國立臺灣大學電機資訊學院電機工程學研究所

碩士論文

Graduate Institute of Electrical Engineering College of Electrical Engineering and Computer Science National Taiwan University Master Thesis

利用 Wi-Fi CSI 做精細室內定位

Exploiting Wi-Fi CSI for Fine-Grained Indoor localization

蘇揚鈞

Yang-Chun Su

指導教授:黃寶儀 博士

Advisor: Polly Huang, Ph.D.

中華民國 102 年7月

July, 2013

國立臺灣大學碩士學位論文

口試委員會審定書

利用 Wi-Fi CSI 做精細室內定位

Exploiting Wi-Fi CSI for Fine-Grained Indoor localization

本論文係蘇揚鈞君(學號 R00921050)在國立臺灣大學電機工程 學系完成之碩士學位論文,於民國 102 年 07 月 02 日承下列考試委員 審查通過及口試及格,特此證明。

口試委員:



系主任 (簽名)



誌謝

能夠順利完成這篇論文,首先要感謝Polly,在我在NSLab的這段期間裡, 給我機會,讓我在錯誤中學習做事的方法,給我機會,讓我學習在面對研究 上的困難時不迷失方向,學會獨立思考並解決問題,最重要的是在我陷入盲 點而不自覺的時候,提醒我哪些地方出現問題而能回到正軌。

特別感謝 Junction 在我研究遇到瓶頸時,給予我適當的建議和可能的思考方向,讓我能夠跳脫原有的思考框架,進而找出解決的辦法。此外,也特別感謝 Tsung-Yun、Yu-Ting、Meng-Lin 在一同奮鬥的這段期間,除了程式撰寫的概念外,也不吝提供各方面的幫助。

最後,感謝我的家人、朋友、以及所有 NSLab 的學長姊們,在我感到無助和難過的時候,給予我鼓勵和支持。

2



摘要

最近,利用 Wi-Fi 做室內定位的技術變得更加引人關注,因為利用廣泛被佈 置的 Wi-Fi 系統可以減少系統在硬體建置的負荷,此外,有先前的研究驗證使用 Wi-Fi 在正交分頻調變(Orthogonal frequency-division multiplexing, OFDM) 運作下 的精密評估資訊,也就是通道狀態訊息(Channel State Information, CSI),用以作為 位置指紋比傳統以接收訊號強度(RSSI)更具代表性。這篇論文中,將會分享關於利 用學校已建置的 Wi-Fi 系統去實作一個以通道狀態訊息為基礎的定位系統。

我們的系統包含了「指紋資料庫」和「位置評估系統」兩個部分。因為在我們的測試環境中可以明顯觀察到多個不同的通道狀態訊群,所以我們利用 K-means 演算法分群,並保留每個探勘點的多個指紋在「指紋資料庫」中。

因為精密的通道狀態訊息是以高維度的向量呈現,所以我們利用一種統計模

型R平方數 (R-square value) 來做指紋比對。除了單一通道狀態訊息封包的比對測 試方法之外,文中也提供一個可以達到更高定位精準度的多通道狀態訊息封包的 比對測試方法。同時為了避免受到易受位置影響的通道狀態訊息而導致的誤判,亦 提供了一個可行的權重投票估計機制。最後我們顯示系統的評估結果,也可以看到 我們的系統表現比傳統利用接收訊號強度實作的系統突出。

關鍵字 — Wi-Fi, 精密室內定位, 通道狀態訊息



Abstract

Nowadays, Wi-Fi-based indoor localization techniques have become attractive, because widely deployed Wi-Fi system could reduce the overhead of infrastructure. Moreover, some prior works argue that Wi-Fi OFDM-based fine-grained estimation data, Channel State Information (CSI), is more representative than traditional RSSI as location fingerprints. In this paper, we shared the experience of utilizing the school built-in Wi-Fi system to build a CSI-based localization system.

Our system includes "Fingerprint Database" and "Localization System." Due to multiple obvious CSI clusters could be observed in our testbed, we utilize K-means algorithm to retain multiple fingerprints for each survey point in our "Fingerprint Database". Because fine-grained CSI are high-dimension vectors, a statistical module (i.e., R-square value) is proposed for fingerprint comparison. Not only Single-CSI comparison, but also a Multiple-CSI comparison testing method is also proposed, which reaches higher accuracy. To reduce the location misjudgment caused by location sensitive CSI, a feasible weighted voting estimation process is also proposed. Finally, we evaluate our system in our testbed and show our system outperforms traditional RSSI-based localization system.

臺

Key words — Wi-Fi, Fine grained Indoor Localization, Channel State Information



Contents

口試委員審定書1
誌謝2
摘要3
ABSTRACT
CONTENTS7
LIST OF FIGURES
LIST OF TABLES

СНАРТ	ER 1 INTRODUCTION	15
СНАРТ	TER 2 RELATED WORK	
СНАРТ	TER 3 CHANNEL STATE INFORMATION	25
3.1	Preliminaries	25
3.1.1	OFDM Technology	25
3.1.2	Frequency-Selective Fading	26
3.1.3	Channel State Information Grouping	27
3.2	CSI Observation	28
3.2.1	CSI Vector	
3.2.2	R-Square Value	29
3.2.3	Stability Experiment	
СНАРТ	ER 4 FINGERPRINTING DATABASE	
4.1	Fingerprint Database Generation	
4.1.1	Cluster Strategy Decision	

4.1.2 CHAP	2 Fingerprint Database PTER 5 IMPLEMENTATION	.39 41	
5.1	Channel State Information Collection	.41	
5.1.1	1 Hardware Setup	.41	
5.1.2	2 Software Setup	.42	
5.2	Infrastructure	.43	
5.3	Localization Algorithm	.44	
5.3.1	1 Clustering & Data Processing Block	.45	
5.3.2	2 Fingerprint Comparison Block	.46	
5.3.3	3 Weighted Voting Estimation	.49	
CHAPTER 6 EVALUATION			
6.1	Experimental Scenario	.51	
6.1.1	1 Experiment Environment	.52	
6.1.2	2 Experiment Methods	.54	
6.1.3	3 Evaluation Matrix	.55	

