Deep-Fading Hole Avoidance for Secure Region Detection using Channel State Information

Jihwan Suh, Yongjae Yoo, Jeongyeup Paek, and Saewoong Bahk

***Abstract*—One of the critical challenges in many wireless systems is the *deep fading hole* problem where signals interfere destructively to create an abrupt change in signal amplitude due to multipath fading. In this work, we tackle this challenge in the context of *secure region detection (SRD)* problem. Specifically, we propose *SHARD*, a novel *hole avoidance* technique that analyzes channel state information (CSI) to significantly improve the accuracy of CSI-based SRD. *SHARD* identifies potential *fading holes* in CSI amplitude, and removes an unreliable portion of data while utilizing the remaining unaffected part. To compensate for the loss of information, we define *phase-distance* for reliable use of time-varying CSI phase, and neighboring reference points are utilized for accurate matching. Our real-world experiments show that *SHARD* can achieve a near-perfect 99.96% true-negative ratio (successfully rejecting devices not in the secure region) and an excellent true-positive of 98.01% for practical usage, significantly better than state-of-the-art prior work. We believe our ideas can be generalized to many RF-based localization systems to mitigate the *deep fading hole* problem and improve their accuracy.**

***Index Terms*—Channel state information, deep fading hole, finger-printing localization, Internet of things, secure region detection.**

I. INTRODUCTION

INTERNET of Things (IoT) is one of the most prominent technology trends that have emerged in recent years, and a vast number of IoT devices have infiltrated our personal lives. Accordingly, the devices create, store, and use private information, and therefore it is necessary to ensure that the devices are used only under proper authorization or authentication. Authorizing the user is one way to achieve this. However, the sheer number of IoT devices often hinders this approach. A more scalable and effective way to solve this problem is to allow activation and participation only to devices located in their predefined 'secure' regions. The problem of verifying whether a device is within their designated region or not for authentication/authorization purposes without obtaining an absolute position is called *secure region detection* (SRD) [1].

SRD is a new problem similar to the localization problem, but different in nature. The main difference is that many existing indoor localization techniques assume the use of pre-acquired data in all areas where the device may potentially exist unless completely out-of-range. This assumption is inappropriate for SRD because the device can be located where no data collection has been made for. For example, it should not be necessary to fingerprint the whole floor or room when a secure region is a single desk. Since it is costly and sometimes impossible to obtain all data from practically unbounded non-secure region, SRD should work with data obtained only from the predefined secure region without negative reference points.

One of the common challenges in many wireless systems, including SRD and indoor localization, is the *deep fading hole* problem [2]–[4]. At a certain physical location, the amplitude of a signal is significantly smaller than its surroundings due to destructive interference from multipath fading, resulting in a *deep fading hole*. The signal obtained at a hole not only has low received signal strength (RSS) compared to its surroundings, but also exhibits a significant change in channel state information (CSI). In other words, CSIs obtained at two sub-centimeter apart points can show a big difference in reality unlike what theoretical channel models suggest. The hole makes it difficult to use fingerprinting techniques (e.g., [5]) which rely heavily on the assumption that signal signatures at nearby locations will be similar [6], [7].

Furthermore, phase information has been difficult to utilize for fingerprinting so far because it varies non-deterministically over time. Although the phase change that occurs as the signal passes through the channel should be constant, the phase measured by a device changes due to asynchronous phase locked loop (PLL) clock at transceivers [8], [9]. Thus, the CSI phase obtained from the same location may be different at different times, making it difficult to use for fingerprinting.

In this paper we propose *SHARD*, a CSI-based *secure region detection* scheme with *hole avoidance*[1]. *Hole elimination* identifies potential *deep fading holes* in CSI amplitude not to be deceived by them, and removes unreliable portion of data while utilizing the remaining unaffected part. Furthermore, CSI *phase-distance* is defined for reliable use of time-varying phase information to improve accuracy. Additionally, to compensate for the loss of information due to hole elimination, a *1-2-4 matching* technique is employed to utilize neighboring reference points if applicable and beneficial.

We implement *SHARD* on commercial Wi-Fi NICs and

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2021-0-02048) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation),

Manuscript received September 28, 2022; approved for publication, October 10, 2022. This paper is specially handled by EIC and Division Editor with the help of three anonymous reviewers in a fast manner.

J. Suh, Y. Yoo, and S. Bahk are with the Department of Electrical and Computer Engineering and INMC, Seoul National University, Seoul 08826, Republic of Korea, email: {jhsuh, yjyoo}@netlab.snu.ac.kr, sbahk@snu.ac.kr.

J. Paek is with the Department of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea, email: jpaek@cau.ac.kr.

S. Bahk is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2022.000046

[1] The name of our proposed scheme, S'HA'RD, embeds *hole avoidance* (HA) inside *secure region detection* (SRD).

1229-2370/22/$10.00 © 2022 KICS

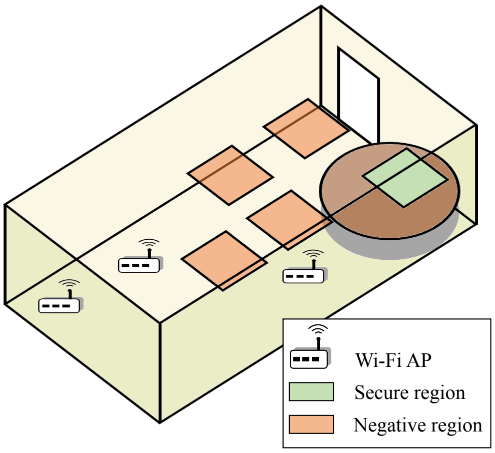


Fig. 1. Example: Secure region on a table (green square area). An user device can be unlocked only when in that region. Negative (non-secure) region (orange square area) are used for evaluation purposes.

evaluate its performance through real-world experiments to show that *SHARD* achieves an excellent true negative (TN) of 99.96% while maintaining a true positive (TP) of 98.01%, which means that devices outside the secure region are near-perfectly rejected (TN) while legitimate devices are adequately authenticated (TP). We also compare *SHARD* against SWORD [1], a recent state-of-the-art SRD scheme, to show that *SHARD* achieves significantly better performance on the same testbed wireless environment.

The contributions of this paper are as follows;

- We demonstrate the existence of *deep fading holes* and how they adversely affect wireless signature detection.
- We propose two novel hole avoidance techniques to address the problem, *hole elimination* and *phase-distance* estimation. We then design *SHARD*, a novel secure region detection scheme that overcomes the deep fading hole problem.
- We implement *SHARD* on a commercial Wi-Fi device for evaluation through real-world experiments, and compare it with a recent state-of-the-art SRD scheme to show its improved performance.

The remainder of this paper is organized as follows: We discuss the related work and our motivation in Section II. Then, Section III presents the design of our proposed scheme *SHARD*. We evaluate *SHARD* through real-world experiments in Section IV, and conclude in Section V.

