2018

中国Wi-Fi探针技术白皮书

Chinese Wi-Fi Sensor Technology White Paper

上海无线通信研究中心
SHANGHAI RESEARCH CENTER FOR WIRELESS COMMUNICATIONS

芝麻科技
zhimatech.com

发布时间

2018.4.5

目录

I. WiFi探针基础原理

1.1 什么是mac地址

MAC（Medium/Media Access Control）地址，用来表示互联网上每一个站点的标识符，采用十六进制数表示，共六个字节（48位）。其中，前三个字节是由IEEE的注册管理机构RA负责给不同厂家分配的代码(高位24位），也称为“编制上唯一的标识符”（Organizationally Unique Identifier），后三个字节(低位24位)由各厂家自行指派给生产的适配器接口，称为扩展标识符（唯一性）。一个地址块可以生成224个不同的地址。MAC地址实际上就是适配器地址或适配器标识符EUI-48。[1] MAC（Media Access Control，介质访问控制）地址，也叫硬件地址，长度是48比特（6字节），由16进制的数字组成，分为前24位和后24位：前24位叫做组织唯一标志符（Organizationally Unique Identifier，即OUI），是由IEEE的注册管理机构给不同厂家分配的代码，区分了不同的厂家。后24位是由厂家自己分配的，称为扩展标识符。同一个厂家生产的网卡中MAC地址后24位是不同的。

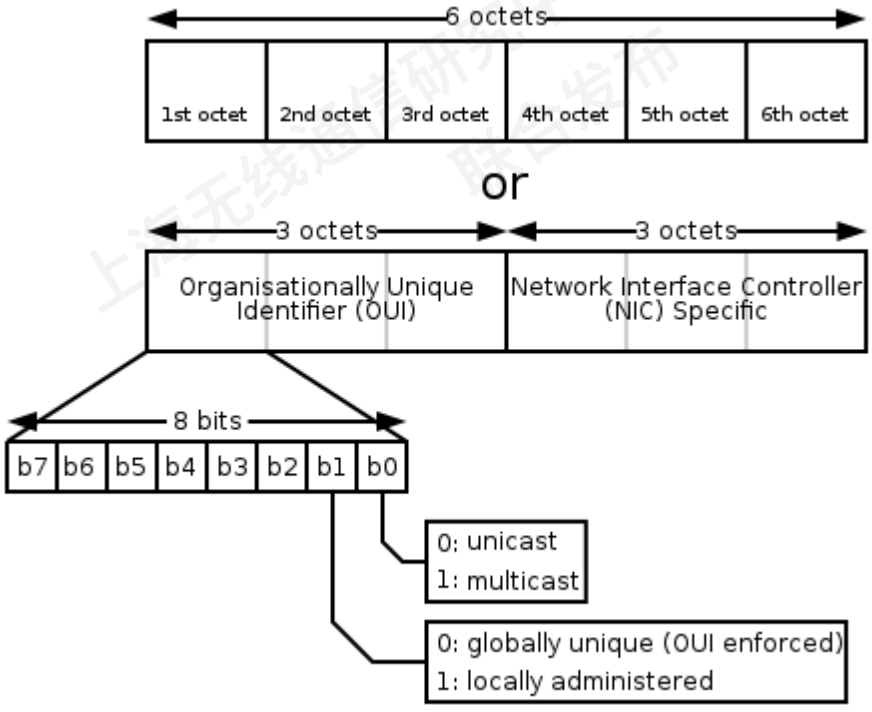

图1.1 Mac地址报文格式

MAC地址对应于OSI（Open System Interconnect开放系统互连）参考模型的第二层数据链路层，工作在数据链路层的交换机维护着计算机MAC地址和自身端口的数据库，交换机根据收到的数据帧中的“目的MAC地址”字段来转发数据帧。

图1.2 OSI网络七层模型

网卡的物理地址通常是由网卡生产厂家烧入网卡的EPROM（一种闪存芯片，通常可以通过程序擦写），它存储的是传输数据时真正赖以标识发出数据的电脑和接收数据的主机的地址。也就是说，在网络底层的物理传输过程中，是通过物理地址来识别主机的，它一定是全球唯一的。比如，著名的以太网卡，其物理地址是48bit（比特位）的整数，如：44-45-53-54-00-00,以机器可读的方式存入主机接口中。以太网地址管理机构(除了管这个外还管别的)（IEEE）（IEEE：电气和电子工程师协会）将以太网地址，也就是48比特的不同组合，分为若干独立的连续地址组，生产以太网网卡的厂家就购买其中一组，具体生产时，逐个将唯一地址赋予以太网卡。形象地说，MAC地址就如同我们身份证上的身份证号码，具有全球唯一性。

图1.3 iOS 设备（左）和安卓设备（右）的mac地址示例

1.2 什么是Wi-Fi探针技术?

