Pervasive and Mobile Computing 23 (2015) 89–103

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Fast track article

CSI-MIMO: An efficient Wi-Fi fingerprinting using Channel State Information with MIMO

Yogita Chapre [a,c,*], Aleksandar Ignjatovic [a], Aruna Seneviratne [b,c], Sanjay Jha [a]

[a] School of Computer Science & Engineering, University of New South Wales, Sydney, Australia
[b] School of Electrical Engineering & Telecommunications, University of New South Wales, Australia
[c] Networks Research Group, NICTA, Sydney, Australia

ARTICLE INFO

Article history:
Available online 9 July 2015

Keywords:
Indoor localization
Wi-Fi fingerprinting
Channel State Information
RSSI
MIMO

ABSTRACT

The paper presents a novel Wi-Fi fingerprinting system CSI-MIMO that uses a fine-grained information known as Channel State Information (CSI). The proposed CSI-MIMO exploits the frequency diversity and spatial diversity using Multiple Input Multiple Output (MIMO) system. The proposed CSI-MIMO uses either magnitude or complex CSI based on the mobility in an indoor environment. The experimental performance of CSI-MIMO is compared with Fine-grained Fingerprinting system (FIFS), CSI with Single Input Single Output (SISO) and a simple CSI with MIMO. The experimental result shows a significant improvement with an accuracy of 0.98 m in static environment and 0.31 m in dynamic environment with optimal war-driving over existing systems.

1. Introduction

Localization is playing a vital role in various Location-based Services (LBS) including production, social networking, entertainment, living and public services. Traditionally, LBS is limited to outdoor scenarios and needs further research to implement potential techniques for indoor applications. LBS such as locating an Automated Teller Machine (ATM), tracing a parcel in a post office, location-based mobile adverts and games, and access control demands more accurate indoor localization. The widely used Global Positioning Satellite (GPS) based localization fails indoors due to weak signals, whereas the triangulation technique is less effective due to complex building geometry, mobility of people and multi-path effect. Furthermore, noise and interference due to presence of other radio devices make indoor localization more challenging.

Indoor localization based on wireless signals such as GSM (Global System for Mobile Communications) and Wi-Fi are more popular due to lower cost, ease of access and use of the existing wireless infrastructure. Earlier indoor positioning approaches exploiting infrared, laser and ultrasonic are complex in nature, require higher cost and use bulky devices. Wireless indoor localization uses various signal parameters for location estimation; signal-based Received Signal Strength Indicator (RSSI), time-based Time of Arrival (TOA) and Time Difference of Arrival (TDOA), and angle-based Angle of Arrival (AOA). Among all, RSSI-based Wi-Fi fingerprinting is the most widely used approach. RSSI is a coarse-grained information that represents received signal strength through various paths for a given transmitter–receiver pair. As the distance between the transmitter and receiver increases, RSSI degrades due to attenuation. This phenomenon is utilized in numerous range detection algorithms including deterministic k-Nearest Neighbor in RADAR [1] and probabilistic Bayes rule in Horus [2].

* Corresponding author at: School of Computer Science & Engineering, University of New South Wales, Sydney, Australia.
E-mail addresses: yogitac@cse.unsw.edu.au (Y. Chapre), ignjat@cse.unsw.edu.au (A. Ignjatovic), aruna.seneviratne@nicta.com.au (A. Seneviratne), sanjay@cse.unsw.edu.au (S. Jha).

http://dx.doi.org/10.1016/j.pmcj.2015.07.002
1574-1192/

However, RSSI is highly sensitive to temporal, hardware and environmental factors [3,4], which affects the localization accuracy significantly. To overcome the volatility of RSSI and improve the accuracy of indoor localization, a more robust fine-grained information is needed.

In 802.11 a/b/n networks, the data is modulated using Orthogonal Frequency Division Multiplexing (OFDM). The modulated data is transmitted and received using multiple data carriers known as sub-carriers. Each sub-carrier has a designated frequency and therefore provides frequency diversity across data sub-carriers. Moreover, data is transmitted and received using multiple antennas using Multiple Input and Multiple Output (MIMO) and offers spatial diversity. The transmitted data travels through the channel using multiple paths. At the receiver side, the received data gets attenuated due to scattering, fading and path loss. The channel quality between a transmitter and a receiver is termed as Channel State Information (CSI) and is extracted from PHY layer. The frequency and spatial diversity offered by CSI can be exploited to justify the multi-path effect.

CSI exhibits various prominent features as compared to RSSI offering more unique location signatures. Traditional RSSI presents an aggregated scalar value at the packet level, whereas CSI indicates the channel information at sub-carrier level. In addition, CSI provides complex channel information that includes magnitude and phase information for each sub-carrier, and for each transmit and receive antenna. Multiple sub-carriers attenuate differently when traveled through various traveling paths, thus resulting in variation of amplitude and phase of each sub-carrier. The fine-granularity of CSI in terms of frequency and spatial diversity makes it suitable metric to represent a location uniquely and to enhance the localization accuracy. Recently, the CSI is extractable using the Linux tool using off-the-shelf Intel 5300 Network Interface Card (NIC) [5].

Current wireless technologies use MIMO antenna systems extensively such as IEEE 802.11n, 3GPP LTE, and mobile WiMAX systems. MIMO enhances data throughput under influence of interference, signal fading and multi-path. Moreover, MIMO systems achieves higher data rates using spatial diverse communication links [6].

The prominent features of the CSI is being explored by researchers for Wi-Fi fingerprinting. FILA [7] and PinLoc [8] uses Single Input and Single Output (SISO) aspect of CSI for localization, whereas FIFS exploits spatial diversity of CSI [9]. FIFS aggregates power of all sub-carriers to represent a location and does not include phase profile of CSI. PinLoc creates a location fingerprint as clusters of amplitude and phase of CSI, however it uses only SISO information. We exploit frequency and spatial diversity offered by CSI to define the location fingerprint distinctly. The resulting high dimensional, complex location signature offers challenges of processing MIMO information for each transmitting and receiving antenna stream.

The paper exploit the fine-grained features offered by CSI to derive the location signatures based on the amplitude and phase of each sub-carrier using MIMO. The main contributions of this paper are as follows.

1. We aggregate the CSI over multiple antennas for each sub-carrier and present a novel location signature CSI-MIMO that consists of magnitude and phase difference between subsequent sub-carriers.
2. We propose a complex CSI-MIMO signature for static environments and an amplitude-based CSI-MIMO signature for dynamic environments. To the best of our knowledge, this is the first system proposing variants of CSI based on the mobility in the indoor environment.
3. We create a radio-map using CSI-MIMO for multiple Access Points (AP) to improve the accuracy and reduce the mean distance error.
4. We investigate the impact on the localization accuracy of CSI-MIMO due to the various factors such as size and type of CSI-MIMO fingerprint used, number of APs, and number of training and test samples.
5. We evaluate the performance of CSI-MIMO using the accuracy as a performance metric with other CSI-based localization systems such as FIFS 2×2, a CSI signature with SISO and a CSI signature with MIMO system.
6. We analyze the effectiveness of the proposed location signature using deterministic k-nearest neighbor (kNN) and probabilistic Bayes Rule algorithms.

The rest of the paper is organized as follows. In Section 2 we briefly present the related work on Wi-Fi fingerprinting using various location signatures. Section 3 introduces preliminaries including RSSI, OFDM, Channel Impulse Response and CSI. Section 4 explains our proposed CSI-MIMO system and the methodology in detail. It includes the system architecture, experiment setup and the CSI-MIMO fingerprint generation. In Section 5, we evaluate the effectiveness of the proposed CSI-MIMO signature for static and dynamic environments. We also evaluate our experimental results for various impact factors including number of APs, war-driving samples and type of localization algorithms. Further, we compare our proposed CSI-MIMO system with other CSI-based fingerprinting systems. Conclusions and directions for the future research are presented in Section 6.

