HUMAN IDENTIFICATION USING WiFi SIGNAL

SUBMISSION DATE: 26.12.17

SUBMITTED BY:

Md. Nafiul Alam Nipu (13201006)
Souvik Talukder(13201061)
Department of Computer Science and Engineering

Supervisor:

Amitabha Chakrabarty, Ph.D
Assistant Professor
Department of Computer Science and Engineering

Co-Supervisor:

Md. Saiful Islam
Lecturer
Department of Computer Science and Engineering

Declaration

We, hereby declare that this thesis is based on results we have found ourselves. Materials of work from researchers conducted by others are mentioned in references.

Signature of Supervisor

Amitabha Chakrabarty, Ph.D
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Signature of Authors

Md. Nafiul Alam Nipu(13201006)

Souvik Talukder(13201061)

Acknowledgement

We are grateful to our supervisor Dr. Amitabha Chakrabarty, Assistant Professor of the School of Computer Science and Engineering of BRAC University for his immense support and guidance without which this thesis was not possible. We are also thankful to Md. Saiful Islam, Lecturer of the School of Computer Science and Engineering of BRAC University for sharing his experience and providing us with ideas on our topic. We also want to thank the BRAC IT for allocating a PC for us, our work might have undone without that. Also to our parents and friends for their help and support. Lastly, we are thankful to all the faculty members of BRAC University who have inspired and motivated us throughout the entire undergraduate program.

Abstract

There have been a large number of methods already exists to identify human(e.g.,face recognition, gait recognition, fingerprint identification, etc.). Channel State Information(CSI) obtained from Wifi chipsets already has proven to be a efficient for detecting humans uniquely. We are presenting a system which can identify human uniquely and we are showing that Wifi signal can be used for identifying humans. We are working on the channel properties of a communication link which describes how a signal propagates from the transmitter to receiver and represents the combined effect. Each of the individuals have unique gait and also it is proven. Therefore, for that every human would have distract signal uniquely in the same Wifi spectrum. Our system will analysis the Channel State Information(CSI) to acquire unique features of an individual which will allow us to identify a human precisely. We have used two separate algorithms with an accuracy of 95% to 84% in Decision Tree and 97.5% to 78% in Random Forest between a group of 2 to 5 people. We propose that this technology can be used in office or in smart homes for security reasons as it is allowing us to identify humans.

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Human identification (i.e. activity recognition) is the core technology that enables a wide variety of applications such as health care, smart homes, fitness tracking, and building surveillance. Among the traditional technologies that detect human presence, Passive Infrared(PIR), cameras, CO2 sensors and radio waves are popular approaches[1,2]. However, PIR sensors experience low range, require multiple sensors to achieve a reliable detection and detection accuracy is low for static persons. Technologies such as radar, ultrawideband SDR-based solutions are reliable but expensive. Again, camera systems require line of sight and sufficient ambient light. Audio and visual approaches also give rise to significant privacy concerns. Wearable sensors require people to wear some extra sensors which may cause some discomfort.

WiFi techniques are being used in our day to day life. With every foot-step of ours we get to see wireless devices starting from our homes,offices even in the rural places. They invisibly fill the air with a spectrum of Radio frequency(RF) signals. When a person started walking through this places, they propagate the signals and

we will be able to see changes on the WiFi spectrum. If we examine this propagation closely with Channel State Information(CSI) it is possible to identify basic human gaits.

In our work we have implemented a system that follows the above technique. As each person's walking style, gait, body shape is different from one another, each person will impact different changes on the WiFi spectrum and we will see different unique perturbations. Therefore, it is possible to identify a person uniquely by examining these perturbations using Channel State Information(CSI). We carefully examine the CSI data and extract signals from different frequency bands and then select appropriate features that allow us to uniquely identify a person. In our approach, it was noticed that every technique does not always work for a given scenario. It differs due to the selected features of the data or even size. We observed some of the techniques that works very efficiently with a scenario. we are also emphasizing on machine learning. Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. In addition, Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. We have used machine learning

classifiers for identification process. The efficiency of each algorithm researched on the data is shown in the later sections.

The techniques used for identification are Boosted Decision tree & Random Forest.

1.1 Motivation

There are number of ways to identify human but most of them are cost effective. In some of the cases it is seen that the computation cost goes higher. In the first place we had the thought in mind that how we can make it cost efficient and moreover, how we can improve the accuracy. Then when we get the idea of identifying human by using WI-Fi signals after researching we find out that still there are some algorithms which are not used for this. We had a thought in mind that why these algorithms are not used and if it is possible to uniquely identify humans with this algorithms. Then again if we talk about security system and if we think about smart cities it is a common need to identify the right people. Without it all of this will be like in a threat. After knowing about how human body propagates signals and how CSI data describes this propagation we started to work on a system for the implementation of a system which can uniquely identify humans. We have chosen Boosted Decision Tree for the implementation as no further work has done with

this method and for comparison we have used Random Forest because it works well with discrete datas.

1.2 Thesis contribution

The objective of this thesis is to implement a system which can identify Humans uniquely so that we can use it for security purpose, as the existing systems are not broadly used for computational costs, dependability, accuracy and other issues. As per the technology is developing towards smart systems it is just the matter of time that our whole system will be digitalize. So, at that time it will be a must thing that we have to identify people with our systems and technology. Our system need to be smart enough to identify accurate persons for all the sectors and as per needed. For doing all this thing our system will be necessary.

1.3 Methodology

There are many algorithms that can be used to study data. Those we are using are as follows:

1.3.1 Boosted Decision Tree

Boosted Decision Trees (BDTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

Classification is an unsupervised learning used to predict the class of objects whose class label is unknown. It is used for creating classification rules by means of decision trees from a given data set. Decision tree is used as a prognostic model. C4.5, C5.0, CART, ID3 are methods for building decision trees. It is an extension of the basic ID3 algorithm [4]. It is simply understandable and interpretable. Even the non-technical people are able to understand DT models after a brief explanation.

Furthermore, important insights can be generated based on experts describing a situation and their preferences for outcomes. It also allows the addition of new possible scenarios. Similarly, it helps to decide most exceedingly terrible, best and expected qualities for various situations

Decision is of three types and one of them is Boosted Decision tree. Boosted trees Incrementally building an ensemble by training each new instance to emphasize the training instances previously mis-modeled. A typical example is AdaBoost. These can be used for regression-type and classification-type problems. Boosted Decision Trees are used for machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

1.3.2 Random Forest

The random forest [22](Breiman, 2001) is an ensemble approach that can also be thought of as a form of nearest neighbor predictor. Ensembles are a divide-and-conquer approach used to improve performance. The main principle behind ensemble methods is that a group of "weak learners" can come together to form a "strong learner". The figure below provides an example. Each classifier,

individually, is a “weak learner,” while all the classifiers taken together are a “strong learner”.

The data to be modeled are the blue circles. We assume that they represent some underlying function plus noise. Each individual learner is shown as a gray curve. Each gray curve (a weak learner) is a fair approximation to the underlying data. The red curve (the ensemble “strong learner”) can be seen to be a much better approximation to the underlying data.

1.3.3 Trees And Forest

The random forest starts with a standard machine learning technique called a “decision tree” which, in ensemble terms, corresponds to our weak learner. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets.

The random forest combines trees with the notion of an ensemble. Thus, in ensemble terms, the trees are weak learners and the random forest is a strong learner.

Chapter 2

Related Works

For identifying humans uniquely many researches has done. Also many has used CSI data sets for their researches. The more the documented data the more it can be manipulated in predictions. As machine learning is being used here. Many of them has got high accuracy and success with their researches.

Jin Zhang_y, Bo Weiz, Wen Hu_y, Salil S. Kanhere, in their work has shown for the first time WiFi signals can also be used to uniquely identify people [6] . The have proposed a system called WiFi-ID that analyses the channel state information to extract unique features that are representative of the walking style of that individual and thus allow them to uniquely identify that person. They implemented WiFi-ID on commercial off-the-shelf devices. Also they have conducted extensive experiments to demonstrate that other system can uniquely identify people with average accuracy of 93% to 77% from a group of 2 to 6 people, respectively. They envisage that this technology can find many applications in small office or smart home settings.

