CSI-MIMO: Indoor Wi-Fi Fingerprinting System

Yogita Chapre*‡, Aleksandar Ignjatovic*, Aruna Seneviratne†‡ and Sanjay Jha*
*School of Computer Science & Engineering, University of New South Wales, Sydney, Australia
†School of Electrical Engineering & Telecommunications, University of New South Wales, Australia
‡Networks Research Group, NICTA, Sydney, Australia
Email: {yogitac, ignjat, sanjay}@cse.unsw.edu.au, {yogita.chapre, aruna.seneviratne}@nicta.com.au

***Abstract*—Wi-Fi based fingerprinting systems, mostly utilize the Received Signal Strength Indicator (RSSI), which is known to be unreliable due to environmental and hardware effects. In this paper, we present a novel Wi-Fi fingerprinting system, exploiting the fine-grained information known as Channel State Information (CSI). The frequency diversity of CSI can be effectively utilized to represent a location in both frequency and spatial domain resulting in more accurate indoor localization. We propose a novel location signature CSI-MIMO that incorporates Multiple Input Multiple Output (MIMO) information and use both the magnitude and the phase of CSI of each sub-carrier. We experimentally evaluate the performance of CSI-MIMO fingerprinting using the k-nearest neighbor and the Bayes algorithm. The accuracy of the proposed CSI-MIMO is compared with Fine-grained Indoor Fingerprinting System (FIFS) and a simple CSI-based system. The experimental result shows an accuracy improvement of 57% over FIFS with an accuracy of 0.95 meters.**

I. Introduction

Localization is the most important aspect in various Location-based services (LBS). LBS includes location-based mobile advertising, parcel tracing in a post office, discovering nearest Automated Teller Machine (ATM), weather services and location-based games. In an outdoor environment, Localization can be more accurately achieved using the Global Positioning System (GPS), but GPS signals are too weak to localize in indoor environment. Alternatively, indoor localization based on wireless signals such as Wi-Fi and GSM are more popular due to open access and lower cost. Wireless indoor positioning uses various location estimation approaches such as signal-based Received Signal Strength Indicator (RSSI), Signal Strength Difference (SSD), time-based including Time Of Arrival (TOA), Time Difference Of Arrival (TDOA) and angle-based Angle of Arrival (AOA). The RSSI-based fingerprinting is the most widely used approach among all. RSSI basically represents the received signal strength at a receiver, which is an average signal value traversed through various paths for a given transmitter-receiver pair. As the distance between the transmitter-receiver increases, the RSSI value degrades due to the attenuation. This phenomenon is exploited in various range detection algorithms, including RADAR [1] and Horus [2] that uses deterministic k-nearest neighbor (kNN) and probabilistic Bayes Rule respectively. The RSSI has proven to be a highly volatile metric for indoor localization[3][4]. The main reasons for volatality are highlighted below . Firstly, RSSI at a given location is the average of the signals received through various paths (multi-path effect) and have temporal variance. Secondly, RSSI fluctuates significantly in the presence/mobility of people and change of the environment such as building layout and type of building material. Finally, RSSI varies considerably based on the hardware used, including the type and orientation of the antenna, sampling rates and quantization bins of the Network Interface Card (NIC).

The sensitivity of RSSI to various spatial, temporal, hardware and environmental factors affects the accuracy of localization significantly. Hossain et al. [5] proposed SSD, which is derived from RSSI to deal with the hardware dependence of RSSI. SSD presents the hardware heterogeneity when the devices used for the training and testing are different. However, the effect of spatial, temporal and environmental parameters are not explored. We believe that, to achieve a more accurate localization, there is a need to extract a fine-grained information that provides temporal stability and frequency diversity to detect the multi-path.

In 802.11 a/g/n networks, the data is transmitted and received using Orthogonal Frequency Division Multiplexing (OFDM). An OFDM uses multiple data carriers called sub-carriers with multiple transmit and receive antennas. The data to be transmitted is modulated on multiple sub-carrier frequencies and transmitted simultaneously and received using multiple data sub-carriers. The received data is affected due to scattering, fading and power loss with distance. This data is exploited to determine the channel quality between the transmitter and the receiver. The PHY layer information about channel quality commonly known as Channel State Information (CSI) can be used due to its frequency diversity (OFDM sub-carriers) and spatial diversity (multiple antennas). The CSI, fine grained physical layer information is extractable using off-the-shelf Intel 5300 NIC [6]. In contrast to RSSI, which provides an aggregated value for a sample, CSI estimates the channel for each sub-carrier. Moreover, CSI presents amplitude and phase information of each sub-carrier for each transmit and receive antenna stream. Multiple sub-carriers traveling along different paths, will have different scattering due to multi-path, thus resulting into unique amplitude and phase for each sub-carrier. Recently wireless technologies use Multiple-input, multiple-output (MIMO) antenna systems extensively such as IEEE 802.11n, 3GPP LTE, and mobile WiMAX systems. The advantage of this technique is the enhanced data throughput even under the influence of interference, signal fading, and multi-path. The MIMO-communication system can accomplish higher data rates than a SISO system by exploiting

978-1-4799-3780-6/14/$31.00 ©2014 IEEE

the spatial diverse communication link [7]. We exploit this feature to create a unique location fingerprint and build a radio-map. This facilitates in designing an indoor positioning system with better accuracy.

