed with light sources, we studied the feasibility of two different feature sets: Light-only and Union of Wi-Fi and light. For the light-only feature set, we train the models to show the extent to which LocID can work in scenarios of less presence of Wi-Fi APs in non-smart environments. Then, the evaluation impact of the union feature sets of Wi-Fi and light fingerprints are estimated and compared to the light-only feature set using only one day of the data set. Second, we studied the evaluation impact of the whole ten-days data set and the device diversity property under the condition of a short-term period. The models were daily retrained on a portion of a day’s data set and then tested on the remaining data set of the same day to show whether the LocID performance changes from day to day or not, as well as from one device to another. Third, we evaluated the performance of LocID against unseen fingerprints in which the models were trained using one-day data set of the first week and tested using fingerprints collected over the remaining days in the other four weeks to study the impact of LocID’s performance consistency over the long-term period.

C. Results

Here, we aim to answer the following questions.

1) Q1: How effective is the baseline light feature set compared to the union of both light and Wi-Fi feature sets?

2) Q2: How effective is the LocID system over the whole ten-days data set and device diversity when models are trained and tested over short-term period (using the same-day data)?

3) Q3: How effective is the LocID for consistency over the long-time period (several weeks)?

To answer the questions, we used two types of machine learning algorithms: SVMs and KNNs to learn the location-identity models. We set the two classifiers’ parameters as follows: for SVM, we found radial basis function (RBF) works best with γ = 0.01 and C = 10 and for KNN, we used the minkowski distance as the distance metric with the optimal parameter of n_neighbors as 3.

1) Impact of Different Feature Sets: We first evaluated LocID using the baseline light fingerprints since light sources are installed and available in all kinds of indoor environments. We trained models with light-only feature set that we got from the LSS algorithm for every two neighbor areas (one is trustworthy and the other is untrustworthy). During learning, we selected a one-day size of the data set that was collected in three periods of times (M, A, E) and randomly divided the locations’ data of the three trustworthy areas (AoI_1, AoI_2, and AoI_4) and the untrustworthy area (AoI_3) into 75% for training and evaluated with the remaining 25% using KNN and SVM classifiers. As shown in Table III, we measured the FNR, FPR, and F1 values for every two neighbor areas and then computed the overall results by averaging all areas and day periods. The results show that the SVM classifier provides slightly better results than KNN with 88.2% F1 score and 7.3% FNR; however, the FPR reached 14.9%.

Following that, we repeated the same evaluation setting but by considering the union of both light and Wi-Fi feature sets when α = 5 and β = 50 are used. Table IV shows the FNR, FPR, and F1 values for every two neighbor areas and the overall results using the same set of the trained and tested locations. The results are highly improved in which the SVM classifier still provides better accuracy than KNN with zero FNR, 1.85% FPR, and F1 enhanced to 99%. Thus, this stage of evaluation shows the effectiveness of crowdsourcing both feature sets for better performance over the course of one-day size of the data set.
 TABLE III
EVALUATION OF THE LIGHT-ONLY FEATURE SET FOR A ONE-DAY DATA SET USING TWO CLASSIFIERS

<table>
<thead>
<tr>
<th>Tested Environment</th>
<th>Trained locations</th>
<th>Tested locations</th>
<th>Period</th>
<th>KNN Classifier</th>
<th>SVM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Location Set 1</td>
<td>Location Set 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AoL1, AoL3)</td>
<td>31</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>(AoL2, AoL3)</td>
<td>27</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(AoL4, AoL3)</td>
<td>25</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Overall [%]</td>
<td></td>
<td></td>
<td></td>
<td>7.3</td>
<td>18.5</td>
</tr>
</tbody>
</table>

 TABLE IV
EVALUATION OF THE UNION FEATURE SETS (WI-FI AND LIGHT) FOR A ONE-DAY DATA SET USING TWO CLASSIFIERS

<table>
<thead>
<tr>
<th>Tested Environment</th>
<th>Trained locations</th>
<th>Tested locations</th>
<th>Period</th>
<th>KNN Classifier</th>
<th>SVM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Location Set 1</td>
<td>Location Set 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AoL1, AoL3)</td>
<td>31</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>(AoL2, AoL3)</td>
<td>27</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>(AoL4, AoL3)</td>
<td>25</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Overall [%]</td>
<td></td>
<td></td>
<td></td>
<td>7.3</td>
<td>18.5</td>
</tr>
</tbody>
</table>

Fig. 9. Performance evaluation of the light-only feature set for the ten-days data set. (a) F1-score. (b) False negative rate (FNR) and false positive rate (FPR).