Mustafa Aljumaily who is a student of Department of Electrical Engineering and Computer Science in The University of Tennessee, Knoxville mentioned that, having accurate detection, recognition, and classification of human activities is still a big challenge that attracts a lot of research efforts due to the problems related to the human body parts and the difficulty in detecting their actions accurately, human clothes and their negative effect on the detection accuracy, and the surrounding environment conditions [7]. Vision based human activity analysis using computer vision and machine learning as the original solution is still having its limitations that are related to the inability to detect whatever happening behind the walls or in the dark places and the uncomfortable feeling of people with cameras everywhere. Recent advances in the wireless technology gave new solutions to tackle these problems as it has been proven that the movement of human body parts will affect the channel state information (CSI) of wireless signals in the indoor environments. The advantages of this technique is the ability to work in dark, non-line of sight detection and working without affecting human daily life activities like the cameras do in the vision-based systems. In his survey paper he has provided an overview of the basics and applications of wireless CSI

and the recent research efforts to utilize the it in track, detect, and classify humans actions and activities in the indoor environments.

Tong Xin, Bin Guo, Zhu Wang, Mingyang Li, Zhiwen Yu from School of Computer Science, Northwestern Polytechnical University, Xi'an, P. R. China , has proposed an approach for human identification [8], which leverages WIFI signals to enable non-intrusive human identification in domestic environments. It is based on the observation that each person has specific influence patterns to the surrounding WIFI signal while moving indoors, regarding their body shape characteristics and motion patterns. The influence can be captured by the Channel State Information (CSI) time series of WIFI. Specifically, a combination of Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT) and Dynamic Time Warping (DTW) techniques is used for CSI waveform-based human identification. They had implemented the system in a 6m*5m smart home environment and recruited 9 users for data collection and evaluation. Experimental results indicate that the identification accuracy is about 88.9% to 94.5% when the candidate user set changes from 6 to 2, showing that the proposed human identification method is effective in domestic environments.

Human identification is being researched for almost a decade. [7] and [8] all use video cameras to record people walking and extract patterns from images. [8] and [7] both capture the silhouette of persons and extract the gait motions of persons for identifications. While the above camera-based approaches achieve good accuracy in identifying individual's they could be considered to be too intrusive (from the perspective of privacy) for use in offices and homes. Moreover, their results are dependent on good lighting conditions. Other works exploit fingerprints [9], iris [10] or Sclera [11] biometrics for identifications. The accuracy of these approaches is higher than using video cameras. However collecting biometric data often makes users uncomfortable and limit their applications. Moreover, several researchers have demonstrated that such biometric systems can be faked [12], [13] and [14] use wearable sensors and derive signatures that are unique to an individual's activity, which in turn can be used to identify people. However, these methods require that the sensors must be worn in a specific manner to ensure accurate operations. Moreover, not everyone is comfortable wearing such sensors on their body at all times. Our work is different in that it leverages the existing WiFi infrastructures for human identifications without the need for additional sensors and intrusive monitoring. Recently many works use wireless radio for

human body sensing. [15] considers the micro Doppler detected by radars for human gait recognitions. However the utilisation of radars is limited due to high costs and regulations. [16], [17] and [18] utilise channel state information (CSI) reported from WiFi cards for human activity recognition. Unlike RSSI which provides the coarse information about the received signal strength, CSI data contain rich information from every wireless sub-channels. Various human activities such as sitting, walking and running create unique perturbations in the CSI data and can thus be used to recognise these activities [19]. In [20], the authors propose a CSI-speed model that establishes a relationship between the CSI variations and the speed of human movement. CSI data has also be used to recognize hand gestures [19] and keystrokes [21] typed on a keyboard.

Chapter 3

System Implementation and Design

In this chapter we present the implementation and design of our proposed system. We start by our problem statement. Then, we present an overview of CSI(Channel State Information) data and we lay out a detailed description of our system.

3.1 Overview of CSI Data

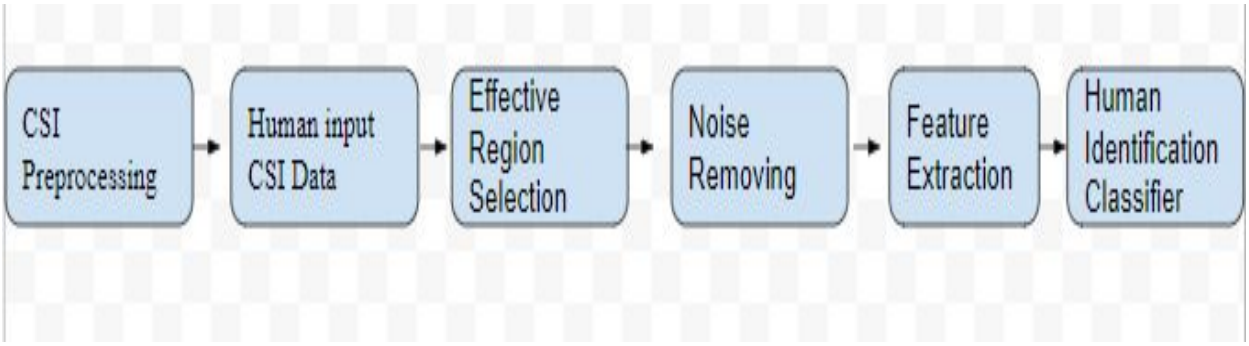


Fig 3.1: System Architecture

In wireless communication, Channel State Information(CSI) describes how a signal propagates from transmitter to receiver. In CSI, there are many subcarriers which

contains amplitude and phase information. Therefore, CSI can capture the combined effects of multiple wireless phenomena like scattering, fading, power decay with distance, shadowing.

$$H(f_k) = \left\| H(f_k) \right\| e^{j<H(f_k)} \text{ ……..(1)}$$

Here $H(f_k)$ is the CSI at the subcarrier with central frequency of f_k , and $\left\| H(f_k) \right\|$ and $< H(f_k)$ denote its amplitude and phase [31], respectively.We have focused only on amplitude in our work.

Most modern off-the-shelf WiFi devices support the IEEE 802.11n/ac standard and include multiple antennas for MIMO communications. These devices can operate 2.4GHz and 5GHz and employ OFDM at the PHY layer. The WiFi NICs continuously monitor the frequency response of the OFDM subcarriers as CSI[30]. Several studies on human identification[6][8][27] have been conducted using CSI which proves that human identification is possible using CSI data.

Let X and Y be the frequency distribution of transmitted and received signals. The two signals are related by the expression below:

$$Y = H \times X + n \text{ ……..(2)}$$

Here, H is the channel frequency response(CFR) for a carried frequency measured at a certain time and n is the noise vector [32].

Let Tx and Rx be the number of transmitting and receiving antennas. Let Sc is the number subcarriers of a certain channel width. Thus the total number of CSI time series is $Tx \times Rx \times Sc$

3.2 Detailed Description Of the System

3.2.1 CSI Preprocessing

Before taking the input data we have to process our system so that it can obtain CSI values. CSI values can be obtained from COTS WIFI network interface cards (NICs) (such as Intel 5300 [25] and Atheros 9390 [26]). In our work we have used Intel 5300 NIC card as the receiver on Desktop PC and TP-Link Router as the transmitter. Our NIC card has 3 antennas and Router has 1 antenna. We use 20 MHz channel width so we get 30 subcarriers information. Therefore, our total number of time series is $1 \times 3 \times 30$ which means we are using a SIMO communication.

For getting CSI, CSI tool is built on Intel 5300 using a custom modified firmware. We first build and install the modified wireless driver on our linux operating

system. Then we installed the modified firmware. Finally, data rate is to be defined to get the CSI data. As linux is the free source it is easy to modify the kernel and drivers. Therefore, we have used Linux operating system Ubuntu 14.0.4

Firstly, for maintaining more control to the interface card we disabled the Network Manager from controlling the interface card and configure it using command-line utilities such as *iw* and *iproute2.*

Then, to obtain CSI data we modified the wireless driver and build the modified driver for our existing kernel version which is 3.16.7. After that, we installed the modified driver into our module updates directory. We also obtained the CSI Tool supplementary materials for our experiment. Finally, we built *log_to_file,* a command line tool that writes CSI obtained via the driver to a file.