The CSI-based localization has been explored in [8] and [9], however, this study uses the Single Input Single Output (SISO) aspect of CSI. Xiao et.al. [10] used the spatial diversity aspect, however, the location fingerprint is represented in terms of power in a sample that is aggregated over all the sub-carriers. The uniqueness of location fingerprint needs to be defined to distinguish between adjacent locations. The consideration of MIMO for CSI-based location fingerprint provides a high dimensional and more defined location signature. However, the usage of MIMO for each transmitter-receiver antenna stream or aggregation over all streams is challenging. Earlier demonstrated approaches like FILA [8], FIFS [10] and PinLoc [9] used CSI information either by averaging power across multiple sub-carriers or clustering amplitude and phase of all sub-carriers respectively. In this paper, we attempt to use both the frequency diversity and spatial diversity by combining multiple sub-carriers and multiple transmit and receiver antennas to generate the location fingerprint.

The main contributions of this paper are summarized as follows.

1) We aggregate the CSI over multiple antennas for multiple sub-carriers and present a novel location signature CSI-MIMO consisting of magnitude and phase deviation between subsequent sub-carriers.
2) We create a radio-map using CSI-MIMO for multiple Access Points (APs) to improve the accuracy and reduce the distance error.
3) We compare the accuracy of CSI-MIMO with other CSI-based localization systems, including earlier proposed FIFS and a simple CSI signature.
4) We investigate the impact on the localization accuracy due to the various factors such as the number of training and test samples, device used in the localization phase and type of algorithm.
5) We analyze the effectiveness of the proposed location signature using k-nearest neighbor (kNN) and FIFS-probabilistic (Bayes Rule) approach.

The rest of the paper is organized as follows. In Section II, we briefly present the related work on Wi-Fi fingerprinting using various location signatures. Section III introduces preliminaries including RSSI, SSD, OFDM and CSI. Our proposed system and methodology are explained extensively in Section IV. It includes system architecture, experiment setup and fingerprint generation technique. In Section V, we compare our proposed system CSI-MIMO with FIFS. We also evaluate our experimental results for various impact factors, including war-driving samples, device used for localization and type of localization algorithms. Conclusions and directions for the future research is presented in Section VI.

II. Related Work

Researches on fingerprinting for indoor localization have exploited various approaches, including radio frequency (RF) signal strength, Channel Impulse Response (CIR) and the most recently CSI.

A. RF signal strength

Wi-Fi signal strength, such as RSSI is being exploited for more than a decade. Bahl et.al. proposed the first Wi-Fi fingerprinting technique RADAR[1]. It collects RSSI from multiple APs and stores the mean value of RSSI in a radio-map. During the localization phase, observed fingerprint is matched against the radio-map. Using the kNN deterministic algorithm, RADAR localizes with an accuracy of 3 meters. To enhance the RADAR performance, Youssef et.al. [2] used probabilistic approach in Horus, which uses a joint clustering technique using a common set of APs to estimate the location. It reduces the computation overhead and provides localization accuracy of 2.1 meters. As discussed earlier, RSS is highly unreliable. Hossain et.al. [5] addressed the hardware heterogeneity using the SSD signature. This approach showed accuracy improvement, when devices used for training and testing are different. However, the sensitivity of SSD to environmental and temporal parameters is unexplored.

Radio Frequency Identification (RFID) approaches such as spotON [11] and LANDMARC [12] exploit RSSI measurements from active RFID tags by using lateration and kNN estimation techniques. However, later requires extra infrastructure to cover the floor-plan.

B. Channel Impulse Response

CIR-based location fingerprinting has been recently proposed over RSSI-fingerprinting to improve the accuracy. It represents a location with measured CIR as a unique fingerprint. Nerguizian et.al [13] proposed a unique CIR that is collected at a location using a channel sounder and spectrum analyzer operating at a bandwidth of 200 MHz. It enhances the accuracy by using Artificial Neural Network (ANN) but devices used for measurements are very costly and bulky. Also, performance is not evaluated for varying bandwidth.

C. Channel State Information

Most recently, CSI fingerprinting using PHY layer fine grained information is being proposed in [9]. This approach exploits the CSI with the location granularity to improve the accuracy. It utilizes the frequency diversity, but doesn't consider the spatial diversity of the CSI. Another CSI-based fingerprinting FIFS [10] leverages both the frequency and spatial diversity and uses a correlation filter with the probability algorithm, however it represents the location by aggregating the power of the sub-carriers.

Our approach, CSI-MIMO uses both frequency and spatial diversity. We exploit the amplitude and phase information of all the sub-carriers to uniquely represent a location. We employ both the kNN and probabilistic methods with varying training and test samples for localization.