狭义上来说，WiFi探针特指WiFi设备通信过程中的一种信号帧，即Probe Request，这种类型的报文专门用来请求终端（笔记本、智能手机等能够连接WiFi网络的设备）周围的WiFi信号，然后由放出WiFi信号的设备如无线路由器、无线AP等设备用Probe Response报文给予回复。因为探针报文中会携带终端的WiFi芯片mac地址，以及交互时间等信息，因此在现在的商业场景中，WiFi探针技术大多被用来记录顾客在商业环境（如商场、超市等）中的行为轨迹，而由于不论Probe Request还是Probe Response都发生在顾客的智能终端打开WiFi开关但尚未连接上WiFi网络之前，也就是说处在正常的网络应用（如QQ和微信等）的传输都还没有开始的阶段，因此WiFi探针技术并不能抓取到用户上网的内容亦或是手机内的隐私内容，仅仅只能通过智能终端的移动轨迹和停留时长等信息来分析该用户在商业环境的行为，同时也说明，没有打开WiFi开关的智能设备是没有办法被探针设备捕捉到的。目前，WiFi探针技术被较为广泛地应用在客流统计与分析以及精准营销等方面。

图1.4 Wi-Fi探针设备工作原理

1.3 探针采集的数据包含哪些内容

不同的厂商采集的内容不同，根据手机不同的状态能采集到的内容也不尽相同，但都会包含如下几部分内容：探针mac地址、探测时间、探针抓到的智能终端的mac地址、智能终端的信号强度（用RSSI表示），如下图：

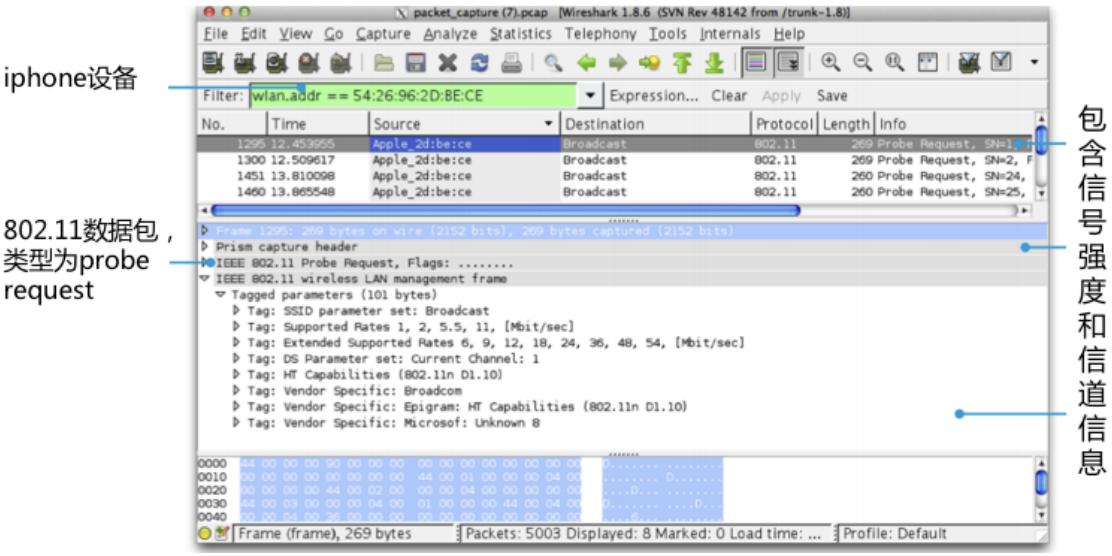

图1.5 探针设备采集信息示意图

1.4 什么是随机MAC

苹果公司在IOS8中引入入的随机mac特性以后,谷谷歌和微软也相继在自自己己的最新手手机操作系统中推出了mac地址随机化的算法。随着新版操作系统市场占有率的逐步提高高,基于mac地址的大数据分析业受到的影响也越来越明显。从芝麻科技数据平台统计的随机mac情况来看,在IOS8推出随机mac特性之前,数据平台上抓到的真实mac占比在95%以上。而在IOS8发布后,这一指标迅速下降,目前基本稳定在 39%左右,其余61%左右的mac地址几几乎全为随机化的mac地址。如下图，随机mac均为Local属性，其Universal/Local Bit会置位为1。