The above procedure is done in Ubuntu Terminal. We have followed the instructions above from Daniel Halperin's work on CSI data[25][28][29]

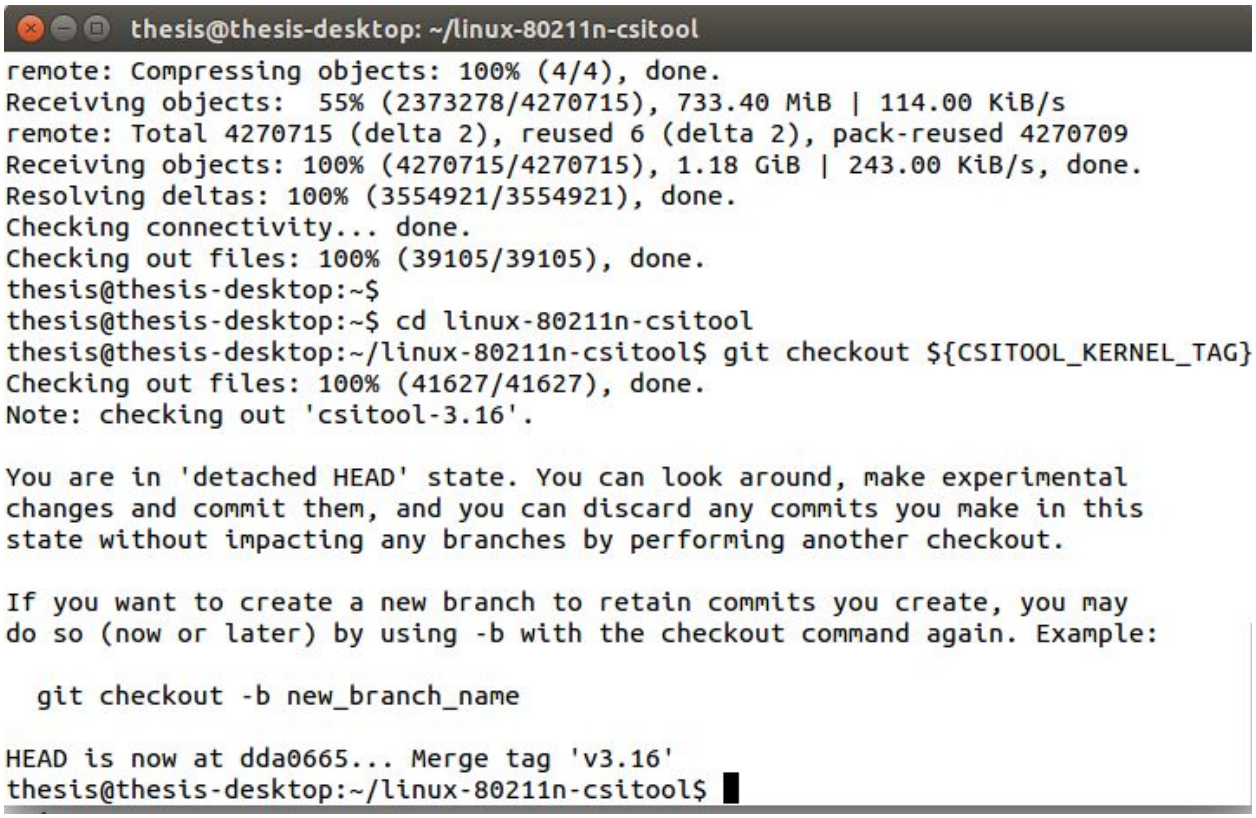


Fig 3.2:A Portion of CSI Driver Installation Process On Terminal

3.2.2 Human Input CSI Data

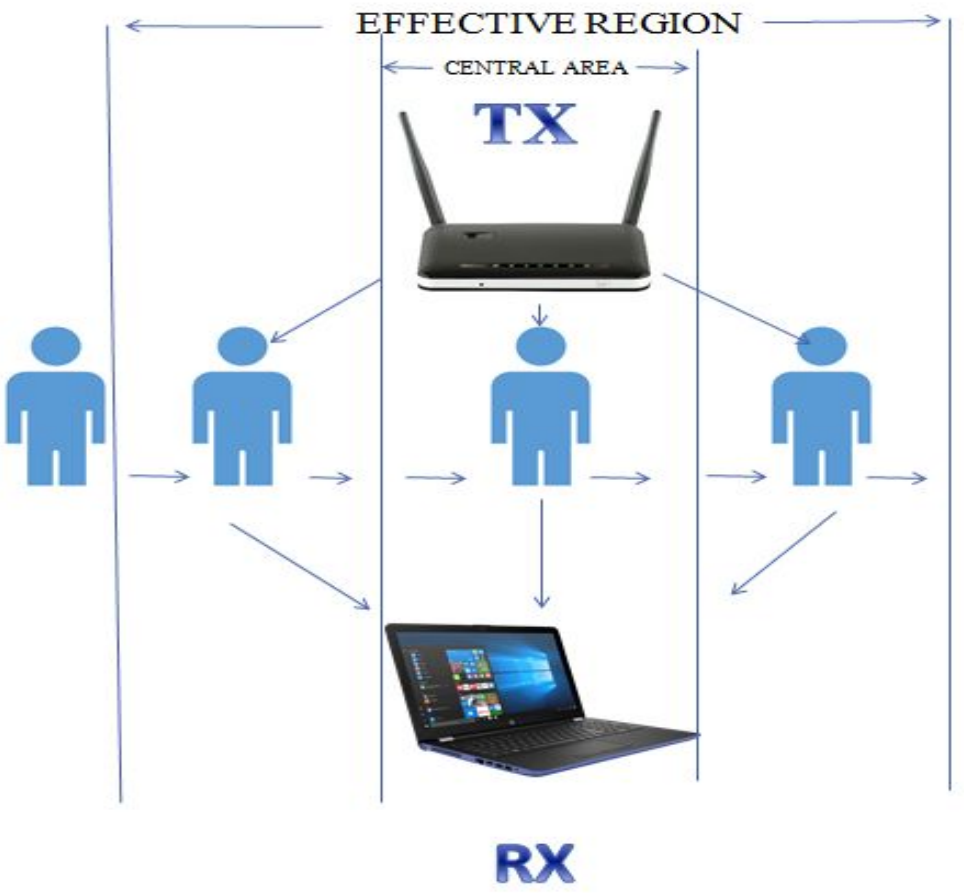


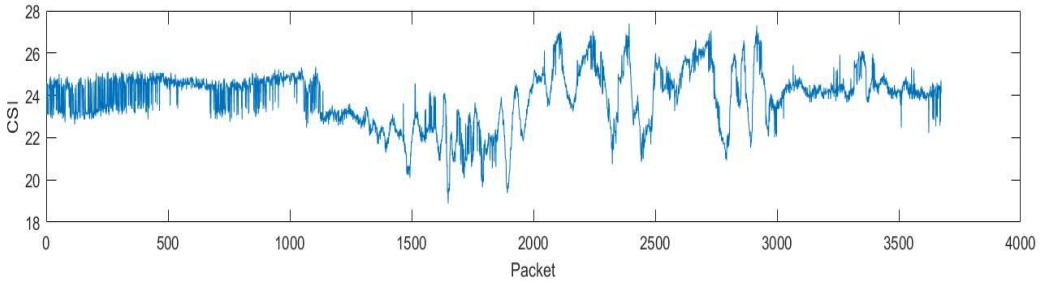
Fig 3.3: Operational Scenario of Our System

After creating the environment for taking CSI data, we took the Human Input Data for Identification purpose. Before taking the data we created a scenario like shown

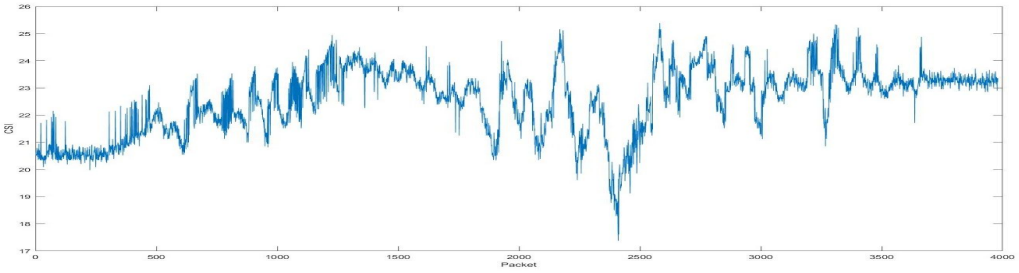
in the above figure. As mentioned above, our system consists of two devices - a desktop PC contained with an Intel 5300 NIC card which works as the receiver(Rx) and TP-LINK TL-WR740N, a WiFi access point(AP) which serves as the transmitter(Tx). As the transmitter has 1 and receiver has 3 antennas, we get $1 \times 3 \times 30 = 90$ data streams per packet. It means that we are getting 30 subcarriers information for each Rx antenna. The two devices were approximately 180 cm far from each other and they were 80 cm above from the ground. The process has taken place in an opened room. There were furnitures and other objects were available in the room. It is also worth mentioning that other WiFi APs were fully functional during the whole process.

The key idea here is the transmitter sends packets and the receiver receives it via the antennas. When a person works walks through the path between the Tx and Rx the signal gets obstructed due to human body and we can see a change in the signal. As gait, walking style, body shape of each human being is different the change in the signal will also vary from person to person. Thus it possible to identify human being analysing those signals.