Fig. 10. Performance evaluation of the union feature sets (Wi-Fi and Light) for the ten-days data set. (a) F1-score. (b) False negative rate (FNR) and false positive rate (FPR).

2) Impact of Whole Data Set and Device Diversity: The results from the previous stage motivate us to evaluate LocID’s performance to answer the Q2 when the short-term period of evaluation is considered. To do this, we trained the models using every 75% of a day’s data set and tested the models with the remaining 25% for all ten days. Here, we aggregated the evaluation results under the two profiles (i.e., AoI-wise and Period-wise) to provide final FNR, FPR, and F1 values over a day. Figs. 9 and 10 show the results using light-only and union feature sets, respectively, on Galaxy Note5 smartphone. Similar to the previous stage, the union feature sets of light and Wi-Fi provide higher results across all ten days of fingerprints using both KNN and SVM classifiers. Specifically, the F1 scores of the union feature sets for most of the tested days at the three periods (M, A, E) are ranging from 92% to 100% compared to the F1 scores of the light-only that range from 78% to 94%.

For FNR, light-only provides FNR values less than 20% except for day 6 and day 7 which reached 30% when using the SVM classifier. These results are not feasible compared to those achieved using union feature sets where FNR values are improved mainly to less than 5% for some days, more than 5% for only day 1, day 2, and day 7, and zero value for the remaining days. This emphasizes the robustness of the union feature sets than the light-only in identifying trustworthy locations’ fingerprints correctly and the ability to avoid frustrating users when unlocking smartphones. Similarly, for FPR, light-only provides FPR scores up to 20% using the KNN classifier, while the SVM classifier provides better performance for all ten-day data up to about 15%. However, using the union fingerprints, FPR is improved to less than 2% for most days, especially using the SVM classifier which shows much lower FPR scores (showing zero scores in some days) compared to the KNN classifier.
As known that the sensitivity of built-in sensors depends on the type of smartphone due to the manufacturing aspects, we show the extent to which readings collected from various smartphones will affect the LocID performance. Therefore, we repeated the evaluation of this stage with the whole ten-days data set collected using another smartphone device (Galaxy S8). Table V shows the overall effectiveness of LocID against device diversity using light-only and union feature sets. Overall, LocID provides slightly better performance with the Galaxy Note5 device than the Galaxy S8 using the union feature sets. Specifically, for Galaxy Note5, the KNN classifier provides the best FNR as 2.9%, while the SVM classifier provides the best FPR as 1.7% and the highest F1 score as 98.3%. However, Galaxy S8 device shows a slight drop in performance with 3.7% FNR, 2.4% FPR, and 97.7% F1 score when using the SVM classifier. Because of our methodology that does not rely on the raw but on the top-ranked (stable and reliable) measurements, the results indicate no big difference in the performance and show the feasibility of LocID usage on diverse smartphones.

3) Consistency Over Long-Term Period: Here, we study the effectiveness of performance consistency over several weeks in which the models are trained on a one-day data set size of the first week and tested using unseen fingerprints of other days’ data sets during the remaining weeks. Note that since results from the previous two stages have shown that union feature sets provide better performance than the light-only feature set, we only considered the union feature sets of light and Wi-Fi fingerprints for this evaluation. In detail, we train the models to learn location fingerprints using 75% of the data set size collected in \( D_1 \) of the weeks \( W_1 \). Then, the models are tested using 25% of the unseen data set collected in other days of the following weeks (\( W_2, W_3, W_4, \) and \( W_5 \)).