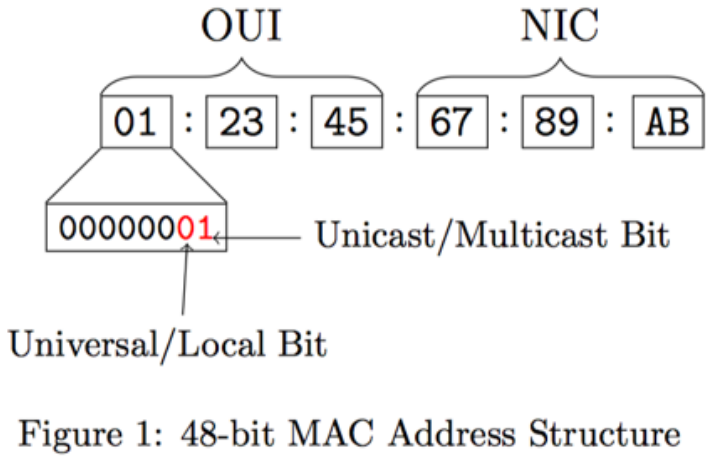

Figure 1: 48-bit MAC Address Structure

图1.6 随机MAC的识别规律示意图

苹果 -苹果从IOS8开始引入在Probereq包中使用随机mac的特性,起初仅在黑屏情况下使用随机 mac,IOS9开始在亮屏状态下也使用随机mac。目前IOS8以上的苹果设备已经超过全部苹果设备的 80%,苹果手机是随机mac最主要的来源。从芝麻科技数据平台的数据来看,每天抓取到的全部真实mac地址中, 有14%左右为苹果手机的mac地址。

安卓-安卓从6.0版本开始,在硬件和驱动支持的手机上使用随机mac进行背景搜索。但由于安卓版本碎片化严重,6.0版本发布至今已经一年多,占有率仅有20%左右。2016年发布的7.0版本在随机mac 功能上没有变化。

windows -微软从Windows10开始引入入mac地址的随机化功能,开启该功能同样需要硬件和驱动支支持。 与其他系统不同的是,windows10不仅在Probereq包中使用用随机mac,其在连接不同的网络时使用的也是不同的mac地址。对同一个无线网络来说,它使用的mac地址是固定的,但对于不同的无线网络来说,它使用的是不同的mac地址。mac地址的算法如下:

addr = SHA-256(SSID , macaddr , connId , secret)

其中,ssid是无线网络的名称,macaddr是手机的真实mac地址,connid在手机第一次连接该 ssid时生成,当该ssid被从手机的ssid列表中删除并重新添加时会改变。secret是手机初始化时随机生成的。

综上所述，Windows10的mac地址随机化算法更有效的避免了真实mac地址的泄露，但是由于windows phone的市场占有率还不高，所以其对数据分析业务影响暂时较小，安卓设备的碎片化程度严重，经过测试目前国内主流的华为、三星、OPPO、VIVO和小米均没有采用mac地址随机化的方式，因此目前来看，随机mac量最大的还是iOS设备。这就使得一些客户在了解到有随机mac的存在以后产生了一定的困惑，是不是iPhone在没连WiFi的时候就一直发随机mac而完全不发真实mac了呢？这会不会导致iPhone的mac就完全抓不到了呢？其实在真实的场景中，iOS设备也并不是一直都发随机mac的，如果某个携带iOS设备的顾客路过某个之前连接过的WiFi信号（比如主流一二线城市里比较主流的WiFi信号：CMCC、ChinaNet、Starbucks等），他的手机会尝试与这些信号进行连接，在这个过程中使用的就是真实mac。而如果此时附近有探针设备，就可以抓到这位顾客的iOS设备的真实mac地址了。再加上现在有些探针设备的厂家会内置一些诱导算法，也会增加真实的iOS设备mac地址的抓取几率，因此整体上来说，随机mac的出现大约使得iOS设备的抓取率下降了50%，而并不是完全抓取不到。

II. WiFi探针数据处理方式的历史沿革

WiFi探针数据最早被用来做WiFi定位，包括对RFID标签的定位和对带WiFi功能的智能终端的定位，然后在后续的演进过程中出现了利用探针数据以客流统计为目的的纯WiFi探针，相比用无线AP做WiFi定位的解决方案来说成本更加低廉也更容易被商家接受。

2.1 WiFi定位

WiFi定位即无线AP对于支持WiFi功能的智能终端的定位，该定位包括对正常接入网络的WiFi终端的定位和对非法AP的定位，是指根据AP收集的周围环境中的无线信号强度信息定位终端位置的技术。AP将收集到的周围环境中终端发射的无线信号信息上报给定位服务器，定位服务器根据得到的无线信号强度信息与AP位置，计算出终端的位置信息，然后再通过显示设备展现给用户。

基于WiFi的无线定位技术可以给客户带来明显的价值。相对于其他定位系统，具有成本低、应用场景不受限制等优势。基于定位技术的增值应用，如室内导航、精准广告推送等为商家带来额外的商业价值。

2.1.1 WiFi定位基本原理

① 1. AP都有全球唯一的MAC地址, 无线AP在一段时间内是不会移动的
② 2. 扫描并收集周围的AP信号, 可以获取到AP广播出来的MAC地址
③ 3. 设备将标示AP的数据发送到位置服务器,服务器检索出每一个AP的地理位置
④ 4. 结合每个信号的强弱程度，时间或者角度,计算出设备的地理位置并返回到用户设备