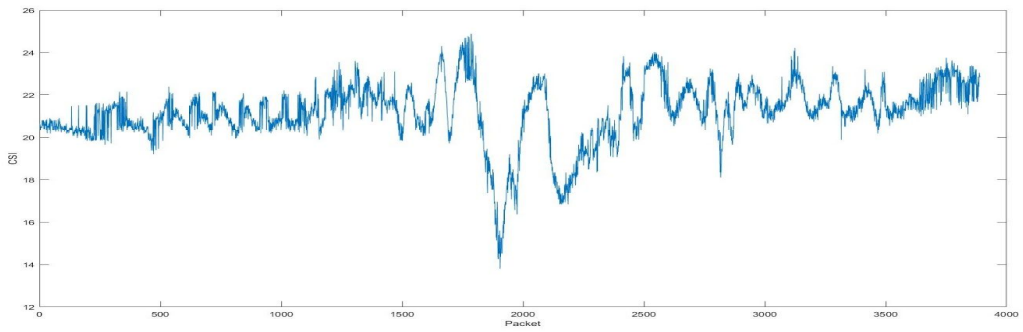
Below figures illustrates the signal received to the interface card due to human distraction. In the figures, signals received to a certain antenna's certain subcarrier information is showed.



(a)



(b)



(c)

Fig 3.4(a,b,c): Human Input Signal Data for Certain Subcarrier

So, to take human input we gathered 5 persons and ask them to walk through the path between the Tx and the Rx. Thus we collected human input CSI data. We sent 1000 packets per second from the Rx to Tx. To do that we used ping command. A notable thing to remember here that pinging must be continued throughout the whole process of collecting datasets. We continued to ping whether there is human walking or not. We have asked each person to walk between the path 20 times. Thus, we collected 20 walking samples for each human for our experiment. We have asked one person to walk at a time and collected his sample. After that, we have asked another person to do the same.

Though CSI data gives phase information and amplitude values, we are going to use amplitude values only for our work. The analysis of phase information will be conducted in our future work.

3.2.3 Effective Region Selection

When we asked people to work through the path between Tx and Rx, we suggested to start working from a certain distance so that we can achieve two goals. One is to get the actual walking sample of the person and the other is to get the values of the effective region shown in the Figure 3.3. Particularly, we are interested in effective region's data. Again, the effective region where Rx and Tx are connected directly is called the central area. Because in this region the human motion caused the most impact in the WiFi spectrum. Thus, this region is the best possible way to identify human uniquely.

Before selecting the effective region we need to consider some scenarios. First of all, about the duration of the effective region in the time series. If the the duration is too low then we will miss a lot of important information. Again, if the duration is too long then a lot of misinformation will be added in the signal. In both case, for our experiment it has to prohibited.

Secondly, we need to determine the startpoint and endpoint of the effective region accurately. A error in calculation may result vital decline in identification.

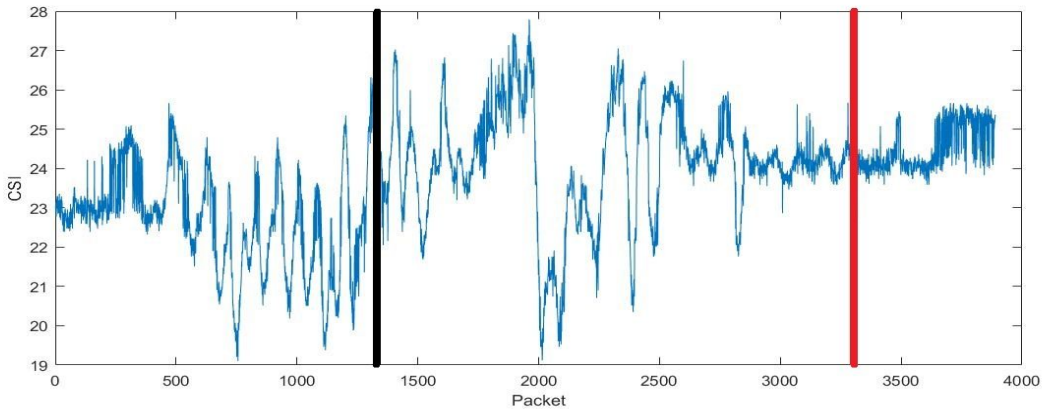
Considering this two challenges and partially motivated from [6], we applied an algorithm to determine the start and end point of the effective region.

In our approach, first we took a person's human sample. Then, we considered values from one antenna as the effective region is same for all the antennas. After that, we partitioned the whole sample to short frames with 50 packets per frame. Then, we calculated the energy of each frame. To calculate the energy, we considered each short frame one by one. In each frame there are 50 packets where in each packet there are 30 subcarrier values. For each packet, we then calculated the sum of squared value of the 30 subcarrier values. By taking the average value we get the energy of a packet. Following this procedure we calculated energy for all 50 packets in a frame. By taking the mean value we can get the energy of a frame. Thus, we calculated energy for all frames of the sample.

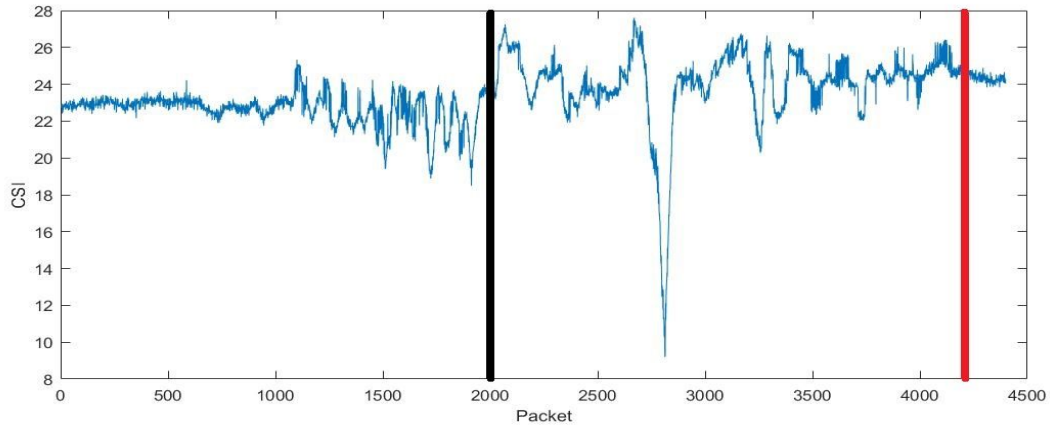
After getting the energy for all frames, we calculated the mean energy of the whole sample. For finding the start point, we checked whether the energy of a frame is greater than the mean energy of the whole sample or not. If for a certain frame, the energy is greater than the mean energy of the whole sample, we consider that the

frame is in effective region. From that we can determine the packet number(i.e. time) of the startpoint of the effective region.

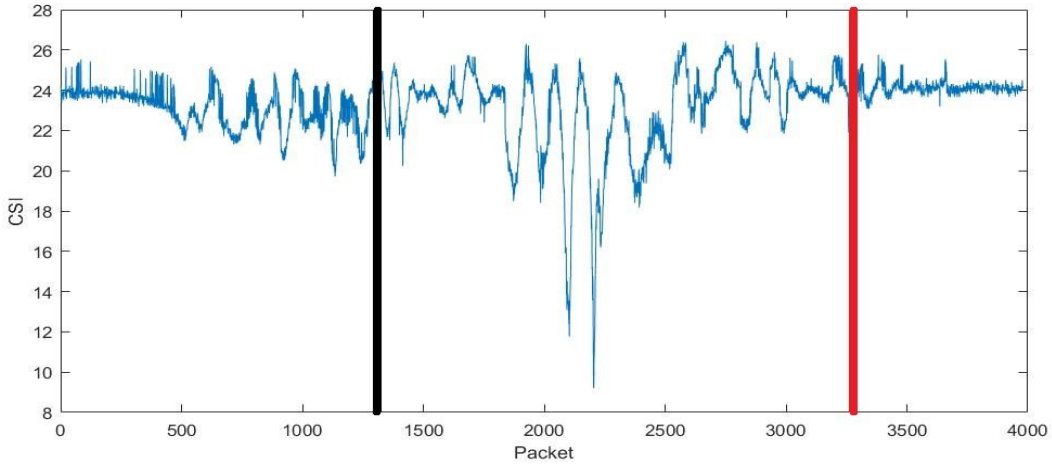
For endpoint we took half of the whole sample length and then sum with the startpoint. Thus, we get the endpoint. From our observation, We assume that effective region is at least half of the whole sample due to our experimental procedure for taking the data input. Thus, we get the startpoint and endpoint of a sample. After selecting the effective region, we used the data for further implementation.