Figs. 11 and 12 show the evaluation results considering two smartphones using two classifiers. Specifically, the F1 score of Note5 smartphone ranges from around 90% to 100% while Galaxy S8 ranges from above 88% to slightly more than 96%. For FNR and FPR scores, Galaxy Note5 on average provides scores less than 6% using both SVM and KNN classifiers. The best performance reported for FNR and FPR was zero values when tested using the fifth week’s data, while when tested using the fourth week’s data, the performance drops of FNR and FPR reach 9%. Similarly, for Galaxy S8 device, FNR scores provided by the SVM classifier are better than those provided by the KNN classifier for all tested weeks; the highest reported FNR was only 3.2%. Also, we got almost similar FPR values provided by two classifiers; the best values were reported during the second and fifth weeks (less than 2%) while the worst values of FPR were during the third and fourth weeks (up to 9%).

Finally, we averaged all results of this stage in Table VI to show the overall impact of the “consistency over long-term period” on the evaluation of our system. In general, the SVM classifier shows F1 and FNR values better than those using KNN for the two smartphones. However, the KNN classifier provides better FPR values than those using SVM for two smartphones. Specifically, the SVM classifier shows an F1-score of 94.4% and FNR of 2.7%, whereas the KNN classifier shows an FPR of 3.7%. We hypothesize that the drop in performance happens not only because of the long time gap between training and testing periods, it also depends on other factors, such as network properties and weather conditions that vary from time to time. Also, the slight decrease in performance over the long-term period can be enhanced by retraining the models periodically whenever the identification errors increase over time. Eventually, our results evaluated under different scenarios and impacts show the promised feasibility of our proposed LocID system for providing secure and usable smartphone unlocking feature in fine-grained and adjacent indoor areas using location data.
D. Identification Results

During the previous experiments, we evaluated the performance of LocID when $N$ number of scans was collected at each location. In this experiment, we use one-scan data size as a test set to perform a suitable tradeoff between system performance and usability in which users indeed need to only collect one-scan of location’s fingerprint whenever they unlock the smartphones. In other words, we aim to evaluate the effectiveness of the models when tested using only one-scan of the location’s fingerprint to perform practical LocID for unlocking smartphones.

Note that the requirement for more than one-scan for identification will corrupt the system’s usability since Wi-Fi scanning API on smartphones provides one scan every 3 s. To this end, we show the extent to which our system is effective when models are tested using only one-scan data. Specifically, each model is trained in the same manner as in the previous experiments using the data set of Day 1: Week 1 of both Wi-Fi and light fingerprints. Furthermore, this process is done for locations at trustworthy and untrustworthy areas and during three periods (Morning, Afternoon, and Evening). After the training process, the models are evaluated using only one-scan of data collected from trustworthy and untrustworthy locations. The data used for the training and testing were collected on different days in the five consecutive weeks as follows: training-day data are $\{ (D_1 : W_1) \}$ and testing-days data are $\{ (D_2 : W_2), (D_3 : W_3), (D_8 : W_4), (D_{10} : W_5) \}$. The reason behind selecting the test days in separate weeks is to study the models’ performance in maintaining the accuracy throughout the long-time use for identification. We used ROC curves to compute the AUC (area under the curve of the ROC), which indicates how models perform classification problems under all possible thresholds and results in different true positive (TPR) and false positive (FPR) rates. Fig. 13 shows the ROC curves with AUC values using only one-scan test fingerprints of both trustworthy and untrustworthy locations.

Evening) to evaluate time consumption for LocID. It is worth noting that the data collection time of Wi-Fi fingerprints for each scan on Android devices takes 2–3 s.1 Table VII shows the authentication delay of our method under the two smartphones. The majority of time consumption comes from the preprocessing stage (location-identity association and LSS algorithm), which occurs only once during the offline phase. Overall, during the identification phase, LocID only requires an average delay of 281 and 277 ms of the two smartphones, respectively. This implies that LocID can identify locations to lock/unlock smartphones timely.

Memory Usage: To develop trained machine-learning models for being useful in real world, it is necessary to ensure the accessibility of LocID’s Android App on smartphones to make online predictions and provide practical locking/unlocking service. However, when deploying machine-learning models, it is important to monitor the memory usage of the LocID system on smartphones. Specifically, the size of the LocID

E. Other Performance Considerations

This section analyzes the LocID performance on other aspects, such as the authentication delay, memory usage, and power consumption on two devices, Galaxy Note5 and Galaxy S8.