2.1.2 WiFi定位的基本框架介绍

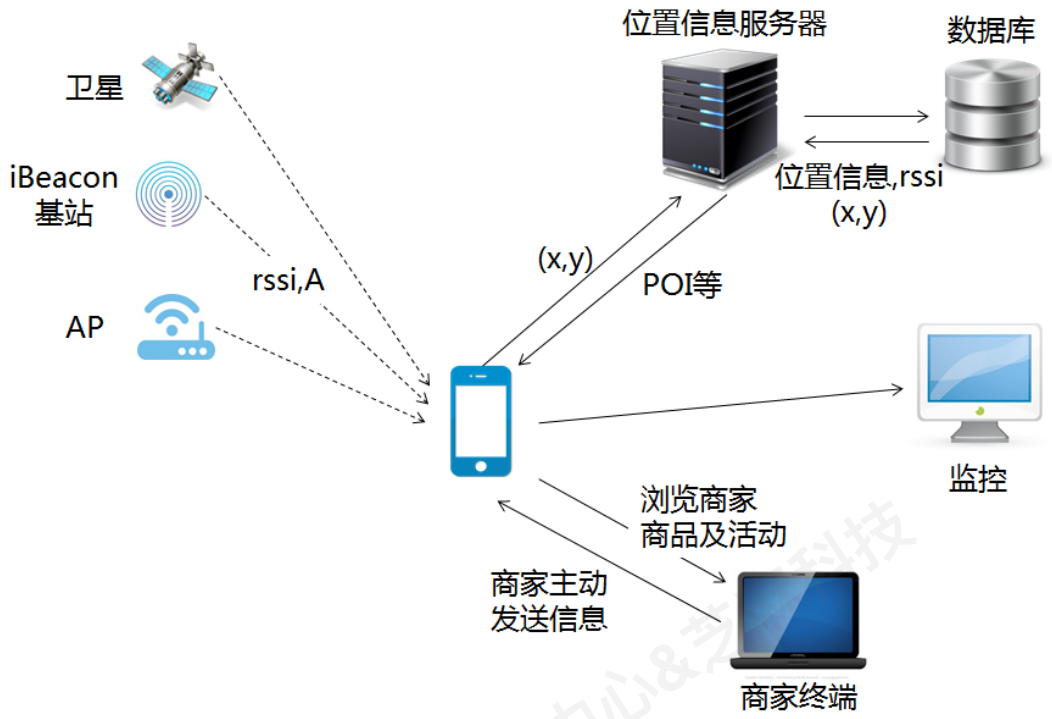

图2.1 Wi-Fi定位的基本框架介绍

2.1.3 室内定位精度

上海无线通信研究中心采用的WiFi定位技术，室内定位精度90%的情况下在3米内，采用超声波定位精度更是可以达到厘米级。

2.1.4 适用性广泛

无线定位技术越来越多的需求出现在室内，如室内导航、动线采集、冷热区图等。而传统的定位系统如GPS、蜂窝系统等在室内的定位效果并不能满足场景需求。而WiFi定位恰好可以弥补GPS和蜂窝系统在这方面的不足。

2.1.5 应用开发广泛

基于WiFi的无线定位技术可以开展丰富的增值业务，如室内导航，利用定位技术做停车场的反向寻车；广告推送应用，商家可以利用无线定位技术精准地给消费者推送广告以吸引其到店；商业价值分析，如商场通过对定位历史数据的分析，获得顾客在不同店面的停留时间，从而实现对不同店面的客流量的多维度统计，为租金政策提供辅助决策信息。

2.2 WiFi探针客流统计

2.2.1 基于WiFi定位的客流统计方法

根据上文提到的内容，WiFi定位的一大主流用法就是用来做客流统计。尤其是一些大型的商场环境（品牌连锁门店一般不会考虑在所有门店实现WiFi定位），在给来访顾客提供WiFi覆盖的同时，顺便就可以完成客流人数的统计。但是由于商场的WiFi部署出于成本考虑往往并不会考虑对商铺的覆盖，即基本上所有的AP都会被部署在过道等公共区域，这就造成对商场的客流统计基本上只能维持在整场客流，或者最多到楼层级客流情况的统计，而没有办法到商铺级别的统计。

基于无线AP的定位精度理论上是比纯探针的方式要高，但是有一个目前看来很“重”的需求就是必须要智能手机登WiFi终端连接上商场的WiFi网络，如果不连接，则其定位精度并不会比纯探针的方式高很多，而目前随着各大运营商以及互联网运营商流量资费的大幅度降低，一个主要的趋势就是愿意在商场连WiFi的顾客越来越少，据测算，即使在上海非常繁华地段的一个高端商场，用了全套思科设备搭建的带定位功能的WiFi网络，每天愿意连接的顾客也占不到其总客流量的10%，这也导致商场无线AP定位方案的总体性价比较低。