(a)



(b)



(c)

Fig 3.5(a,b,c): The Startpoint and Endpoint of Effective Region

3.2.4 Noise Removing

After finding the effective region our challenge is to remove the noise.The CSI signal data gained through the Interface Card is noisy. Again, when there is no

person walking through the path CSI data will capture ambient noise from other RF transmissions in the vicinity. These datas are not useful and must be removed for better accuracy and identification. Therefore, we need to remove the noise to get a better result.

We have used Butterworth Low Pass Filter for filtering our signal and removing high frequency noise. Butterworth is used for noisy signals where the goal is to have as flat a frequency response as possible in the passband. According to [8], the frequency of the variations in CSI time series caused by human walking is around 10 Hz. Therefore, we have used a cutoff frequency of 10 Hz. We have used second order filter.

We have used MATLAB for processing our input data for identification. In matlab we can create a butterworth filter by using the following command:

[b,a] = butter(n,wn,ftype);

Here, n is the order number of the filter, wn is the cutoff frequency, ftype is to determine whether it is a low pass or high pass filter. In our case, n = 2, wn = 10, ftype = LOW.

b and a are transfer function coefficients. For, digital filter the transfer function is expressed in terms of b and a as

$$G(z) = \frac{B(z)}{A(z)} = \frac{b_0+b_1z+b_2z^M+...+b_Mz^M}{a_0+a_1z+a_2z^M+...+a_Nz^N} \quad \text{.....(3)}$$

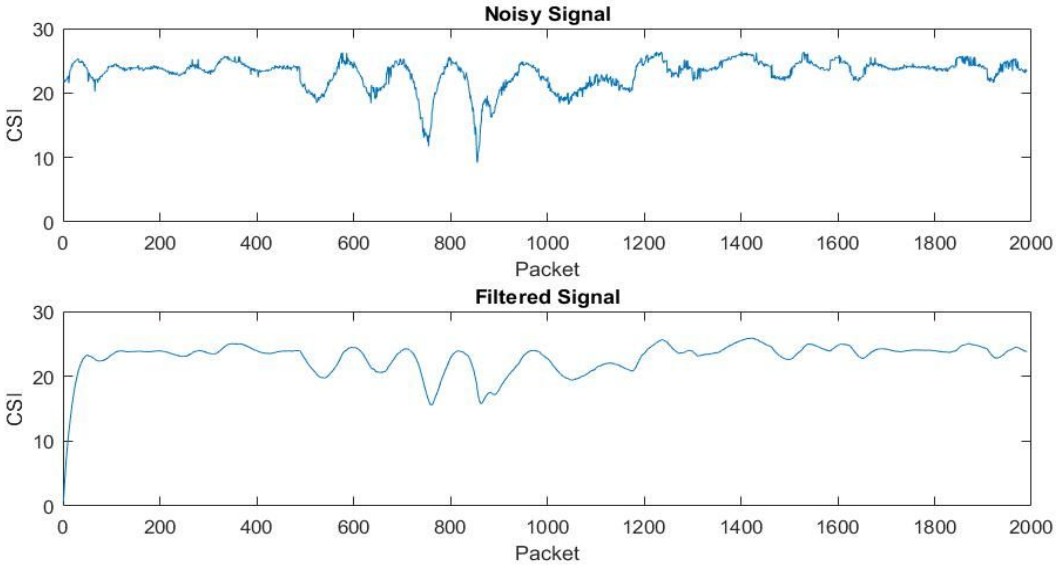
For example, we want to apply the filter on a data x , then we need to execute the following command:

y = filter(b,a,x);

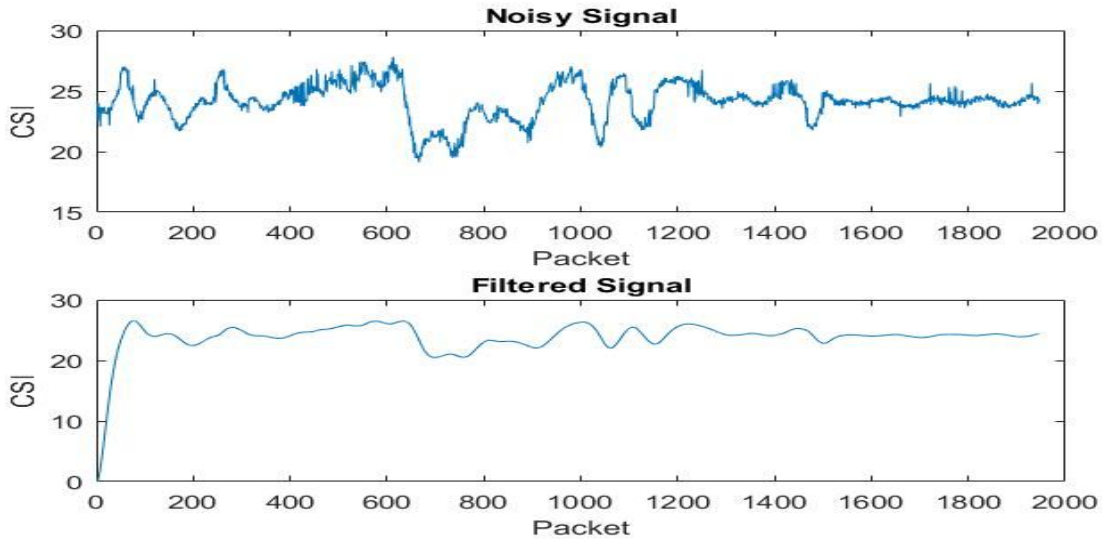
The above command calculates through following equation:

$$y_j = \sum_{i=1}^{N} b_i x_{j-1} + \sum_{i=1}^{M} a_i y_{j-1} \quad \text{.....(4)}$$

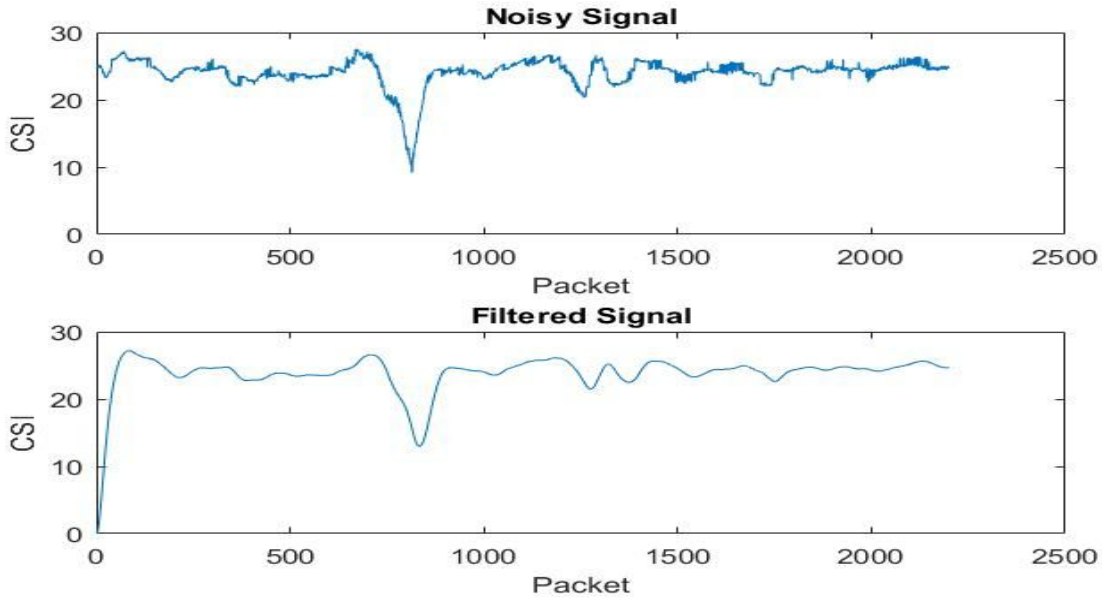
The following figures (Figure 3.6 (a,b,c)) shows the filtering results.