II. Related Work and Motivation

In this section, we first introduce prior work in the literature that discusses techniques to overcome multipath fading. Then, we motivate our work by showing how significantly the *deep fading hole* problem impacts SRD performance.

A. Related Work

Multipath fading adversely affects communication and localization in many wireless systems. Since it is mostly due to reflections off the floor, ceiling, walls, as well as objects, it is challenging if not impossible to predict nor estimate in practical settings. Thus, instead of attempting to eliminate multipath fading, various techniques have been proposed to overcome it.

For example, Bhatti *et al.* [10] proposed an outlier detection scheme using machine learning to eliminate outlier signal data for indoor localization with IoT. Ock *et al.* [3] identified from an electronic shelf labeling system that, almost always, a few e-tags (out of a thousand) are disconnected from the gateway even at a close distance due to deep fading holes. To avoid deep fading holes and improve system reliability, they used multiple radios since multiple channels are less likely to experience deep fading at identical locations. Maric [11] constructs frequency hopping patterns for cellular systems to minimize bit errors caused by selective fading. For similar reasons, Bluetooth and IEEE 802.15.4 standards both have built-in frequency hopping techniques to mitigate the impact of a few bad channels or deep fading holes.

For indoor localization, Xu *et al.* [12] used RSSI measured at several random adjacent points to avoid deep fading holes. Zhang *et al.* [2] derived lower bound for time-of-arrival based localization error in the presence of fading. Xie *et al.* [13] localized a moving person through analyzing fading CSI pattern upon moving objects. Huang *et al.* [14] proposed a localization method in which the errors caused by multipath fading are estimated and compensated in a vector space. Bao *et al.* [15] proposed a time-difference-of-arrival based localization scheme that uses a new coding scheme to effectively combat fading. Luo *et al.* [16] augmented Bluetooth low energy (BLE) beacons to a Wi-Fi-based localization system at places where it is difficult to distinguish Wi-Fi signals between two different reference points.

Regarding SRD, the work by Yoo *et al.* [1] proposes SWORD which uses one-class classification machine learning technique to detect secure regions. However, their main focus is in devising a classification model that works with positive reference points only, and they did not discuss the deep fading hole problem that inevitably occur in indoor environments. Furthermore, their SRD performance (both TP and TN) is significantly lower than that of *SHARD*.

B. Problem and Motivation

The simplest form of SRD would be to compare the signal strength measured at current location to those in pre-acquired fingerprint database, similar to RSSI fingerprinting-based indoor localization [5]. If it is similar enough (within a certain threshold) to one of the positive reference points (RPs) in the database, then it is regarded as being in the secure region. We will call this the *'baseline SRD'*. The key difference with localization is the existence and scope of negative RPs during the learning/setup phase which could be none for SRD.

To demonstrate the existence of deep fading holes, we conducted a preliminary experiment using the baseline SRD scheme (on the same setup as Section IV) and looked for false negative (FN) points, the testing points that are considered to be outside the secure region although they are within the secure region. We use one Wi-Fi AP and set the secure region

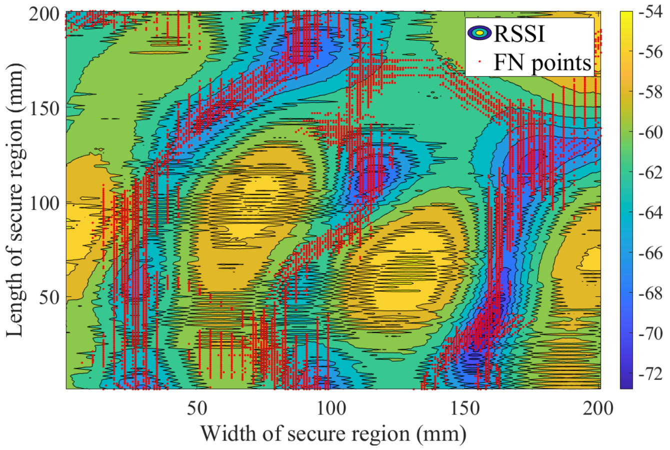


Fig. 2. RSSI contour map within the secure region (real measurements), together with false negative (FN) points (red dots) from the baseline fingerprinting method.

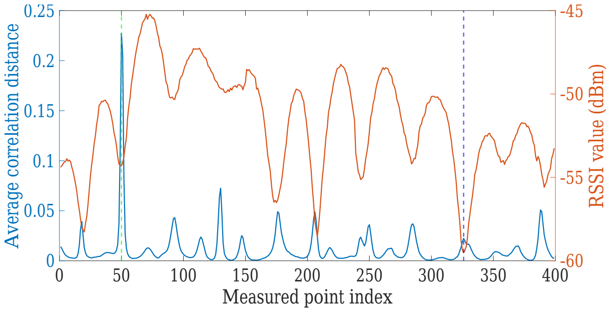


Fig. 3. Deep fading hole problem (real measurements); RSSI and correlation distance change abruptly and non-continuously even within a very short physical distance. It is observed much more often than expected.

to be a two-dimensional 200 × 200 mm^2 area on a table (as shown in Fig. 1). All points at intervals of two millimeters are set as reference points, and separate data obtained at intervals of one millimeter are used as test input data.

Fig. 2 plots the color map of RSSI measurements together with the FN points (red dots) determined by the baseline SRD. The first thing to note from the figure is that, despite being within a small area (200 × 200 mm^2), RSSI values vary significantly and non-monotonically. An area with RSSI of -72 dBm is only a couple centimeters apart from an area with -54 dBm, and there are areas with low RSSI in between two areas with high RSSI. This result cannot be explained using analytical fading models for wireless signals, and these are the areas where deep fading holes are occurring due to multipath. Secondly, the result shows that FNs (red dots) usually occur when RSSI is low. These observations explain why most RSSI/CSI fingerprinting-based localization or SRD schemes perform miserably in many indoor environments.

To analyze this phenomenon deeper, we conducted another experiment. We collected CSI values at 400 points; all points are on a straight line, and adjacent points are two millimeters apart. Afterwards, we calculated the average *correlation distance* (defined in Section III-A) of CSI amplitude at each point with adjacent two points.

Fig. 3 plots the measured average RSSI and *correlation distance* of CSI amplitudes at all 400 points. At a certain point, RSSI is significantly and abruptly smaller than its surroundings, and the CSI amplitude tends to also change significantly at those points (e.g., as shown with dotted lines in Fig. 3). These are the points which we define as *deep fading holes*, and we are showing through experiments that these holes occur much more frequently than one may imagine. Thus, without handling these deep fading holes, no fingerprinting based schemes would be able to perform at satisfactory level. To this end, the goal of this work is to design a mechanism to detect deep fading holes and mask the problem, which we discuss next.

