



计算机网络与通信  
Computer Network

# Wi-Fi alarm

基于 Wi-Fi CSI 和机器学习的环境异常感知

Environment anomaly perception based on Wi-Fi CSI and machine learning

课程项目报告

Version 1.0.8

Written by [计算机网络某课程设计小组]

2020 年 1 月 3 日

# 目录

<b>1</b>	<b>绪论 .....</b>	<b>3</b>
1.1	项目背景及意义 .....	3
1.2	国内外相关研究与应用现状 .....	3
1.3	本项目主要工作 .....	4
<b>2</b>	<b>Wi-Fi CSI 相关技术理论 .....</b>	<b>5</b>
2.1	Wi-Fi CSI 概述 .....	5
2.2	Wi-Fi CSI 特性分析 .....	6
2.3	Wi-Fi CSI 的采集 .....	7
2.4	Wi-Fi CSI 数据信息分析 .....	8
<b>3</b>	<b>概要设计 .....</b>	<b>10</b>
3.1	问题描述 .....	10
3.2	Wi-Fi alarm 系统模型 .....	10
<b>4</b>	<b>详细设计 .....</b>	<b>11</b>
4.1	CSI 信息采集与实时分析 .....	11
4.1.1	硬件环境 .....	11
4.1.2	软件环境 .....	12
4.1.3	训练数据采集 .....	13
4.2	CSI 数据预处理与特征选取 .....	13
4.2.1	CSI 数据的预处理 .....	14
4.2.2	特征的选取 .....	15
4.3	模型训练与预测 .....	16
4.3.1	机器学习算法的选择 .....	16
4.3.2	支持向量机 SVM 分类 .....	17
<b>5</b>	<b>实验结果与分析 .....</b>	<b>17</b>
5.1	CSI-tool 通信测试 .....	17
5.2	不同环境下的 SNR 分析 .....	19
5.3	RSSI 与 CSI 稳定性对比分析 .....	19
5.4	数据处理实时性分析 .....	20
5.5	环境状态实时监测结果分析 .....	21
<b>6</b>	<b>总结与展望 .....</b>	<b>22</b>
6.1	总结 .....	22
6.2	展望 .....	23
	<b>参考资料 .....</b>	<b>23</b>
	<b>附录一：网卡驱动配置与 Linux 802.11n CSI tool 安装过程 .....</b>	<b>24</b>
	<b>附录二：项目分工 .....</b>	<b>26</b>
	<b>附录三：项目工程各源代码作用说明 .....</b>	<b>26</b>

# 1 绪论

## 1.1 项目背景及意义

由于 Wi-Fi 终端的广泛使用和物联网技术的持续发展，Wi-Fi CSI 信号感知这一方向成了物联网领域近年的研究热点。一方面，传统的室内感知系统（如超宽带 UWB、信号强度指示 RSSI）存在缺陷：超宽带 UWB 技术依赖特殊设计的信号或硬件设备，难以应用于日常生活；而 RSSI 信号只反映多径波形叠加的总幅度，涵盖信息极少不足以支撑更进一步的需要。另一方面，人们日常生活中越来越需要环境感知技术，尤其是无穿戴的无源感知技术，来实现易用性和普适性的需要，但对于通信设备的感知应用这一命题也极具挑战性。幸运的是，Wi-Fi 路由器发送的信道状态信息（Channel State Information, CSI）具备良好的多径叠加的幅值、相位特性，同时 Wi-Fi 路由器的广泛应用也符合其普适性的要求。基于以上，Wi-Fi CSI 信号无源感知这一研究领域应运而生。

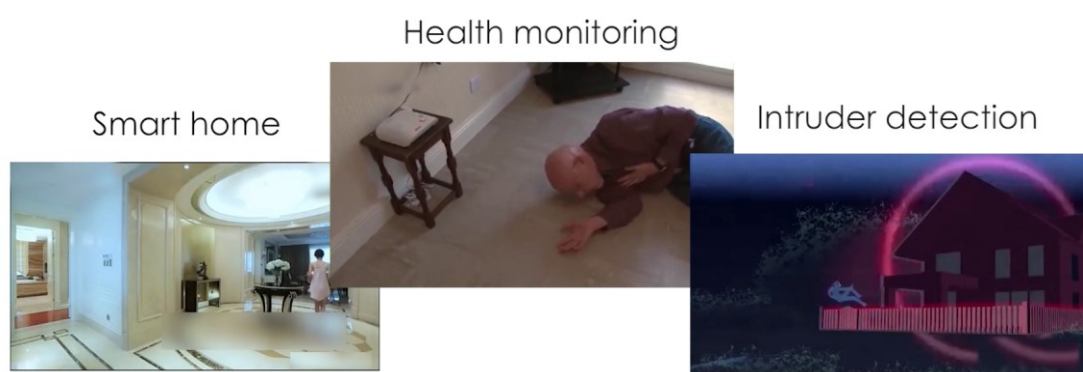


图 1.1 Wi-Fi CSI 信号感知应用场景

## 1.2 国内外相关研究与应用现状

2010 年，Halperin 等在 SIGCOMM 会议上，提出利用 Intel 5300 无线网卡获得 CSI 的测量工具 Linux 802.11n CSI-Tool，并将其开源，此后麻省理工学院、华盛顿大学、斯坦福大学等相关研究人员开始着手在商用 Wi-Fi 设备上提出更多的 Wi-Fi 感知应用，如室内定位、人体检测等。2012 年，Xiao 等提出一种设备无关的检测系统 FIMD，充分考虑时间稳定性和频率多样性，实现高检测率和抗窄带干扰，获得更好的检测效果。FIMD 采用不同时刻 CSI 信息的相关性对移动目标进行检测，其中的 CSI 幅值信息可以用来对静态、动态环境进行区分。2013 年，Zhou 等提出一种全向被动式检测系统 Omni-PHD，采用全向天线技术，针对

博物馆展览等实现小区覆盖被动式人体检测。2014 年, Qian 等提出一种被动式人体检测系统 PADS, 主要利用的是 CSI 幅值和相位信息。通过对原始 CSI 数据预处理, 提取特征信息, 再采用分类算法可有效检测人体不同动作。Wang 等提出的 E-eyes 系统, 根据静止动作和行走活动两类行为对 CSI 幅值的影响不同, 采用移动方差阈值法来区分行走和静止两种不同行为。Melgarejo 等提出的 Gestures Recognition 系统, 利用有向天线技术实现细粒度的手势识别, 采用动态时间规整算法将动态指纹与静态指纹数据库进行匹配, 从而识别不同手势。Han 等提出一种设备无关的室内人体摔倒检测系统 Wi-Fall, 分析 CSI 幅度变化与人体不同动作之间的关系, 将动态下的特征视为异常, 对因人体动作所导致的 CSI 变化提取特征信息, 采用局部异常因子 LOF 和支持向量机 SVM 算法, 对摔倒行为和预先设定好的四种行为进行区分。2015 年, Ali 等提出一种按键识别系统 Wi-key, 由于不同按键所引起的多径失真不同, 根据这种特性来识别按键。当用户按键时, 其手指移动遵循某种特定的形式和方向, 并形成某种特殊的模式, 根据这种模式所记载的 CSI 时间序列, 可用 CSI 不同波形对特定按键进行检测。2016 年, Wang 等实现的 Wi-Hear 系统首先利用特殊定向天线来捕获嘴唇的位置, 然后再根据反射回来的信号监测嘴唇变化, 识别发出的字母等动作。2017 年, Wang 等提出一种实时, 非接触, 低成本且精确的室内跌倒检测系统 RT-Fall, 允许用户自然而持续地活动, 主要利用 CSI 相位信息和相位差来区分人体摔倒及类似行为, 并比较 CSI 的幅度和相位变化, 将摔倒和类似摔倒行为进行区分, 实现日常生活中人体摔倒检测。在四种不同室内场景中的实验结果表明, RT-Fall 性能优于 Wi-Fall, 平均灵敏度提高 14%, 检测率提高 10%。

### 1.3 本项目主要工作

基于 Wi-Fi CSI 信息的环境异常监测, 基本原理是无线信号传输过程中经反射、绕射、折射等多径效应之后会携带反映环境状态信息, 通过对其信息进行预处理及特征提取, 然后再使用分类算法即可完成对环境状态的监测。本项目主要以信道状态信息为基础, 实现对室内环境状态的实时监测, 这里的环境状态指环境中是否有人活动, 其主要工作如下:

#### (1) 利用普通商用 Wi-Fi 设备获取信道状态信息

从当前常用的 Wi-Fi 设备获取特征信息, 包含反映室内环境状态信息的信道状态信息, 对其进行采集, 并分析其特性, 选择可以反映室内环境状态的数据流信息, 作为特征提取的关键信息, 为环境状态监测提供分类依据。