2. Related work

Conventionally, Wi-Fi fingerprinting consists of two phases: Training/Off-line phase and Testing/Localization phase. In the training phase, data at various distinct locations called as Reference Points (RPs) is collected over a duration. The collected N data samples are used to derive a location signature, which is then stored in a database called as radio-map along with the coordinates of the location. In the testing phase, data is collected at an unknown location called as a Test Point (TP), a location fingerprint is created in similar way as in the training phase and is matched against the radio-map to infer the location.

The indoor localization exploits various wireless signals: traditional RSSI, Channel Impulse Response (CIR) and the most recent CSI. We categorize various Wi-Fi fingerprinting methods based on the used signal.

2.1. Radio Frequency (RF) signal strength

Localization techniques using radio frequency signal strength such as Radio Frequency Identification (RFID) were proposed during early stages of wireless indoor localization. The spotON [10] used RSSI from active RFID tags and estimated the location using trilateration. LANDMARC [11] localized using k-Nearest Neighbor but required extra infrastructure to cover entire floor-plan.

Bahl et al. presented the first Wi-Fi fingerprinting technique RADAR [1] utilizing the existing wireless infrastructure such as wireless Access Points. RADAR used the RSSI for multiple APs and stored the mean value of the RSSI as a location fingerprint in a radio-map. Using deterministic k-Nearest Neighbor algorithm, RADAR provided an accuracy of 3 m. Youssef et al. adapted the probabilistic approach in Horus [2] by using joint clustering algorithm using common set of APs for localization. Horus showed an accuracy improvement of 2.1 m over RADAR with lower computations. However, the sensitivity of RSSI remained challenging. To deal with the hardware dependence of RSSI, Hossain et al. proposed Signal Strength Difference (SSD) signature, which is derived from RSSI. SSD showed improvement in accuracy even when devices used for training and test phases are different. However, the temporal and environmental sensitivity of SSD remained unanalyzed.

2.2. Channel Impulse Response

Channel Impulse Response as a location fingerprint is proposed as an alternative metric over RSSI. Nerguizian et al. [12] proposed an unique Channel Impulse Response as a location fingerprint in a mining environment. The accuracy was enhanced by using Artificial Neural Network (ANN) but it required costly and bulky measurement devices. Also, the system performance using dynamic bandwidth was not analyzed. Jin et al. proposed location fingerprints as a vector of approximated Channel Impulse Response (ACIR) amplitudes [13]. The kernel function was stored in the radio-map with ACIR vectors that posed the computation overhead. Furthermore, the analysis was entirely based on the simulation rather than realistic test-bed experimentation.

Patwari et al. [14] presented location distinction algorithm using Channel Impulse Response amplitude as a temporal link signature to detect the change in transmitter location. The measurement set used lower bandwidth and shorter path lengths and have not analyze its trade-off on the detection performance. To improve the detection rate and reduce the false alarm rate, Zhang et al. [15] proposed a complex temporal signatures for location distinction.

2.3. Channel State Information

CSI as a fine-grained information from PHY layer has been exploited for the Wi-Fi fingerprinting to overcome the complexity of the localization systems that used Channel Impulse Response based signatures. Sen et al. [8] proposed the PinLoc, the first CSI-based Wi-Fi localization. It used the amplitude and phase of Channel Frequency Response (CFR) clusters as a location signature and demonstrated 89% mean accuracy. However, PinLoc exploited the SISO aspect of CSI. Xiao et al. [9] in FIFS utilized the spatial diversity along with the frequency diversity of CSI for Wi-Fi fingerprinting. FIFS used a location signature, which is the power summation over independent sub-carriers and a probability correlation filter that resulted in improved accuracy over Horus.

The proposed CSI-MIMO exploits both the frequency diversity and spatial diversity of CSI. We generate a complex location signature using amplitude and phase information of all the sub-carriers that represents a location distinctly. We evaluate the performance using deterministic k-Nearest Neighbor and probabilistic Bayes rule for various impact factors such as size of CSI-MIMO signature, training and test samples, and number of APs.

3. Preliminaries

This section briefly discuss various location signatures including RSSI, Signal Strength Difference (SSD), Channel Impulse Response (CIR) and CSI from the literature. We also overview the fundamental OFDM system that forms basis for the CSI signature.

3.1. Received Signal Strength Indicator

The conventional RSSI-based location signature uses N RSSI samples collected at a location for an AP and is represented as

$$RSSI = [s_1\ s_2\ s_3 \ldots s_N]_{1\times N}\,. \tag{1}$$

The dimensions of RSSI-based location signature for K APs in a radio-map is $K \times N$. The researchers have derived various location signatures using RSSI. SSD is proposed to provide the hardware heterogeneity when devices used for training and test phases are not identical. Mobile devices with a variety of NICs offer different quantization bins and sampling rates that results in different values of the RSSI. This RSSI difference contributes to increase in the estimation error, when different devices used for localization. SSD represents the signal strength difference from two APs. For K APs, it is represented as

$$SSD = [s_{11} - s_{21}\ s_{12} - s_{22} \ldots s_{1N} - s_{2N}]_{K-1\times N}\,. \tag{2}$$

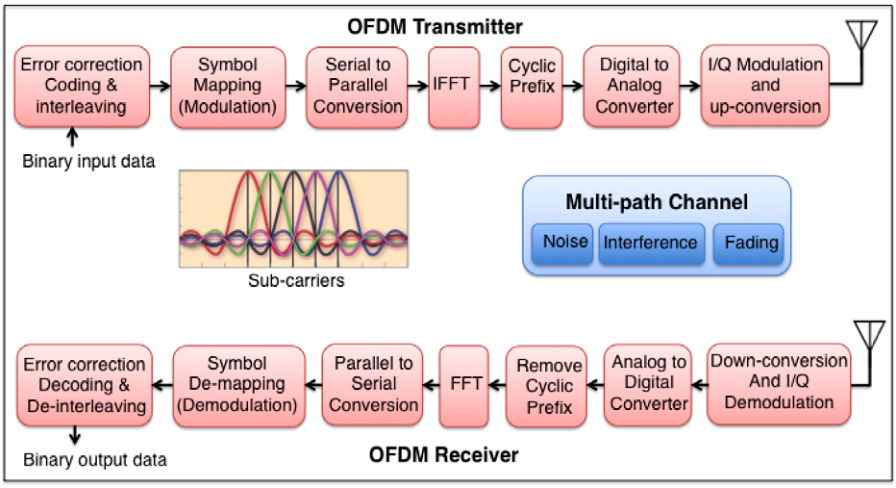

Fig. 1. An OFDM transceiver.

3.2. Orthogonal Frequency Division Multiplexing

A basic block diagram of a OFDM system consists of a transmitter and a receiver as shown in Fig. 1.

The data to be transmitted is modulated and converted into N data sub-carriers. To convert the N sub-carriers from frequency domain to time domain, an Inverse Fast Fourier Transform (IFFT) is performed. Further, digital data is converted into analog using Digital to Analog Converter (DAC). For transmission in the multi-path channel, an in-phase and the quadrature-phase modulation is performed. At the OFDM receiver side, received data is down-converted using in/quadrature demodulation and passed through a Low-Pass Filter (LPF). The filtered signal is converted into digital form using an Analog to Digital Converter (ADC) and then N point Fast Fourier Transform (FFT) is performed. The transmitted data traveled in a multi-path channel degrades due to noise, interference and fading. Each sub-carrier experiences different fading, resulting in unique amplitude and phase of each sub-carrier. The distinctiveness of amplitude and phase information of each sub-carrier is an efficient way to represent a location uniquely.