2.2.2 基于探针的WiFi客流统计方法

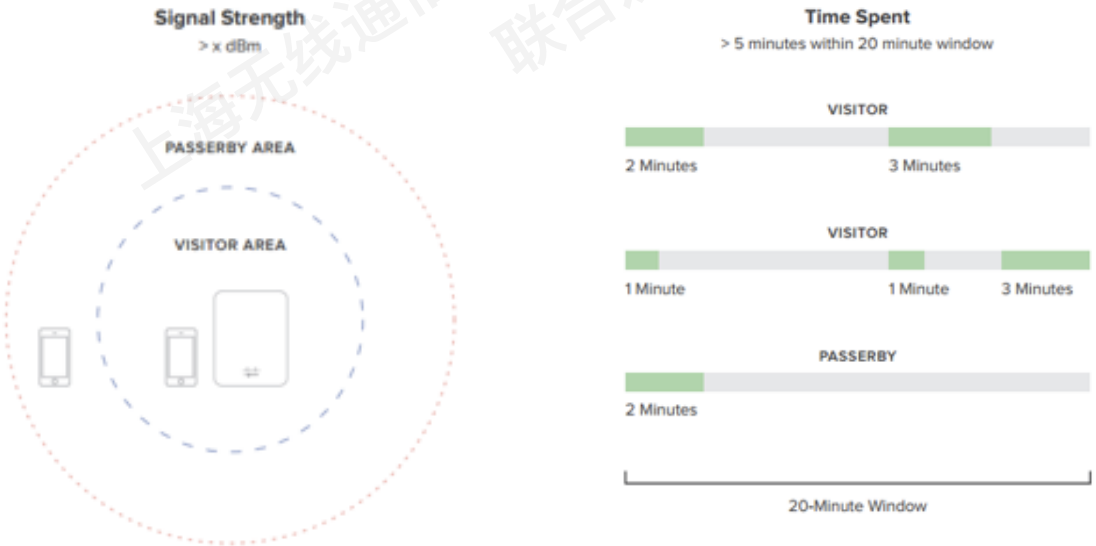

图2.2 目前主流的传统WiFi客流统计

目前另一类主流的基于WiFi探针的客流统计方法，是以WiFi探针为中心，不同的信号强度为半径，画两个圈，内圈叫做“入店区域”，入店区域半径约等于店铺宽度，外圈叫做“客流区域”，客流区域的半径约大于店铺宽度一定范围。其中在入店区域范围内的设备，还会

考虑其停留时长，但不同的厂家在这一块有不同的算法，有的厂家是在20分钟的窗口内，累计停留时长达到5分钟，即算入店，有的厂家则是持续停留时长达到一定的时间即算入店，这个“持续停留时长”可以根据不同的商家业态进行动态调整。但是相对来说，这种方式的统计方法还是属于固定阈值的统计方法，即入店区域和客流区域的判定阈值（即信号强度，用RSSI表示）相对固定，对于一些形状规整的店铺来说，还可以有比较理想的统计结果，而对于一些异形店铺则会有些力不从心。