6.2	Performance Evaluation	* * * *
0.2		C C C C C C C C C C C C C C C C C C C
6.2.1	1 Self-Testing	
6.2.2	2 Daily Traces Localization Evaluation	
6.2.3	3 Weighted Voting Estimation Evaluation	61
6.3	RSSI-based Localization System Comparison	62
6.3.1	1 RSSI-based Localization Mechanism	62
6.3.2	2 Performance Comparison	63
6.4	Comparison between our system and FIFS	64
6.4.1	1 FIFS System vs. our system	64
6.4.2	2 FIFS System Evaluation	65
CHAP	PTER 7 CONCLUSION & DISCUSSION	
REFEF	RENCE	



List of Figures

Figure 3-1 Frequency-Selective Fading could detect by subcarriers27
Figure 3-2 Stability Experiment Result
Figure 3-3 (a) CSI vector plot, shifting left (-)/right (+) from reference location. (b) CSI
comparison result of (a). (c) CSI comparison result, moving toward
(-)/backward (+) from reference location
Figure 4-1 Multiple patterns of CSI vectors could be observed (a) Collect by laptop
antenna 1 (b) Collect by laptop antenna 1 and 2
Figure 4-2 Cluster Number Testing Algorithm
Figure 4-3 Possible Cluster number
Figure 4-4 Data format of fingerprint database40

Figure 5-1 Intel Wi-Fi Wireless Link 5300 NIC & Lenovo ThinkPad X200
Figure 5-2 AP Location Map of 6 th floor in BL Hall, Mark by green star
Figure 5-3 System Architecture
Figure 5-4 Single-CSI Comparison Method
Figure 5-5 Cross-examination of Multiple-CSI Comparison Method. The larger
accumulated R-square value combination (i.e., Red combination $V1_{11}+V1_{22}$
and Blue combination $V1_{12}+V1_{21}$) would retain in vector
Figure 6-1 Line-of sight Corridor Section (Scenario 1)
Figure 6-2 Non-line-of-sight L-Shape Corridor section (Scenario 2)
Figure 6-3 Experiment Setting
Figure 6-4 Self-testing Result
Figure 6-5 Single-CSI Localization Result (Scenario 1)
Figure 6-6 Multiple-CSI Localization Result (Scenario 1)
Figure 6-7 Single-CSI Localization Result (Scenario 2)
Figure 6-8 Multiple-CSI Localization Result (Scenario 2)60
Figure 6-9 Efficiency of Weighted Voting Process (a) Single-CSI (b) Multiple-CSI62
Figure 6-10 Localization Performance RSSI vs. CSI





List of Tables

Table 3-1 Inte	l Wi-Fi Link 5300 NIC.	Grouping Mode Ng=2	2



Chapter 1 Introduction

Precise indoor localization is a necessary technique for many applications in daily life, such as personalized advertisements [17], health caring [18][19] and emergency detection [20]. There are many indoor localization systems have been proposed. However, pre-deployed specific infrastructure or expensive hardware devices are needed for many early systems.

To reduce the infrastructure overhead, Wi-Fi-based fingerprinting technique is one of main approaches through using existing deployment. Wi-Fi system is one ubiquitous system that deployed in indoor environment, due to its low-cost and open access properties. Therefore, Wi-Fi-based RSSI fingerprinting systems have become more attractive. However, RSSI is coarse measurement of the received signal power, that can't provide other environment information for localization. Moreover, detected RSSI could be similar in separated location and varying largely at the same location. Hence, large localization error is hard to degrade.

Recently, some works argue that utilizing an OFDM-based fine-grained estimation result (i.e., Channel State Information) is a better way and with potential to achieve higher accuracy of indoor localization. Channel State Information (i.e., CSI) is reported from PHY layer of NIC (i.e. Wireless network card), which represents the channel response properties of communication link over all different frequency subcarriers. Because frequency-selective fading effect could be represent by high-dimension CSI vectors, using CIS vectors as fingerprints of location is more representative than RSSI. We would go through preliminaries of OFDM and CSI in section 3.1.

In this paper, we share the experience of utilizing the built-in Wi-Fi system in 6th floor of Barry Lam (BL) Hall in National Taiwan University to build a CSI-based fingerprinting localization system. This system could be mainly divided into two main parts, one is "Fingerprint Database" and the other one is "Localization System". The mechanism of "Localization System" could further divide into two main processing

blocks, one is "Fingerprint Comparison" block and the other one is "Weighted Voting Estimation" block.

Our fingerprint database retains multi-fingerprints for each survey point due to multiple CSI clusters could be usually observed in our testbed. In this work, a testing algorithm is proposed to decide a proper clustering strategy for utilizing K-means clustering algorithm. Base on the observation of testing result, statistical mean vectors of two main clusters at each survey point would be used as fingerprints. The detail of fingerprint database generation would be discussed in section 4.1.

Because fine-grained CSI vectors are high-dimension, we proposed a statistical module (i.e., R-square value) for fingerprint comparison. The R-square value would represent the level of the observed outcome (i.e., testing CSI) replicated the model (i.e., fingerprint vector). In this work, we proposed not only a Single-CSI comparison testing method, but also a Multiple-CSI comparison testing method, which reaches higher localization accuracy.

Through the observation of prior experiments of stability, we find that CSI vectors are stable enough at the same location for long time testing. However, CSI could vary largely, even only slight offset from survey point center. Hence, the weighted voting estimation process is needed for our localization system. We utilize multiple APs in our testbed to select nearest neighbors as candidates and calculate the statistical R-square value for final judgment. About the implementation detail of our whole system would be described in Chapter 5.

In Chapter 6, we evaluate our system in three main parts. The first part is about self-testing to show the potential of CSI for localization. The second part is about the performance of our proposed localization algorithm based on three daily traces to show the stability of our system. The third part is about the necessity of weighted voting estimation process. To study the benefit of using CSI, we also compare the performance of our system with a common RSSI-based localization system.

Finally, conclusion and discussion would be presented in Chapter 7.



Chapter 2 Related Work

There are large amount of research that investigate and implement indoor localization system during this two decades. Many early localization systems are based on special infrastructure and hardware, such as Cricket Location Support System [1]. This system uses ultrasonic to get time-of-flight measurements of each location. These measurements could be used to provide high accuracy result, but pre-deployed specific infrastructure is required. Moreover, these devices cost are too expensive for general users. To meet certain constraints, LANDMARC [2] selected RFID tags and readers to implement their system. However, it requires densely deployment yet restricts expandability. To reduce the overhead of infrastructure, RSSI (Received signal strength indicator) fingerprinting localization using existing deployment (i.e. access point) is one of main approaches widely adopted in this topic.