III. Design of *SHARD*

SHARD is a hole avoidance scheme for solving the SRD problem which includes the hole elimination, phase-distance estimation, and 1-2-4 matching techniques.

A. Baseline Fingerprinting-based SRD

Before discussing the design of *SHARD*, we first describe the baseline fingerprinting concept that are commonly used in many SRD and indoor localization works. The baseline fingerprinting technique measures some form of *similarity* between the CSI amplitude at a reference point and that at a test input point. Then, it determines whether the input is near the corresponding reference point using the similarity and a certain threshold. The key design factors are what 'similarity' measure to use and how to determine or derive the thresholds.

We measure the similarity between CSI amplitudes using a metric called *correlation distance* which is widely used for clustering in various applications [17]–[21]. Correlation distance is defined as $1 - r_{xy}$, where r_{xy} is the Pearson's correlation coefficient given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad (1)$$

where n is the length of the sequence, and $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i, respectively.

When baseline fingerprinting is applied to the SRD problem, it is infeasible to optimize the threshold to satisfy a desired FP since negative (non-secure) region is unbounded[2]. Therefore, the only option is to set a threshold that satisfies only a desired TP using the pre-acquired training/learning data. A practical approach for deriving appropriate thresholds is discussed in Section III-F.

[2] The main difference between SRD and finger-printing based localization is that the latter assumes the use of pre-acquired data in all areas where the device may potentially exist unless completely out-of-range. This assumption is inappropriate for SRD because the device can be located where no data collection has been made for. Since it is costly and sometimes impossible to obtain all data from practically unbounded non-secure region, SRD should work with data obtained only from the predefined secure region without negative reference points.

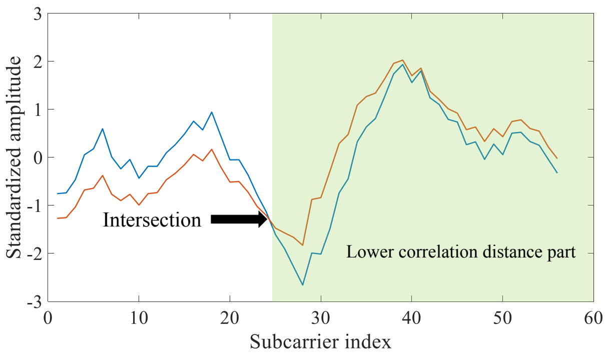


Fig. 4. CSI amplitude (for 56 subcarriers) at two points that are two millimeters apart. Although the shapes of the sequences look similar, their correlation distance may diverge significantly.

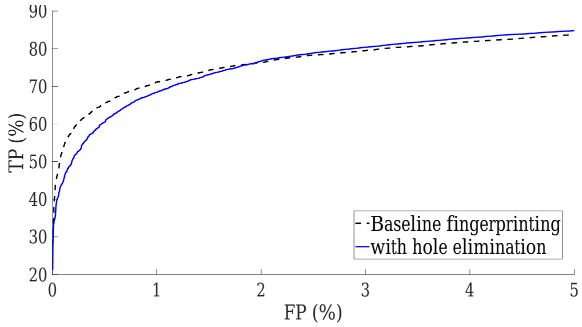


Fig. 5. ROC curves for baseline SRD and the SRD with hole elimination. Hole elimination by itself has trade-off; Despite eliminating misleading (hole) information, we need additional mechanism to compensate for the loss of information.

B. Hole Elimination

Fingerprinting techniques work under the assumption that the CSIs obtained from two physically nearby locations have high similarity [22]. However, our real-world measurements in Fig. 3 have shown that the similarity can be low at deep fading holes even if the physical distance is small. To devise a method to detect and avoid deep fading holes, we analyze the CSI amplitudes at the holes in detail.

Fig. 4 plots the CSI amplitudes (for 56 subcarriers[3]) at two sample points two millimeters apart. The shapes of the two data lines look somewhat similar, but the correlation distance between the two lines is approximately 0.17 which is much larger than common 2 mm-spacing cases where the average distance of 20,000 cases is 0.0078. We identified that the correlation distance is larger at a hole even when most parts of the amplitude distribution are similar. In other words, if the correlation distance between two points is large, one of the two points is likely to be, and can be regarded as being, at a hole.

Therefore, we propose *hole elimination* which eliminates parts of CSI amplitude data with high correlation distance, and uses only the remaining parts for fingerprinting. When the CSI amplitudes at the reference points are given, the correlation distances for all pairs of adjacent points are calculated. Then, if the average is bigger than a certain threshold, the point is suspected to be a deep fading hole. The threshold is calculated as a value which causes half of the reference points to be considered as potential holes that must be examined further (e.g., 0.02 from our dataset). For example in Fig. 4, the correlation distance between the sequences for the last 30 subcarriers (27 to 56) is 0.0243, which is significantly better(lower) than the 0.17 of the whole sequence.

However, if we calculate the correlation distance between two too short sequences, the distance may look small regardless of the actual physical distance. In other words, reducing the length of sequence decreases the uniqueness of each reference point and increases the probability of false positives (FP) in fingerprint matching; i.e., there could be multiple points, positive or negative, with similar subsequences. To prevent such a problem, we set the minimum length of a sequence as 75% of the subcarriers (i.e., 42) and calculate the distance only for sequences whose length is not less than the minimum. Finally, when testing an input point, *hole elimination* is applied to the input point if and only if the target reference point has been determined as a potential hole.

Fig. 5 plots the receiver operating characteristic (ROC) curves of the baseline SRD and the SRD to which *hole elimination* is applied. It can be seen that, as we adjust the threshold based on the desired TP, *hole elimination* improves FP performance in the higher TP region while not in the lower TP region. When the detection threshold (explained later) is set tightly (i.e., conservative judgement on secure region), since a lot of positive data not at the hole has already been well-judged as TP, the increase in FP caused by information loss is greater than the increase in TP due to the effect of *hole elimination*. Conversely, when the threshold is more relaxed, similar negative data are already judged to be positive due to information loss, so the effect of *hole elimination* continuously leads to an increase in TP, and the performance is reversed. Thus, we need a way to compensate for loss of information from *hole elimination*, especially for lower FP cases which is regarded as more important metric than higher TP for security purposes.

C. Phase-distance Estimation

When RF signal is transmitted, the phase change that occurs as the signal experiences the channel should be constant theoretically. However, CSI phase information has been difficult to utilize in fingerprinting techniques so far because it varies significantly over time due to several practical reasons such as the asynchronous PLL clock at transceivers [8], [9]. One of the goals of this work is to devise a technique to utilize this time-varying CSI phase information for more reliable and accurate fingerprint matching. If this is possible, we can not only compensate for the loss of information from hole elimination, but also provide additional information for improved fingerprinting that can be used in a variety of localization systems.