2.2.3 基于大数据和机器学习算法的进店顾客识别算法

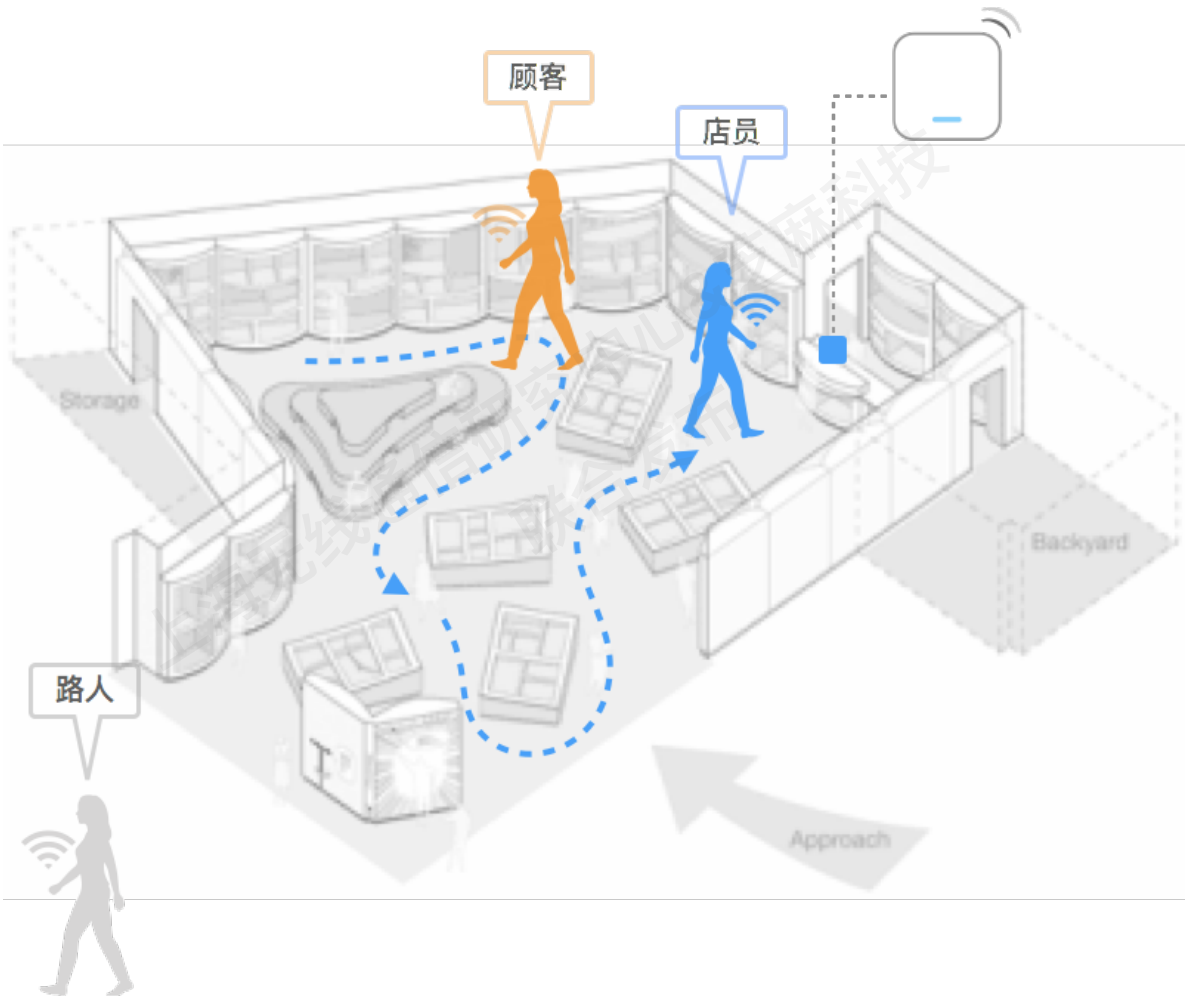

图2.3 基于机器学习算法的顾客进店判断模型

该方法是采集店铺里店员的mac地址和移动轨迹数据作为正向样本，利用机器学习算法学习其样本特征，并对模型进行训练从而生成进店判断模型，然后利用该模型对探测到的人群进行进店判断，该模型会持续学习并不断迭代从而最大限度提升顾客进店判断的准确率。该方法不但可以使得商场能够以极低的成本实现准确率不低于WiFi定位的店铺级客流统计，

同时也可以大大改善一直困扰品牌连锁门店由于店铺形状问题而导致WiFi客流探测不够准确的情况。

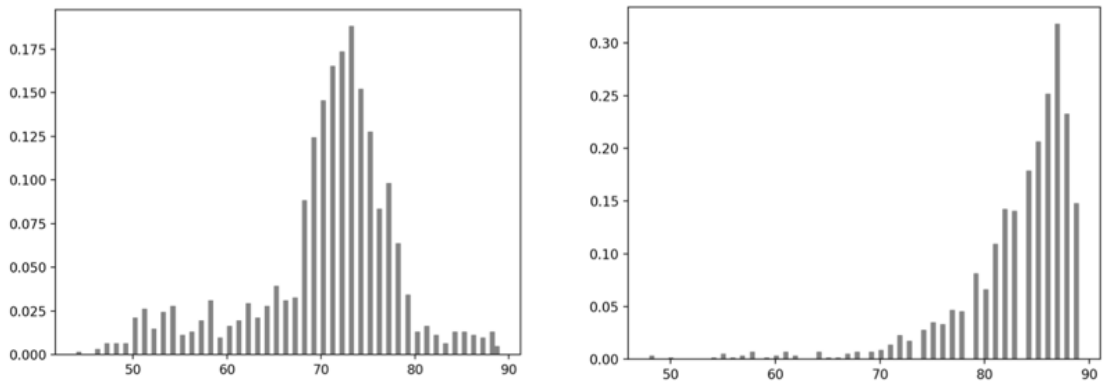

图2.4 顾客与路人信号强度示意图

2.3 探针数据的采集精度如何?