RADAR [3] is a well-known work, which divided localization into two phases. In the training phase, it collects RSSI vectors as fingerprints to build a radio map. In the tracking phase, it uses the received RSSI vector as a signature to compare with location fingerprints in radio map. Then it introduces the k-NN algorithm to help to estimate location of the target.

Horus [4] is another popular RSSI-based localization system that also includes the offline phase and online phase. It uses Bayes' theorem and defines a clustering module to achieve higher accuracy in estimated results. In addition to this, many other mathematical models and theory [12][13][14][15] have been applied to RSSI-based localization system and achieve room level or meter level accuracy.

Although those available approaches could provide enough location accuracy for specific applications, they still do not satisfy many other practical applications. Furthermore, performance of RSSI fingerprinting system still affected by density of existing deployments [10] and bounded by the nature restriction [11].

Recently, one toolkit [5], based on Linux operating system, and used to record Channel State Information (CSI) from Intel Wi-Fi Link 5300 wireless NIC, has been released. The CSI packet not only involves simple RSSI data, but also contains detail channel response properties at the level of OFDM (Orthogonal frequency-division multiplexing) subcarriers. Based on CSI, many advanced technique issues could be evaluated, such as the impact of channel fading, multi-user MIMO effect and external interference. These available environment features could be used and with potential for achieving higher accuracy in localization.

There is a brand new localization system, called PinLoc [7], using statistics fine-grained CSI of mean and variance to create fingerprints database. Because their system focus on spot localization, they compute the correlation value between CSI tracing data and each location fingerprint to decide if target device is in specific spot area or not.

In addition, they also verified that CSI vector could preserve a statistical structure over time, CSI vary in enough small granularity, and be different in each location with high probability.

Hence, inspired by their work and the potential benefit of using CSI for

localization, we implement an offline CSI-based fingerprinting localization system for exploring. It uses school building built-in Wi-Fi system to reduce the overhead of infrastructure. Unlike PinLoc system uses a Roomba-mounted laptop to records multiple CSI fingerprints at different location in each spot area through war driving process, we only statistics CSI vectors at each survey point. During tracking phase, our system uses CSI vectors from each AP to point out nearest neighbor candidate locations and through weighted voting estimation process to given an estimation result. The whole estimate algorithm expands our system potential for target tracking and further more applications.

There are another two new localization systems based on fine-grained CSI, called FILA [8] and FIFS [9]. Unlike PinLoc system, utilizing CSI vectors naively, they based on their observation of subcarrier frequency-selective fading feature to transfer CSI raw data to effective CSI value for localization.

FILA localization system could be divided into three major divisions, (1) CSI processing, (2) Calibration and (3) Location Determination. In the first division, because Fine-grained CSI could help provide information of frequency-selective fading in the environment, the authors introduced a methodology to compute the effective CSI

value from the fine-grained CSI as an environment signature. In the second part, they introduced a propagation model with two environment parameters, which is used to transfer the effective CSI value to distance. This calibrated distance would be used for location determination part.

However, distance model accuracy highly depends on environment parameters. Distinguishing different survey points with same distance from AP could also be another problem. Although, we also utilize multiple APs for localization in our work, our system use weighted voting estimation process as a substitution.

FIFS is an expanded work from FILA. Because Intel 5300 wireless NIC could support multiple input multiple output (MIMO) mechanism, this system utilized the fine-grained CSI of multiple antenna pairs of access point and NIC. Their system proposed an effective CSI vector transformation model and introduces a probability model for position estimating. They also compare their result with the famous RSSI-based fingerprinting work, Horus. It could be clearly seen that FIFS outperforms Horus. Although FIFS works under SISO mechanism, it still reaches higher accuracy than Horus.

Our work focus on MIMO mechanism for localization, due to multiple CSI vector

clusters could be usually found at one location in our test bed. Not like FIFS creates each fingerprint by averaging CSI over different antennas, our system uses K-means clustering algorithm to find main CSI vector clusters at each location to create multiple fingerprints database. Finally, we design an algorithm for localization and evaluate our system in Chapter 6.

臺



Chapter 3 Channel State Information

3.1 Preliminaries

The Channel State Information (CSI) is an estimation report that represents the channel response properties of transmission link. A complex CSI vector would be used to illustrate the channel properties of each antenna combination and each complex value (i.e. CSI value) is a report of an independent subcarrier. This subcarrier level measurement is based on OFDM technique.

3.1.1 OFDM Technology

Orthogonal frequency-division multiplexing (OFDM) technology is widely used in

IEEE802.11 a/g/n for more effectively transmit data. In OFDM system, a wideband channel is divided into multiple narrowband sub-channels, which are named subcarriers. Subcarriers are spaced at 3.125 KHz intervals and used an Inverse Fast Fourier Transform (i.e., IFFT) module to generate composition waveform for transmission. The frequency domain model of each subcarrier could be represented by:

$$Y(f_i) = H(f_i) X(f_i) + N, \quad i = 1, ..., n$$

 f_i represents the frequency of the i_{th} subcarrier and *i* stands for different index of subcarrier. $Y(f_i)$ and $X(f_i)$ are received and transmitted signal. *H* is a complex value vector, that contributes by channel state information value of all subcarrier channels (i.e., $H(f_i)$). *N* is the environment noise.

3.1.2 Frequency-Selective Fading

Fading is a phenomenon of attenuation over signal propagation. The phenomena are often frequency-selective in multipath environment. Coarse measurement RSSI could only illustrate average fading result. On the other hand, the fine-grained CSI could detect this feature by subcarriers. Therefore, the CSI value of each subcarrier could be very different, although they are collected at the same place. We transfer complex CSI vector to CSI magnitude vector to illustrate the fading feature. The selective fading phenomenon could be shown by Figure 3-1[16]. Signal attenuates differently in different frequency band.