### (2) 对来自信道状态信息的数据特征进行提取

首先,因多径效应的影响会造成无线信号传播的变化,对采集到的含有较多环境噪声等干扰的信道状态信息,提出提取主信道信息的预处理方法,提取出可以明显反应室内环境状态的特征信息。

### (3) 基于信道状态信息的实时环境监测方法研究

根据提取到的环境状态特征,从实时检测到的 Wi-Fi 信号中感知环境状态。

### (4) 基于 Wi-Fi CSI 的环境状态分析

通过分析信道状态信息因环境变化而引起的异动,分析这些变化信息的特征,选取机器学习算法,利用特征提取、分类等方法实现环境状态监测。

## 2 Wi-Fi CSI 相关技术理论

无线电信号不仅可以用于传输数据,还可以用来感知环境,在室内环境下,信号发射端产生的无线电波经由折射、反射、散射等多条路径传播,在信号接收端处形成多径叠加信号,多径叠加信号受其传播物理空间的影响,携带反映环境特征的信息。这里所说的环境是信号传播的物理空间,既包括人的因素(是否有人以及人的位置、特征、姿势、动作等),也包括其他外物的因素。

无线信号在传播过程中调制了环境信息,那么如何从接收信号中解调这些环境信息呢。例如:如果手机接收到的 Wi-Fi 信号较弱,可能时由于手机距离无线路由器较远;而如果手机接收到的 Wi-Fi 信号强度骤降,很可能是手机进入了某些特定的封闭空间,如电梯。用接收信号强度推断接收端与发射端的距离或者表征接收机的位置特点是传统无线感知采用的重要方法之一。近二十年里也涌现了大量利用信号强度信息感知环境特征的应用,其中比较具有代表性的是接收信号强度指示(Received Signal Strength Indicator, RSSI)和信道状态信息(Channel State Information, CSI)。

### 2.1 Wi-Fi CSI 概述

信道状态信息在无线通信领域,就是通信链路的信道属性。它描述了信号在每条传输路径上的衰弱因子,即信道增益矩阵中每个元素的值,如信号散射(Scattering),环境衰弱(fading, multipath fading or shadowing fading),距离衰减(power decay of distance)等信息。

在刻画多径传播中,无线信道常用信道冲击响应(Channel Impulse Response, CIR)建模。

在线性时不变的假设下，CIR 可以表示为：

$$h(\tau) = \sum_{i=1}^N \alpha_i e^{-j\theta_i} \delta(\tau - \tau_i)$$

其中， $\alpha_i$ ， $\theta_i$ ， $\tau_i$  分别为第  $i$  条路径的幅度衰减、相位偏移和时间延迟， $N$  为传播路径总数， $\delta(\tau)$  为狄克拉脉冲函数。式中的每一项从时域上表示了一条传播路径的幅度、相位和时延。由于多径传播在频域上表现为频率选择性衰落，因而也可以通过信道频率响应（Channel Frequency Response, CFR）刻画多径传播。CFR 包括幅频响应和相频响应，在无限带宽的条件下，CFR 和 CIR 互为傅里叶变换。

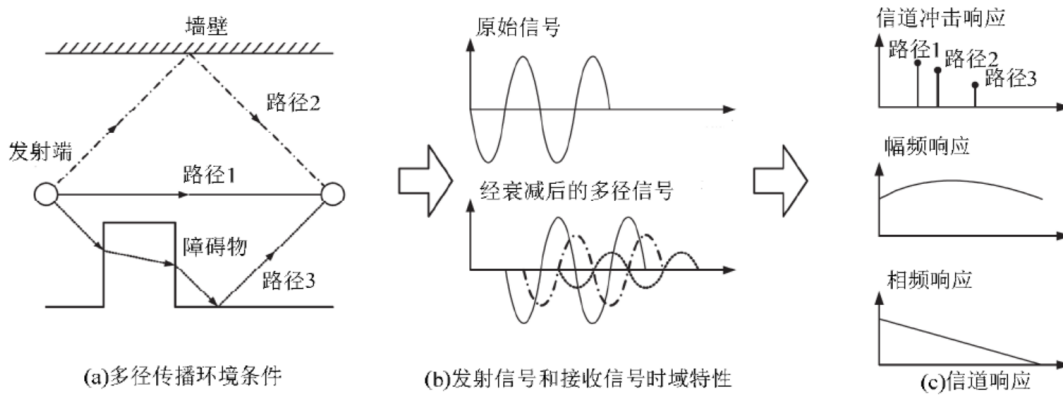


图 2.1 多径传播环境条件下时域、频域特征示意图

精确测量 CIR/CFR 需要专业的信道测量仪器，例如网络矢量分析仪。而自 2010 年以来，研究人员通过修改固件，使得在普通的 Wi-Fi 设备上也能以信道状态信息 CSI 的形式获取一个采样版本的 CFR。具体而言，利用兼容 IEEE 802.11 a/g/n 的无线网卡即可从每个接收数据包获取一组 CSI，每组 CSI 代表一个正交频分复用（Orthogonal Frequency Division Multiplex, OFDM）子载波的幅度和相位。

## 2.2 Wi-Fi CSI 特性分析

由于 CSI 作为物理层信息，包含了诸多介质访问控制（Medium Access Control, MAC）层不可见的信道信息。一方面，CSI 可从一个数据包中同时测量多个子载波的频率响应，而非全部子载波叠加的总体幅度响应，从而更加精细地刻画频率选择性信道。另一方面，CSI 既可测量每个子载波的幅度，还可测量每个子载波的相位信息。CSI 将单值的 RSSI 扩展至频域，并且附加了相位信息，从频域上为无线感知提供了更为丰富、细粒度的信道状态信息。

信号状态信息 CSI 与接收信号强度指示 RSSI 都反映信道质量的好坏。RSSI 的强弱在

一定程度上反映了信道的好坏，众多通信技术如 RFID，FM，GSM，Wi-Fi，ZigBee 等均可在终端设备上获取 RSSI 信息，以根据当前信道质量调整通信策。但与 RSSI 相比，CSI 信息具有如下优势：

(1) CSI 相较于 RSSI 更加细粒度。CSI 可以体现出子载波的频率信号，而 RSSI 只能反应子载波叠加后信号的强度；CSI 在一定程度上刻画了多径传播，RSSI 只是反应了多径叠加的总幅度。我们可以形象的把 RSSI 比作一束白光，那么 CSI 即可看作以正交频分复用为棱镜色散出的光谱，每一束单色光（对应正交频分服用中不同的子载波）都呈现了不同频率下多径传播的幅度和相位。

(2) CSI 比 RSSI 更具有稳定性和可靠性。制约 RSSI 稳定性和可靠性的根本因素是 RSSI 测量的是信号多径传播的叠加效果，并不能逐一区分多条信号传播路径。RSSI 会因为信号多径传播引起的小尺度阴影衰落而不再随传播距离增加单调递减，从而限制测距精度。而 CSI 对于不同的传播环境可呈现不同的子载波幅度和相位特征；而对于相同的传播环境，CSI 的整体结构特征更容易保持相对稳定。

## 2.3 Wi-Fi CSI 的采集

采用 Intel 5300 网络接口卡(NIC)作为无线收发器记录信道状态信息，与只能获取到 RSSI 的其他网卡不同，支持 802.11n 标准的 5300 网卡可以在 Wi-Fi 信号的发送和接收方之间记录信道状态信息，其能提供 OFDM 子载波的信号幅度和相位。每个天线流的 CSI 由 30 个可读子载波组成。每个组是一个包含两个复数形式的正交信号的 OFDM 子载波。WL5300 由两个不同带宽的工作模式，一种模式使用 20MHz 信道和 56 组子载波，另一种模式使用 40MHz 信道和 114 组予载波。30 个可读 CSI 均匀的分布在这 56 个或 114 个组中。