2.3.1 探测范围

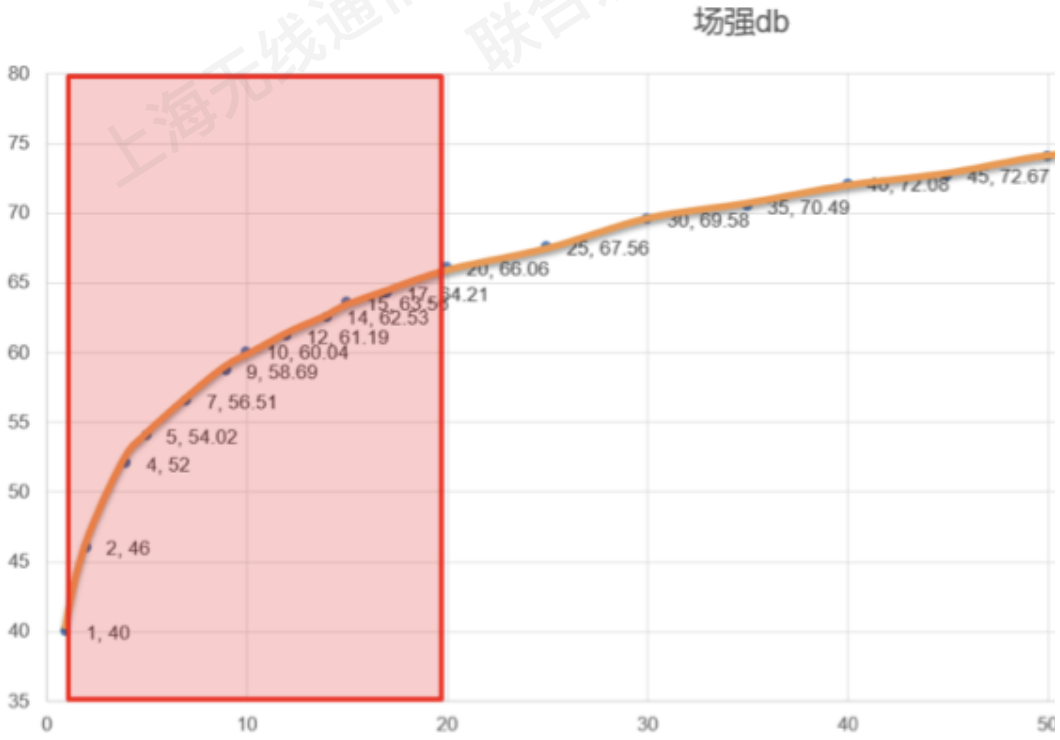

图2.5 距离影响下的信号强度曲线图

我们知道，无线信号的强度随着传播距离的增加而衰减，接收方与发送方离得越近，则接收方所收到的信号强度就越强；反之如果接收方距离发送方越远，则接收方收到的信号就越弱。根据测量接收到的移动终端信号强度和已知的无线信号在空气中传播时的自由空间衰减模型，可以大致估算出收发双方之间的距离，根据上图可以看出，在信号强度大于-65dBm的范围内，信号强度的衰减与距离的关系成强相关关系：即每一个dB的信号强度，大约对应5米距离，而在超出20米的范围以后，随着距离的增加，信号强度的衰减就没有那么明显，比如在25米到30米的时候，距离差了5米，而信号强度可能才衰减了2个dB。因此基于这个原理，探针的有效探测范围其实就在20米以内，再加上其实在探针的环境中，手机等智能终端是信号发送方，探针是信号接收方，而智能终端的发射功率往往很有限，因此这个距离还要再打些折扣，一般在15米左右。

2.3.2 采集率

采集率是指在一定的时间范围内，在探针探测范围内或指定探测阈值范围内，探针设备能抓取到的设备数量占该范围内所有设备数量的比例。

目前从采集率来看，如果与摄像头比较（尽管由于统计口径的不同这种对比并不科学），大概相当于摄像头采集客流的60-70%左右。

III. WiFi探针数据的应用

3.1 门店应用

绝大多数的消费者购物的行为顺序都遵循“逛街—进店—挑选—试用—决策—付款—离店”的规律，而交易数据仅仅只记录了“付款”这一个时间点发生的动作，但是在“付款”动作发生之前，也就是交易数据产生之前，店铺的经营者并不知道消费者到底什么时候来的？他考虑了多久才决定下单？有多少人进店之后经过反复挑选却没有下单就离开了？而当客流数据与交易数据打通之后，我们可以以一个更全面的视角观察每一个消费者，很多过去模棱两可的问题都可以以数据的形式表现出来，例如：

- 消费者下单次数和到店次数的比例是怎样的?
- 哪些区域的商品转化率更高?
- 在一次营销活动结束后，到底有多少人被影响后到店，这些人中哪些人发生了交易而哪些人又离开了?
- 消费者平均进店多久后才会发生购买行为?
- 每个收银台前的平均等待时间是多少?
- 上文提到的购物行为顺序的漏斗模型，每个环节的时间点和人群规模是多少？哪个环节的损耗最大?

基于WiFi探针的客流数据就可以很好解决以上经营中的问题，助力门店运营与营销，同时还可以帮助品牌进一步解答如下问题：

- 门店客流量怎么样？成交转化到底有多少?
- 新老顾客占比如何？回头客有多少?
- 顾客如何构成？他们有什么特征?
- 如何更精准触达？如何监测有效性?