3.3. Channel Impulse Response

In OFDM system, channel estimation comprises of a vector of N complex elements representing channel in frequency domain with N sub-carriers. For the channel between a transmitter and a receiver, Channel Impulse Response in frequency domain is denoted by $h(\tau)$

$$h(\tau) = \sum_{l=1}^{L} \alpha_l e^{j\phi_l} \delta(\tau - \tau_l) \tag{3}$$

where α_l is the amplitude and ϕ_l is the phase of the lth multi-path component, τ_l is its time delay, L is total number of multi-path and $\delta(\tau)$ is the Dirac delta function.

The time domain channel response $h(\tau)$ is known as Channel Impulse Response and can be derived by taking the Inverse Fast Fourier Transform (IFFT) of the channel frequency response.

$$h(t) = IFFT(h(\tau)). \tag{4}$$

3.4. Channel State Information

In a wireless network, the channel quality of the communication link is known as CSI. It represents the phenomenon of scattering, fading and power decay of the transmitted signal. An IEEE 802.11a/g/n network uses OFDM, where data is transmitted using orthogonal sub-carriers at distinct frequencies. A MIMO system in a narrow-band flat fading channel can be given by

$$y_i = Hx_i + N_i \tag{5}$$

where y_i and x_i are the received and transmitted signal vectors respectively, and H denotes the channel matrix and N_i is the noise vector.

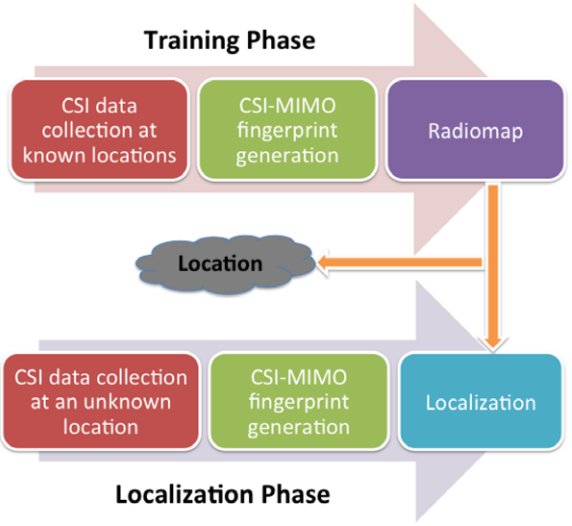

Fig. 2. Block diagram of CSI-MIMO system.

A known training sequence $x_1, x_2, \ldots, x_n$ is transmitted and the channel matrix H is estimated at the receiver side. The received signal y_i is given by

$$Y = [y_1, y_2, \ldots, y_n] = HX + N. \tag{6}$$

Hence, the channel matrix H can be estimated by

$$\hat{H} = \frac{Y}{X} \tag{7}$$

which is the fine-grained PHY layer complex information for N sub-carriers and is represented as

$$H = [H_1, H_2, H_3, \ldots, H_N]^T \tag{8}$$

where

$$H_i = |H_i| \, e^{j\sin\left(\angle H_i\right)}. \tag{9}$$

For MIMO systems with multiple transmit and receive antennas, H_i has dimensions $p \times q$ and can be represented as

$$H_i = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1q} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2q} \\ \cdots & \cdots & \cdots & & \cdots \\ h_{p1} & h_{p2} & h_{p3} & \cdots & h_{pq} \end{bmatrix}. \tag{10}$$

Thus, the dimensions of the channel matrix H is $p \times q \times N$ where h_{pq} is the complex value representing the amplitude and phase information of each sub-carrier.

4. CSI-MIMO system

The architecture of the proposed CSI-MIMO system, the test-bed for the experimentation, the methodology to generate the CSI-MIMO location fingerprint and the localization algorithms is briefly explained.

4.1. CSI-MIMO system architecture

Similar to conventional Wi-Fi fingerprinting, CSI-MIMO comprises of two phases: Training phase/Off-line phase and Testing/Localization phase and is shown in Fig. 2.

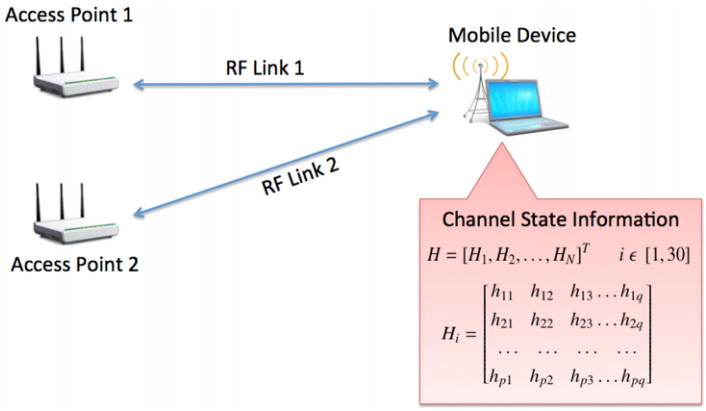

Fig. 3. Experiment setup of CSI-MIMO system.

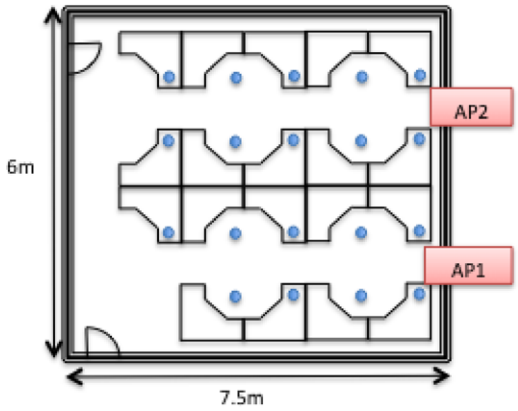

Fig. 4. Floor-plan of the computer lab—static environment.

4.1.1. Training phase

In training phase, a mobile device is placed at a known location referred as Reference Point (RP). The CSI data is collected at various reference points in a given floor-plan for all the deployed APs as shown in Fig. 3.

Samples received from an AP represents the complex CSI over N sub-carriers for the number of transmit and receive antennas of an AP and a mobile device respectively. A unique CSI-based fingerprint is generated by using amplitude and phase information of each sub-carrier, which is presented in further sections of the paper. The fingerprint is finally stored in a radio-map along with the location coordinates of each reference location. We war-drive the entire floor-plan of the test-bed to build the CSI-MIMO fingerprinting database. Finally the training data collected using various training locations is used localization phase for location estimation. We considered the reference points and training points as similar and used interchangeably in the paper.

4.1.2. Localization phase

During localization phase, the CSI data at an unknown location often referred as a Test Point (TP) is collected using a mobile device. A CSI-MIMO test fingerprint is generated for the test point in a similar manner as the training phase. To estimate the location, the generated test fingerprint is matched against the stored CSI-MIMO radio-map. The location estimation can be carried out either by the mobile device or by a third party such as server using various localization algorithms. The paper uses both the deterministic k-Nearest Neighbor and probabilistic Bayes rule algorithms for localization.

4.2. Experiment setup

The experiment uses Intel Linux 802.11n tool [5] that extracts CSI data using off-the-shelf Intel 5300 NIC. Wireless APs (TP-LINK TL-WA901ND) acts as transmitters and are fixed at a location. The mobile device is actually a laptop equipped with Intel 5300 NIC. We have collected data in two distinct environments: a computer lab and postgraduate research lab with floorplan as shown in Figs. 4 and 5 respectively.