Since several papers [8], [9] have shown that the form of the change is linear, we define *phase-distance* to represent

[3] The IEEE 802.11 standard defines 64 subcarriers with 20MHz bandwidth: 48 for data, 4 pilot, and 12 virtual. However, commercial Atheros chip provides measurement from only 56 subcarriers. Intel allows 30.

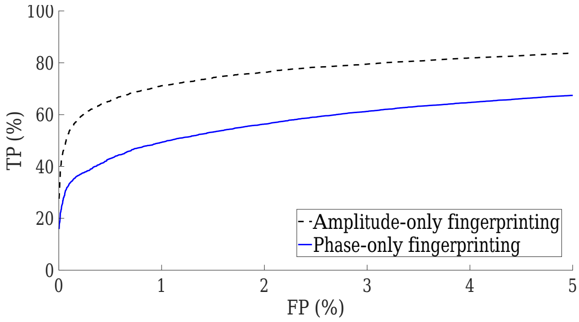


Fig. 6. ROC curves for baseline SRD using ‘amplitude-only’ and SRD using ‘phase-only’. ‘phase-only’ delivers insufficient performance on its own.

the similarity in phase similar to the correlation distance for amplitude. The measured phase $\phi_{k,t}$ of subcarrier k at time t is given by

$$\phi_{k,t} = \theta_k + \lambda_t k + \beta_t, \tag{2}$$

where θ_k is the ‘phase change’ of subcarrier k caused by the channel, and λ_t and β_t are the subcarrier-dependent and -independent phase errors at time t, respectively.

To define a new similarity measure between two phases, we examine the difference between two measured phases which is given by

$$\phi_{k,t} - \phi'_{k,\tau} = \theta_k - \theta'_k + (\lambda_t - \lambda'_\tau)k + \beta_t - \beta_\tau. \tag{3}$$

Applying linear regression to (3), the best fit of $\phi_{k,t} - \phi'_{k,\tau}$ can be written as

$$\phi_{k,t} - \phi'_{k,\tau} \sim \Lambda k + B. \tag{4}$$

The ‘phase change’ may work similarly to the amplitude. In other words, if the two points where the phases are measured are close, $\theta_k - \theta'_k$ are expected to be small, and therefore $\phi_{k,t}$ and $\phi'_{k,\tau} + \Lambda k + B$ should be similar like the amplitude. Now, we define *phase-distance* as the correlation distance between $\phi_{k,t}$ and $\phi'_{k,\tau} + \Lambda k + B$. We define a similarity measure that can be used for phase so that fingerprinting technique can be applied on CSI phase.

However, even if the actual shapes of the two phases are different, the compensation via linear regression has the effect of decreasing uniqueness, thus a system using only phase has lower performance than using amplitude. Fig. 6 plots the performance of baseline SRD using either only-amplitude or only-phase. When FP is 5%, the TP difference between the two techniques is about 16.29%, indicating a large performance gap. In other words, it can be seen that using only the phase results in much worse performance compared to using amplitude only.

Therefore, we use phase information together with, and in addition to, amplitude. If the correlation distance between input amplitude and reference amplitude is sufficiently small `OR` *phase-distance* between input phase and reference phase is sufficiently small, then the input data point is determined to be within the secure region. Also, if the correlation distance between input amplitude and reference amplitude is moderately small `AND` *phase-distance* between input phase and reference phase is moderately small, then the input data point is determined to be within the secure region. Thus, we have a total of four thresholds, two for amplitude and two for phase, where the two for each type represent ‘sufficiently small’ and ‘moderately small’. Said differently, two thresholds for an `OR` operation and two thresholds for an `AND` operation exist, and we define them as $Th^{\mathbf{OR}}_{\text{amp}}$, $Th^{\mathbf{OR}}_{\text{pha}}$, $Th^{\mathbf{AND}}_{\text{amp}}$, and $Th^{\mathbf{AND}}_{\text{pha}}$.

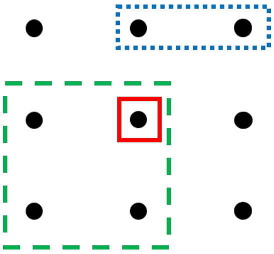


Fig. 7. The 1-2-4 matching technique considers neighboring reference points for similarity matching in order to compensate for loss of information when subcarrier elimination is necessary at a hole. Depending on the number of eliminated subcarriers, either 1 or 2 or 4 RPs are utilized.

D. 1-2-4 Matching

SHARD’s *1-2-4 matching* technique works in conjunction with the *hole elimination*. It not only compensates for information loss but also uses information interactively for better performance.

If an input point is being compared to a reference point not at a hole, then only one reference point is considered for matching in the same way as the baseline SRD scheme. However, when the input point is being compared to a reference point that may be at a hole, then *hole elimination* is applied and a subset of the subcarriers are compared for correlation distance. If over 75% of the subcarriers in the amplitude sequence is similar (42 or more subcarriers), we regard the input to resemble the two adjacent reference points as the blue dotted line in Fig. 7. Furthermore, if half (50%) or more are similar, adjacent square-shaped four reference points are compared as the green dashed line in the figure. By utilizing the neighboring reference points, we can reduce the ambiguity due to loss of information caused by deep fading holes.

E. Putting It All Together – SHARD

Finally, *SHARD* combines hole elimination, phase-distance estimation, and 1-2-4 matching techniques to enable accurate SRD with hole avoidance. First, *SHARD* filters the input test data using the correlation distance threshold which achieves 99% TP on pre-acquired learning data. Empirically, the threshold value is ~0.05, defined as $Th^{\mathbf{filter}}_{\text{amp}}$. The reason for not filtering at 100% is because the relaxed threshold also increases FP, and about ~1% of the whole region is the trade-off point for *SHARD* to rescue and correctly judge the positive data in deep holes.

Then, *SHARD* attempts to determine whether the input test data is from the secure region using CSI amplitude only. If unsure, it uses phase-distance as a second criterion, which can be seen as an `OR` operation. Then, it checks whether both amplitude and phase satisfy the thresholds for an `AND` operation. If not determined to be within the secure region, *SHARD*