Figure 3-1 Frequency-Selective Fading could detect by subcarriers

3.1.3 Channel State Information Grouping

Intel Wi-Fi Link 5300 wireless NIC is operated at 20MHz high throughput mode (HT mode). Base on the standard 802.11n MAC protocol the wideband channel is divided into total 56 subcarriers. To reduce the size of CSI report [6], Intel's implementation is grouping 56 subcarriers into 30 groups. Because NIC only reports these 30 grouped subcarriers, complex CSI vector would cover 30-group subcarrier data not all 56-subcarrier data in our work. The reported subcarrier ID of each grouping

		_,		
BW	Grouping Ng	Ns	Carriers for which matrices are sent	
	1	56	All data and pilot carriers: -28, -27,2, -1, 1, 2,27, 28	学"回回"
20 MHz	2	30	$\begin{array}{c} -28, -26, -24, -22, -20, -18, -16, -14, -12, -10, -8, -6, -4, -2, -1, \\ 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 28 \end{array}$	
	4	16	-28,-24,-20,-16,-12,-8,-4,-1,1,5,9,13,17,21,25,28	

mode is shown in Table 3-1, which refers from the Intel 802.11n specification in 2009.

Table 3-1 Intel Wi-Fi Link 5300 NIC, Grouping Mode Ng=2.

3.2 CSI Observation

Before we design and implement a localization system, we set up one Wi-Fi AP in

our laboratory and collected some CSI vectors from that AP to analyze CSI properties.

3.2.1 CSI Vector

Original CSI vector is a complex number vector that could be represented as

$$V_{CSI} = [v_1, v_2, \dots, v_i, \dots, v_N], i \in [1,30],$$

where each subcarrier v_i could be defined as

$$v_i = |v_i|e^{jsin(\angle v_i)}$$

where $|v_i|$ is the magnitude and $\angle v_i$ is the phase response of i_{th} subcarrier. In our system, we translate all complex number vector (i.e., V_{CSI}) to real magnitude vector (i.e.,

 $|V_{CSI}| = [|v_1|, |v_2|, ..., |v_i|, ..., |v_N|]$) and treat magnitude vector as "CSI vector".

3.2.2 R-Square Value

R-square (\mathbb{R}^2) value, which is a statistics coefficient generally used to determine goodness of fit of a statistical model. In our system, we exploit \mathbb{R}^2 value of the testing CSI vectors between the reference base CSI vectors (fingerprint) to determine if these two CSI vectors are similar enough. The mathematic function could be represented as:

$$R^{2}(R,T) = 1 - \frac{SS_{err}}{SS_{tot}}, \quad i = 1, ..., 30$$

R is reference CSI vector and T is testing CSI vector

 $SS_{tot} = \sum_i (R_i - \bar{R})^2$ is the total sum of square of reference vector. $SS_{err} = \sum_i (R_i - T_i)^2$ is the residual sum of square of testing and reference vector.

If two CSI vectors are similar enough, the R^2 value would be very close to 1. On the other hand, the R^2 value could be a negative value.

3.2.3 Stability Experiment

If CSI vector is suitable for localization, the selective-frequency channel fading phenomena should be stable enough, which means CSI vector should exist a statistical structure characteristic at each location. In addition, this CSI statistically structure should be very different from each other at different location. Hence, we explored the property of CSI vector by the following two experiments.

In the first experiment, we collect 100 CSI vectors at the same location (i.e., 4 meter away position from AP on line of site) for continuous five days and compute the mean magnitude of CSI vectors as the feature of each day.

To examine the stability of CSI vector, the feature of the first day would be used as reference base to compare R^2 value of the other days. In Figure 3-2, it could be observed that the R^2 value has a slightly decreasing trend varying over time. This trend might cause by many personal objects or furniture in the crowded laboratory being moved around every day, so the fading phenomena also slightly changed. Although the R^2 values decrease slightly, it still high enough to identify the location of laptop. In addition, our localization area is in the corridor and almost all hardware infrastructures are fixed, therefore, the disturbance toward CSI vectors could be reduced.



Figure 3-2 Stability Experiment Result

The second experiment is for testing the granularity of CSI vector. We place our laptop at reference location (i.e. the same as the first experiment) and collect 200 CSI vectors as reference base. Then we collect testing CSI vectors through moving laptop (toward/backward Wi-Fi AP) and shifting laptop (left/right) from the reference location center by gradually increasing the distance in between. At each testing location, we compute the mean vector over 200 CSI vectors as the feature of that location for test.



Figure 3-3 (a) CSI vector plot, shifting left (-)/right (+) from reference location. (b) CSI comparison result of (a). (c) CSI comparison result, moving toward (-)/backward (+)

from reference location.

We plot CSI vectors at the reference location and at each testing location in (a) of Figure 3-3, and show the R^2 value between each CSI mean vector and the reference mean vector in (b). The variation of CSI mean vector could be very large, even though the testing location is still very close to the reference location (i.e., 1cm ~ 2cm offset). The R^2 value could even decrease to negative value. This observation can also be found in (c), if we slightly decreasing or increasing the distance between laptop and Wi-Fi AP.

Based on the above observation, CSI vector is too sensitive to location offset and

this feature might cause some misjudgment of localization. Our system would introduce a weighted voting estimation process to reduce the probability of misjudgment.

臺



Chapter 4 Fingerprinting Database

4.1 Fingerprint Database Generation

After studying CSI vector properties, we collected CSI vector at each survey point in our testbed and observed the CSI vector again to decide how to generate fingerprint database for localization.

Through plotting all 200 CSI vectors (i.e., there is an example plot, Figure 4-1(a)) collected from one antenna of laptop at the same time in testbed, obviously multiple CSI vector patterns could usually be found. This phenomenon might due to the deployed APs in our testbed is MIMO system supported (i.e., with two transmission antennas). Small spatial difference of transmission antennas could cause different fading result

[21][22]. Hence, if we use two antennas of laptop to collect CSI vectors, doubled obvious CSI vector patterns could be observed in Figure 4-1(b). In the following test and implementation, we only utilize one antenna of laptop to collect CSI vectors.

4.1.1 Cluster Strategy Decision

To implement a more precise localization system, before creating fingerprint database for localization, clustering CSI vectors is necessary to extract the representative features for each position.