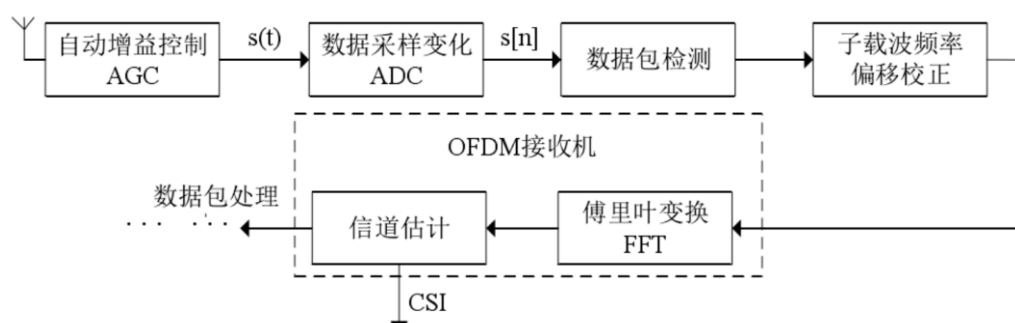


图 2.2 基于 802.11n 标准的 CSI 获取流程

在本项目的系统中, 5300 网卡安装在台式机上, 同时该台式机安装了 Ubuntu12.04 32 位系统。CSI-tool 由于采用了开源 iwlmwifi 无线驱动程序, 因此可以启用调试模式, 从而允许 NIC 将 CSI 报告给主存储器。

在安装了相应的硬件和软件设备之后, 通过连接无加密的 AP (本项目采用一个普通商用无线路由器作为 AP), 然后用 ping 命令在电脑和路由器之间建立会话, 就能处理收到的数据包以及记录相应的 CSI。从所有的数据包记录的 CSI 将被写入一个文件中, 这个文件可以通过定制的 MATLAB 脚本读取以进行处理。首先, 由于原始 CSI 报告采用二进制文件格式, 因此利用 MATLAB 中的 mex 编译 C 程序读取二进制 CSI 数据。处理结果的解压格式是在单元中压缩的 n 个结构。有 n 个正确接收的数据包, 就会对应于有相等 n 个结构的 CSI 分组, 每个分组包含天线参数以及原始 CSI 值。然后使用这些参数来计算经校正的原始 CSI 值的归一化 CSI 值。在每个正确接收的数据包中, 不仅有 CSI 值而且接收数据包的天线参数都保存在报告中。其数据格式将在下一节进行分析。

## 2.4 Wi-Fi CSI 数据信息分析

本项目采用两天线无线路由器, 以及三天线无线网卡, 而每对收发天线可以获得 30 个子载波, 因此对于每个数据包, 实际上应该为  $2 \times 3 \times 30$  的复数形式, 共得到 180 维组成的 CSI 数据矩阵, 获得的 CSI 数据结构如图 2.3 所示。

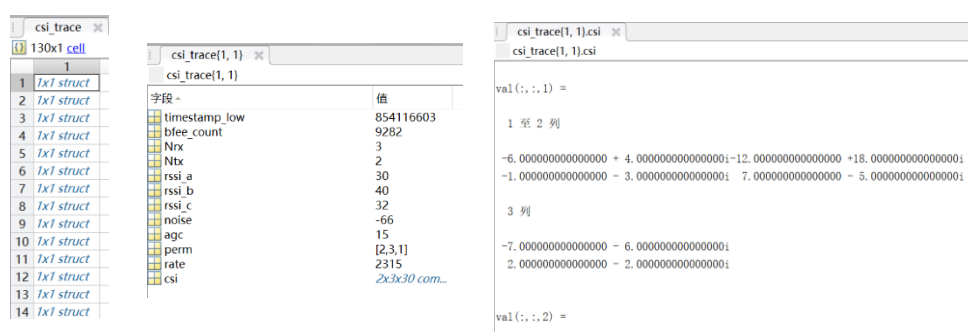


图 2.3 CSI 的数据格式

从图 2.3 所示 CSI 的数据结构形式可以看出, 在所有接收到的数据包文件中, 首先使用 read\_bf\_file 函数来对接收到的数据包进行读入操作, 进一步得到 csi\_trace 文件, 将提取出的每个数据包存储在 csi\_entry 结构体中, 以数据包 csi\_trace{1,1} 为例。每个数据包中包含有收发天线数、RSS 值、每个数据包发送的速率、噪声以及信道状态信息 CSI 等, 通过这些数据信息可以反映无线链路的主要性能, 具体信道状态信息的数据信息以及所表达的属性, 如表 2.1 所示。



表 4.1 信道状态信息的数据特性及其属性

数据特性	属性
timestamp_low	无线网卡 1MHz 时钟的低 32 位,每 4300 秒或者 72 分钟更新一次
bfee_count	总的波束成形测量次数, 由驱动记录后发送给用户空间
Nrx	无线网卡用来接收数据包的天线 (Intel 5300 通常为 3)
Ntx	发送数据流的天线数量 (本文采用 2 天线无线路由器发送)
rss_i_a, rss_i_b, rss_i_c	分别对应每个发射端到接收端之间无线网卡测量的 RSSI,单位是 dB
noise	噪声
agc	自动增益控制
perm	表示接收天线对得到信号进行处理时的射频链路过程, 如[1 2 3]表示天线 A 对应链路 A, 天线 B 对应链路 B, 天线 C 对应链路 C
rate	每个数据包的发送速率, 是一个 16 位数, 如 0x8463
csi	信道状态信息, 为 $N_t \times N_r \times 30$ 的复数矩阵

利用 `get_scaled_csi` 函数从 `csi_entry` 中提取出 CSI 信息, CSI 为一个  $2 \times 3 \times 30$  的复数矩阵, 矩阵中总共有 30 维数据, 表示 OFDM 下的 30 个子载波信息, 每一维数据中包含有 6 个复数, 分别代表的是天线 1、2、3 上的子载波的实部和虚部, 可以反映接收端天线上的 RSS 和 CSI 值分布情况, 如表 2.2 所示。由表 2.2 可知, CSI 数据信息以  $2 \times 3$  的复数矩阵形式呈现, 每根天线上包含 30 个子载波信息, 描述了无线信号传输过程中的振幅和相位等信息。

天线 1						天线 2						天线 3					
子载波	...	子载波		子载波		子载波	...	子载波		子载波		子载波	...	子载波		子载波	
1		30				1		30				1		30			
实部	虚部	...	...	实部	虚部	实部	虚部	...	...	实部	虚部	实部	虚部	...	...	实部	虚部

由上述可知, 所采集到的 CSI 为多维矩阵, 需要对其进行降维以便后续处理, 每个数据包中都包含 6 个不同数据流, 而每个数据流又由 30 个子载波构成, 表示为二维矩阵的形式, 如计算公式如下所示:

$$CSI = \begin{bmatrix} H_{1,1} & H_{1,2} & \dots & H_{1,30} \\ H_{2,1} & H_{2,2} & & H_{2,30} \\ \vdots & & \ddots & \vdots \\ H_{6,1} & H_{6,2} & \dots & H_{6,30} \end{bmatrix}$$

在 CSI 矩阵中, 每个子载波均以复数形式存在, 即表示为  $a + bi$ 。根据复数相关运算法则, 可以求得该复数的模值为  $\sqrt{a^2 + b^2}$ , 幅角值为  $\theta = \arctan \frac{b}{a}$ 。因此, 可将该值作为  $H(f_k) = \|H(f_k)\| e^{j\sin(\angle H(f_k))}$  的振幅和相位, 就是说可以通过 CSI 复数矩阵, 得到每条链路上各子载波的振幅和相位信息, 进一步提取其他特征。

## 3 概要设计

### 3.1 问题描述

本项目解决的问题描述如下：给定一个常见的具有安防需求的室内无线网络环境，比如居民房、档案室、会议室等，这些环境中已经部署安装了用于网络通信的无线网络基础设施，包括无线接入点、无线设备等。无线接入点通常为能够发射无线信号的无线路由器，也可以是具备无线网络共享功能的设备，比如智能手机、笔记本电脑以及配备无线网卡的台式电脑等。无线设备是指能够通过无线网络发送和接收数据，并且具备一定数据处理能力的设备，常见的智能手机、笔记本电脑以及配备无线网卡的台式电脑等都属于这类无线设备。我们需要解决的问题便是充分利用现有的这些无线接入点和无线设备等基础设施，不借助于其它特殊设备，实现对该环境中异常“目标”的感知。“目标”的定义为能够干扰无线信号并且具备移动能力的对象。为了简化问题，本文特指在无线网络环境中进行各种活动的人。

本项目问题的提出基于以下事实：一个目标，比如人，如果出现在一个充满无线信号的环境中，将不可避免的对无线信号产生干扰，即使这个人并未站在信号发射器和接收器的 LOS 路径上遮挡信号。这些信号包括很大频段范围，如在无线网络里普遍使用的 2.4GHz 和 5GHz 的 Wi-Fi 和 Wi-Max。通过对无线传播特性的了解，我们可以知道一个人出现在无线网络环境中之所以会影响无线信号，一方面是由于人体会吸收无线信号，另一方面是由于多径效应。