3.2 商场应用

随着商业地产行业的快速发展，2016年全国新开业的大中型商业项目多达511家，商业总体量超过4600万平方米，且该数据尚不包含各类专业市场及商业面积在3万平方米以下的小型商业项目。整体看来，全国新开业购物中心体量庞大且同质化严重。利用WiFi探针数据以及相应的大数据算法，以场内客流为本质对商场架构进行调整，利用大数据工具以客流为基础，以对顾客的精细化运营，打造智慧商场突破同质化并提升用户体验为目标，一些基于

WiFi探针数据应用的互联网公司提出了一些新的应用方向，如商场内无线AP及探针数据的混合处理，基于场内顾客逛店行为产出标签，基于行为协同过滤算法打造场内顾客系统推荐算法，搭建顾客生命周期价值智能识别模型等等。

图3.1 探针的商场场景应用

IV. 设备+设备管理平台的重要性

企业级WiFi网络架构经过多年的发展，已经从最初的胖AP经过硬件AC+AP发展到了现在的云AC及有线无线一体化架构。架构演进之路本身也揭示了其实人们对于WiFi网络的需求远不仅仅是无线接入那么简单，其背后需要有与之相配套的基础网络层、管理控制层以及用户认证层等完整的企业网架构来支撑。

时至今日，一个品牌其连锁门店动辄上千，在每个门店都配备专业IT人员显然是不现实的，但是安装在成百上千家门店的设备却还需要总部的IT人员进行管理，实时了解这些设备的运行状态，并在必要的时候通过远程的方式对这些设备进行必要的操作，这是一个统一的设备管理平台的重要性不言而喻。

探针设备可以说脱胎于无线AP，因此一个好的探针设备，在硬件上除了需要具备堪比企业级AP的7*24全年不间断运行的能力（甚至在某种程度上来说，由于探针往往属于业务系统，因此其不间断运行能力还要高于非业务AP），还需要能够低成本地非常简单地进行部署，最好是即插即用，并且必须支持集中管理、远程设置以及持续不断地更新迭代。想要满足这些需求就必须要有一个强有力的云端平台来做支撑，管理人员需要可以通过管理平台实时掌握运行在千里之外的探针设备的运行状态，有设备出现问题要能够及时接到通知。

此外，因为涉及到业务系统，如何保证数据传输过程中的安全性，也是不可忽视的问题，从这个层面来说，一个从系统底层架构开始就打好安全基础，从数据的存储、传输到应用全程都有安全措施保障的平台，同时具备稳定的迭代周期，严格的权限划分和管理，灵活的可扩展性（10台设备到10000台设备随时可扩展，业务0中断）等等，也是在这个时代不可或缺的平台级功能。

关于上海无线通信研究中心

上海无线通信研究中心是2003年11月，中国科学院上海微系统与信息技术研究所、上海市科学技术委员会、东南大学和长宁区共同发起，成立了隶属于上海市科委的事业单位。

中心主要从事宽带无线移动通信关键技术和新一代无线移动通信系统集成测试研发及标准化工作，多次承担国家、中科院和上海市重要科研项目，取得了多项有影响的成果。作为移动通信国家级合作基地，开展了广泛的国内外技术合作。

为提升在下一代无线通信领域自主创新能力，服务国家重大科技项目，服务无线移动通信技术产业发展需求，经过几年的发展，中心已经形成了一支由国内外知名学者、博士、硕士组成的百余人高层次研发团队，设有“通信与信息系统”博士、硕士培养点。 中心通过挖掘和整合内部科技创新资源，向社会全面开放测试仪器设备，目前拥有价值6000万的仪器设备，200余项发明专利，100余项标准化提案。同时构建了多种测试服务平台为通信、电子、计算机、工控等科研领域的用户提供了专业的测试服务与解决方案，为新产品的预言及早期开发提供了支撑。

希望了解研究中心更多信息及服务，敬请访问官方主页 http://www.shrcwc.org/index.html。

关于芝麻科技

芝麻科技（全称南京芝麻信息科技有限公司）成立于2013年，是国内领先的线下数据技术服务公司，专注于通过线下数据的采集、分析及应用，为品牌门店和购物中心等商业客户提供数据化运营管理服务，助力实体商业做更智慧的生意。

作为国内首批开发Wi-Fi感知设备并结合数据挖掘将其应用在零售领域的领军者， 芝麻科技拥有该项技术在中国的国家发明专利，开发并运用设备管理云平台及客流分析平台等产品以提高门店管理效率； 2015年率先实现线上线下数据联通，并与阿里巴巴联合发布零售智能数据服务；芝麻科技已形成大规模的线下数据库并发布线下数据管理平台， 涉及基础客流分析，消费者画像，精准营销DMP等多个维度的应用服务；后与南加州大学成立数据科学实验室，深度研究线下数据的商业应用，推出针对购物中心的数据化运营服务。

目前，芝麻科技已经在全国覆盖10万余消费生活场所，覆盖中国约7亿部移动设备，日处理数据百亿条。服务宝洁，欧莱雅，欧时力，赫基，劲霸，北汽，福特，碧桂园，静安嘉里中心，佳兆业商业，虹悦城等国内外知名企业及商业中心。

希望了解芝麻科技更多信息及服务，敬请访问官方网站 www.zhimatech.com，欲了解更多企业动态请扫码关注公众号。