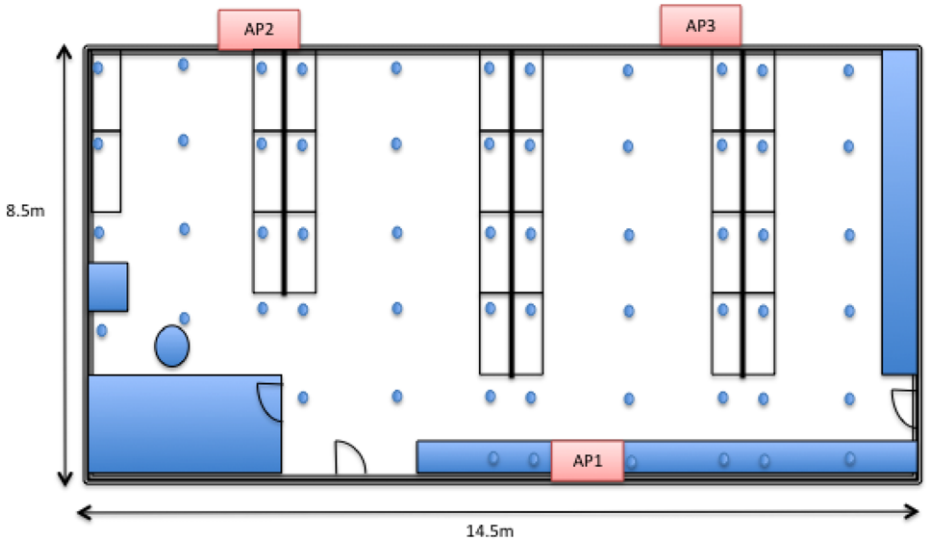

Fig. 5. Floor-plan of the research lab—dynamic environment.

Table 1
Summary of the experimentation test-bed and data.

Parameter	Computer lab	Research lab
Size	6 m × 7.5 m	14.5 m × 8.5 m
Training locations	19	52
Test locations	19	52
Access points	2	3
Distance between RPs	1.5–2 m	1–1.5 m
Maximum training samples	1000	1000
Maximum test samples	100	100
Users present	3	20
Mobility	Static	Dynamic

The two test-beds are different floor areas and have different mobility. Based on the defined training and test locations, we have sampled the data for each of the deployed AP. The summary of the experimental data for each of the test-bed is presented in Table 1.

The two test-bed environments are categorized into static and dynamic based on the mobility, which depends on the number of people present and the frequency of their mobility during experimentation. The test-bed represented by computer lab was almost static with 3 people sitting at their place most of the time during experimentation. In contrast, the research lab was fully occupied with the people at every desk locations with the frequent mobility such as moving in and out of the place and therefore was a typical dynamic environment.

In regards to the Access Point, we have placed sufficient number of APs i.e. two access points in computer lab and three in the research lab and defined number of training and testing locations depending on the size of the floor-plan. The location of the deployed APs is fixed so that received signal at each of the training location receives from all the APs is strong enough. The training locations are equally placed along each of the dimensions of the floor-plan, however the distance between training locations in x dimension is different than in y dimension of the floor-plan and hence the distance between RPs is defined. This non-uniformity of the distance between the training locations is due to the limitations of the test-bed having unequal desk locations.

4.3. CSI-MIMO fingerprint

The CSI data collected at a location is a multidimensional information representing N sub-carriers containing complex values for MIMO. In IEEE 802.11 a/b/n network, OFDM uses 48 data sub-carriers. Intel Linux tool [5] reports CSI for 30 sub-carriers. Eq. (8) represents a CSI sample collected at a location. For an antenna stream with $p = 1$ and $q = 1$, the CSI value for each sub-carrier is represented by Eq. (9). Our proposed location fingerprint CSI-MIMO is basically an amplitude difference vector and a phase difference vector of sizes 1×29 and 1×30 respectively. We utilize the complex CSI of all the 30 sub-carriers to build a unique location fingerprint. The steps to generate fingerprint using the collected data is as follows:

1. For each sample, we aggregated the CSI value of each sub-carrier over MIMO. The resulting aggregated CSI per sub-carrier is reduced to the vector of dimension 1×30.

$$H_{avg} = \sum_{m=1}^{p} \sum_{n=1}^{q} h_{mn}. \tag{11}$$

2. We extract the amplitude and phase values from the aggregated CSI, which is a complex vector. We unwrapped the phase information to ensure the inclusion of multiples of 2π in the phase response.

$$H_{avg\text{-}amp} = \left|H_{avg}\right| \tag{12}$$

$$H_{avg\text{-}ph} = \angle H_{avg}. \tag{13}$$

3. Further, we subtracted amplitude and phase values from subsequent sub-carriers and to reduce the fingerprint size to 1×29 and 1×30 respectively. We retained the phase of first sub-carrier as the initial phase.

$$H_{amp_i} = \left|H_{avg\text{-}amp_i}\right| - \left|H_{avg\text{-}amp_i+1}\right| \tag{14}$$

$$H_{ph_i} = \angle H_{avg\text{-}ph_i} - \angle H_{avg\text{-}ph_i+1}. \tag{15}$$

4. Next, we averaged the amplitude and the phase difference over N samples to get amplitude and phase difference mean as the location signatures.

$$H_{amp\text{-}diff} = \sum_{i=1}^{N} H_{amp_i} \tag{16}$$

$$H_{ph\text{-}diff} = \sum_{i=1}^{N} H_{ph_i}. \tag{17}$$

5. Finally, we generated a complex signature by using both amplitude difference mean and phase (unwrap) difference mean at a location as shown in Eq. (18).

$$H_{diff} = \left[H_{amp\text{-}diff}\ H_{ph\text{-}diff}\right]. \tag{18}$$

6. The generated CSI-MIMO fingerprint H_{diff} represents the fingerprint for a location using an AP. Therefore, for X number of APs, the fingerprint is represented as

$$H_{diff} = \left[H_{diff\ 1}\ H_{diff\ 2}\ \cdots\ H_{diff\ X}\right]. \tag{19}$$

We evaluate two versions of the proposed CSI-MIMO fingerprint: CSI-MIMO with an amplitude difference $H_{amp\text{-}diff}$ and a complex CSI-MIMO H_{diff} with magnitude and phase difference for its performance evaluation in static and dynamic indoor environments.

4.4. Temporal stability and spatial diversity of CSI-MIMO

The applicability of the CSI to facilitate more accurate localization is due to its better temporal stability and spatial distinction, and is demonstrated in PinLoc [8,16]. We evaluate the temporal stability and spatial separation of our proposed CSI-MIMO signature.

Fig. 6 shows the CSI-MIMO complex signature for two distinct locations. We have collected 1000 samples from an AP at two locations separated by one meter. We generated the CSI-MIMO fingerprint for both the locations as mentioned in Section 4.3. The graph shows the CSI-MIMO complex signature for 1000 samples along with the mean of all the samples. It can be seen that the shape of the channel frequency response is preserved for each of the location indicating the temporal stability of the CSI-MIMO signature.

Fig. 7 shows the sample mean of CSI-MIMO for both the locations for a better visibility. It can be clearly seen that the two locations are very well separated. Also, the shape of the channel frequency response (CFR) is distinct for each of the location. The two locations are distinctly separated and distinguishable using an unique CSI-MIMO fingerprint.

4.5. CSI-MIMO localization algorithms

For localization of a mobile device, we exploit our proposed CSI-MIMO fingerprint using both the algorithms: the deterministic k-Nearest Neighbor and probabilistic Bayes rule. We have assumed data distribution as Gaussian i.e. observations are real values falling between two real limits.