### 3.2 Wi-Fi alarm 系统模型

根据前文的问题描述可以知道，本项目是对无线网络环境中异常目标的出现进行监测，因此将本项目系统形象地命名为“Wi-Fi alarm”，意为 Wi-Fi 报警器。

结合应用环境特点和需要解决的问题，一个 Wi-Fi alarm 系统由以下三部分组成：（1）信号发射器（Transmitter），比如在典型的无线网络环境中的无线路由器，以及具有热点功能的智能手机、笔记本电脑等；（2）监测点（Monitoring Point, MP），比如笔记本电脑等；（3）应用服务器（Application Server, AS），用于处理和分析监测点采集到的数据。在本项目实验中，信号发射器为无线路由器，标记为 AP；监测点为配备了无线网卡的台式电脑，标记为 MP；应用服务器为本项目实验所使用的笔记本电脑，标记为 AS。

本系统的整体框架如下图所示：

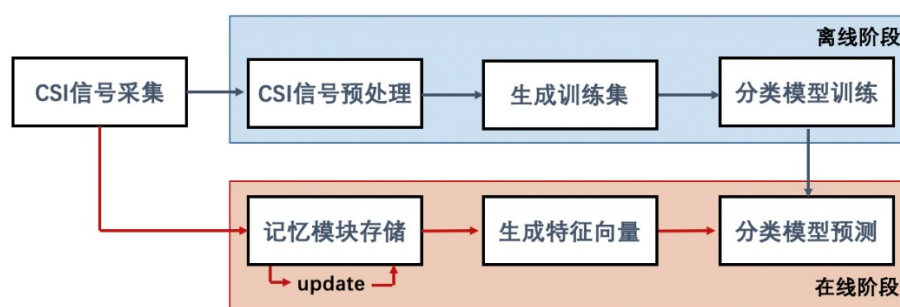


图 3.1 Wi-Fi alarm 系统整体框架

整个系统的工作流程分为离线阶段和在线阶段：

（1）离线阶段：主要工作内容是在室内环境中 AP 和 MP 部署完成后，MP 实时采集信号信息，包括 CSI 数据、环境状态等，然后对这些信号信息进行处理并从中提取特征，利用机器学习算法训练出分类模型。

（2）在线阶段：主要工作内容是 MP 实时采集无线信号信息，将信号发送给 AS。AS 利用训练好的分类模型对信号进行实时预测，并根据预测值给用户相应的反馈。

在下一部分的内容中将详细阐述本项目各功能模块的具体实现。

## 4 详细设计

### 4.1 CSI 信息采集与实时分析

由于无线网络技术的快速发展，项目只需常见的 Wi-Fi 设备等基础设施，但是为提取物理层更细粒度的信道状态信息，需要对作为接收端的计算机进行包括操作系统、无线网卡驱动等特殊设置，才能采集得到可用的 CSI 信息。然而，目前市面上获取 CSI 信息的设备除了软件无线电等，普通的无线网卡暂时不具备获取 CSI 数据信息的能力，只有少数可以满足 CSI 的提取，如 Intel 5300 等部分无线网卡。因此，可以采用开源项目 Linux 802.11n CSI-Tool 测试平台，对物理层的 CSI 数据进行采集，并提供后续的处理支持。

#### 4.1.1 硬件环境

主要包括：支持 802.11n 标准的无线路由器；配置 Intel 5300 网卡的计算机；普通笔记本电脑。

（1）支持 802.11n 标准的无线路由器：作为信号发射器 AP，发射无线信号。

本项目选用市面上常见的 MERCURY MW313R 无线路由器,支持 2.4GHz、5GHz 并发,支持 IEEE 802.11n 标准。实验主要采用 2.4GHz, 157 无线信道, 20MHz 带宽的信号。

(2) 配置 Intel 5300 网卡的计算机: 作为监测点 MP, 接收无线信号, 并对其中的 CSI 信息进行提取和转发。

本项目选用安装了 Intel 5300 网卡的 GIGABYTE 台式计算机。

(3) 普通笔记本电脑: 作为应用服务器 AS, 对接收端转发的 CSI 信号进行实时处理。

本项目选用 ThinkPad S3 笔记本电脑。



(a) Intel Wi-Fi link 5300



(b) MERCURY MW313R 无线路由器

图 4.1 硬件环境所需设备

## 4.1.2 软件环境

(1) 配置 Intel 5300 网卡的计算机: 作为本项目的采集软件平台, 首先需要搭建特定的系统——限定内核版本范围的 Linux 系统, 这里安装的是 Ubuntu 12.04 版本; 然后需要修改内核驱动和网卡驱动; 最后安装支持 IEEE 802.11n 协议的 Linux 802.11n CSI-Tool 的 CSI 采集工具。(详细配置与安装过程见“附录一”)

本项目 MP 运行的程序为项目目录下的 netlink 工程, 对其进行 CMake 之后可得到可执行文件 log\_to\_file 和 log\_to\_socket。其中 log\_to\_file 可以将 CSI 信息打包成文件, 采集频率和采集时间可以人为控制; log\_to\_socket 可以将 CSI 信息通过 TCP socket 实时发送给分析方, 发送包的频率可以通过参数控制。

(2) 普通笔记本电脑: Linux 802.11n CSI-Tool 采集工具也提供 MATAB 平台下的处理技术, 用于对采集到的原始 CSI 数据信息进行处理, 本项目在其基础上编程实现了预期功能。做为 AS 的笔记本电脑首先要安装 MATLAB, 这里安装的是 MATLAB R2018b。

本项目 AS 运行的程序为项目目录下的 svm\_train 工程和 realtime\_wifi\_alarm 工程。其中 svm\_train 工程离线运行, 用于 SVM 分类器的训练; realtime\_wifi\_alarm 工程实时运行, 通

过 TCP socket 接收实时 CSI 信息并进行处理，之后做为模型的输入得到预测值，并实时显示在界面中（监测到有人背景显示红色，无人时背景显示绿色）。

### 4.1.3 训练数据采集

选取的实验场所是 XXX 活动室。活动室内由会议桌、椅子、书架、值班台等杂物，实验环境较为复杂，符合实际室内的场景要求。我们选取门口较为开阔的方形区域作为实验监测环境，作为 AP 的无线路由器和作为 MP 的台式电脑放置在监测环境对角线位置，相距约 6m。

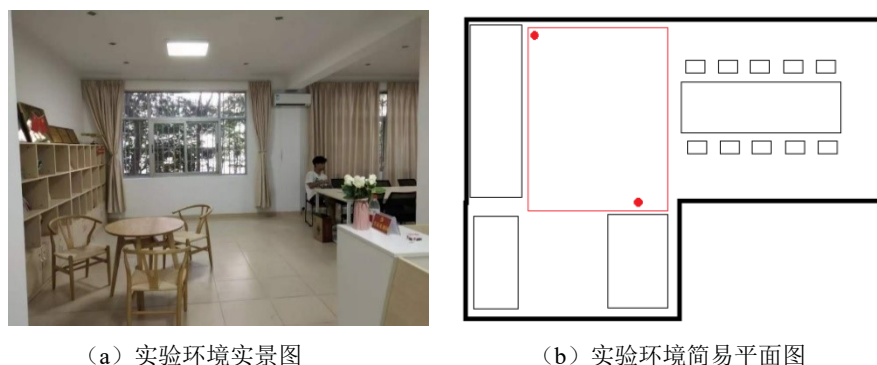


图 4.2 本项目实验环境

采集数据主要分为空环境、环境内有人慢走、环境内有人快走三种类型的数据。为保证训练数据充足，在不同时间的空环境下采集共 12 组数据，环境内有人慢走、环境内有人快走各采集 3 组数据，每次采样时长为 30s，总共采集 18 组数据。



图 4.3 训练数据采集

## 4.2 CSI 数据预处理与特征选取

通过对原始数据结构分析可知，CSI 信息主要包括两部分：幅值信息和相位信息。但是因为未处理的相位的随机性和不可预测性，选取 CSI 的幅度作为特征比较简单。但 CSI 的原始幅值信息因受环境及物理干扰的影响，在时域范围内含有较多的噪声，会对环境的监测

以及算法实现复杂度造成不利影响。因此，需要对原始的幅值信息进行预处理。

### 4.2.1 CSI 数据的预处理

本项目对 CSI 数据预处理的核心思路在于提取环境主信道信息。基于本项目的实验环境，环境监测区域对角线为 AP 到 MP 的直线传播路径，可将其称之为本次实验环境中无线信号传播的主信道。主信道有如下两个特点：(1)主信道是直线传播路径，受环境信号散射、环境衰弱等因素的干扰很小，具有能量大、噪声少的优点；(2)主信道受环境状态影响最大，适合作为分类特征。基于主信道信息上述两点优势，本项目对原始 CSI 信息进行滤波，提取主信道信息。