In our work, we design an algorithm by making use of the statistically different CSI vector structures at each survey point to decide a proper cluster strategy for fingerprint generation. This algorithm is based on K-means clustering algorithm to cluster. To handle empty cluster problem caused by K-means, we choose to create a new cluster consisting of the one observation farthest from its centroid.


Figure 4-1 Multiple patterns of CSI vectors could be observed (a) Collect by laptop antenna 1 (b) Collect by laptop antenna 1 and 2

To decide a proper cluster strategy by our testing algorithm, which shows in Figure 4-2, we select the testing initial value of possible cluster number N=10. Then this algorithm would treat each CSI vector as one different node and cluster them into former decided possible number of groups. Afterward, our system would calculate mean vector of different cluster as feature and through computes the R^2 value to decide if the possible cluster number is suitable. If there exists any R^2 value from arbitrary pair of mean vectors higher than a threshold (i.e. empirically set 0.5), then subtract one from the value of possible clustering number.



Figure 4-2 Cluster Number Testing Algorithm

Iterate the above process until no R^2 value higher than threshold, then records currently possible cluster number. If the possible cluster number N=2 and the R^2 value is higher than threshold, then record possible cluster number N=1.

We examine 13 survey points from 4 different APs in our survey area and show the result of all combinations in Figure 4-3. About 90% test cases experience 2 to 4 clusters.

Then we analyzed one possibly observed CSI distribution when our testing algorithm decided to cluster them into 3 cluster groups. CSI vectors are still mainly clustered into 2 main groups and the probability of the smallest group contains less than 25 CSI vectors out of total 200 CSI vectors with probability over 80%.



Figure 4-3 Possible Cluster number

Those CSI vectors, which are clustered into the smallest group, are with high

chance interfered seriously by people activities and contain large noise. Because we did not prevent people activities during CSI collection, they can open and close the metal door if they want to enter their laboratory or walk through the corridor. Therefore, we set cluster number N = 3 for clustering in our system but only choose the largest two main clusters to create fingerprints and filter those noisy CSI vectors.

4.1.2 Fingerprint Database

Based on the K-means clustering result, we calculate the mean CSI vector of the largest two main clusters individually. Although, the receiving ratio of CSI vectors of which are classified into the second largest cluster is smaller than the largest one, the probability is high enough and could not be ignored in our test bed.

Therefore, we add a "Fingerprint ID" column (FP ID) in database to represent both fingerprints. In other words, each AP would provide two fingerprints at each survey position, one with FP ID=1 and the other one with FP ID=2. Figure 4-4 shows the data format of fingerprint database. The numbers above each block represents how many columns in each data row. The "AP ID" block is used to identify different AP and the "Location" block stores the coordinate x-value and y-value on the map of testbed. Each



Figure 4-4 Data format of fingerprint database



Chapter 5 Implementation

5.1 Channel State Information Collection

Currently, a Linux 802.11n CSI toolkit has been released in ACM SIGCOMM in 2011. This toolkit contains patched Intel close source firmware, open-source (i.e., iwlwifi) wireless driver and user space tools to log CSI. To retrieve CSI from the PHY layer, we setup a laptop and put it on a wooden chair to collect Channel State Information at each survey point.

5.1.1 Hardware Setup

Due to the released CSI toolkit is a hardware specific driver, we purchased an off

shelf Wi-Fi card (i.e., Intel Wi-Fi Wireless Link 5300) and mounted it on a compatible laptop (i.e., Lenovo ThinkPad X200). This hardware is show in Figure 5-1.

There are three antennas for MIMO radios on IWL5300 NIC, but only two built-in Wi-Fi antennas on our laptop. We use one built-in WWAN antenna as the 3rd antenna for IWL5300 NIC. To avoid the different structure of antenna affecting exploiting localization result, we would not use CSI collected from the 3rd antenna. Our system currently only consider SISO link CSI for localization, and we base on CSI collected from the 1st antenna on NIC.



Figure 5-1 Intel Wi-Fi Wireless Link 5300 NIC & Lenovo ThinkPad X200

5.1.2 Software Setup

For using user space utilities to log CSI from NIC, some Linux kernel header functions are necessary. We follow the installation instructions to patch Linux-based operating system (i.e., Ubuntu 10.04LTS with 2.6.36 kernel.) kernel and install necessary headers.

In addition, since IWL5300 driver always attempts to associate to the Wi-Fi AP with the largest signal strength, we write shell scripts to ask NIC to associate to the AP with specific MAC address, and automatically dump the CSI from AP for offline processing.

5.2 Infrastructure

For students can surf the Internet conveniently, numerous Wi-Fi APs have been deployed in every floor of school building currently. To reduce the overhead of infrastructure, we utilized this school building built-in Wi-Fi system to collect CSI for localization system. For safety issues, we could not get specific location map and MAC address map of built-in Wi-Fi APs from computer center, but we get a roughly AP location map of 5th to 7th floor of Barry Lam (BL) Hall in National Taiwan University.



Figure 5-2 AP Location Map of 6th floor in BL Hall, Mark by green star

Although our test bed is in 6th floor of BL Hall, our laptop could also detect Wi-Fi signal from AP in 5th and 7th floor. We use Wi-Fi analyzer app on Android-based mobile device (i.e., Samsung Note2) to help to find more specific location of Wi-Fi AP in 6th floor (i.e., show in Figure 5-2) and create MAC address map.

5.3 Localization Algorithm

In order to obtain localization result, an effective localization algorithm is needed. Our localization algorithm could be divided into a former processing block and two main blocks. The "Clustering & Data Processing" block is a former processing block, which use to transfer the tracking phase CSI raw data to testing CSI vectors. The first



Figure 5-3 System Architecture

5.3.1 Clustering & Data Processing Block

"Clustering & Data Processing" block contains "Clustering" block and "Data Processing" block. The "Clustering" block clusters the input CSI vectors into 3 clusters and sends CSI vectors belonging to the largest two clusters to the "Data processing" block. The other detail of clustering mechanism has been mentioned in section 4.1. The "Data processing" block would output the mean magnitude CSI vector as feature for each receiving cluster.

The only difference between "Clustering & Data Processing" block (1) and (2) in Figure 5-3 is the number of input CSI raw vectors. The former needs 200 CSI vectors at each survey point for generating fingerprints and the later only needs 20 CSI vectors at each survey point for generating trace data.