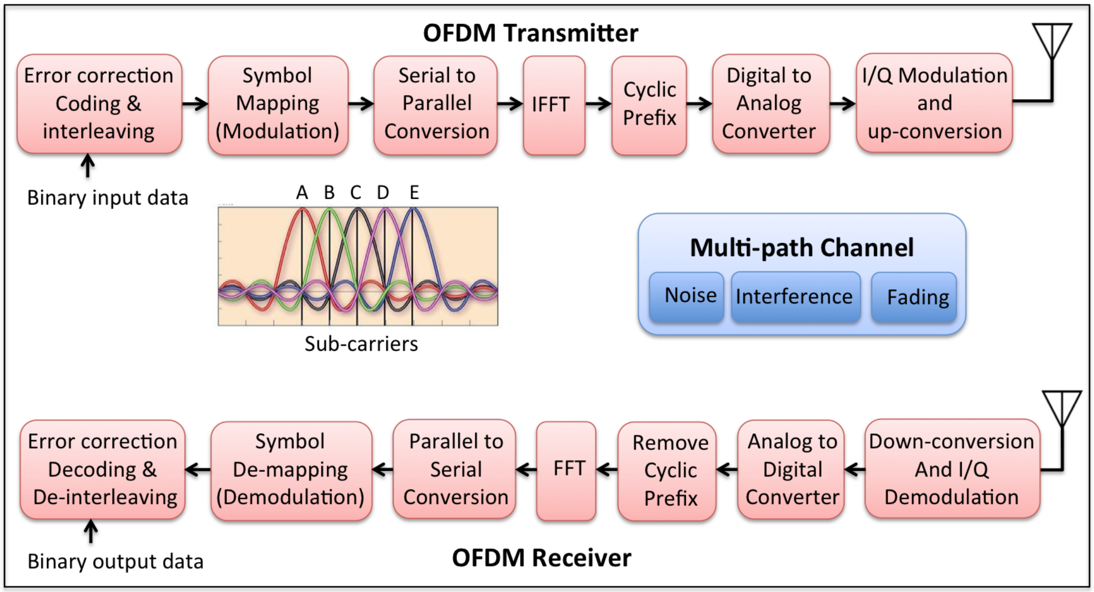

Fig. 1. OFDM Transceiver Block Diagram

III. Preliminaries: Notation and Terminologies

In this section, we discuss briefly about various location signature, such as RSSI, SSD along with CSI. We also overview the OFDM technology and the fundamentals of CSI, which forms the basis of a location signature.

A. Received Signal Strength Indicator

Traditional RSSI-based location signature consists of N number of RSSI samples from an AP at a location. Thus, for K APs the dimension of the location signature is $K * N$ as shown in Eqn. (1). The location fingerprint, usually stores the mean value of RSSI samples from an AP in the radio-map. Another scheme SSD used the signal strength difference of the samples from two different APs. Thus SSD-based signature is of $k - 1 * N$ dimensions as shown in Eqn. (2). Since mobile devices are built with a variety of NIC, it provides different quantization bins and sampling rates. Hence, the RSSI reported by each of the mobile devices varies substantially[3]. This results higher estimation error if the device used for training and localization phase are not identical.

$$RSSI = \begin{bmatrix} ss_{11} & ss_{12} & ss_{13} \dots ss_{1N} \\ ss_{21} & ss_{22} & ss_{23} \dots ss_{2N} \\ \dots & \dots & \dots \quad \dots \\ ss_{K1} & ss_{K2} & ss_{K3} \dots ss_{KN} \end{bmatrix} \quad (1)$$

$$SSD = \begin{bmatrix} ss_{11} - ss_{21} & ss_{12} - ss_{22} \dots ss_{1N} - ss_{2N} \\ ss_{11} - ss_{31} & ss_{12} - ss_{32} \dots ss_{1N} - ss_{3N} \\ \dots & \dots \quad \dots\dots \\ ss_{11} - ss_{K1} & ss_{11} - ss_{K2} \dots ss_{11} - ss_{KN} \end{bmatrix} \quad (2)$$

B. Orthogonal Frequency Division Multiplexing

An OFDM system consists of a transmitter and a receiver as shown in Fig. 1. An OFDM transmitter modulates the data to be transmitted and data is then converted from serial to parallel N sub-carriers. Inverse Fast Fourier Transform (IFFT) is performed on each sub-carrier data to convert it to time domain and a cyclic prefix is added. This N dimensional sub-carrier data is converted to analog data using Digital to Analog Converter (DAC). Further, the analog data is modulated using sine-cosine signal, and then a guard band is added to the combined signal and finally it is transmitted using the transmit antenna. At OFDM receiver end, the received analog data is demodulated using sine-cosine signal and passed through a Low Pass Filter (LPF). To convert the signal back to the frequency domain, first the analog data is converted into digital and the cyclic prefix is removed. Later, a N point Fast Fourier Transform (FFT) is performed on the data. Finally, to retrieve the information, data is converted into serial for demodulation.

C. Channel State Information

CSI in a wireless network represents the channel properties of the communication link between a transmitter and a receiver. This not only provides the insights on how the signal travels from the transmitter to the receiver, but also provides the joint effect of scattering, fading and power decay with the distance. The information is calculated at receiver side and called as CSIR (CSI at the receiver). This received signal is then quantized and fed back to the transmitter. The channel estimation at the transmitter is called at CSIT (CSI at the transmitter). The wireless LAN (WLAN) interfaces, e.g. IEEE 802.11a/g/n use OFDM technique to transmit the signals using orthogonal sub-carriers at distinct frequencies.