Authentication Delay: Since this work is based on a client–server architecture design, the authentication delay is defined as the time interval of two stages: 1) fingerprint sensing time on smartphone and 2) time of preprocessing, training, and classification processes on the server. In total, it consists of the time a user collects location fingerprints on the device, data preprocessing, training, and classification on the server. We followed the same evaluation settings explained in Section V-D with the two adjacent areas AoI_1 (20 trusted locations) and AoI_3 (22 untrusted locations) and repeated experiments for ten times during the three periods (Morning, Afternoon, and Evening) to evaluate time consumption for LocID. It is worth noting that the data collection time of Wi-Fi fingerprints for each scan on Android devices takes 2–3 s.1 Table VII shows the authentication delay of our method under the two smartphones. The majority of time consumption comes from the preprocessing stage (location-identity association and LSS algorithm), which occurs only once during the offline phase. Overall, during the identification phase, LocID only requires an average delay of 281 and 277 ms of the two smartphones, respectively. This implies that LocID can identify locations to lock/unlock smartphones timely.

Memory Usage: To develop trained machine-learning models for being useful in real world, it is necessary to ensure the accessibility of LocID’s Android App on smartphones to make online predictions and provide practical locking/unlocking service. However, when deploying machine-learning models, it is important to monitor the memory usage of the LocID system on smartphones. Specifically, the size of the LocID
Android application on Galaxy Note5 and Galaxy S8 devices is 11.06 and 3.38 MB, respectively. However, the size of trained SVM and KNN models is 4.7 and 8.2 kB, respectively (see Table VII). The average memory usage of the LocID application when deploying on a smartphone is almost very little when compared to other applications. This emphasizes the usability of the LocID system for real-world deployment on smartphones.

**Power Consumption:** Location sensing of LocID is triggered only when the user is residing inside his/her registered secure area and wants to unlock the smartphone. If a user is in a nonprivate environment, e.g., outside the building, LocID’s processing needs will be negligible (i.e., no location sensing). However, LocID may still be occasionally triggered for a long period of time, e.g., the user is staying in his/her private location for couple of hours, while leaving the phone on a table. Therefore, we use Android Battery Historian\(^2\) to profile LocID’s energy consumption. We charged the battery of Galaxy Note5 and Galaxy S8 devices to its full capacity with a 3000 mAh and kept the smartphones’ services running normally while the LocID application was continuously sensing Wi-Fi and light fingerprints for an hour. We then collected battery logs from smartphones and visualized battery historian report using docker toolbox which displayed the battery level when the phones were in the discharging state. The power estimates consumed by LocID are only 5.2% (156 mAh) and 1.44% (43.2 mAh), see Table VII, in an hour for Galaxy Note5 and Galaxy S8, respectively. LocID is still acceptable for daily locking/unlocking use especially with help of the triggering methods, i.e., lightweight sensors such as Accelerometer which provides activity detection to avoid unnecessary Wi-Fi scanning and save battery.

**F. Limitations and Future Work**

In this work, some issues remain to be explored as follows.

**Identification Operation Design:** Since this work is aimed only to examine the feasibility of the proposed idea as a classification system, we used client/server implementation to transfer the location data between the client (user’s smartphone) and the server. However, a local on-device identification design is a preferred choice because of two reasons. First, in client/server design, transferring location data through network communication is unsafe and can be interrupted due to malicious activity or connection and system failure. Second, recent smartphones have expanded in storage and computational resources as well as the rapid development of the deep learning tools on devices (e.g., Tensorflow lite), both enabling trained deep learning models to be deployed and run locally on smartphones. We assume a client/server for our experiments and leave on-device local identification for future work.

**Generality of the System:** In this work, we collected the data set and conducted the experiments inside only one building at our university. However, this may not be sufficient to ensure the generality of our system. Therefore, we may need to implement more experiments on other buildings with different structure geometry and network nodes availability.

Position of Smartphones: For simplicity, during the data collection process, we used fixed position and orientation of smartphones (were put on tables, chairs, and sofa, with the screen facing up orientation). However, this may not be sufficient to simulate all practical implementations. To understand the requirement for LocID, we plan to extend our work by collecting more location fingerprints with other scenarios of user mobility and smartphone orientations and investigate its impact on the system performance.