Algorithm 1 *SHARD*

```
Input: $(CSI_{\mathrm{amp}}, CSI_{\mathrm{pha}})$
$\mathbb{S}$: a coordinate set of all reference points
$f_c$: correlation distance between two amplitudes
$f_c^i$: correlation distance between two amplitudes with minimum length $i$ hole elimination
$f_p$: phase-distance between two phases
for coordinate $(i, j)$ in $\mathbb{S}$ do
    if $f_c(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{filter}}$ then
        if $f_c(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ or
            $f_p(\mathrm{CSI}_{\mathrm{pha}}, \mathrm{CSI}_{\mathrm{pha}}^{(i,j)}) < Th_{\mathrm{pha}}^{\mathbf{OR}}$ then
            $IN \leftarrow 1$
            break
        else if $f_c(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{AND}}$ and
            $f_p(\mathrm{CSI}_{\mathrm{pha}}, \mathrm{CSI}_{\mathrm{pha}}^{(i,j)}) < Th_{\mathrm{pha}}^{\mathbf{AND}}$ then
            $IN \leftarrow 1$
            break
        else if coordinate $(i, j)$ is in the hole and
            $f_c(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{AND}}$ then
            if $f_c^{42}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ and
                $(f_c^{42}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i-1,j)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ or
                $f_c^{42}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j-1)}) < Th_{\mathrm{amp}}^{\mathbf{OR}})$ then
                $IN \leftarrow 1$
                break
            else if $f_c^{28}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ and
                $f_c^{28}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i-1,j)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ and
                $f_c^{28}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i,j-1)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ and
                $f_c^{28}(\mathrm{CSI}_{\mathrm{amp}}, \mathrm{CSI}_{\mathrm{amp}}^{(i-1,j-1)}) < Th_{\mathrm{amp}}^{\mathbf{OR}}$ then
                $IN \leftarrow 1$
                break
            end if
        end if
    end if
end for
if $IN == 1$ then
    CSI input is from secure region.
else
    CSI input not is from secure region.
end if
```

executes hole elimination and 1-2-4 matching to rescue the data points that fall in holes within the secure region. Finally, if the test point passes none of these criteria, then it is regarded as a location outside the secure region. Algorithm 1 concisely summarizes the decision process of *SHARD* as a combination of baseline SRD, phase-distance, hole elimination, and 1-2-4 matching. In addition when multiple Wi-Fi APs are used, *SHARD* follows the principle of *majority rule* to improve accuracy even further under such scenarios.

Since *SHARD* compares the input point and each reference point, the maximum complexity is proportional to the number of reference points in the secure region. Negative points experience the maximum complexity, whereas positive points can suffer from 0 to maximum complexity depending on its location in the secure region.

F. Threshold Selection

Negative point training data cannot be used to optimize the thresholds according to the basic concept of SRD problem in which there is no data from unbounded exterior of the secure region. This means, thresholds must be determined only from TP of positive training data without FP results. Nevertheless, to get a sense of FP performance according to threshold configurations, we examined ∼100 million uncorrelated CSI pairs. We calculate the correlation distance of all amplitude pairs, and then plot the empirical cumulative distribution function (eCDF) in Fig. 8. It shows that the distances are roughly uniformly distributed between 0 and 2. This means that small exports of thresholds reduce FP linearly by the same ratio; i.e., in order to make FP smaller, a smaller threshold should be set. Nevertheless, since we do not have negative data for FP calculation during training process, we set the thresholds according to desired TP.

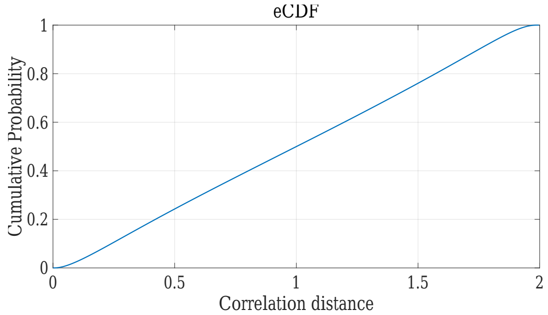


Fig. 8. eCDF of CSI amplitude correlation distance for 100 million data point pairs.

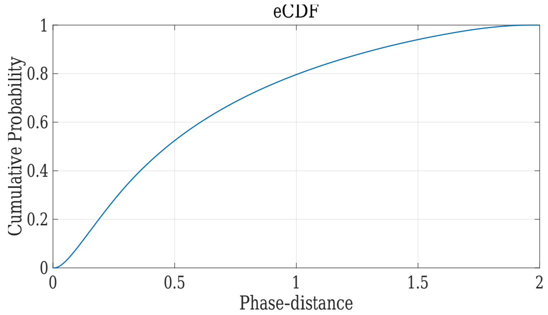


Fig. 9. eCDF of CSI phase-distance for 100 million data point pairs.

First, we set the desired TP to 80% when using a single AP, and set $Th_{\mathrm{amp}}^{\mathbf{AND}}$ to a value that achieves about 55% by baseline SRD because the result of `AND` operation on two schemes achieving 55% is around 80% $(1-(1-0.55)^2 \approx 0.8)$. $Th_{\mathrm{amp}}^{\mathbf{AND}}$ is also used as the threshold in Subsection III-D because satisfying the 2–4 condition can be seen as similar to the `AND` operation. Similarly, we set $Th_{\mathrm{amp}}^{\mathbf{OR}}$ to a value that achieves about 90% by baseline SRD because `OR` operation of two 90% schemes is around 80% $(0.9^2 \approx 0.8)$.

In the case of phase, we also examine the eCDF of phase-distances as plotted in Fig. 9. It shows that when the threshold is low, even a small increment in the threshold increases FP more than in the case of amplitude. Therefore, we set $Th_{\mathrm{pha}}^{\mathbf{AND}}$ and $Th_{\mathrm{pha}}^{\mathbf{OR}}$ so that the cumulative probability of them matches the cumulative probability of $Th_{\mathrm{amp}}^{\mathbf{AND}}$ and $Th_{\mathrm{amp}}^{\mathbf{OR}}$. When the number of APs increases, the threshold values of the amplitude increase with the same ratio as the increase in the number of APs, and in the case of phase, they are set so that the same cumulative probability is maintained.

TABLE I
EXPERIMENT SETTINGS.

Measurement settings	Value
Wireless chipset	Atheros 9380
Number of $\{$AP, station$\}$	$\{3, 1\}$
Wi-Fi PHY	802.11n
Measurement time per each test point	1–2 s
Office size	$6.5 \times 3.0\,\text{m}^2$
Size of each region	$800 \times 800\,\text{mm}^2$
Distance between RPs	4 mm
Distance between positive points for testing	1 mm
Distance between negative points for testing	8 mm
Number of RPs	51×51
Number of positive points for testing	201×201
Number of negative points for testing	$101 \times 101 \times 4$

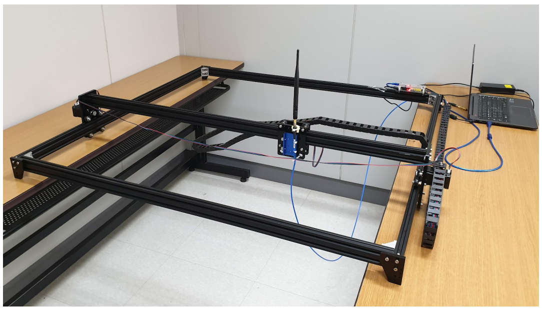


Fig. 10. Equipment for millimeter-accurate fine-grained controlled Wi-Fi CSI measurements.

IV. Evaluation

This section describes the experiment setup and the evaluation of *SHARD*'s SRD performance. Recall that in a SRD application, rejecting devices outside the secure region (TN) is as important as admitting devices within the secure region (TP). Accidentally authorizing devices in non-secure region (FP) can be a critical security vulnerability.