In narrow-band flat fading channel, a MIMO system is represented as

$$y_i = Hx_i + N_i \tag{3}$$

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

To estimate the channel matrix H, a known training sequence also called as the pilot sequence is transmitted and channel response H is measured at the receiver side. Considering the pilot sequence is $x_1, x_2, \ldots, x_n$, then the combined received signal y_i over $i = 1, 2, \ldots, n$ is given by

$$Y = [y_1, y_2, \ldots, y_n] = HX + N \tag{4}$$

Therefore, channel matrix H can be estimated by

$$\hat{H} = \frac{Y}{X} \tag{5}$$

and it represents the fine grained PHY layer information over multiple sub-carriers with $p \times q \times s$ dimensions, where p and q are the number of transmit and receive antennas respectively and s denotes the total sub-carriers. The channel matrix H is represented as

$$H = [H_1, H_2, \ldots, H_N]^T \tag{6}$$

where

$$H_i = |H_i| \, e^{jsin(\angle H_i)} \tag{7}$$

For MIMO systems, H_i have $p \times q$ dimensions and is represented as

$$H_i = \begin{bmatrix} h_{11} & h_{12} & h_{13} \ldots h_{1q} \\ h_{21} & h_{22} & h_{23} \ldots h_{2q} \\ \ldots & \ldots & \ldots \quad \ldots \\ h_{p1} & h_{p2} & h_{p3} \ldots h_{pq} \end{bmatrix} \tag{8}$$

where h_{pq} are complex numbers representing the amplitude and phase of each sub-carrier for an antenna stream. The total number of spatial streams equals

$$rank(H) \leq min(p, q) \tag{9}$$

Thus, for each group of sub-carrier, the channel matrix is of dimension $p \times q$. In IEEE 802.11 a/g/n, there are 48 data sub-carriers. The Linux driver for Intel 5300 card provides CSI for 30 sub-carriers[14].

IV. CSI-MIMO System & Methodology

In this section, we briefly explain the CSI-MIMO training/fingerprinting phase and testing/positioning phase. We also describe our experimental testbed, devices used for data collection, the methodology to generate the CSI-MIMO location signature and finally the positioning algorithms used for location estimation.

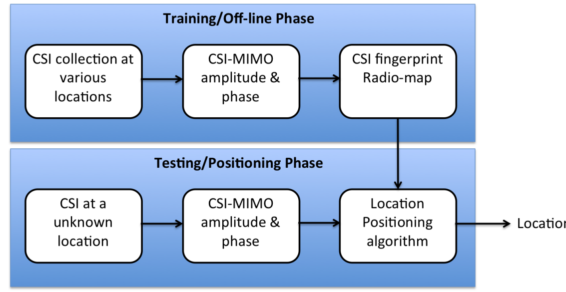

Fig. 2. CSI-MIMO system

A. Training/Fingerprinting Phase

Fig. 2 shows the CSI-MIMO System architecture. During the training phase, a mobile device collects CSI data using off-the-shelf Intel 5300 hardware. Data packet received from an AP is basically the raw CSI data over multiple sub-carriers (30 out of 48 data carriers) for number of transmit and receive antennas of an AP and a mobile device respectively. We generate a unique fingerprint based on amplitude and phase of each sub-carrier. Since, each sub-carrier fades differently, we subtract the amplitude and phase information of subsequent sub-carriers that represent a CSI-MIMO fingerprint at a location. We war-drive the entire testbed to collect the CSI at various locations and generate the CSI-MIMO fingerprint and store it in a CSI-MIMO fingerprinting database.

B. Testing/Positioning Phase

For Positioning, CSI data at an unknown location is collected and it is processed using the method of training phase to create a location fingerprint. It is compared against the stored fingerprints in the radio map to estimate the unknown location. We utilize both amplitude and phase information of CSI-MIMO over all 30 sub-carriers for localization. In this paper, we have used both the deterministic (k-nearest neighbor) and probabilistic (maximum likelihood estimation) algorithms to compare the results with FIFS.

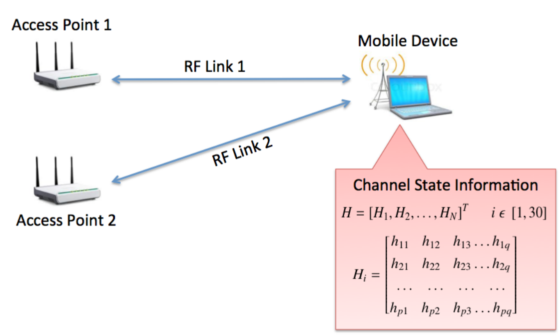

Fig. 3. CSI experiment setup

C. Experiment Setup

In our experiment, the CSI data is collected using the Linux 802.11n tool [14] installed on a mobile device. Two Wireless

APs (TP-LINK TL-WA901ND) are mounted at a fixed location as transmitters. The CSI data collection was carried out using a Dell latitude E4300 and HP 8530p laptops. Both the laptops are equipped with Intel 5300 NIC.

Fig. 4 shows the floor-plan of our testbed, which is basically a $6m \times 7.5m$ lab environment. We have collected data at 19 different locations. The distance between various reference locations varies from 1.5 m to 2 m. The test area was occupied by 8 people performing usual activities sitting and moving in and out frequently. Data measurements were carried out during weekdays. During data collection, at each location, a user was present continuously during working hours.