**Assessment for Adversarial Attacks:** Location information is highly sensitive and could be exploited to predict (or even identify) a person’s trusted locations and movements. For example, SmartLock provided by Google could pose serious threats on users because their smartphones automatically stay unlocked within a wide-range of registered trusted locations/areas. That is, whenever an attacker is in close proximity to a secure area, the stolen smartphone will bypass the standard lock screen and freely access the phone. Therefore, the LocID system resists the attack by reducing the range of locking/unlocking area and evaluating authentication within fine-grained adjacent and close locations. In contrast, designing the LocID module using machine-learning techniques requires addressing security assessments where adversaries may launch advanced attacks to fool the authentication model. The system may receive two possible attacks (but not limited to).

1. **Model Stealing and Reverse Attacks:** Fingerprint-based location authentication solutions with client–server settings often hold the database on a server that compares with online location fingerprints. This makes it vulnerable to various attacks and leaks of the user’s location privacy. For example, the attacker aims to infer the data set used to build the authentication model and use a few queries to generate an approximation model that closely matches the target one. However, many proposed privacy-preserving fingerprint-based location solutions may mitigate this attack by taking advantage of the secure two-party computation (STPC) techniques [52], [53].

2. **Reply Attacks:** Location fingerprints rely completely on user-specified information, which leads to possible replay or faking attacks. An attacker may generate adversarial examples and manipulate input data to fool or bypass the authentication model, i.e., leading the model to generate wrong output and misclassification [54].

Note that all the considerations discussed in this section could inspire future works in the context.

**VI. Conclusion**

We have presented a novel framework, LocID, that investigates providing a fine-grained and usable unlocking smartphone service by exploiting location information within adjacent indoor areas. In addition, LocID includes a mathematical technique named LSSs that analyzes both radio signal characteristics and light intensity measurements to create stable and reliable feature vectors for location-identity learning. Our evaluation results on a real-world data set collected
from 68 trustworthy and untrustworthy locations distributed in four adjacent areas, demonstrate that LocID effectively classifies the locations' fingerprints in terms of both area-wise and period-wise settings over short-term and long-term approaches. Furthermore, we show that LocID can model location data when tested using only one scan of readings to simulate actual smartphone identification. In future work, we plan to extend the work by developing an authentication system with practical requirements and deployments, such as on-device training and prediction, exploiting more features (e.g., cellular and magnetometer); and maintaining model accuracy throughout the use of a system.

REFERENCES


[10] P. Davidson and R. Piché, “A survey of selected indoor positioning techniques (e.g., cellular and magnetometer); and maintaining model accuracy throughout the use of a system.”


[50] Aishwarya Ram Vinay received the B.S. degree from the Department of Computer Science Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India, in 2016, and the master’s degree from the Department of Electrical and Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea, in 2021. She is currently working as an Applications Engineer with Synopsys, Seongnam, South Korea.

Mohsen Ali Alawami received the B.S. and M.S. degrees from the Department of Electronics and Communications Engineering, Cairo University, Giza, Egypt, in 2010 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea, under the direction of H. Kim.

His research interests are focused on security engineering, location determination and positioning systems, mobile security with machine learning and deep learning, and wireless communications.

Hyounghick Kim received the B.S. degree from the Department of Computer Science Engineering, Sungkyunkwan University, Suwon, South Korea, in 1999, the M.S. degree from the Department of Computer Science, KAIST, Daejeon, South Korea, in 2001, and the Ph.D. degree from the Computer Laboratory, University of Cambridge, Cambridge, U.K., in 2012.

He is an Associate Professor with the Department of Computer Science and Engineering, College of Software, Sungkyunkwan University. After completing his Ph.D., he worked as a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. He previously worked for Samsung Electronics as a Senior Engineer from 2004 to 2008. He also worked as a Distinguished Visiting Researcher with Data61, CSIRO, Sydney, NSW, Australia, from 2019 to 2020. His current research interests are focused on usable security, blockchain, security vulnerability analysis, and data-driven security.