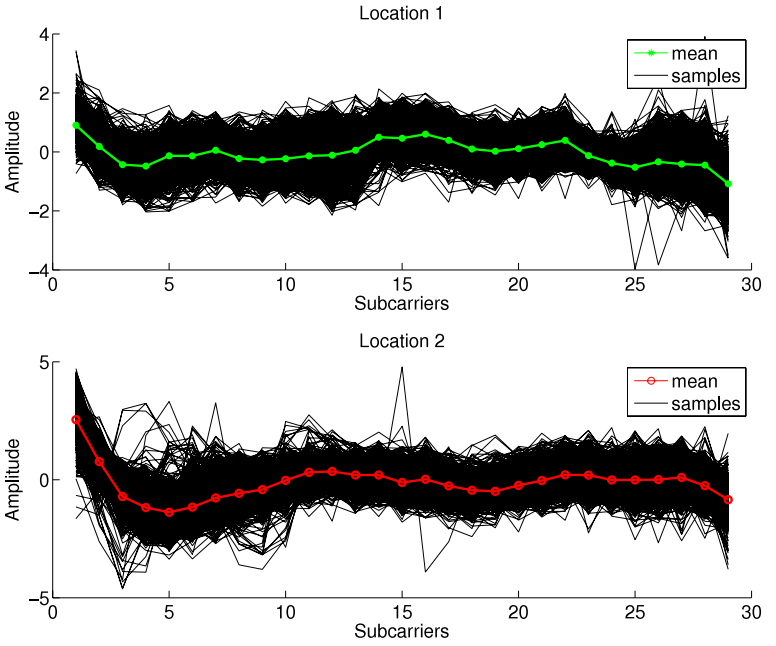

Fig. 6. Temporal stability with CSI-MIMO.

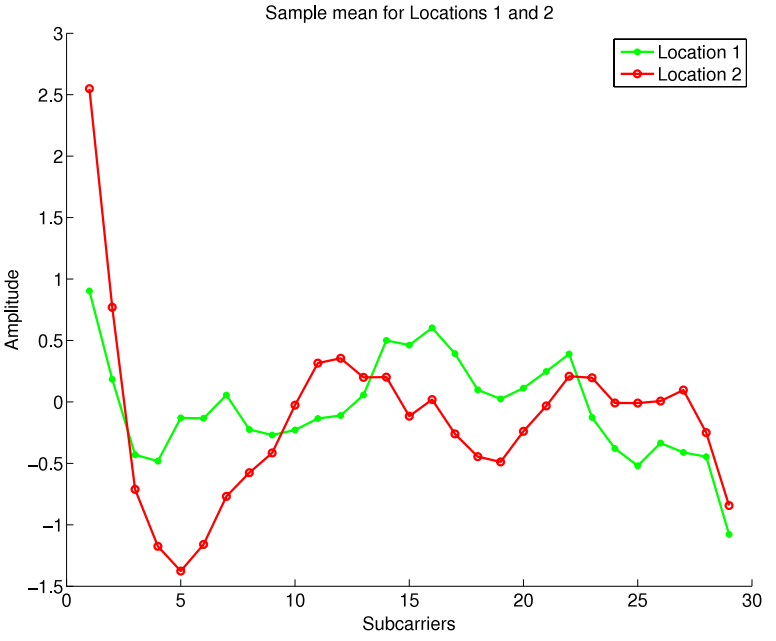

Fig. 7. Spatial separation with CSI-MIMO.

4.5.1. Deterministic k-Nearest Neighbor

The deterministic k-Nearest Neighbor algorithm uses an Euclidean distance to calculate the distance between a test fingerprint and each training fingerprint in the radio-map. It averages k minimum distances to finally give an estimated value. In this paper, we consider $k = 1$ and only use minimum Euclidean distance as location estimation, which is represented by Eq. (20). Consideration of k for more than one neighbor is not explored for CSI-MIMO and might be considered as the direction for future research.

$$E_{dist} = \sqrt{\sum^{K} (H_{train} - H_{test})^2}. \tag{20}$$

4.5.2. Probabilistic Bayes rule

We localize the mobile device using Bayes rule and calculate the prior probabilities using Pearson correlation coefficient [9]. The location l is estimated by maximizing the posterior probability.

According to the Bayes rule, the posterior probability $P(l_i|H_{test})$ is given by,

$$P(l_i|H_{test}) = \frac{P(l_i)\,P(H_{test}|l_i)}{\sum_{1}^{i} P(l_i)\,P(H_{test}|l_i)} = \frac{P(l_i)\,P(H_{test}|l_i)}{P(H_{test})} \tag{21}$$

where, H_{test} is the test fingerprint for K APs, H_{train} is a fingerprint from the radio-map, $P(l_i)$ is the prior probability and $P(H_{test}|l_i)$ is likelihood.

The prior probability $P(l_i)$ of the test fingerprint on each training fingerprint in the radio-map is calculated as follows.

$$P(l_i) = \frac{\rho_{H_{test}H_{train}}}{\sum_{i=1}^{L} \rho_{H_{test}H_{train}}} \tag{22}$$

where L is total training locations in the radio-map and $\rho_{H_{test}H_{train}}$ is the Pearson coefficient that defines correlation between the test and each training fingerprint.

$$\rho_{H_{test}H_{train}} = \prod_{k=1}^{K} \frac{cov\left(H_{test}^{k}, H_{train}^{k}\right)}{\sigma_{H_{test}^{k}}\sigma_{H_{train}^{k}}}. \tag{23}$$

The likelihood of the test fingerprint for given training fingerprint in the radio-map is evaluated as

$$P(H_{test}|l_i) = \prod_{k=1}^{K} P(H_{test}|l_i)\,. \tag{24}$$

Finally, the location is estimated by maximizing the posterior probability $P(l_i|H_{test})$.

$$\hat{l} = \max \sum_{i}^{L} P(l_i|H_{test})\,. \tag{25}$$

5. Experimental data analysis and evaluation

The experimental data is analyzed for two variants of the proposed CSI-MIMO system: an amplitude CSI-MIMO and a complex CSI-MIMO. Also, we analyze the effect of various factors such as number of access points, war-driving samples and type of localization algorithm used. Furthermore, we compare the proposed CSI-MIMO system with other CSI-based fingerprinting systems.

5.1. Type and size of CSI-MIMO fingerprint

We analyze the localization accuracy of the proposed CSI-MIMO fingerprint variants in static and dynamic environments. One of the variant of the proposed CSI-MIMO fingerprint consisting of amplitude and phase difference of all the sub-carriers is been presented in [17].

We evaluate the impact of size and complexity of CSI-MIMO on the localization accuracy using amplitude difference, phase difference and both. We experimentally evaluate the accuracy achieved by the complex CSI-MIMO signature for static and dynamic environments. Initially, for evaluation, we have considered 1000 training and 100 test samples with four combinations: 1000/100, 100/100, 100/50 and 50/50. The impact of war driving on the localization is presented further in this section. The Mean Distance Error (MDE) is used as a performance metric to evaluate the performance of CSI-MIMO. The MDE for three CSI-MIMO variants for static and dynamic test-beds is shown in Figs. 8 and 9.

For static environment such as computer lab, the minimal mean distance error is achieved using a complex CSI-MIMO signature, which consists of amplitude and phase difference as shown in Fig. 8. The static indoor environment that exhibits less mobility provides lower phase fluctuations. The complex CSI-MIMO signature that includes the amplitude variation along with the phase fluctuations presents more defined location signature and therefore separates two locations more distinctly. Addition of phase information to the amplitude information in CSI-MIMO provides an additional level of distinctiveness to the location signature.