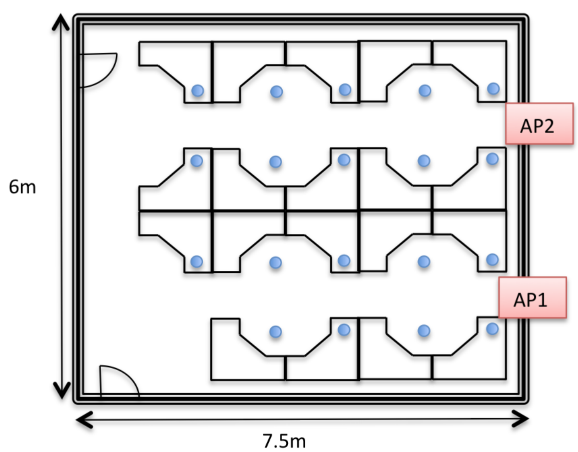

Fig. 4. Floor-plan of the testbed

At each location, we placed one of the laptops and collected the samples for at least 24 hours. Each sample represents the CSI for 30 sub-carriers over the total number of transmit and receive antennas. For each location, we carried out measurements for all the APs. Even though we have used Intel 5300 NIC laptops, the antenna position of each device is different. In such scenario, we ensure that the measurements at a location for all the APs are done using the same mobile device. However to analyze the effect of hardware, we have collected the test data using both the laptops.

D. CSI-MIMO Fingerprint

Each location is represented by multidimensional CSI data consisting of $p \times q$ for MIMO (for p transmit and q receive antennas) and s is the third dimension for 30 sub-carriers. CSI at each location is collected at least for 24 hours, which provides approximately 100,000 samples. Location fingerprint is basically an amplitude difference vector and a phase difference vector of size 1×29. Equation 6 represents a CSI sample collected at a location. For p transmit and q receive antennas, the CSI value for each sub-carrier is represented by Eqn. 8. The steps to generate a fingerprint from the collected data is as follows:

1) For each sample, the CSI value of each sub-carrier is aggregated over MIMO and the resulting aggregated CSI per sub-carrier reduces to the vector of dimension 1×30 and can be represented as

$$H_{avg} = \sum_{m=1}^{p} \sum_{n=1}^{q} h_{mn} \tag{10}$$

2) Next, we calculate the amplitude and phase from the aggregated CSI which is a complex vector. We unwrap the phase information to ensure that the multiples of 2π are included in the phase response.

$$H_{avg-amp} = |H_{avg}| \tag{11}$$

$$H_{avg-ph} = \angle H_{avg} \tag{12}$$

3) Amplitude and phase value is then subtracted for subsequent sub-carriers, thus reducing the fingerprint size to 1×29.

$$H_{amp_i} = |H_{avg-amp_i}| - |H_{avg-amp_i+1}| \tag{13}$$

$$H_{ph_i} = \angle H_{avg-ph_i} - \angle H_{avg-ph_i+1} \tag{14}$$

4) The amplitude and the phase difference is averaged over N samples to get amplitude and phase difference mean as the location signatures.

$$H_{amp-diff} = \sum_{i=1}^{N} H_{amp_i} \tag{15}$$

$$H_{ph-diff} = \sum_{i=1}^{N} H_{ph_i} \tag{16}$$

5) Finally, we create one signature by using both amplitude difference mean and phase (unwrap) difference mean at a location as shown in Eqn. 17.

$$H_{diff} = [H_{amp-diff} \;\; H_{ph-diff}] \tag{17}$$

Using the war-driving data, we generate a radio-map that stores the fingerprints of all the locations for a given floor-plan.

E. CSI-MIMO-based Localization

For positioning of a mobile device, we have used both the deterministic kNN and the probabilistic Bayes rule. We have assumed Gaussian distribution of the data.

1) Deterministic k-Nearest Neighbor: For the deterministic kNN algorithm[1], we select k to be 1 and calculate the minimum Euclidean distance between each location of the radio-map and a test fingerprint using the following equation.

$$E_{dist} = \sqrt{\sum^{K} (H_{train} - H_{test})^2} \tag{18}$$

2) Probabilistic Bayes rule: To estimate the location, we adapt probabilistic model [10] using amplitude and phase difference signatures as shown in Eqn. 17. Let H_{test} be the test fingerprint to be located for K access points and H_{train} is one of the fingerprint from the radio-map. The the location estimation problem is to find the location l that maximizes the posterior probability $P\left(l_i|H_{test}\right)$.

According to Bayes rule,

$$P\left(l_i|H_{test}\right) = \frac{P\left(l_i\right)P\left(H_{test}|l_i\right)}{\sum_1^i P\left(l_i\right)P\left(H_{test}|l_i\right)} = \frac{P\left(l_i\right)P\left(H_{test}|l_i\right)}{P\left(H_{test}\right)} \tag{19}$$

To calculate the prior probability, we use Pearson correlation (similar to FIFS) to find the correlation between the test fingerprint with that of training fingerprints as shown in Eqn. 20.

$$\rho_{H_{test}H_{train}} = \prod_{k=1}^{K} \frac{cov\left(H_{test}^k, H_{train}^k\right)}{\sigma_{H_{test}^k}\sigma_{H_{train}^k}} \tag{20}$$

Higher the correlation coefficient, more likely the test location is close to the training location. The prior probability of test on each training point is given by Eqn. 21.

$$P\left(l_i\right) = \frac{\rho_{H_{test}H_{train}}}{\sum\limits_{i=1}^{L} \rho_{H_{test}H_{train}}} \tag{21}$$

where L is total training locations in the radio-map.