根据 2.1 节的分析，无线信道在时域下用 CIR 建模，即表示为多个冲激函数的叠加，如图 2.1 所示。CFR 为 CIR 的傅里叶变换，即无线信道频率响应，而 CSI 为 CFR 的一个采样版本，由图 2.2 也可看出，获取 CSI 信息前进行了傅里叶变换。基于以上原理，可设计滤波算法流程如下：(1)对原始 CSI 信息进行傅里叶逆变换，得到 CFR 信息；(2)选 CFR 信息中能量最大值的 20%作为阈值进行滤波，滤掉能量小于  $\max \times 20\%$  的部分（剩余部分即为主信道信息）；(3)进行傅里叶变换，将 CFR 还原为 CSI 信息。

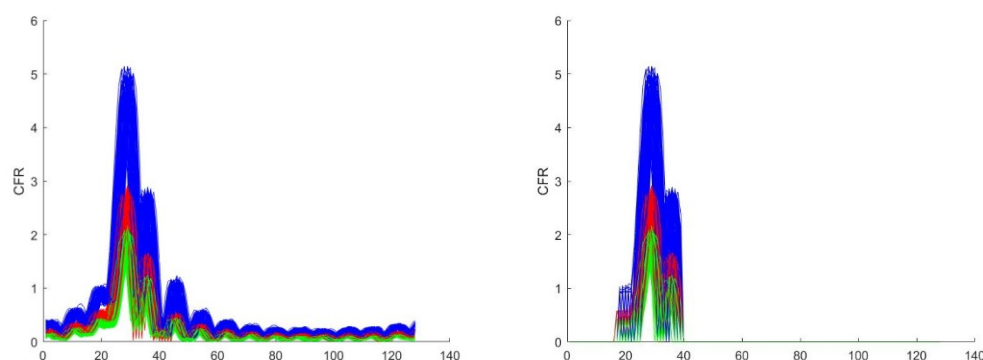


图 4.4 滤波前后的 CFR 信息对比

从图 4.4 可以看出，CFR 信息为多个类冲激函数叠加，其中有明显峰值体现了主信道信息，其余部分为多径效应下的噪声，将其滤除后进行傅里叶变换，得到结果如图 4.5 和图 4.6 所示。可以看出，CSI 幅值有明显平滑，当环境状态为有人时，幅值有明显变化，可作为分类特征。



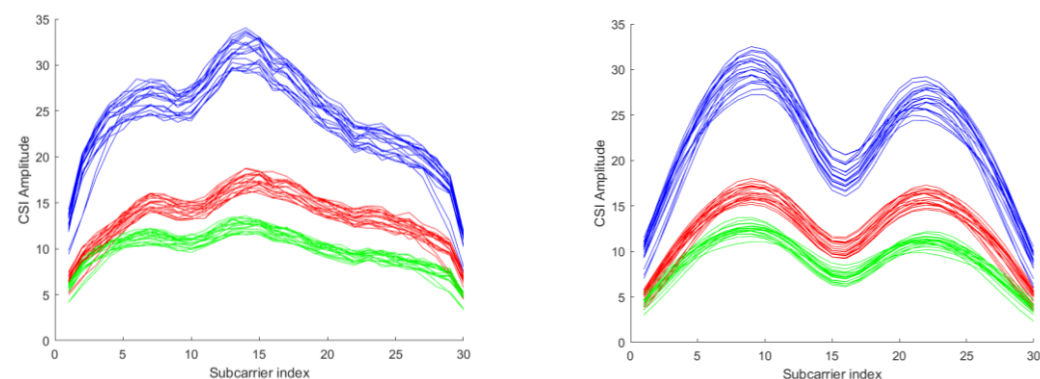


图 4.5 空环境下预处理前后的 CSI 幅值对比

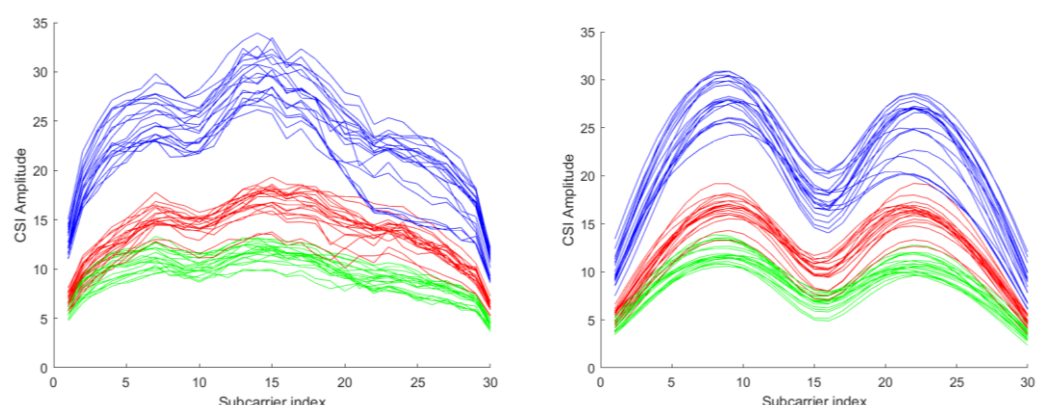


图 4.6 有人环境下预处理前后的 CSI 幅值对比

## 4.2.2 特征的选取

经过数据预处理的信道状态信息，可以提取很多不同的特征，因为未处理的相位的随机性和不可预测性，这里简单起见，选取 CSI 的幅度作为特征。在监测区域内，环境状态发生变化时，幅度等时域特性会发生一定变化。在特定时间窗口内，将提取出的这些特征量，送到分类识别单元，进而可以对环境状态进行监测。

Intel 5300 网卡配有三条天线，即使对于相同的分组接收，三个天线的信道频率响应也有很大不同，这被称为 CSI 的空间分集特性。根据这一特性，采用多天线数据流组成特征有利提高分类精度，所以选取的是一条发送天线，三条接收天线的的数据流，即每个样本有  $1 \times 3 \times 30$  个 CSI 幅度值。

另一个重要的问题是，如何将这些样本组成特征向量作为分类模型的训练输入，这里有两种方案。（1）将单个样本展开成维度为(90,1)的列向量作为特征输入。该方案的好处是实现简单，实时监测时只需对将采集到的单个样本进行预测即可。但该方案没用充分利用 CSI 信息随时间变化的特征信息，可能导致分类精度的降低。另外，由于监测点 MP 与应用服务

器 AS 之间的通信延迟以及 AS 的处理响应速度有限，单样本预测可能会出现排队时延，破坏监测实时性，即可能出现目标进入环境时不能及时反馈的情况。(2) 将连续时间的  $n$  个样本展开成维度为  $(n \times 90, 1)$  的列向量作为特征输入。这样的好处是利用了 CSI 信息随时间变化的特征，并且降低了对实时监测响应速度的要求。理论上分析，方案二更优，下面还会通过实验进行进一步论证。

### 4.3 模型训练与预测

通过前面的步骤得到了样本数据和特征之后，接下来就是分类模型的选择和训练。由前面的问题描述可知，本项目是一个简单的二分类问题，即对环境空与有人两种状态进行分类预测。本项目测试不同的样本输入组合方案，采用多种模型进行训练，然后评估其误差，然后选取误差最小的模型和样本组合方法对数据进行预测。本项目最终使用支持向量机 (Support Vector Machine, SVM) 对数据集进行训练和分类。

#### 4.3.1 机器学习算法的选择

目前，解决分类问题的方法很多，单一的分类方法主要包括：朴素贝叶斯 (Bayes)、决策树 (Decision Tree)、逻辑回归 (Logistic Regression)、K-近邻、人工神经网络、支持向量机 (Support Vector Machine, SVM) 和基于关联规则的分类等。需要对各种分类模型进行预测，评估误差，选取最优的分类模型。另外，根据 4.2.2 节的分析，特征组成的方案对分类模型的效果也有较大影响，因此这里还需要测试不同的样本输入组合方案。

这里对于采集的 18 组数据，每组数据选用 96 个样本值，分别选用单样本组成和  $n$  分别为 2、4、6、8 的多样本组成进行实验。在构造模型之前，要把数据集随机分成训练数据集和测试数据集，一部分用于训练，另一部分用于测试，这里按照 25% 的比例划分训练集与测试集。然后对常见的分类模型进行训练，实验结果如表 4.1 所示。

表 4.1 不同特征组成方案

单特征点 样本数	特征点数量	特征向量维度
1	1728	(90, 1)
2	864	(180, 1)
4	432	(360, 1)
6	288	(540, 1)
8	216	(720, 1)