A. Experiment Setup

We use four laptops with Atheros 9380[4] Wi-Fi NICs operating as three APs and a station, and use the Atheros CSI tool[5] [23] for CSI collection. The experiment environment is a regular university office room. We first set a secure region and four negative regions where the size of each region is $800{\times}800$ mm, and install three APs at fixed locations as shown in Fig. 1.

We measure 200 CSI samples at each reference point which takes approximately 1–2 s, and calculate the average of the first 100 samples for reference data and use the following 100 as testing input. We obtain CSIs at 4 mm intervals for reference points (RPs), at 1 mm intervals for positive testing points, and at 8 mm intervals for negative testing points, which makes the number of positive and negative points approximately equal. This results in using 51×51 reference points for the secure region, 201×201 positive testing points within secure region, and $101 \times 101 \times 4$ negative testing points outside the secure region. The movement of the antenna in millimeter granularity is precisely controlled using an equipment called 'laser engraver' shown in Fig. 10. TABLE I shows the details of the setup.

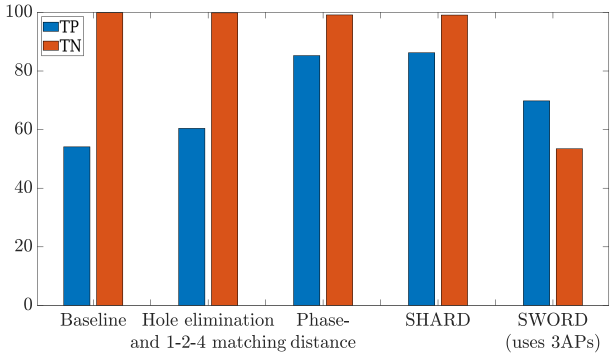


Fig. 11. SRD performance of *SHARD* and its components on a single AP, compared to the baseline finger-printing based SRD and SWORD (with 3 APs).

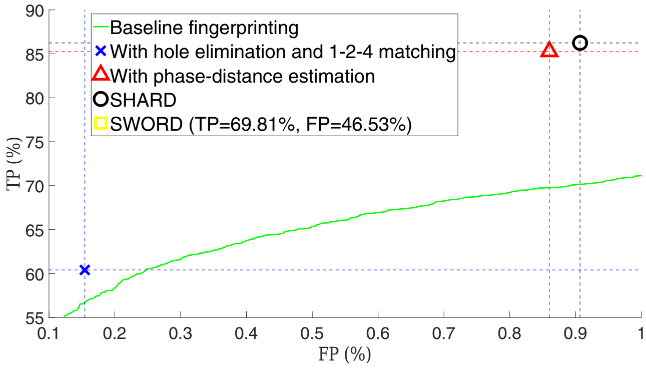


Fig. 12. SRD performance of *SHARD* and its components on a single AP, together with that of SOWRD and the ROC curve of baseline SRD.

For comparison, we also implemented SWORD which is a state-of-the-art SRD scheme. We use the same settings and parameters as in [1]. Therefore, SWORD always uses 3 APs while *SHARD* works with 1 AP.

B. SHARD Performance with Single AP

We first compare the performance of (1) baseline SRD, (2) SRD with hole elimination and 1-2-4 matching, (3) SRD with phase-distance estimation, (4) *SHARD* with all three techniques combined, and (5) SWORD [1]. Fig. 11 plots the TP and FP of each scheme on a single AP scenario except for SWORD which requires three APs. It shows that *SHARD* significantly improves the TP performance compared to the baseline SRD while maintaining similar FP ($1{-}$TN) performance. The figure also shows that the three techniques of *SHARD* all contribute to the improvement[6].

Furthermore, *SHARD* significantly outperforms SWORD despite using only a single AP while SWORD uses three. In fact, the performance of SWORD was lower than what was presented [1]. This is because for SWORD, the size and parameters of the neural network must be adjusted according

[4] Supports IEEE 802.11a/b/g/n, 3-stream 11n MIMO, with PCIe interface.

[5] Available at https://wands.sg/research/wifi/AtherosCSI/

[6] Hole elimination, one of our key ideas, must go in conjunction with 1-2-4 matching to cope with information loss for performance improvement.

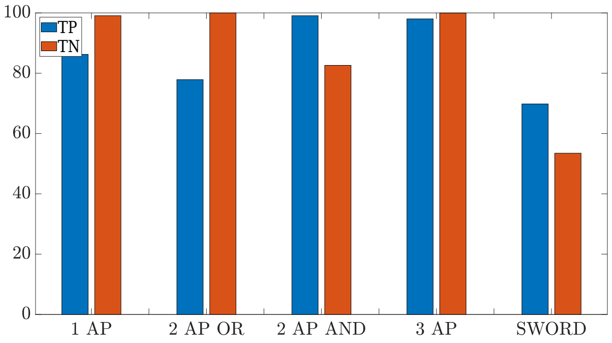


Fig. 13. *SHARD* performance on multiple APs.

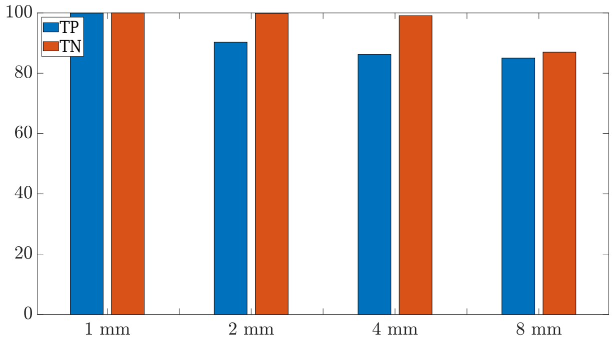


Fig. 14. *SHARD* performance with different RP spacing intervals.

to the amount of data and the size of secure region. That is, SWORD requires fine-tuning per environment where as *SHARD* learns from data.

To examine the trade-off between FP and TP improvement, Fig. 12 plots the ROC curve. It clearly shows that *SHARD* outperforms baseline SRD with a 16.11% improvement in TP (86.24%), and its three techniques are essential components for the improvement. For example, for the same FP, TP with phase-distance is 85.27% compared to 69.76% of the baseline SRD. The data point for SWORD is not visible within the range of the figure because of its poor FP performance, which is insufficient to be used practically.

C. Impact of the Number of APs

In general, using additional Wi-Fi APs will improve fingerprinting performance since it provides more information for higher probability of data uniqueness. It mitigates the deep fading hole problem as well since two antennas physically separated by a distance greater than half-wavelength are uncorrelated [24], and at Wi-Fi's 2.4 GHz band, half-wavelength is about 12 cm which is far smaller than distances between two APs in typical deployments.