The research lab presents a typical dynamic indoor environment with more mobility. In contrast to static environment, the phase information varies drastically in dynamic environment. This affects the separation between two locations by degrading the uniqueness of the CSI-MIMO location fingerprint. In such scenario, when amplitude information itself is

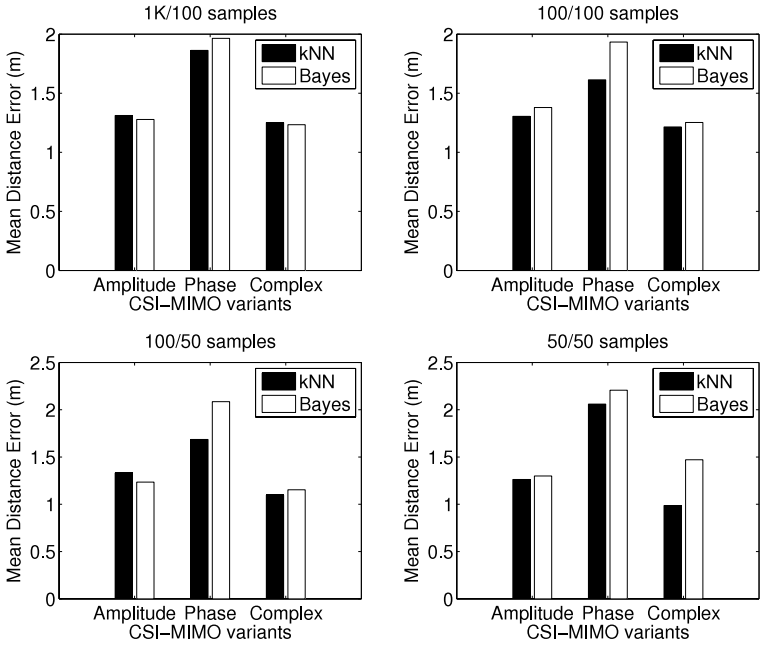

Fig. 8. MDE of CSI-MIMO variants in a static environment.

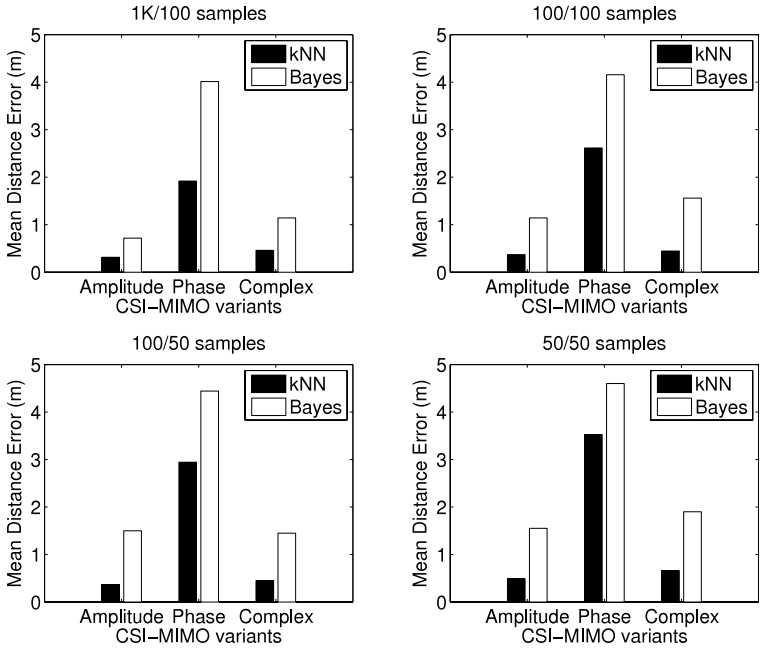

Fig. 9. MDE of CSI-MIMO variants in dynamic environment.

exploited for localization, better results are achieved without relying on the phase values. Fig. 9 shows that the minimal mean distance error is achieved using amplitude variant of CSI-MIMO for all the sample combinations using both the localization algorithms.

Table 2 shows the summary of the MDE achieved using the two variants of CSI-MIMO for static and dynamic indoor environment.

For static environments, the complex CSI-MIMO signature achieves the lowest mean distance error using k-Nearest Neighbor and Bayes algorithms. Deterministic K-Nearest Neighbor and Probabilistic Bayes approaches provides the lowest MDE of 0.98 m and 1.15 m using 50/50 and 100/50 war-driving sample combinations respectively. This analysis

Table 2
Summary of results for CSI-MIMO amplitude and complex variants.

Test-bed	War-driving samples	k-Nearest Neighbor		Bayes	
		Amp	Complex	Amp	Complex
Static (computer lab)	1K/100	1.3119	1.2512	1.2782	1.2331
	100/100	1.3045	1.2146	1.3791	1.2520
	100/50	1.3307	1.0968	1.2350	1.1530
	50/50	1.2620	0.9873	1.2996	1.4711
Dynamic (research lab)	1K/100	0.3147	0.4596	0.7175	1.1416
	100/100	0.3710	0.4464	1.1424	1.5599
	100/50	0.3667	0.4506	1.4986	1.4491
	50/50	0.4864	0.6587	1.5515	1.9005

demonstrates that the number of samples required for the localization is a crucial factor in determining the performance of the localization system and is investigated in later part of the section.

In contrast to static environment, an amplitude CSI-MIMO gives better accuracy of 0.31 m and 0.71 m using 1K/100 war-driving samples for k-Nearest Neighbor and Bayes in a dynamic environment respectively. However, dynamic environment requires more war-driving samples as compared to static environments. The dynamic environment possesses mobility that leads to severe signal fluctuations due to multi-path effect. More war-driving samples provides better estimate of the signal and helps in more accurate localization. The CSI-MIMO with its amplitude variant localizes more accurately than its complex variant due to significant phase variations in dynamic environment. Therefore, the CSI-MIMO amplitude variant provides an advantage of lower dimensions and complexity of the location signature.

In summary, a complex CSI-MIMO works more efficiently in static environments with less war-driving and an amplitude variant of CSI-MIMO performs better in complex environments. Henceforth the rest part of the paper uses a complex CSI-MIMO for static environment and an amplitude CSI-MIMO for dynamic environment.

5.2. Effect of impact factors

The effect of various experimental parameters such as number of deployed APs, number of training and test samples, and type of the localization algorithm on the localization accuracy of CSI-MIMO is investigated. The deterministic k-Nearest Neighbor and probabilistic Bayes rule are used to analyze the performance of CSI-MIMO.

5.2.1. Number of APs

Wi-Fi fingerprinting approach for indoor localization requires adequate number of APs to cover the entire floor-plan to localize efficiently. The number of APs for the localization is the crucial factor to determine the accuracy of the localization. Location fingerprint for optimum APs provides more uniqueness to the signature. Depending on the floor-plan size, we deployed sufficient number of APs effectively and ensured that it covers the entire floor-plan. We have considered various combinations of number of APs to evaluate the optimum APs required for more accurate location estimation: AP 1 represents first AP, AP 1&2 represents the combination of first and second AP and AP 1, 2&3 represents the combination of first, second and third AP. The results are presented for only selected combination of access points that provides significantly comparison.

Fig. 10 shows the Mean Distance Error (MDE) of CSI-MIMO for various number of APs using both the localization algorithms.

It is clearly seen that the mean distance error achieved using only one AP is the highest for all sample combinations when using both the algorithms in a static and a dynamic environment. As more number of APs included for localization, the MDE decreases. The minimal MDE is reached for two and three APs in static and dynamic environment respectively and therefore considered as optimal APs for localization in each of the environment. Inclusion of more number of APs in the location signature provides more features to the location signature and distinct separation of the location signatures. However, there is a convergence for the number of access points that provide no more improvement in the localization accuracy and reaches a saturation value. This intended behavior is not explored in the paper.

5.2.2. Number of training and test samples

We analyze the MDE achieved using CSI-MIMO for different combinations of training and test samples using optimal APs i.e. three APs for research lab and two APs for computer lab. Empirically, we evaluate the optimal war driving data that achieves minimal distance error in both experiment environments. We have considered maximum 1000 training and 100 test samples with four combinations: 1000/100, 100/100, 100/50 and 50/50.

Fig. 11 shows the MDE for four war driving sample combinations k-Nearest Neighbor and Bayes algorithm.

For dynamic environment, the sample combination of 1K/100 provides minimal MDE for both k-Nearest Neighbor and Bayes. However, the combination of samples varies for static environment that achieves lowest MDE for both the localization algorithms.