The likelihood of $P\left(H_{test}|l_i\right)$ is evaluated as

$$P\left(H_{test}|l_i\right) = \prod_{k=1}^{K} P\left(H_{test}|l_i\right) \tag{22}$$

We finally estimate the mean and variance of training data and calculate the posterior probability using Prior probability $P\left(l_i\right)$ and likelihood $P\left(H_{test}|l_i\right)$. The unknown location is estimated from the posterior.

$$\hat{l} = \max \sum_{i}^{L} P\left(l_i|H_{test}\right) \tag{23}$$

V. Experiment Results

In this section, we evaluate the performance of the CSI-MIMO for various training and test samples. We compare our results with FIFS and the basic CSI signature, which comprises of amplitude and phase information over all the sub-carriers as represented by Eqn. 6. Furthermore, the performance of CSI-MIMO is analyzed using kNN and Bayes algorithm adapted similarly in FIFS.

In our experiment, we have sampled training data over 24 hours and collected approximately 100000 samples at a location. For positioning phase, we have collected 1000 samples at an unknown location using different mobile devices. We assess the CSI-MIMO, based on the impact of war-driving data, the device used for the positioning phase and type of localization algorithm.

A. War-driving Data and Device for Positioning

The accuracy of CSI-MIMO for various combinations of training and test samples is evaluated in this section. We have considered 100, 1000, 10000 training samples, and 100 and 1000 test samples. We estimate the number of training and test data required for accurate localization empirically. Set 1 and 2 are the test sets with the same training and test device, whereas Set 3 uses a different mobile device in training and test phase.

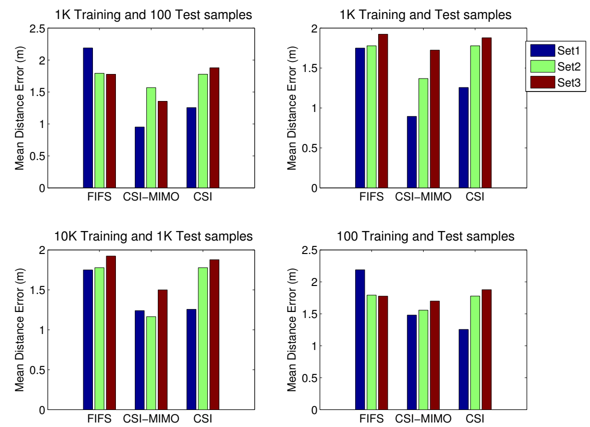

Fig. 5. Mean distance error for various test-sets using kNN.

The mean distance error for various test sets using the kNN algorithm is shown in Fig. 5. The minimum distance error of 0.95 m is achieved for 1K training and 100 test samples. The distance error of proposed CSI-MIMO system is mostly less as compared to FIFS and the simple CSI for all the test sets.

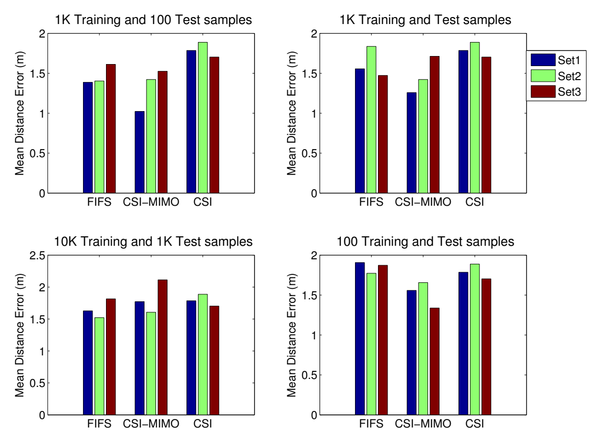

Fig. 6. Mean distance error for various test-sets using Bayes.

The mean distance error of various test sets using Bayes rule is shown in Fig. 6. The maximum accuracy achieved using Bayes rule for 1K training and 100 test samples is 1.02 m, which is slightly lower than kNN.

Our proposed CSI-MIMO provides better accuracy mostly for all the samples combinations using kNN and Bayes. However, FIFS performs better than proposed CSI-MIMO using Bayes algorithm for 10K/1K samples. FIFS uses summation of the power in all the sub-carriers as a location fingerprint. Over large number of samples, the statistics of the samples such as mean and variance provides better probability and thus resulting in more accurate localization.