表 4.2 不同模型与不同特征组成方案比较

(a) 正确率分析						
	1	2	4	6	8	Ave.
Tree	91.0	95.3	90.5	91.7	87.5	91.2
Liner Discriminant	99.3	100.0	95.8	97.6	98.6	98.3
Logistic Regression	100.0	100.0	94.0	94.8	97.2	97.2
SVM	100.0	100.0	100.0	100.0	100.0	100.0
KNN	98.1	99.2	97.5	99.0	97.7	98.3
Ave.	97.7	98.9	95.6	96.6	96.2	
(b) 预测速度 (obs/sec)						
	1	2	4	6	8	Ave.
Tree	20000	16000	14000	2500	1000	10700
Liner Discriminant	19000	17000	10000	1900	1000	9780
Logistic Regression	7600	7500	3500	900	440	3988
SVM	17000	14000	12000	1300	1200	9100
KNN	9300	7800	9200	1700	1300	5860
Ave.	14580	12460	7940	1660	988	

从表 4.2 所示的实验结果可以看出：在模型的选择上，支持向量机 SVM 表现最优，测试精度达到 100%。在单特征点样本数的选择上，每个特征点 2 个样本的选择使得测试精度高达 98.9。最终选用双样本值组成特征值，用 SVM 进行模型训练和预测。

### 4.3.2 支持向量机 SVM 分类

支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面，SVM 可以通过核方法进行非线性分类。

由于特征是由临近时序的样本值组成的，因此，在实时监测阶段，需要设置一个记忆模块，用以保存临近的样本值，每次预测时取出最近样本值按特征组成方法组成特征进行预测，系统将预测结果实时显示在界面上。

## 5 实验结果与分析

### 5.1 CSI-tool 通信测试

测试数据文件采集和实时环境监测两种模式下的通信情况。

#### (1) 数据文件采集（离线阶段）

监测点 MP 向路由器发送 PING 请求，从而获取包含 CSI 的数据包，用于后续解析，如图 5.1 所示。

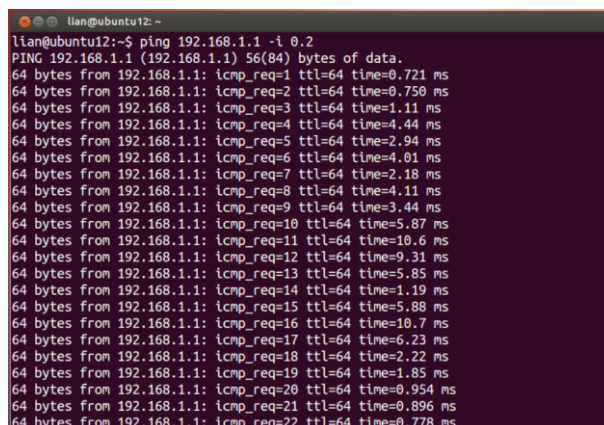


图 5.1 监测点 MP 向路由器发送 PING 请求

运行 `log_to_file [filename.dat]`，监测点从采集到的数据包中解析出 CSI，保存至 `filename.dat` 文件。运行时命令行输入如图 5.2 所示。

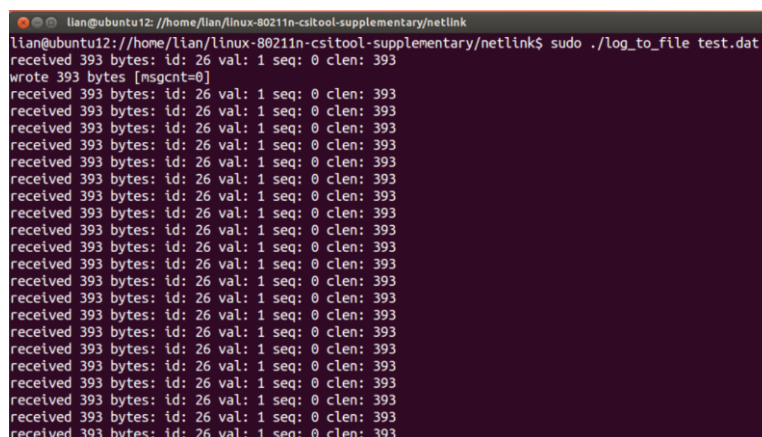


图 5.2 采集本地训练数据

### (2) 实时环境监测 (在线阶段)

监测点 MP 向路由器发送 PING 请求获取包含 CSI 的数据包过程与 (1) 中相同, 如图 5.1 所示。

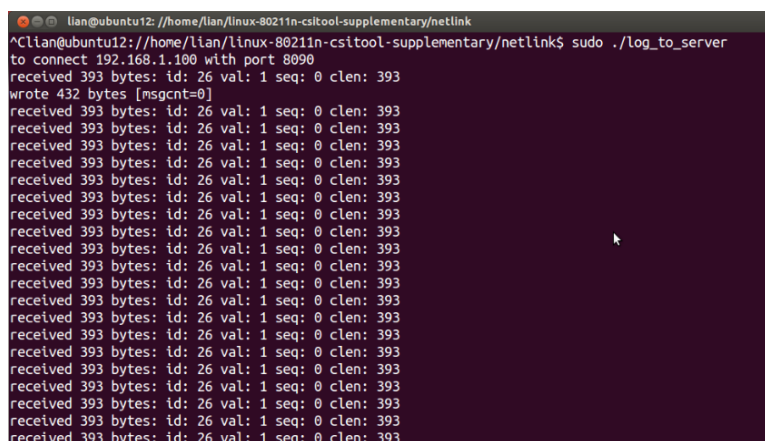


图 5.3 实时环境监测

运行 `log_to_sever`，监测点从采集到的数据包中解析出 CSI，并实时内容传送给应用服务器 AS，命令行输出信息如图 5.3 所示。将发送速率调整得当，则 AS 可以实时反应 CSI 变化。

## 5.2 不同环境下的 SNR 分析

通过对空环境和有人环境中的信噪比（SIGNAL-NOISE RATIO，SNR）测试分析，每个天线上各数据流的 SNR 在一定程度上可以反映环境有人无人状态，如图 5.3 所示。

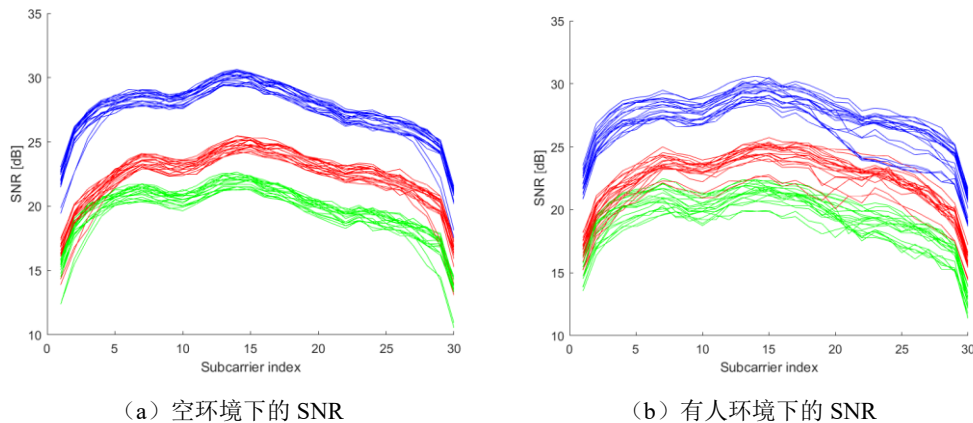


图 5.4 空环境和有人环境下的 SNR 对比

从图 5.3 (a) 可以看出，不同子载波下的 SNR 不同，当环境为空时，三天线的 SNR 值基本稳定，并且相同天线上各数据流之间的 SNR 值变化保持基本一致，不同天线上各数据流之间的 SNR 值变化趋势保持相似。从图 5.3 (b) 可以看出，当有人进入环境时，相天线上个数据流之间的 SNR 值变化仍基本一致，而不同天线上各数据流之间的 SNR 变化却不一样。这是因为在空环境和有人环境时，人体对无线信号幅度衰减、相位偏移和时延等影响不同，可将 SNR 作为环境监测的依据。

## 5.3 RSSI 与 CSI 稳定性对比分析

为验证 RSSI 易受多径效应影响的缺陷，CSI 在多径环境下更稳定，对 RSSI 和 CSI 的信号特性进行对比实验，并对其特性做出分析。首先，在静止且无干扰的室内环境，分别采集 30s 的 RSSI 和 CSI 数据。然后，对 CSI 中某一天线的 30 个子载波计算其幅值的均值，并计算 RSSI 的均值。利用采集到的 CSI 数据分别提取出 RSSI 和 CSI 的幅值信息，观察分析两者的区别。由于多径效应的影响，RSSI 受此影响较大，而 CSI 则呈现子载波特性，受此影响较小，如图 5.4 所示。