Because we would evaluate the performance of using single CSI vector for localization, the "Data Processing (3)" block in Figure 5-3 transfers the CSI raw data from each AP to corresponded CSI magnitude vector for using Single-CSI to localization.

5.3.2 Fingerprint Comparison Block

We design two different testing methods for fingerprint comparison. The first method only uses single CSI vector from each AP for localization at a time and the second method is using multiple CSI vectors collected from several packets sent by each AP at a time. About these two different methods, we would evaluate if using multiple CSI vectors could help to achieve higher localization accuracy. Again, our methods would utilize 4 different APs in each map scenario, which are marked in Figure 6-1 and Figure 6-2.

(1) Single-CSI Comparison Method:

This method exploits single CSI vector from each AP to estimate location in each round. Our system based on AP ID to create R-Square vectors (i.e., V1, V2, ... etc. in Figure 5-4) for each testing CSI vector. By collecting all R^2 values (i.e., V1₁, V1₂, ... etc. in Figure 5-4) between testing CSI vector and fingerprints that with the same AP ID. Afterward, our system sends R-Square matrix (i.e., mixed the R-square vectors of each AP) to the "Weighted Voting Estimation" block.



Figure 5-4 Single-CSI Comparison Method



(2) Multiple-CSI Comparison Method:

Instead of using single CSI vectors each time, Multiple-CSI comparison method uses 20 CSI vectors from each AP for each localization rounds. Before comparison, our system feeds 20 CSI vectors sent from each AP into the former described "Clustering & Data processing" block to get two mean vectors as the features of these 20 CSI vectors.





Figure 5-5 Cross-examination of Multiple-CSI Comparison Method. The larger

accumulated R-square value combination (i.e., Red combination $V1_{11}+V1_{22}$ and Blue

combination $V1_{12}+V1_{21}$) would retain in vector.

Because there are two fingerprints stored for each survey point, our system would do cross-examination by pairing two fingerprints and two testing features up, which shows by red lines and blue lines in Figure 5-5. Then, our system computes the accumulated R² value of each combination and retains the larger one in the R-Square vector. Afterward, merge all R-Square vectors for each AP and send the result R-Square matrix to the "Weighted Voting Estimation" block.

5.3.3 Weighted Voting Estimation

Due to noises from surrounding environment and multipath effects, testing CSI vector might be very similar to multiple fingerprints. That is the R^2 value obtained from other locations could be even higher than that from the ground truth. As a result, our system would use multiple APs (i.e., 4 APs) to calibrate the estimation result.

Through selecting three largest values from each R-Square vector, different APs could point out three nearest neighbor locations individually. Therefore, this block could generate a table, which contains 12 nearest neighbor survey points (i.e., a survey point ID used to identify every survey point) and the corresponding R^2 values.

Our system would base on the survey point (location information) to accumulate

all corresponding R^2 values and choose the survey point with the largest accumulating

 R^2 value as the estimation result.





Chapter 6 Evaluation

6.1 Experimental Scenario

We evaluate our proposed localization system in a line-of-sight corridor section (i.e., the first scenario). This scenario is examined in a straight corridor section, so laptop could collect CSI packets from all APs without any blocking. In the following sections, we would show the self-testing result at first and examining Single-CSI localization (i.e. use Single-CSI comparison method) and Multiple-CSI localization (i.e. use Multiple-CSI comparison method) mechanism performance. In this work, we also compare our result with an RSSI-based localization system and FIFS system.

To study the effect of AP signal blocked by building structure, we also examine our

localization system in a L-shaped section (i.e., the second scenario) so partial Wi-Fi signal from APs might be blocked by the building structure.

6.1.1 Experiment Environment

The first (Line-of-sight) scenario is in a corridor section with 3.45 meters width and 10.8 meters length. We utilize 4 Wi-Fi APs (i.e., Located at the Green Star in Figure 6-1) in this section for localization and divide this section into 13 survey points (i.e. Marked by the Red dots in Figure 6-1). Each point is apart from its neighbor points by 0.9 meter. We collect 200 CSI vectors at each survey point to generate fingerprints.



Figure 6-1 Line-of sight Corridor Section (Scenario 1)

The second (Non-line-of-sight) scenario includes one corner of the corridor. We also utilize 4 Wi-Fi APs (i.e., Located at the Green Star in Figure 6-2) in this L-shape section for localization and divide this section into 12 survey points (i.e., Marked by the Red dots in Figure 6-2). The distance between each survey points is the same as the first scenario.



Figure 6-2 Non-line-of-sight L-Shape Corridor section (Scenario 2)

6.1.2 Experiment Methods

To evaluate our system performance, we do site survey and generate three traces in four different days for each scenario. We place our laptop on a wooden chair (i.e., 35cm x 35cm x 45cm, show by Figure 6-3) and move the laptop to each survey point manually. Through executing shell scripts, the NIC could automatically associate to each AP by mac address table continuously and log the received CSI.



Figure 6-3 Experiment Setting

Because the school Wi-Fi AP blocks the mechanism of pinging AP itself, our laptop pings Google DNS server 8.8.8.8 to collect CSI vectors from AP. In addition, due to the limit of this specific NIC driver, minimum ping time interval is 200ms.

6.1.3 Evaluation Matrix

There are two matrix would be used in our evaluation. The first matrix is the "Zero-error percentage", which means the estimation result is the same as the data collection ground truth. The second matrix is "Average 80-percentile error."

6.2 **Performance Evaluation**

We evaluate our system by three main parts. The first part is about self-testing to show the potential of CSI for localization. The second part is about the performance of our proposed localization algorithm based on three traces collected in three different days to show the stability of our system. The third part is about the necessary of weighted voting estimation process.

We also compare our result with a simple RSSI-based localization system to show the value of using CSI for localization.

6.2.1 Self-Testing

Figure 6-4 is the self-testing result of the first scenario. Although only using Single-CSI localization mechanism for self-testing, the localization system could identify the laptop location accurately. Over 90% estimation results are the same as the ground truth. However, if the estimation results are not accurate, the worst case of the localization error could be larger than 10m. This is due to the fading features are similar in these two different locations.