Algorithm	Train/ Test Samples	FIFS	CSI-MIMO	CSI	Performace Gain over FIFS (set1)
kNN	1K / 100	2.1886	0.9520	1.2563	57%
	1K / 1K	1.7497	0.8947	1.2563	49%
	10K / 1K	1.7497	1.2401	1.2563	29%
	100 / 100	2.1886	1.4804	1.2563	32%
Bayes	1K / 100	1.3869	1.0215	1.7855	26%
	1K / 1K	1.5557	1.2583	1.7855	19%
	10K / 1K	1.6278	1.7711	1.7855	-8%
	100 / 100	1.9058	1.5584	1.7855	18%

TABLE I
MEAN DISTANCE ERROR IN METERS FOR TEST SET 1 USING KNN AND BAYES

Table I presents the mean distance error for various training and test samples using kNN and Bayes algorithm. The maximum performance gain of the proposed CSI-MIMO system over FIFS is 57% and 26% for kNN and Bayes respectively for 1K training and 100 test samples.

Two inferences can be drawn from the above results. Firstly, the accuracy achieved for 1000 training and 100 test samples is higher using both the algorithms. Thus we can conclude that the war driving of 1000 training and 100 test is optimal. Secondly, the minimum distance error of all the test sets for the optimum samples is always lower as compared to other CSI signatures.

B. Type of algorithm

In this section, we try to analyze the effect of different type of algorithms such as deterministic kNN and probabilistic Bayes rule on the localization accuracy.

The mean distance error for kNN and Bayes using various training and test samples is shown in Fig. 7. We have used test set 1 and 2 for further evaluation.

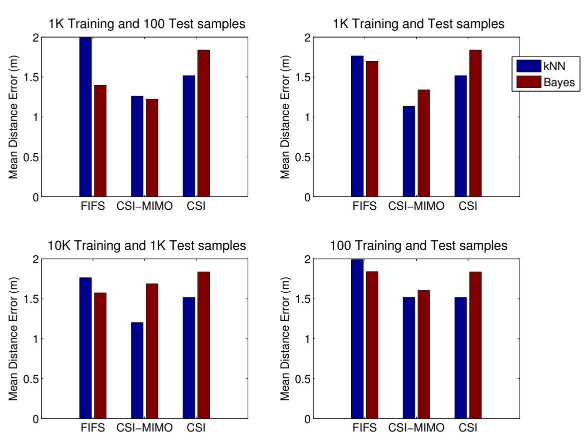

Fig. 7. Mean distance error using same device for kNN and Bayes.

It can be observed that, CSI-MIMO works comparably for both the algorithms. The most prominent feature of CSI-MIMO is that it achieves almost equal accuracy for both deterministic and probabilistic algorithms using optimum war driving.

Table II shows the mean distance error for kNN and Bayes rule. A 37% performance gain is achieved using kNN for 1k training and 100 test samples, whereas Bayes provides 20% gain for 1K training and test samples. The overall minimum distance error of 1.13 m is achieved using kNN for 1K training and test samples. From above results, we can infer that the deterministic kNN provides better accuracy and performs better than FIFS.

Algorithm	Train/ Test Samples	FIFS	CSI-MIMO	CSI	Performace Gain over FIFS
kNN	1K / 100	1.99	1.25	1.51	37 %
	1K / 1K	1.76	1.13	1.51	36 %
	10K / 1K	1.76	1.2	1.51	32 %
	100 / 100	1.99	1.519	1.517	24 %
Bayes	1K / 100	1.39	1.22	1.83	12 %
	1K / 1K	1.69	1.34	1.83	20 %
	10K / 1K	1.57	1.68	1.83	-7 %
	100 / 100	1.83	1.60	1.83	12 %

TABLE II
MEAN DISTANCE ERROR IN METERS FOR TEST SET 1 AND 2 USING KNN AND BAYES.

The CSI-MIMO achieves better performance over FIFS and simple CSI. Our proposed system CSI-MIMO uses all the 30 sub-carriers and thus represents the location at sub-carrier level. It uses a complex signature including amplitude and phase of all the 30 sub-carriers whereas FIFS uses summation of power over all the sub-carriers. This enables CSI-MIMO to localize more accurately.

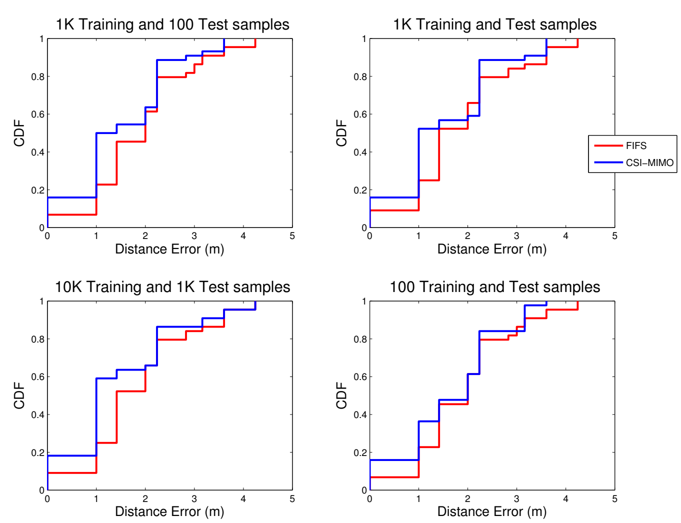

Fig. 8. CDF of FIFS and CSI-MIMO using kNN.