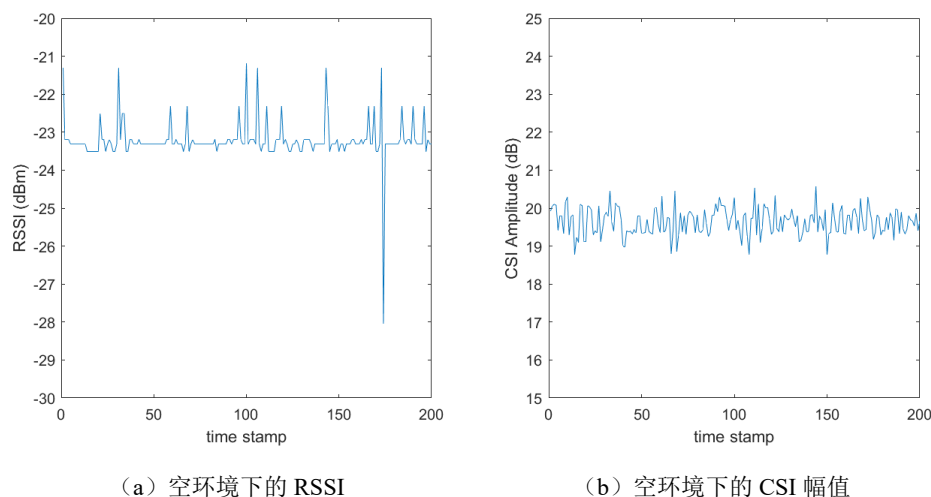


图 5.5 空环境下的 RSSI 和 CSI 幅值稳定性对比

在图 5.4 中所示, 图(a)显示 RSSI 随时间的变化情况, 波动较大, 存在 5dB 以上的波动, 图(b)显示 CSI 随时间的变化情况, 波动较小。因此, RSSI 自身存在的不稳定性因素导致进一步的系统误差, 相对 RSSI 而言, CSI 在静止的室内环境下稳定性更好, 可以进一步用来区分环境状态等感知应用。

## 5.4 数据处理实时性分析

作为检测系统, 实时性的要求极高。如果监测点 MP 的数据发送速率过大, 则会造成应用服务器 AS 接收到的数据过多, 来不及处理。这样会导致当前环境状态改变时, 应用服务器 AS 需要经过一段时间才会做出响应, 这在检测系统中是不允许的。通过修改 CSI 发送程序, 使得检测点 MP 在捕捉到多个包后, 才向应用服务器 AS 发送一次数据, 从而控制数据率。通过减小发送速率, 使得应用服务器 AS 能够与监测点 MP 保持实时同步, 在此基础上, 适当增大发送速率, 使得检测反应较灵敏。

通过调整 SLOW\_MSG\_CNT 调整发送速率, 原程序中 SLOW\_MSG\_CNT 值为 1, 表示监测点 MP 将收到的每个包都向 AS 发送, 将 SLOW\_MSG\_CNT 值调整为 4, 即将发送速率减小为原来的 1/4, 此时发送速率得当, 则 AS 可以实时反应 CSI 变化。

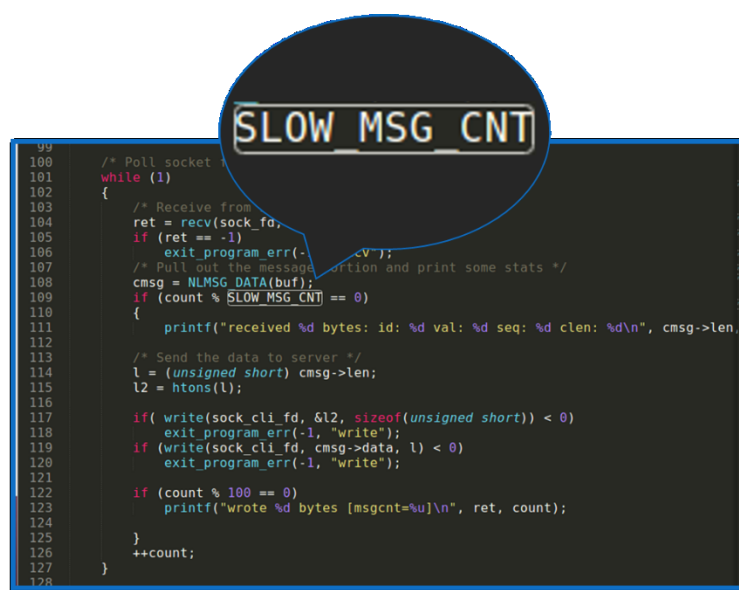
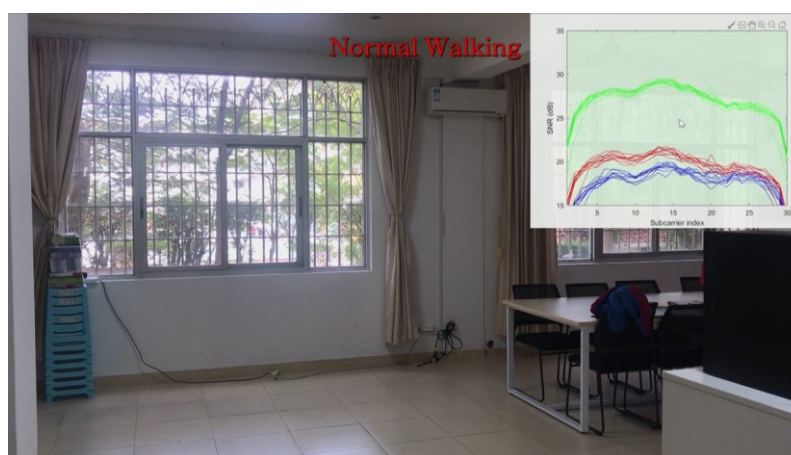


图 5.6 发送速率控制参数

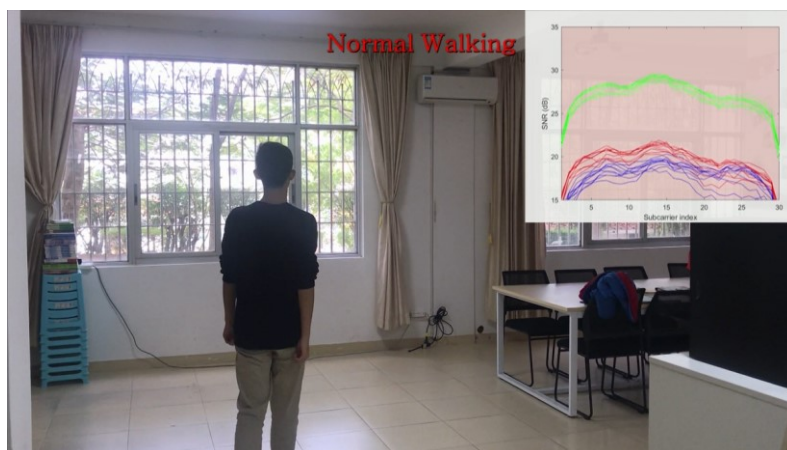
通过调节监测点 MP 向应用服务器 AS 发送数据速率，最终能够实现应用服务器 AS 对当前变化做出实时且灵敏的反应。当目标进入环境后，信噪比指标变化剧烈，检测程序背景迅速由静止状态的绿色变为红色。当目标离开环境后，信噪比指标迅速收敛，检测程序背景迅速变回绿色。

## 5.5 环境状态实时监测结果分析

本实验对四种情况进行测试：无人时的空环境、有人缓慢经过、有人快速经过、有人在环境中活动（时而行走时而静止）。具体实验效果见本项目演示视频。从结果，可以看到本系统可对环境异常（这里特指有人进入环境）作出实时且灵敏的反应。



(a) 空环境



(b) 有人环境

图 5.7 环境状态实时监测结果

## 6 总结与展望

### 6.1 总结

本项目首先对 Wi-Fi CSI 相关技术理论进行探究，比较了 RSSI 和 CSI 的性质特点，并利用现有无线设备和特定版本的无线网卡，通过采集环境内人的活动影响 Wi-Fi 信号的变化，来监测环境异常（本项目中特指有人出现在环境中），并能够实时且灵敏的反应。主要工作包括以下几个方面：

#### （1）利用普通商用 Wi-Fi 设备获取信道状态信息

从当前常用的 Wi-Fi 设备获取特征信息，包含反映室内环境状态信息的信道状态信息，对其进行采集，并分析其特性。

#### （2）对来自信道状态信息的数据特征进行提取

首先，因多径效应的影响会造成无线信号传播的变化，对采集到的含有较多环境噪声等干扰的信道状态信息，本项目使用提取主信道信息的预处理方法，提取出可以明显反应环境状态变化的特征信息。