It is clearly seen that the complex CSI-MIMO location signature in a static environment provides maximum accuracy of 0.98 m using 50/50 samples combination using a deterministic k-Nearest Neighbor. However, a probabilistic Bayes achieves

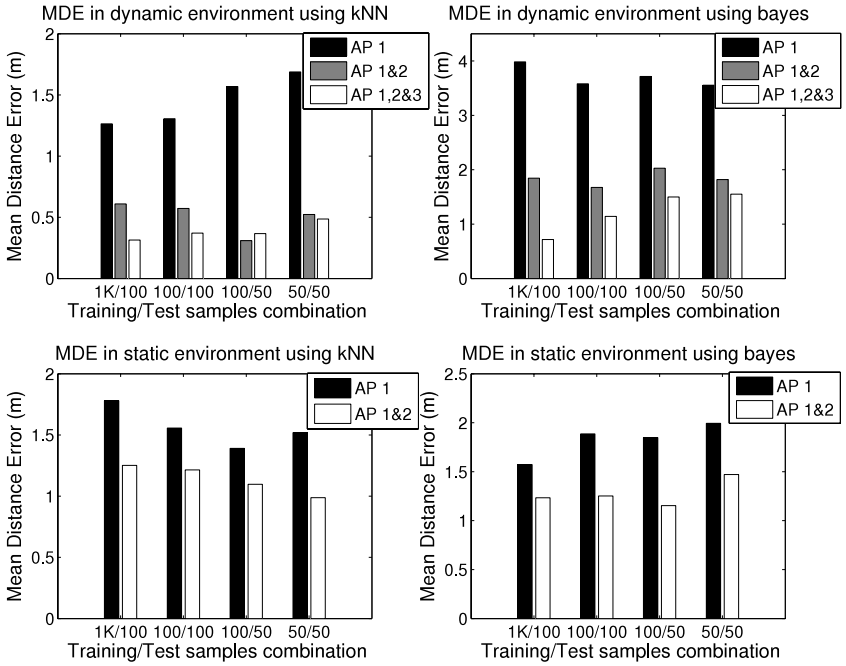

Fig. 10. MDE vs number of APs.

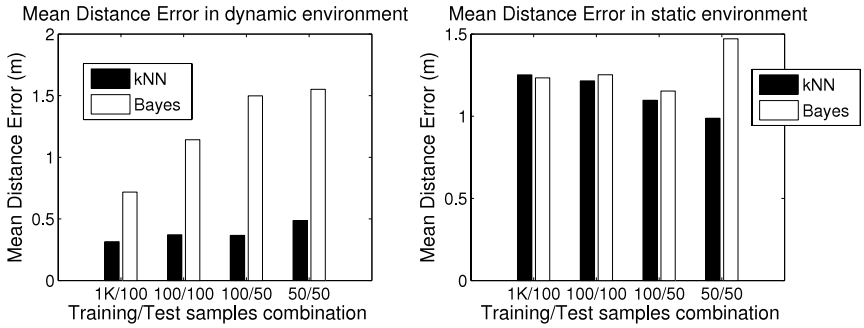

Fig. 11. MDE vs. war driving samples.

Table 3
MDE in meters for various war driving samples.

Test-bed	CSI-MIMO variant	War-driving samples	k-Nearest Neighbor	Bayes
Static (computer lab)	Complex	1K/100	1.2512	1.2331
		100/100	1.2146	1.2520
		100/50	1.0968	1.1530
		50/50	0.9873	1.4711
Dynamic (research lab)	Amplitude	1K/100	0.3147	0.7175
		100/100	0.3710	1.1424
		100/50	0.3667	1.4986
		50/50	0.4864	1.5515

an accuracy of 1.15 m using 100/50 samples combination. Whereas, in dynamic indoor scenario that uses an amplitude CSI-MIMO location signature, maximum accuracy of 0.31 and 0.71 m is achieved using 1K/100 samples combination.

As seen in Table 3 in static scenario, lower samples achieves better accuracy. However Bayes algorithm needs more training samples as compared to k-Nearest Neighbor since the Bayes approach calculates the prior and posterior probabilities using the training data statistics such as mean and standard deviation of the training data. More the samples, better the estimation of the signal statistics. In contrast to static scenario, dynamic environment requires maximum war-driving of 1K/100 to achieve minimal MDE.

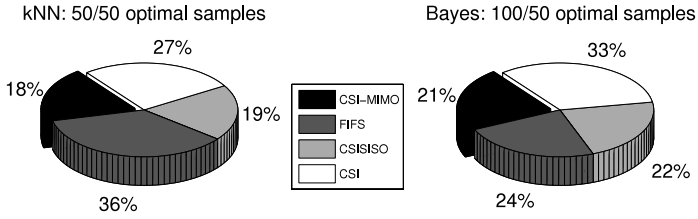

Fig. 12. Performance of CSI-based systems in static scenario.

A static environment possesses more stable channel characteristics than a dynamic environment. The signal degrades significantly due to multi-path effect in dynamic situations and therefore causes severe amplitude and phase fluctuations resulting in scattered CSI information. More war-driving samples helps in estimating the signal statistics efficiently that improves the location estimation using deterministic and probabilistic techniques.

In summary, the sample combination of 1K/100 is optimal for maximum accuracy in dynamic environment, whereas 100/50 and 50/50 war-driving is optimal for static scenarios.

5.2.3. Type of localization algorithm

The performance of our proposed CSI-MIMO is analyzed using k-Nearest Neighbor and Bayes algorithms. We summarize the performance of CSI-MIMO based on the localization algorithm using Tables 2 and 3.

It is evident that CSI-MIMO provides better estimations using both the localization algorithms i.e. deterministic k-Nearest Neighbor and probabilistic Bayes. However, the maximum accuracy is achieved using deterministic approach in both the experimental environments for all the combinations of training and testing data. The deterministic k-Nearest Neighbor achieves maximum accuracy of 0.98 m with 50/50 war-driving and 0.31 with 1K/100 war-driving for static and dynamic environments respectively. However the accuracy achieved using Bayes rule is lower than the deterministic approach with maximum accuracy of 1.15 with 100/50 war-driving and 0.71 m with 1K/100 war-driving both test-beds.

In the static environment, the maximum accuracy achieved using Bayes rule requires more training samples to estimate the probabilities more precisely, which results in more accurate location estimation.

5.3. Comparison of CSI-MIMO with other CSI-based fingerprints

We present the comparison of the performance of CSI-MIMO with other CSI-based fingerprinting systems such as FIFS, a CSISISO and CSI. For better interpretation of the comparison system, we briefly explain the system with their corresponding signature. A CSISISO is a complex CSI for all the 30 sub-carriers with the first element of the MIMO matrix such that $H_i = h_{11}$, whereas CSI uses all the sub-carriers that are averaged over MIMOs with $H_i = H_{avg}$ as represented in Eqs. (9) and (11). More details regarding the location signature of FIFS is presented in [9].

Figs. 12 and 13 shows the percentage MDE achieved by various CSI-based fingerprinting systems for respective optimal samples using k-Nearest Neighbor and Bayes algorithms in static and dynamic indoor environments respectively. The MDE is represented as a percentile of the sum of the mean distance error achieved by each of the CSI-based fingerprinting system including CSI-MIMO for each of the sample combination and localization algorithm and is calculated as

$$\text{MDE}\ (\%) = \frac{\text{MDE}_{\text{CSI-MIMO}}}{\text{MDE}_{\text{CSI-MIMO}} + \text{MDE}_{\text{FIFS}} + \text{MDE}_{\text{CSISISO}} + \text{MDE}_{\text{CSI}}}. \tag{26}$$

In static environment, CSI-MIMO provides better accuracy than other CSI-based fingerprints. Using deterministic k-Nearest Neighbor with optimal samples of 50/50, CSI-MIMO achieves 18% of MDE with 0.98 m. FIFS, CSISISO and CSI provides 27%, 19% and 36% MDE signifying 1.92 m, 1.04 m and 1.44 m respectively. In similar manner, using Bayes rule, with optimal samples of 100/50, CSI-MIMO achieves 21% of MDE with 1.15 m. FIFS, CSISISO and CSI provides 33%, 22% and 24% MDE signifying 1.28 m, 1.18 m and 1.76 m respectively. It is evident that CSI-MIMO performs better than other CSI-based location signatures with lowest MDE. However, it is seen that the CSISISO performs comparably to CSI-MIMO with slight MDE difference of only 1% due to prominent CSI features that provides the frequency diversity and location distinction in environment having less mobility.