To investigate the impact of the number of APs, Fig. 13 plots the TP and TN performance of *SHARD* with varying number of APs. The three APs case outperforms the one AP case for both TP and TN. Specifically, *SHARD* with 3 APs achieves a near-perfect TN of 99.96% with an excellent TP of 98.01% which are sufficient to be used for practical purposes.

However, in the cases of two APs, the tendency is different. The majority rule cannot be applied if two opposite decisions tie, and thus *SHARD* must choose a policy between `OR` operation or `AND` operation. Accordingly, in the `AND` case, a large performance improvement in TN but a large performance reduction in TP is inevitable, and the opposite phenomenon occurs for the `OR` case. Therefore, we recommend using odd number of APs, preferably three than one.

Finally, as shown in Fig. 11, *SHARD* performs significantly better than SWORD under the fair condition of using 3 APs.

D. Impact of RP Interval

We use dense RPs (e.g., 2 mm spacing) within a secure region. The purpose is to obtain sufficient data for accurate classification *without any data from outside the secure region*, which is a key requirement of the SRD problem. Several papers point out that obtaining data for fingerprinting is labor-intensive [25], [26]. Although our RP density within the secure region may seem high, we claim that it is several orders of magnitude more labor intensive to collect data from non-secure region which is practically unbounded.

In order to examine the performance change according to the distance interval between RPs, we run experiments with varying RP intervals. Fig. 14 shows that both TP and TN decrease as the interval increases, which is an intuitive result because the amount of information decreases.

In the 1 mm spacing scenario, *SHARD* achieves TP of 99.95% and TN of 99.995%, which means it can be used practically with one AP at the expense of data collection effort. Conversely, TP of 85.02% and TN of 86.99% are achieved with 8mm spacing, and the performance decreases further as the interval increases. In this situation, pragmatic performance can be achieved using multiple APs, as discussed in the previous subsection.

E. Boundary Result

One of the key features of SRD problem is sharply dividing off secure region from non-secure region for precise classification. Then a natural suspicion would be "what happens at the boundary of a secure region?" Even if we disregard the deep fading hole problem and assume ideal single-path fading model, the boundaries of a secure region would be a challenge for SRD since the signal immediately outside the secure region can be similar to the signal immediately inside.

To examine how *SHARD* distinguishes the secure region from its periphery, we plots both FN and FP points for a 300×300 mm^2 area where only 200×200 mm^2 is the secure region. Fig. 15 plots this result together with the RSSI contour color map. Note that FN points (red dots) are points that are in the secure region but are determined as being outside, and FP points (blue dots) are points that are outside the secure region but are determined as being inside. The figure clearly shows that *SHARD* well separates the secure region sharply with only 5–9 mm of slack margin (see blue dots concentrated around 200 mm for both width and length of secure region).

In addition, it shows hole avoidance removes the effect of deep fading hole since there are only a small number of FN points. The effect of hole avoidance is discussed more in the next section.

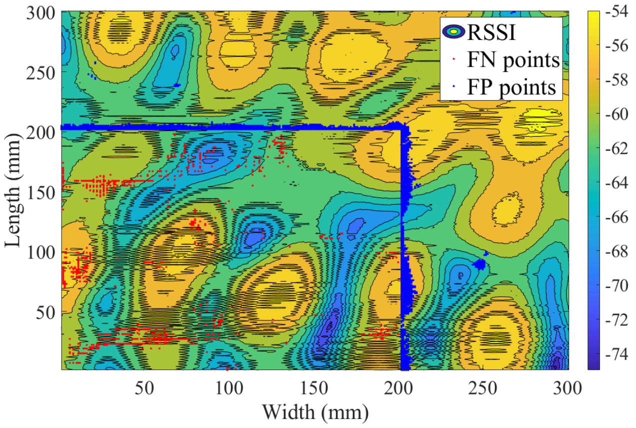


Fig. 15. RSSI contour within and near the secure region (real measurements), together with FN points (red dots) and FP points (blue dots) from *SHARD* with 3 APs.

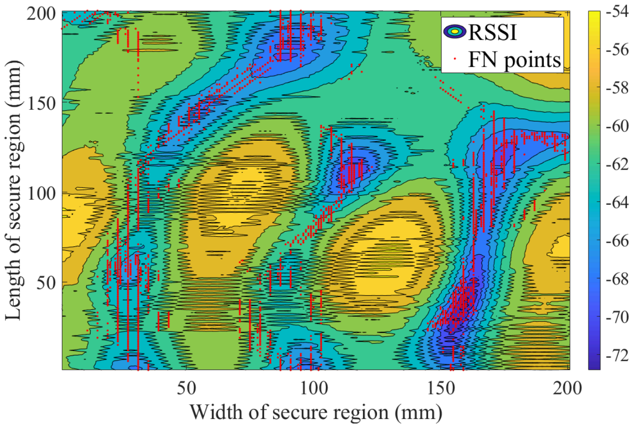


Fig. 16. RSSI contour within the secure region (real measurements), together with false negative (FN) points (red dots) from *SHARD*. There is clear improvement from Fig. 2.

F. Effect of Hole Avoidance

FP result cannot be, and should not be, taken into account when configuring the thresholds for SRD; SRD problem by definition must work without negative RPs. However for evaluation purposes only, we adjust the thresholds through trial-and-error so that FP matches 5% in order to examine how hole avoidance alleviates the effect of deep fading holes.

Fig. 16 plots the FN points from *SHARD* along with the color map of RSSI measurements. Compared to that of baseline SRD previosly shown in Fig. 2, Fig. 16 clearly shows significantly fewer FN points, which is an achievement made by our *hole avoidance* technique.

V. Conclusion

Deep fading hole is one of the critical challenges that many indoor wireless systems face, especially in fingerprinting-based localization. This work proposed *SHARD*, a Wi-Fi CSI based secure region detection scheme with hole avoidance, which addresses the *deep fading hole* problem. We have devised techniques to identify abrupt changes in CSI amplitude, eliminate unreliable data, and utilize time-varying CSI phase for improved detection accuracy. Through our real-world experiments, we have shown that our hole avoidance techniques successfully address the problem and *SHARD* can correctly reject devices outside the secure region with 99.96% accuracy while detecting devices inside the secure region with 98.01%. We believe that this work will advance not only *secure region detection* schemes for security of IoT devices, but can also be generalized to many fingerprinting-based indoor localization schemes.

References

[1] Y. Yoo, J. Suh, J. Paek, and S. Bahk, "Secure region detection using Wi-Fi CSI and one-class classification," *IEEE Access*, vol. 9, pp. 65 906–65 913, 2021.

[2] X. Zhang, C. Tepedelenlioğlu, M. Banavar, and A. Spanias, "CRLB for the localization error in the presence of fading," in *Proc. IEEE ICASSP*, 2013.

[3] J. Ock, H. Kim, H.-S. Kim, J. Paek, and S. Bahk, "Low-power wireless with denseness: The case of an electronic shelf labeling system—design and experience," *IEEE Access*, vol. 7, pp. 163 887–163 897, 2019.