Figure 6-4 Self-testing Result

6.2.2 Daily Traces Localization Evaluation

To evaluate our system, we collect three traces of the two different scenarios and we discuss the first scenario at first. The Figure 6-5 is the localization result of Single-CSI localization mechanism. One could observe that the localization results of three daily traces are very similar. Our localization system could reach about 30% zero error estimations in average and the 80-percentile error mean is around 4.2m.

The large localization error of our system is caused by fingerprint mismatching, since CSI vectors granularity are small and sensitive. Although we collect CSI trace at each survey point with slight offset, the CSI could vary very much. However, the sensitive fading phenomenon at different survey point could also be identified by fine-grained CSI, so around one of third accurate localization could be achieved.



Figure 6-5 Single-CSI Localization Result (Scenario 1)

The result of use multiple CSI vectors for localization in scenario 1 show by Figure 6-6. Our system also achieves about 40% zero error estimation in average and the 80-percentile error mean is around 2.55m. The improvement could come from processing of multiple CSI vectors before fingerprint comparison. Large noise contained in the CSI vector could be filtered. Therefore, some excessive mismatching cases could be reduced.



Figure 6-6 Multiple-CSI Localization Result (Scenario 1)

To study the effect of partial Wi-Fi signal from AP blocked by building structure, we evaluate our system in the second scenario, which mentioned above. The localization results of Single-CSI and Multiple-CSI localization mechanism are shows by Figure 6-7 and Figure 6-8 individually. The error distance calculation is based on Manhattan distance.



Figure 6-7 Single-CSI Localization Result (Scenario 2)

By Figure 6-7, one could observe that the result of Single-CSI localization mechanism in second scenario is very similar to the first scenario. Our localization system could also reach about 30% zero error estimations in average and the 80-percentile error mean is around 4m.



Figure 6-8 Multiple-CSI Localization Result (Scenario 2)

The performance of Multiple-CSI localization mechanism in the second scenario decreases slightly. Our system achieves in average around 30% zero error estimation and the 80-percentile error mean is 3.25m. The slight decreasing performance might cause by sensitive of location feature of CSI, because slight offset of laptop location during collecting trace in different days could not be avoided. However, this result is still better than Single-CSI localization mechanism.

Hence, if laptop still could receive the CSI vectors from AP, the effect of partial Wi-Fi signal blocked by the building structure could be very small.



6.2.3 Weighted Voting Estimation Evaluation

To examine the necessary of weighted voting estimation process, we compare the localization result with and without voting mechanism in the first scenario. Without voting process means that system would estimate location by only selecting the fingerprint with the largest R^2 value.

By observing Figure 6-9, the voting process could help to improve Single-CSI localization mechanism accuracy under 80-percentile about 1m and also help to improve Multiple-CSI localization mechanism accuracy under 80-percentile more than 1m.

However, if original estimation error is large, the effect of voting process would decrease. Because the received CSI vector could be very similar with different fingerprints at different location, therefore those mismatching could not be avoided.



Figure 6-9 Efficiency of Weighted Voting Process (a) Single-CSI (b) Multiple-CSI

6.3 **RSSI-based Localization System Comparison**

To study the benefit of using CSI, we implement a common RSSI-based fingerprinting localization system to compare the performance between RSSI-based and CSI-based localization system.

6.3.1 RSSI-based Localization Mechanism

We utilize the CSI reports that are used to create CSI fingerprints. Through treating the statistical mean value of RSSI, which also contains in the CSI packets, as the feature of each survey point with respect to different APs, so each fingerprint would contain four RSSI features.

During the tracking phase, the RSSI localization system calculates the Euclidean distance between the testing RSSI vector and the fingerprints. Afterward, this system would select the location with smallest distance as the estimate result.

6.3.2 Performance Comparison

To compare RSSI and CSI fairly, we only compare the RSSI localization result with Single-CSI localization mechanism result, which show by Figure 6-10. The CSI based localization system is obviously better than the RSSI based localization system. The fine-grained CSI help to increase 80-percentile localization accuracy around 1m and increase the probability of zero error estimation about 3 times.



Figure 6-10 Localization Performance RSSI vs. CSI

6.4 Comparison between our system and FIFS

In this work, we also implement the up-to-date work FIFS system and use our CSI log for testing to compare the system performance between our system and FIFS system.

6.4.1 FIFS System vs. our system

FIFS system is also an offline localization system, which utilize CSI from MIMO Wi-Fi AP for localization. However, there still exists two main difference between FIFS system and our proposed system.

The first one is about the fingerprint generation method. In FIFS system, they

studied the subcarrier correlation property with the subcarrier spacing and find that the subcarrier correlation decreases as the subcarrier spacing increases. Moreover, the adjacent channels in 802.11n is non-orthogonal and the non-overlap bandwidth is about 5MHz, so they divided the whole 20MHz channel into 4 sub-channels and then averaging CSI over multiple antennas and within each sub-channel to generate fingerprint and test trace. Hence, their fingerprints are 4-dimension only vectors.

The second one is about the localization techniques. Our proposed system is using R-square model and weighted voting technique. However, in FIFS system, they use posterior probability model and weighted average technique for localization. For more detail about their computation method, one could find their proceeding paper [9] in the related work section.

6.4.2 FIFS System Evaluation

To compare FIFS system fairly, we slightly modified FIFS system for evaluation.

First, there is an AP selection mechanism included in FIFS system. This AP selection mechanism helps FIFS system to select 3 APs with largest signal strength

for localization each time. However, our system utilizes 4 fixed APs for localization in each scenario, so we modified FIFS system to always use 4 APs for localization version during evaluation.

Second, they use different amount CSI vectors to generate fingerprint database and trace file. Also for fair comparison, the usage of CSI vectors are also modified to use the same amount CSI vectors as our system for evaluation.

The evaluation result of FIFS system shows by Figure 6-11. Their 80 percentile error is about 4.3m and zero error percentage is about 5%. Compare the localization result of FIFS system to our proposed system using single CSI vector for localization, our system still outperforms than FIFS system in our testbed.



Figure 6-11 System performance of FIFS



Chapter 7 Conclusion & Discussion

Base on the released toolkit, extracting fine-grained CSI from PHY layer of NIC becomes possible. We utilize this toolkit and the existing Wi-Fi system to study using fine-grained CSI for localization. Due to two obvious CSI clusters in our test-bed could be observed, we use K-means clustering algorithm to help to create a multi-fingerprints database. To exploit this multi-fingerprints feature, we not only introduce a Single-CSI localization mechanism, but also introduce a Multiple-CSI localization mechanism, which could improve accuracy further.