The cumulative distribution function (CDF) of localization error for various training and test samples using kNN and Bayes algorithm is shown in Figures 8 and 9.

It can be seen that, for optimum war driving 90% error probability of CSI-MIMO is 2.7 m compared to FIFS with 3.2 m using kNN. Thus providing an improvement of 16% . The median accuracy achieved by CSI-MIMO is 1.5 m, which is better than 2.2 m of FIFS.

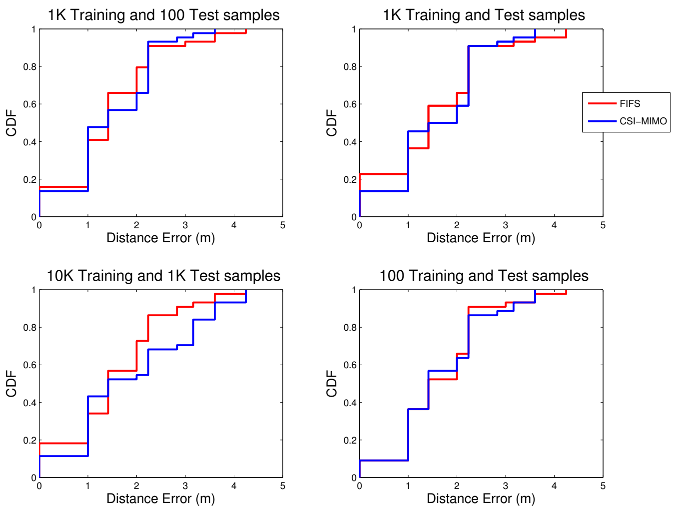

Fig. 9. CDF of FIFS and CSI-MIMO using Bayes.

From above results, it can be concluded that CSI-MIMO achieves better performance compared to FIFS with maximum accuracy using the optimum samples. Furthermore, the impact of the device used for localization and type of algorithm is not significant. Our proposed system is more robust than other CSI-based system and can uniquely represent the location using frequency and space diversity of CSI.

VI. Conclusion

In this paper, the PHY layer information extracted using off-the shelf Intel 5300 card, is utilized to Wi-Fi fingerprinting. A novel location fingerprint using frequency and space diversity of channel state information can be effectively used to represent the location. The sub-carriers representing frequency diversity and MIMO signifying the spatial diversity provide uniqueness to distinguish locations in indoor positioning system. Using both deterministic and probabilistic methods, we evaluate the number of training and test samples to provide better accuracy and minimal war-driving. Our experimental results show that, 57% gain is obtained over state-of-the art FIFS using CSI-MIMO for kNN algorithm. Also, accuracy of 0.95 m is achieved using simple data aggregation over MIMO with optimal war driving.

In this paper, we have aggregated the MIMO data for each sub-carrier, however, it will be challenging to use the individual antenna streams for localization and analyze its complexity in terms of computations. The impact of the dynamic environment including the presence and mobility of people will be interesting to investigate in future.

References

[1] P. Bahl and 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. IEEE*, vol. 2, 2000, pp. 775–784.

[2] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *MobiSys'05: Proceedings of the 3rd international conference on Mobile systems, applications and services*. USA: ACM, 2005, pp. 205–218. [Online]. Available: http://doi.acm.org/10.1145/1067170.1067193

[3] K. Kaemarungsi and P. Krishnamurthy, "Analysis of WLANs received signal strength indication for indoor location fingerprinting," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 292–316, 2012, special Issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119211001234

[4] Y. Chapre, P. Mohapatra, S. Jha, and 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)*, Oct 2013, pp. 304–307.

[5] A. Mahtab Hossain, Y. Jin, W.-S. Soh, and H. N. Van, "SSD: A robust RF location fingerprint addressing mobile devices' heterogeneity," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 65–77, 2013.

[6] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *Proceedings of the ACM SIGCOMM 2010 conference*, ser. SIGCOMM '10. New York, NY, USA: ACM, 2010, pp. 159–170. [Online]. Available: http://doi.acm.org/10.1145/1851182.1851203

[7] D. A. Hall, "Understanding benefits of MIMO technology," *Microwaves and RF*, 2009.

[8] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. Ni, "CSI-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300–1309, 2013.

[9] S. Sen, R. R. Choudhury, B. Radunovic, and T. Minka, "Precise indoor localization using PHY layer information," in *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, ser. HotNets-X. New York, NY, USA: ACM, 2011, pp. 18:1–18:6. [Online]. Available: http://doi.acm.org/10.1145/2070562.2070580

[10] J. Xiao, K. Wu, Y. Yi, and L. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *21st International Conference on Computer Communications and Networks (ICCCN)*, 2012, pp. 1–7.

[11] J.Hightower, R.Want, and 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.*, Feb 2000.

[12] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wirel. Netw.*, vol. 10, no. 6, pp. 701–710, Nov. 2004. [Online]. Available: http://dx.doi.org/10.1023/B:WINE.0000044029.06344.dd

[13] C. Nerguizian, C. Despins, and S. Affes, "Geolocation in mines with an impulse response fingerprinting technique and neural networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 603–611, March 2006.

[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, pp. 53–53, Jan. 2011. [Online]. Available: http://doi.acm.org/10.1145/1925861.1925870