[4] T. Watteyne, S. Lanzisera, A. Mehta, and K. S. Pister, "Mitigating multipath fading through channel hopping in wireless sensor networks," in *Proc. IEEE ICC*, 2010.

[5] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, vol. 2, 2000.

[6] G. Pecoraro, S. Di Domenico, E. Cianca, and M. De Sanctis, "LTE signal fingerprinting localization based on CSI," in *Proc. IEEE WiMob*, 2017.

[7] Q. Song, S. Guo, X. Liu, and Y. Yang, "CSI amplitude fingerprinting-based NB-IoT indoor localization," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1494–1504, 2018.

[8] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1342–1355, 2019.

[9] H. Xue, J. Yu, F. Lyu, and M. Li, "Push the limit of multipath profiling using commodity WiFi devices With limited bandwidth," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4142–4154, 2020.

[10] M. A. Bhatti *et al.*, "Outlier detection in indoor localization and Internet of things (IoT) using machine learning," *J. Commun. Netw.*, vol. 22, no. 3, pp. 236–243, 2020.

[11] S. Maric, "Construction of optimal frequency hopping sequences for minimizing bit errors in selective fading channels characteristic to digital cellular systems," in *Proc. IEEE VTC*, 1993.

[12] C. Xu, B. Firner, Y. Zhang, and R. E. Howard, "The case for efficient and robust RF-based device-free localization," *IEEE Trans. Mobile Comput.*, vol. 15, no. 9, pp. 2362–2375, 2016.

[13] H. Xie *et al.*, "Accelerating crowdsourcing based indoor localization using CSI," in *Proc. IEEE ICPADS*, 2015.

[14] S. Huang and W.-C. Wong, "COMVELOC: A compensation vector-based indoor localization system in WIFI environments," in *Proc. IEEE TENCON*, 2018.

[15] X. Bao, J. Li, and S. Das, "A novel fading-tolerant high-accuracy localization algorithm using distributed space-time block codes," in *Proc. IEEE GLOBECOM*, 2008.

[16] R. C. Luo and T. Hsiao, "Indoor localization system based on hybrid Wi-Fi/BLE and hierarchical topological fingerprinting approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10 791–10 806, 2019.

[17] F. R. Almeida, A. Brayner, J. J. Rodrigues, and J. E. B. Maia, "Improving multidimensional wireless sensor network lifetime using pearson correlation and fractal clustering," *Sensors*, vol. 17, no. 6, p. 1317, 2017.

[18] J. Y. Song, W. Chang, and J. W. Song, "Cluster analysis on the structure of the cryptocurrency market via bitcoin–ethereum filtering," *Physica A: Statistical Mechanics its Applicat.*, vol. 527, p. 121339, 2019.

[19] I. Priness, O. Maimon, and I. Ben-Gal, "Evaluation of gene-expression clustering via mutual information distance measure," *BMC bioinformatics*, vol. 8, no. 1, pp. 1–12, 2007.

[20] G. J. Torres, R. B. Basnet, A. H. Sung, S. Mukkamala, and B. M. Ribeiro, "A similarity measure for clustering and its applications," *Int. J. Electr. Comput. Syst. Eng.*, vol. 3, no. 3, pp. 164–170, 2009.

[21] M. R. Berthold and F. Höppner, "On clustering time series using euclidean distance and pearson correlation," *arXiv preprint arXiv:1601.02213*, 2016.

[22] P. Davidson and R. Piché, "A survey of selected indoor positioning methods for smartphones," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1347–1370, 2017.

[23] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proc. ACM MobiCom*, 2015.

[24] A. Goldsmith, *Wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

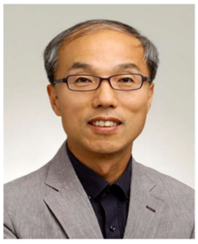
[25] Q. Chen and B. Wang, "FinCCM: Fingerprint crowdsourcing, clustering and matching for indoor subarea localization," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 677–680, 2015.

[26] J. Niu, B. Wang, L. Cheng, and J. J. P. C. Rodrigues, "WicLoc: An indoor localization system based on WiFi fingerprints and crowdsourcing," in *Proc. IEEE ICC*, 2015.

Jihwan Suh is currently a Ph.D. candidate at the School of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea. He received his B.E. degree in the field of Informatics and Mathematical Science from Kyoto University in 2013. His research interests are in the area of security, localization, and 5G with AI.

Yongjae Yoo is currently a Ph.D. student at the School of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea. He received B.S. degree in Electrical and Computer Engineering from Seoul National University in 2019. His research interests are in the area of security and machine learning in wireless networks, Internet of Things and 5G networks.

Jeongyeup Paek received his B.S. degree from Seoul National University in 2003 and his M.S. degree from University of Southern California in 2005, both in Electrical Engineering. He then received his Ph.D. degree in Computer Science from University of Southern California in 2010. He worked at Deutsche Telekom Inc. R&D Labs USA as a research intern in 2010, and then joined Cisco Systems Inc. in 2011 where he was a Technical Leader for the Connected Grid Mesh (CG-Mesh) system in the Internet of Things Group, Connected Energy Networks Business Unit. In 2014, he was with the Hongik University, Department of Computer Information Communication as an assistant professor. Jeongyeup Paek is currently an associate professor at the Department of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea, and also a visiting scholar at University of Southern California, Los Angeles, USA.



Saewoong Bahk received the B.S. and M.S. degrees in electrical engineering from Seoul National University (SNU), in 1984 and 1986, respectively, and the Ph.D. degree from the University of Pennsylvania, in 1991. He was with AT&T Bell Laboratories as a Member of Technical Staff, from 1991 to 1994, where he had worked on network management. From 2009 to 2011, he served as the Director of the Institute of New Media and Communications. He is currently a Professor at SNU. He has been leading many industrial projects on 3G/4G/5G and the IoT connectivity supported by Korean industry. He has published more than 300 technical articles and holds more than 100 patents. He is a member of the National Academy of Engineering of Korea (NAEK). He was a recipient of the KICS Haedong Scholar Award, in 2012. He was President of the Korean Institute of Communications and Information Sciences (KICS). He has been serving as Chief Information Officer (CIO) of SNU. He was General Chair of the IEEE WCNC 2020 , IEEE ICCE 2020 , and IEEE DySPAN 2018. He was Director of the Asia–Pacific Region of the IEEE ComSoc. He is an Editor of the IEEE Network Magazine and IEEE Transactions on Vehicular Technology. He was TPC Chair of the IEEE VTC-Spring 2014, and General Chair of JCCI 2015, Co-Editor-in-Chief of the Journal of Communications and Networks (JCN), and on the Editorial Board of Computer Networks Journal and the IEEE Tran. on Wireless Communications.