(a)



(b)



(c)

Fig 3.6(a,b,c): Filtered Signal After Applying Butterworth Filter

3.2.5 Feature Extraction

To identify human uniquely, we need to extract features that represents each person's gait analysis and walking style uniquely. As mentioned previously we have $1 \times 3 \times 30$ data streams which means for each packets we have 90 data streams. For a sample containing all packets we have extracted statistical features for those 90 data streams from all packets. We considered 5 time domain features - skewness, mean, maximum, kurtosis, median and 2 frequency domain features - energy and highest fft peaks for our human identification process. A definition of each feature is given in the table below

Table 1: Definition of Each Feature

Feature	**Definition**
Skewness	Measuring the asymmetry of the amplitude data around the sample mean.
Mean	Mean is average amplitude value
Maximum	Maximum is the maximum subcarrier amplitude value

	among all packets for all subcarriers
Median	Amplitude value separating the higher half of the data sample from the lower half. It may be thought of as the middle value
Kurtosis	Measuring the peakedness of the CSI amplitude distribution
Energy	Measure of total energy in all frequencies. It is calculated as $E = \sum_{i=1}^{N/2} (magnitude)^2$
Highest FFT Peaks	Largest frequencies in the signal and their magnitude

After extracting all the feature for all the samples we used them to human identification classifier for identification process.

3.2.6 Human Identification Classifier

After processing the input datasets and extracting unique features for every sample we fed the datasets to human identification classifiers. We have use two machine learning tree based classifier - 1. Boosted Decision Tree 2. Random Forest

Before going in the details, we want to address an overview related to our data sets and reasons for choosing these two classifiers. As our approach is in a supervised way, we can use any supervised machine learning algorithms. But we have to sort down because of the non-linearity of our data set. Algorithm like Naive Bayes works very well in linear approach but can not solve nonlinear problems well. For nonlinear problems predictable algorithms like decision tree, random forest work well. Considering this in our mind we chose the above mentioned classifiers.

For our work, we set some parameters for both algorithms. Table 1 and table 2 shows the parameter values of Boosted Decision Tree and Random Forest respectively.

Table 2: Table showing the Decision Tree parameter values

Predicting Class	Identifying Humans

Features	Skewness, Maximum, Mean, Median, Kurtosis, Energy, Highest FFT Peaks
Minimum Number Of Child Node	1
Maximum Number of Child Node	5
Minimum Number Of Cases	5
Maximum Number Of Levels	10

Table 3: Table showing the Random Forest parameter values

Predicting Class	Identifying Humans
Features	Skewness, Maximum, Mean, Median, Kurtosis, Energy, Highest FFT Peaks
Minimum Number Of Child Node	5
Maximum Number of Child Node	100
Minimum Number Of Cases	5

Maximum Number Of Levels	10

After setting the parameter values we used these classifiers for human identification. Chapter 4 gives a in depth analysis about the results achieved from the classifiers.

3.2.7 Experimental Setup

We have used a desktop PC with Intel Link 5300 WIFI NIC as the receiver, which has Intel core-i5 processor and 4Gb of ram and in that Ubuntu 14.0.4 is used as the Operating System.TP-LINK TL-WR740N was producing signals and in the other side the NIC card was used as the receiver and was receiving the distorted signals that was coming after the collision with the people.

We had placed the receiver and transmitter on two parallel surfaces. We did have maintained 180 Cm distance between them. The router was sending 1000 packets/s to the NIC card which was attached with the PC and was used as receiver and the CSI data was measured on ICMP ping packets.

Chapter 4

Result Analysis

As mentioned earlier, we have used Boosted Decision Tree & Random Forest for identifying human uniquely.We got accuracy of 95% in Boosted Decision Tree and 97.5% in Random Forest between two people.

Comparing the accuracy level between two persons Random Forest gave better accuracy but when the group size increases Random forest tends to show lower accuracy than Boosted Decision Tree.

When the group size is 5, Boosted Decision Tree gave 84% accuracy and Random Forest gave 78% accuracy. As smart home office or house consists more than 2 persons, we can say that boosted decision tree gave us better result in identifying human uniquely from more people.

This chapter is organized accordingly. In 4.1, we gave the evaluation and results that we achieved through our work on human identification, 4.2 shows the limitations of our system.

4.1 Evaluation

In this section, we evaluate and compare our results that we achieved from two different classifiers.

Table 4: Accuracy (In Percentage) Of The Used Classifiers

Number of People	Boosted Decision Tree	Random Forest
• 5	• 84	• 78
• 4	• 88.75	• 80
• 3	• 90	• 86.67
• 2	• 95	• 97.5

Table 4 illustrates the accuracy of the classifiers in identifying human. When there are two person the accuracy level is up to 95% in Decision Tree and 97.5% in Random Forest. When the group size increases both classifiers decline in accuracy and show less accuracy as it is tough to identify human uniquely when there are more people available. In this case Decision Tree gave better results than Random Forest. When the group size is 3,4, 5 people Decision Tree gave 90%, 88.75%, 84% accuracy respectively but Random Forest gave 86.67%, 80%, 78% accuracy.

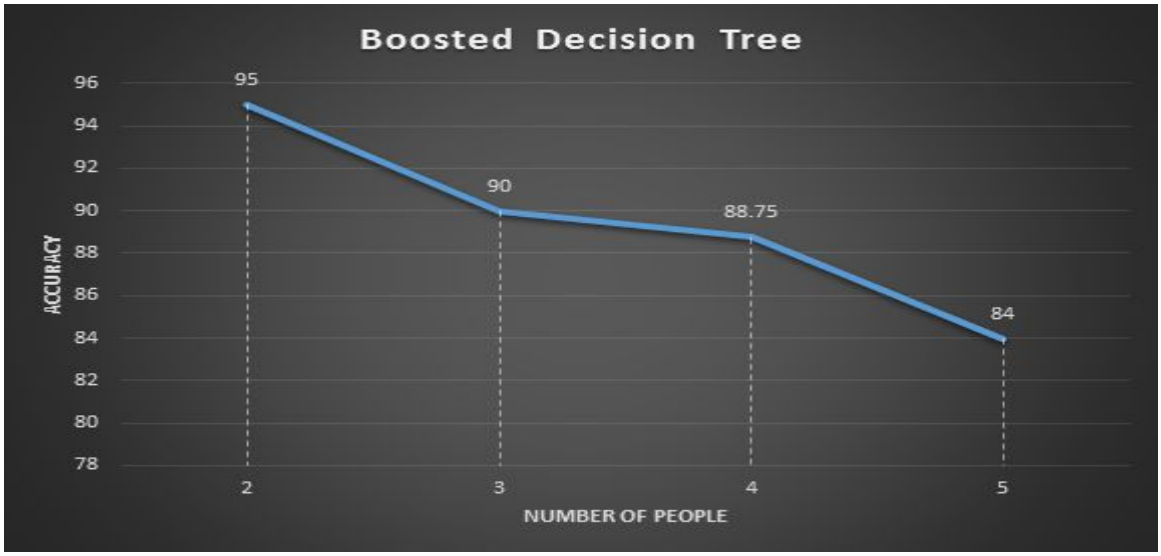


Fig 4.1: Boosted Decision Tree Accuracy Level From Group Size 2-5

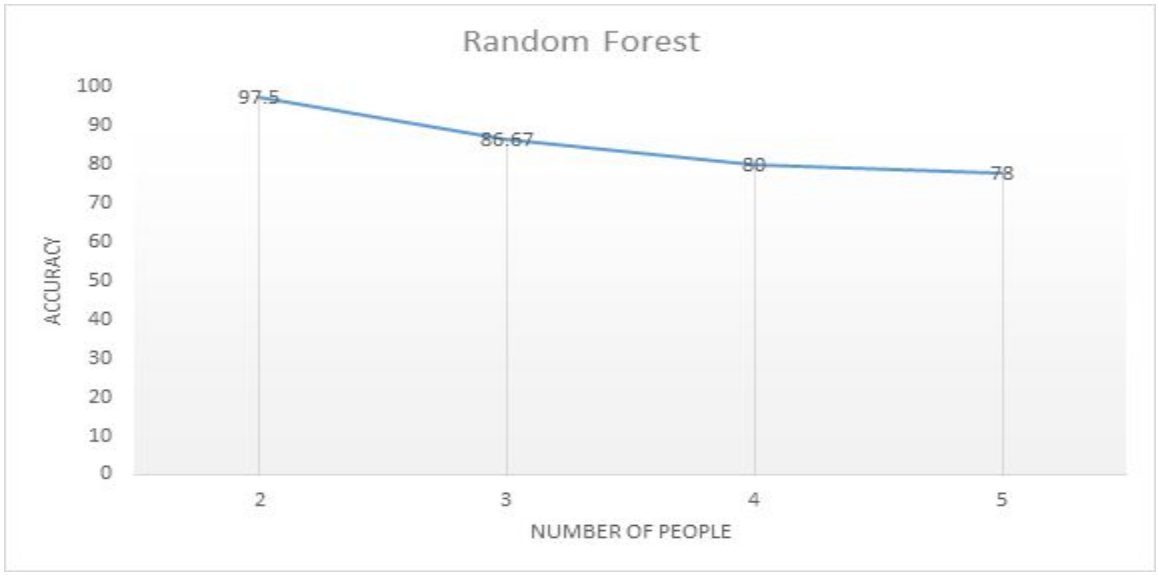


Fig 4.2: Random Forest Accuracy Level From Group Size 2-5

Figure 4.1 and 4.2 illustrate the graph of the accuracy level of Decision Tree and Random Forest. Number of people is given on the X-axis and Accuracy Level is

given on the Y-axis. From both figure we can see that the accuracy declines when the number of people increases but comparatively Decision Tree shows better results and lower difference in accuracy due to the group size. Though Random Forest gave highest accuracy, the difference on accuracy is also high. As in smart home or offices there will be more than 2 people available, we can say that Decision Tree is more accurate than Random Forest.