Actually, high-dimension environment features of CSI helps us to obtain good results when doing self-testing. However, CSI vector is very sensitive to environmental changes and the offset of collecting locations. CSI vector could vary largely, even only small offset from the survey point center. Hence, the localization performance of daily trace could decrease seriously. To address this problem, we provide a weighted voting estimation method by utilizing different APs to reduce the probability of misjudgment and obtain better results.

There are still some limitations could be discussed before making a real-time localization system and these might be possible future works. For example, the CSI collection time function of our system could approximately be represented as:

$$T_{collect} = C_{ap} \times (T_{ping} \times C_{ping} + T_{as})$$

 C_{ap} is total AP number. C_{ping} is total ping times. T_{as} is the AP association time usually about 1~5s. T_{ping} is the shortest ping time interval, limited by the toolkit driver, is 200ms.

Our system needs to associate to four different APs one by one, so if one selects to use our proposed Multiple-CSI localization mechanism, CSI collection process at each survey point is approximately $16s+4T_{as}$ (i.e., $4 \times (20 \times 200(ms) + T_{as})$). The collection time is still too long to use for real-time applications.

In addition, about our proposed localization technique, such as K-means clustering

algorithm and weighted voting method are very simple techniques. For example, we add the weight of each vote together for localization currently. However, this method might suffer from the relationship between the number of survey points and the number of K-NN. Hence, find the improvement method of this technique could also be another issue.

Last but not the least, there are still many hidden features in the fine-grained CSI from MIMO AP, which could be discussed and explored. For example, in this work, we inspired by other related works to utilize magnitude value of CSI value to process the received data. However, this method is also a simple way to utilize CSI. Hence, dig the properties of CSI in detail to explore more effective data processing method and use those hidden features for application are possible improvement directions.



Reference

- [1] Nissanka B. Priyantha, Anit Chakraborty and Hari Balakrishnan, "*The Cricket Location-Support System*," in Proc. of MobiCom, 2000
- [2] Lionel M. Ni, Y. Liu, Y. Lau and A. Patil, "LANDMARC: Indoor Location Sensing Using Active AFID," in Proc. of PerCom, 2003
- [3] M. Youssef and A. Agrawala, "The Hours WLAN Location Determination System," in Proc. of MobiSys, 2005
- [4] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," in Proc. of InfoCom, 2000

- [5] Daniel Halperin, Wenjun Hu, Anmol Sheth and David Wetherall, *"Tool Release: Gathering 802.11n Traces with Channel State Information*" in Proc. of SIGCOMM, 2011
- [6] Souvik Sen, Bozidar Radunovic, Tom Minka and Romit Roy Coudhury, "Spot Localization using PHY Layer Information," in Proc. of MobiSys, 2012
- [7] IEEE Std. 801.11n-2009: Enhancements for higher throughput, http://www.ieee802.org
- [8] Jiang Xiao, Kaishun Wu, Youwen Yi and Lionel M. Ni, "FILA: Fine-grained Indoor Localization," in Proc. of InfoCom, 2012
- [9] Jiang Xiao, Kaishun Wu, Youwen Yi and Lionel M. Ni, "FIFS: Fine-grained Indoor Fingerprinting System," in Proc. of ICCCN, 2012
- [10] Eiman Elnahrawy, Xiaoyan Li and Richard P. Martin, "The Limits of Localization Using Signal Strength: A Comparative Study," in Proc. of SECON, 2004
- [11] Tsung-Han Lin, I-Hei Ng, Seng-Yong Lau, Kuang-Ming Chen and Polly Huang,
 "A Microscopic Examination of an RSSI-Signature Based Indoor Localization System," in Proc. of HotEmNets, 2010
- [12] Mu Zhou, Prashant Krishnamurthy, Yubin Xu and Lin Ma "Physical Distance Vs Signal Distance An Analysis Towards Better Location Fingerprinting," in Proc. of HPCC, 2011
- [13] Andreas Teuber, Bernd Eissfeller and Thomas Pany, "A Two-Stage Fuzzy Logic Approach for Wireless LAN Indoor Positioning," in Proc. of IEEE/ION Position Location Navigat. Symp., 2006
- [14] Mauro Brunato and Roberto Battiti, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," in Prob. of the International Journal of Computer and Telecommunications Networking, 2005
- [15] Roberto Battiti, Alessandro Villani and Thang Le Nhat, "Neural network models for intelligent networks deriving the location from signal patterns," in Proc. of AINS, 2002
- [16] Radio Design for MIMO Systems with an Emphasis on IEEE 80211, http://sites.ieee.org/
- [17] Xiaofan Jiang, Chieh-Jan Mike Liang, Kaifei Chen, Ben Zhang, Jeff Hsu, Bin Cao and Feng Zhao, "Design and Evaluation of a Wireless Magnetic-based Proximity Detection Platform for Indoor Applications," in Proc. of IPSN, 2012

- [18] Chun-Chieh Hsiao, Yi-Jing Sung, Seng-Yong Lau, Chia-Hui Chen, Fei-Hsiu Hsiao, Hao-Hua Chu and Polly Huang, "Towards Long-Term Mobility Tracking in NTU Hospital's Elder Care Center," in Proc. of PerCom, 2011
- [19] Alessandro Redondi, Marco Tagliasacchi, Matteo Cesana, Luca Borsani, Paula Tarrio and Fabio Salice, "LAURA - LocAlization and Ubiquitous monitoRing of pAtients for health care support," in Proc. of APLEC, 2010
- [20] Rui Zhang, Fabian Höflinger and Leonhard Reindl, "Inertial Sensor Based Indoor Localization and Monitoring System for Emergency Responders," in Proc. of IEEE Sensors Journal, 2013
- [21] Hajime Suzuki, Thi Van Anh Tran and Iain B.Collings, "Characteristics of MIMO-OFDM Channel in Indoor Environments," in Proc. of EURASIP Journal, 2007
- [22] Jishu DasGupta, Karla Ziri-Castro and Ron Addie, "*Time variation characteristics* of MIMO-OFDM broadband channels in populated indoor environments," in Proc. of LAPC, 2011