#### （3）基于信道状态信息的实时环境监测方法研究

根据提取到的环境状态特征，从实时检测到的 Wi-Fi 信号中感知环境状态。通过调整相关参数，使系统兼具实时性、灵敏性。

#### （4）基于 Wi-Fi CSI 的环境状态分析

通过分析信道状态信息因环境变化而引起的异动，分析这些变化信息的特征，采用多样本组合的特征组成方式，使用选取 SVM 进行训练，实现环境状态的判别。



## 6.2 展望

Wi-Fi CSI 作为反映环境状态的稳定的、细粒度的信息，其功能十分强大，是当前物联网领域的研究热门。本项目从最基础实验环境开始搭建，限于学期时长，无法更加深入利用 Wi-Fi CSI 做难度更大的工作。但本项目具有很好的可扩展性，从任务目标、环境搭建、数据处理、训练模型等方面都可以做更加深入的工作：

（1）任务目标：本项目仅实现最简单的有人无人监测，后续可以延续这个思路进行人体姿态、室内定位等工作；

（2）环境搭建：本项目使用台式机作为监测点，实验环境搭建难度大，后续可考虑笔记本电脑外接网卡；

（3）数据处理：本项目利用理论原理，仅提取出主信道信息，对于其他信息利用不足，后续可研究其他滤波降噪方法，提高 CSI 信息利用率；

（4）训练模型：本项目由于任务相对简单，采用 SVM 作为训练模型，后续面向更加复杂的任务，可以考虑使用深度学习的方法。

## 参考资料

- [1] 杨 铮 刘云浩.Wi-Fi 雷达：从 RSSI 到 CSI[J].中国计算机学会通讯,2014,10(11):55-60.
- [2] Qian K , Wu C , Yang Z , et al. PADS: Passive Detection of Moving Targets with Dynamic Speed using PHY Layer Information[C]// Proceedings of the 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2014.
- [3] Xuyu Wang<sup>†</sup>, Lingjun Gao<sup>†</sup>, Shiwen Mao<sup>†</sup>, et al. DeepFi: Deep Learning for Indoor Fingerprinting Using Channel State Information[C]// 2015 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2015.
- [4] 岳国玉. 基于 WiFi 信号的室内人体动作检测研究及应用[D].
- [5] 郑耀. 基于 WIFI 信号的室内定位技术研究[D].
- [6] 邓晓华. 基于 CSI 的被动式室内定位与目标计数方法研究[D]. 杭州电子科技大学, 2014.

## 附录一：网卡驱动配置与 Linux 802.11n CSI tool 安装过程

### 1、操作系统的安装与软件源配置

- a) 采用的操作系统为 Ubuntu12.04 64 位，点击[下载链接](#)进入下载页面，选择



ubuntu-12.04.5-alternate-amd64.iso

Ubuntu 系统安装过程比较简单，在此不再赘述。

- b) 将 Ubuntu 系统软件源更换为国内源，推荐使用网易(163)源。  
c) 执行命令 `sudo apt-get update`

### 2、安装依赖

- a) 执行  

```
sudo apt-get -y install git-core kernel-package fakeroot build-essential
ncurses-dev
```

 如果出错考虑换源问题。  
 b) 执行 `sudo apt-get -y install libnl-dev libssl-dev`  
 c) 执行 `sudo apt-get -y install iw`

### 3、下载并编译内核

- a) [下载地址](#)，在 Ubuntu 系统 Home 目录下，解压刚刚下载的文件，解压命令：  

```
tar -xvf intel-5300-csi-github-master.tar.gz
```

  
 b) 切换至解压后的文件夹下：  

```
cd intel-5300-csi-github
```

  
 c) 执行命令 `make oldconfig` 此处需要注意，过程需要多次确认，一直按住回车键即可。  
 d) 执行命令 `make menuconfig` 会弹出一个窗口，什么也不需要做，Save 并 Exit 即可，这一步主要作用是生成一个 `.config` 文件。  
 e) 执行命令 `make -j5`，整个流程中最为耗时的一步，我的机器大约需要运行 20 分钟。5 这个参数可以理解为编译的速度，根据机器性能可以适当增大或减小。  
 f) 执行命令 `sudo make install modules_install`，安装相关模块，耗时十五分钟左右。这一步执行以后刚开始可能会弹出 `couldn't load...` 等字样，这是因为还没有执行 14 步，加载不出是正常的。  
 g) 执行命令 `sudo make install` 把 13 步遗留的问题补上。然后再执行一次 `sudo make install modules_install`  
 h) 以此执行下述四条命令  
 i. 

```
sudo mkinitramfs -o /boot/initrd.img-`cat
include/config/kernel.release` `cat include/config/kernel.release`
```

  
 ii. `make headers_install`  
 iii. 

```
sudo mkdir /usr/src/linux-headers-`cat include/config/kernel.release`
```

  
 iv. 

```
sudo cp -rf usr/include /usr/src/linux-headers-`cat
include/config/kernel.release`/include
```

至此，内核编译完成，重启系统后引导界面会出现新的启动项 4.2.0 版本



#### 4、替换固件

本质就是将 iwlwifi-5000-2.ucode.sigcomm2010 这个文件复制到系统的固件目录下，让其发挥作用。依次执行下述四条命令即可：

- i. `git clone git://github.com/dhalperi/linux-80211n-csitool-supplementary.git`
- ii. `for file in /lib/firmware/iwlwifi-5000-*.ucode; do sudo mv $file $file.orig; done`
- iii. `sudo cp linux-80211n-csitool-supplementary/firmware/iwlwifi-5000-2.ucode.sigcomm2010 /lib/firmware/`
- iv. `sudo ln -s iwlwifi-5000-2.ucode.sigcomm2010 /lib/firmware/iwlwifi-5000-2.ucode`

替换固件工作完成。

#### 5、编译收数文件

依次执行 `cd ~/linux-80211n-csitool-supplementary/netlink` `make`

#### 6、改驱动和收数环节

将/home/csi/intel-5300-csi-github/drivers/net/wireless/iwlwifi/dvm/main.c 中 1045 行：

`priv->connector_log=iwlwifi_mod_params.connector_log;`

改为

`priv->connector_log=1;`

修改驱动之后需要重新 install，即 `cd intel-5300-csi-github`

然后依次执行

- i. `sudo make -j3`
- ii. `sudo make install modules_install`
- iii. `sudo make install`
- iv. `sudo update-grub`

#### 7、收数方案（AP 模式）

a) 将安装好的电脑作为 client，连接一台没有密码的路由器（有密码的是连不上的）。

执行

```
sudo rmmod iwldvm iwlwifi mac80211 cfg80211
```

这一步如果报错 `...is in use by ...`，说明想要卸载的模块在被其他模块使用。

例如 `cfg80211 is in use by ath9k`，那我们需要使用 `sudo rmmod` 命令先卸载 `ath9k`。

b) 执行

```
sudo modprobe iwlwifi connector_log=0x1
```

连接上路由器 wifi 后，我们可以查看其 connection information，在 Default Route 栏获取路由器的 IP 地址。假设为 192.168.1.1。使用 `ctrl+alt+T` 新建另一个终端，并执行指令

`ping 192.168.1.1 -i 0.5` (0.5 为发包间隔，使用 `sudo` 权限可以获取更短的发包间隔)

我们的电脑会不停地 ping AP，这时就可以测 CSI 了。

c) 在原终端打开 `log_to_file` 收数，进入 `/linux-supplementary/netlink`，执行命令

```
sudo ./log_to_file test.dat
```

可以看到输出了 `wrote 393 bytes received 393 bytes:id:26 val:1 seq:10 clen:393` 等字样，收数成功！可以打开 `netlink` 目录查看收到的 `dat` 文件。

## 附录二：项目分工

	姓名	学号	组内主要分工
组长			项目统筹、方案设计
组员			实验环境搭建、实时系统构建
			文献阅读、制作演示视频
			机器学习部分
			数据采集与处理

## 附录三：项目工程各源代码作用说明

项目主页：<https://github.com/liuminghui233/Wi-Fi-alarm>

项目目录结构及相关源代码作用：

项目子目录	作用	关键代码文件
netlink	数据采集	log_to_file.c
matlab	数据研究分析	csi_analysis.m csi_timeseq.m
svm_classifier	机器学习模型训练	get_batch_data.m get_dataset.m SVMtrainClassifier.m
realtime_wifi_alarm	实时监测系统	realtime_analysis.m