Table 5 and 6 illustrate the classification matrix for different classifiers and Figure 4.3 and 4.4 illustrate the 3D bar chart of the classification matrix.

Table 5:Classification Matrix of Boosted Decision Tree

	Class predicted Adib	Class predicted Nipu	Class predicted Tanil	Class predicted Souvik	Class predicted Rakib
Observed Adib	19.0		1.0		
Observed Nipu	1.0	15.0	2.0	2.0	
Observed Tanil	1.0	1.0	16.0	2.0	
Observed Souvik		1.0	3.0	16.0	
Observed Rakib	1.0			1.0	18.0

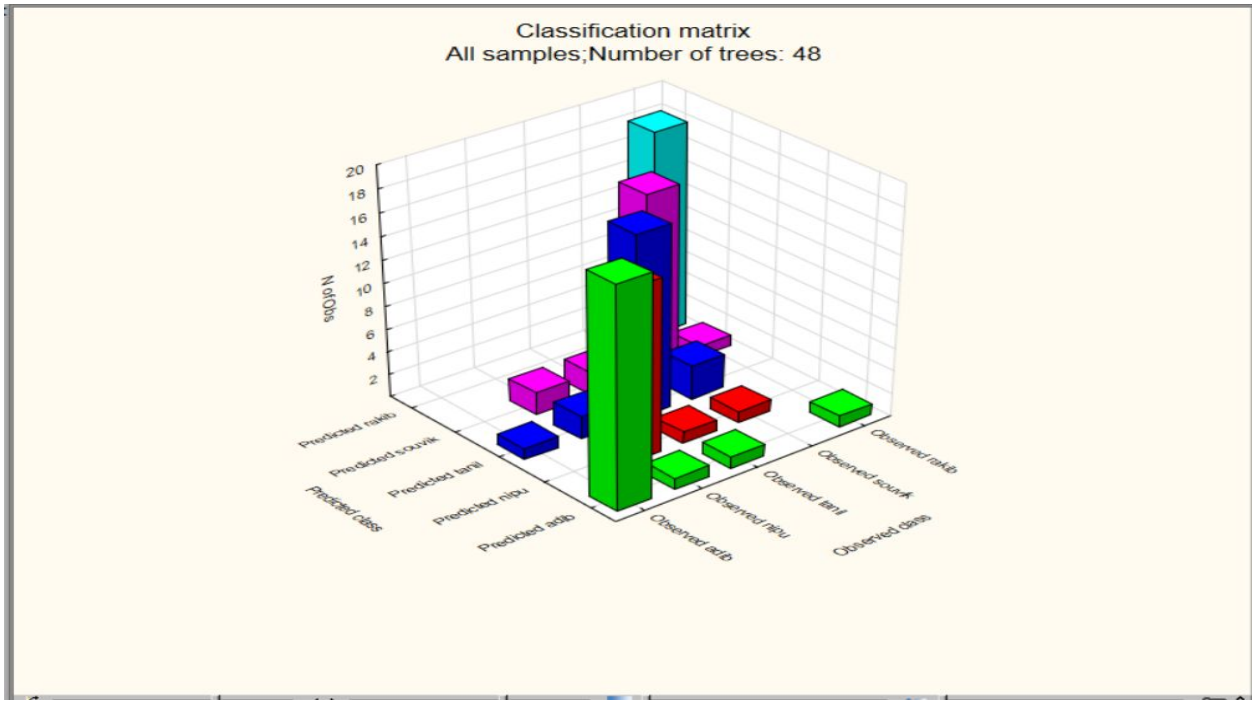


Fig 4.3: 3D Bar Diagram of Classification Matrix - Boosted Decision Tree

The above figure shows the classification result of Boosted Decision Tree. We have selected five individuals for our work and we have named them as Adib, Nipu, Tanil, Souvik and Rakib. As mentioned earlier, we have collected 20 samples of each individual's walking pattern consisting of 100 samples. Then, we extracted features from these samples. For classification we fed all these datasets to the classifier for identification. As Figure 4.3 illustrates Boosted Decision Tree accurately identified 19 datasets of Adib, 15 datasets of Nipu, 16 datasets of Tanil, 16 datasets of Souvik and 18 datasets of Rakib from 20 datasets of each individual.

It means that Boosted Decision Tree accurately identified 84 times from 100 datasets and it predicted wrong 16 times resulting in 84% accuracy. Figure 4.3 also shows out of 20 classes of each individual how many classes are predicted right and how many classes are predicted wrong and also if wrong which individual the classifier guessed.

Table 6: Classification Matrix of Random Forest

	Class predicted Adib	**Class predicted Nipu**	**Class predicted Tanil**	**Class predicted Souvik**	**Class predicted Rakib**
Observed Adib	15.0	4.0	1.0		
Observed Nipu		19.0		1.0	
Observed Tanil	1.0	4.0	10.0	5.0	
Observed Souvik		2.0		17.0	1.0
Observed Rakib	1.0		1.0	1.0	17.0