The proposed CSI-MIMO achieves ever better performance in dynamic scenario. The lowest MDE of 3% and 4% with 0.31 m and 0.71 m is achieved for optimal samples of 1K/100 using both the localization algorithms. In contrast, other CSI-based fingerprints provides poor performance with the range of 22%–38% MDE. It is clearly seen that in dynamic environments, CSI-MIMO performs outstandingly well.

6. Conclusion

In this paper, we exploited the fine-grained PHY layer CSI information for Wi-Fi fingerprinting. A novel location fingerprint CSI-MIMO is proposed offering frequency and spatial diversity of CSI that represents a location distinctly. We

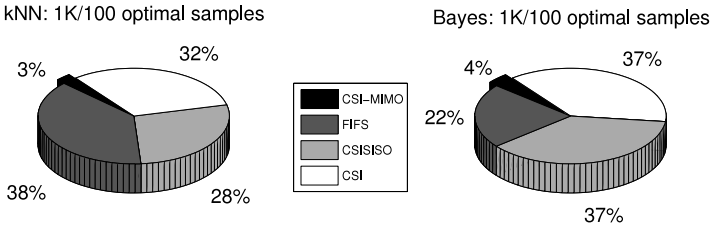

Fig. 13. Performance of CSI-based system in dynamic scenario.

have investigated the effect of CSI-MIMO variants on the localization accuracy in static and dynamic indoor environments. Based on the mobility pattern of indoor scenarios, we have proposed amplitude and complex signature variant of the CSI-MIMO. Our experimental results showed that dynamic environments required more war-driving of 1K/100, whereas static environment needed war-driving of 100/50 or 50/50 samples. Moreover, the maximum accuracy of 0.98 and 0.31 m is achieved using deterministic k-Nearest Neighbor algorithm in static and dynamic environment.

In this paper, we have utilized a high-dimensional, complex CSI signature as a location fingerprint. The proposed signature CSI-MIMO may result in more complexity and computations when deployed for larger indoor environments. The optimization of CSI-MIMO signatures to a smaller yet distinct location fingerprint will be quite challenging.

Acknowledgment

This work has been funded in part by the EC via grant agreement no. 611659 for the AU2EU project.

References

[1] P. Bahl, V. Padmanabhan, RADAR: An in-building RF-based user location and tracking systems, in: Proceedings INFOCOM00': Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol 2, IEEE, 2000, pp. 775–784. http://dx.doi.org/10.1109/INFCOM.2000.832252.
[2] M. Youssef, A. Agrawala, The Horus WLAN location determination system, in: MobiSys'05: Proceedings of the 3rd International Conference on Mobile Systems, Applications and Services, ACM, USA, 2005, pp. 205–218. http://dx.doi.org/10.1145/1067170.1067193. URL http://doi.acm.org/10.1145/1067170.1067193.
[3] K. Kaemarungsi, P. Krishnamurthy, Analysis of WLANs received signal strength indication for indoor location fingerprinting, Pervasive Mob. Comput. 8 (2) (2012) 292–316. http://dx.doi.org/10.1016/j.pmcj.2011.09.003. special issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing. URL http://www.sciencedirect.com/science/article/pii/S1574119211001234.
[4] Y. Chapre, P. Mohapatra, S. Jha, A. Seneviratne, Received signal strength indicator and its analysis in a typical WLAN system (short paper), in: IEEE 38th Conference on Local Computer Networks, LCN, 2013, pp. 304–307. http://dx.doi.org/10.1109/LCN.2013.6761255.
[5] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Predictable 802.11 packet delivery from wireless channel measurements, in: Proceedings of the ACM SIGCOMM 2010 Conference, SIGCOMM'10, ACM, New York, NY, USA, 2010, pp. 159–170. http://dx.doi.org/10.1145/1851182.1851203. URL http://doi.acm.org/10.1145/1851182.1851203.
[6] D.A. Hall, Understanding benefits of MIMO technology, Microwaves and RF.
[7] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, L. Ni, CSI-based indoor localization, IEEE Trans. Parallel Distrib. Syst. 24 (7) (2013) 1300–1309. http://dx.doi.org/10.1109/TPDS.2012.214.
[8] S. Sen, R.R. Choudhury, B. Radunovic, T. Minka, Precise indoor localization using PHY layer information, in: Proceedings of the 10th ACM Workshop on Hot Topics in Networks, HotNets-X, ACM, New York, NY, USA, 2011, pp. 18:1–18:6. http://dx.doi.org/10.1145/2070562.2070580. URL http://doi.acm.org/10.1145/2070562.2070580.
[9] J. Xiao, K. Wu, Y. Yi, L. Ni, FIFS: Fine-grained indoor fingerprinting system, in: 21st International Conference on Computer Communications and Networks, ICCCN, 2012, pp. 1–7. http://dx.doi.org/10.1109/ICCCN.2012.6289200.
[10] J. Hightower, R. Want, G. Borriello, SpotON: An indoor 3D location sensing technology based on RF signal strength, Technical Report, Univ. of Washington, Dep. of Comp. Science and Eng., Seattle, WA, Feb.
[11] L.M. Ni, Y. Liu, Y.C. Lau, A.P. Patil, LANDMARC: Indoor location sensing using active RFID, Wirel. Netw. 10 (6) (2004) 701–710. http://dx.doi.org/10.1023/B:WINE.0000044029.06344.dd. URL http://dx.doi.org/10.1023/B:WINE.0000044029.06344.dd.
[12] C. Nerguizian, C. Despins, S. Affes, Geolocation in mines with an impulse response fingerprinting technique and neural networks, IEEE Trans. Wireless Commun. 5 (3) (2006) 603–611. http://dx.doi.org/10.1109/TWC.2006.1611090.
[13] Y. Jin, W.-S. Soh, W.-C. Wong, Indoor localization with channel impulse response based fingerprint and nonparametric regression, IEEE Trans. Wireless Commun. 9 (3) (2010) 1120–1127. http://dx.doi.org/10.1109/TWC.2010.03.090197.
[14] N. Patwari, S.K. Kasera, Robust location distinction using temporal link signatures, in: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, MobiCom'07, ACM, New York, NY, USA, 2007, pp. 111–122. http://dx.doi.org/10.1145/1287853.1287867. URL http://doi.acm.org/10.1145/1287853.1287867.
[15] J. Zhang, M.H. Firooz, N. Patwari, S.K. Kasera, Advancing wireless link signatures for location distinction, in: Proceedings of the 14th ACM International Conference on Mobile Computing and Networking, MobiCom'08, ACM, New York, NY, USA, 2008, pp. 26–37. http://dx.doi.org/10.1145/1409944.1409949. URL http://doi.acm.org/10.1145/1409944.1409949.
[16] A. Bhartia, Y.-C. Chen, S. Rallapalli, L. Qiu, Harnessing frequency diversity in Wi-Fi networks, in: Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom'11, ACM, New York, NY, USA, 2011, pp. 253–264. http://dx.doi.org/10.1145/2030613.2030642. URL http://doi.acm.org/10.1145/2030613.2030642.
[17] Y. Chapre, A. Ignjatovic, S. Jha, A. Seneviratne, CSI-MIMO indoor Wi-Fi fingerprinting system, in: IEEE 39th Conference on Local Computer Networks, LCN, 2014.