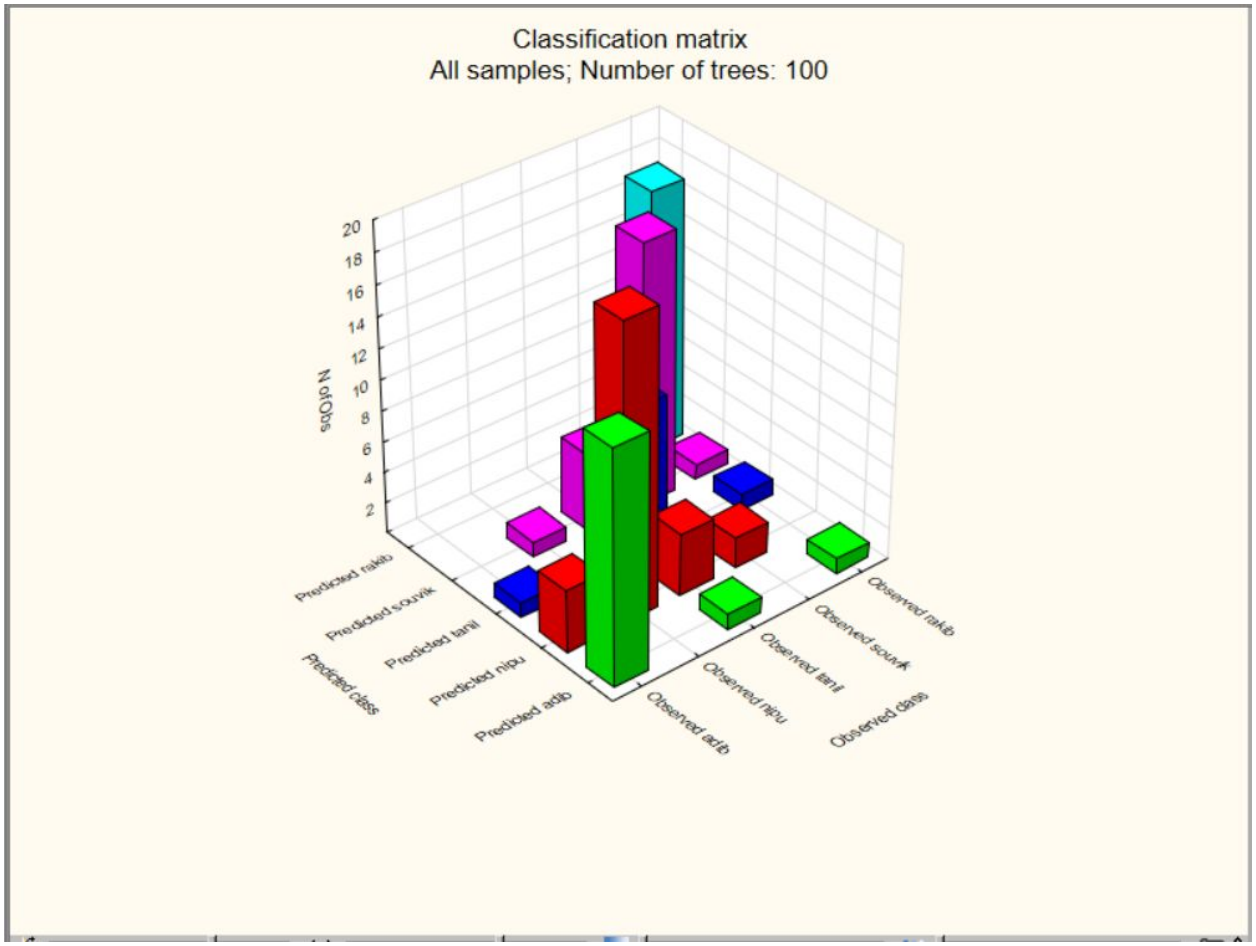


Fig 4.4: 3D Bar Diagram of Classification Matrix - Random Forest

As Figure 4.4 illustrates Random Forest accurately identified 15 datasets of Adib, 19 datasets of Nipu, 10 datasets of Tanil, 17 datasets of Souvik and 17 datasets of Rakib from 20 datasets of each individual. It means that Random Forest accurately identified 78 times from 100 datasets and it predicted wrong 22 times resulting in 78% accuracy. Figure 4.4 also shows out of 20 classes of each individual how many classes are predicted right and how many classes are predicted wrong and

also if wrong which individual the classifier guessed. The results which are shown in figure 4.3 and 4.4 are for group size 5 as our main focus is implementing a system where we can identify more people.

4.2 Limitations

Like every other system our system also has some limitations. Limitations of our system is given below:

- CSI data is sensitive to environment and a change in the environment will change in the measurement of the CSI data. Therefore, during taking the samples our environment had to be remain static.
- The system tends to show lesser accurate result when the group size increases as the classifiers have to identify human uniquely from more human samples.
- For our work, we used some fixed deployment of AP and PC. We believe that this does not pose any issues. We also took our walking sample in a predefined path. However, we did not consider other factors such as walking with stretcher, wheel-chair, back-pack, injured limbs etc.

Chapter 5

Conclusion and Future Work

In our work, we are proposing a WiFi based device free human identification system which can identify particular humans uniquely. Every individuals have different body shape and gait. For this reason when anyone walks through the effective region which is defined by us for our research purpose they will distract the signal which is being sent by the router. These distractions in the CSI (Channel State Information) data is observed thoroughly for identifying particular human being. Our system has achieved 84% to 95% if Boosted Decision Tree is used & 78% to 97.5% if Random Forest is used for the computation for a group of 5 to 2 people. From this we have realised our system can be used in a particular or we can say small spaces like office area or home space.

From the preliminary results we have observed that our system can be useful for the security system of any offices or smart homes. We have the thought in mind that have the chance to implement our system for bigger purpose. Still now we did

not get the chance to this system for big spaces. As we have to follow some certain criterias for data collecting. If we think about a smart city then we have to work broadly for the implementation. Again for now we do not have enough resources to work for bigger spaces. Because the setup for the data collection system for bigger areas probably would be more critical and for now bearing the expenses is difficult for us. We are looking forward to work with any organization or people for implementing our system in bigger and better way. Here, we are working with only a group of 5 people that is why we know that our system works gently for fewer people and small spaces. But for optimizing the the result of huge number of people or bigger space, As example for using this system for smart city or residential area's security system we have to work with our system in a wide-ranging approach.

References

[1] Tero Kivimaki et al. A Review on Device-Free Passive Indoor Positioning Methods. International J. of Smart Home, 08(1):71{94, 2014.

[2] Thiago Teixeira et al. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. ACM Comput. Surv., 5:1{77, 2010.}

[3]L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

[4]John F. Magee. Decision Tree for making Decisions.

[Online].Available: https://www.lucidchart.com/pages/decision-tree[Accessed October 13,2017].

[5] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Elsevier Inc., 2011.

[6]Jin Zhangy, Bo Weiz, Wen Huy, Salil S. Kanhere et al. WiFi-ID: Human Identification using WiFi signal. 2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)

[7]Mustafa Aljumaily et al. A survey on WiFi Channel State Information (CSI) utilization in Human Activity Recognition

[8]Xin, Tong, Bin Guo, Zhu Wang, Mingyang Li, Zhiwen Yu, and Xingshe Zhou. "FreeSense: Indoor Human Identification with Wi-Fi Signals." In Global Communications Conference (GLOBECOM), 2016 IEEE, pp. 1-7. IEEE, 2016

[9] C. Wang et al. Human identification using temporal information preserving gait template. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2164–2176, 2012.

[10] L. Wang et al. Silhouette analysis-based gait recognition for human identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12):1505–1518, 2003.

[11] P. Airey and J. Verran. A method for monitoring substratum hygiene using a complex soil: The human fingerprint. Fouling, cleaning and disinfection in food processing, 2006.

[12] G. Molenberghs et al. Review of iris recognition: cameras, systems, and their applications. Sensor review, 26(1):66–69, 2006.

[13] Z. Zhou et al. A new human identification method: Sclera recognition. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 42(3):571–583, 2012

[14]Hack apple touch id. [Online]. Available:

http://www.theverge.com/2013/9/22/4759128/chaos-computer-club-biometric-hack-apple-touch-id. [Accessed October 11,2017].

[15] J. M¨antyj¨arvi et al. Recognizing human motion with multiple acceleration sensors. In Systems, Man, and Cybernetics, volume 2, pages 747–752. IEEE, 2001.

[16] J. M¨antyj¨arvi et al. Identifying users of portable devices from gait pattern with accelerometers. In Acoustics, Speech, and Signal Processing (ICASSP), volume 2, pages ii–973. IEEE, 2005.

[17] I. Orovi´c et al. A new approach for classification of human gait based on time-frequency feature representations. Signal Processing, 91(6):1448–1456, 2011.

[18] B. Wei et al. Radio-based device-free activity recognition with radio frequency interference. In International Conference on Information Processing in Sensor Networks, pages 154–165. ACM, 2015.

[19] Y. Zeng et al. Analyzing shoppers behavior through wifi signals. In Proceedings of the 2nd workshop on Workshop on Physical Analytics, pages 13–18. ACM, 2015.

[20] L. Sun et al. Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In Proceedings of the 21st Annual International Conference on

Mobile Computing and Networking, pages 77–89. ACM, 2015.

[21] K. Ali et al. Keystroke recognition using wifi signals. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pages 90–102. ACM, 2015.

[22]Breiman(2001). Random Forest An Ensemble Method. [Online]. Available: http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics. [Accessed November 12,2017].

[23] IEEE. Std 802.11n-2009: Enhancements for higher throughput. (Cited on pages 1, 22, 58, 59, 70, 130, and 134.)

[24] Wi-Fi Alliance. Wi-Fi peer-to-peer (P2P) technical specification, version 1.1, 2010. (Cited on pages 1, 135, and 137.)

[25] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11n traces with channel state information. ACM SIGCOMM CCR 41(1):53.

[26] S. Sen, J. Lee, K.-H. Kim, and P. Congdon. Avoiding multipath to revive inbuilding WIFI localization. In Proceeding of ACM MobiSys, 2013, pp, 249–262

[27] Zeng, Yunze, Parth H. Pathak, and Prasant Mohapatra. "WiWho: wifi-based person identification in smart spaces." In Proceedings of the 15th International Conference on Information Processing in Sensor Networks, p. 4. IEEE Press, 2016.

[28] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. 802.11 with multiple antennas for dummies. ACM SIGCOMM CCR, 40(1), January 2010.

[29] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. ACM SIGCOMM, 2010

[30] Z. Yang et al. From rssi to csi: Indoor localization via channel response. ACM Computing Surveys (CSUR), 46(2):25, 2013.

[31]C. Wu, Z. Yang, Z. Zhou, K. Qian, Y. Liu and M. Liu, "PhaseU: Real-time LOS identification with WiFi,"2015 IEEE Conference on Computer Communications (INFOCOM), Kowloon, 2015, pp. 2038-2046. doi: 10.1109/INFOCOM.2015.7218588. [Online].

Available:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7218588&isnumber=7218353. [Accessed November 30,2017].

[32]J. Xiao, K. Wu, Y. Yi, L. Wang and L. M. Ni, "FIMD: Fine-grained Device-free Motion Detection," *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, Singapore, 2012, pp. 229-235.

doi: 10.1109/ICPADS.2012.40. [Onilne].

Available:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6413692&isnumber=6413550