



基于 Wi-Fi 协议的历史轨迹探测系统的研究与实现

王奕钧

(公安部第一研究所, 北京 100044)

摘要:历史轨迹探测对于商业用户行为分析、警务嫌疑人身份落地及重点人布控都有着重要作用和现实意义。目前的探测方式或采用移动客户端静默搜集用户信息,或采用运营商分流数据,或采用摄像头拍摄结合人工分析等方式,这些探测方式存在着采集样本受限、提取有效信息困难、分析效率低下等各种弊端。提出了一种基于 Wi-Fi 协议的历史轨迹探测方式,并给出具体的系统设计与功能实现,为商业分析、警务应用提供一种新的分析用户行为的途径。

关键词: Wi-Fi; 轨迹探测; 探针请求

中图分类号: TP393.4

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016146

Research and implementation of the historical trajectory detection system based on Wi-Fi protocol

WANG Yijun

First Research Institute of Ministry of Public Security, Beijing 100044, China

Abstract: The historical trajectory detection plays a vital role and has the practical significance for commercial user behavior analysis, the identity of the police suspect landing and surveillance of key. Current detection methods use mobile client silence to collect user information, or use operators to shunt data, or shooting combined with artificial analysis. The ways have many disadvantages such as the sample is limited, the effective information is difficult to extract and the analytical efficiency is low. A historical track detection based on Wi-Fi protocol was proposed, and specific system design and function realization were given which provided a new way to analyze the behavior of users.

Key words: Wi-Fi, trajectory detection, probe request

1 引言

历史轨迹探测,即发现目标人或设备的活动区域及移动轨迹,在商业应用中,可以用于统计交通运营状况、接入点商圈分布、用户高频行为等;在警务工作中,可以对追踪嫌疑人位置、布控抓捕嫌疑人提供数据支持。传统的历史轨迹探测方式多采用移动客户端报送数据、摄像头录像监控等方式,但都存在影响实用性的限制。例如:移动客户端

需要不断地更新版本与功能以吸引用户;摄像头录像监控依赖于图像识别技术的发展,对图像清晰度、拍摄角度、周围光线都有较高的要求等。随着无线网络技术的迅速普及,城市中 Wi-Fi 网络接入点几乎遍布城市的每一个角落,不论是咖啡馆、快餐店,还是商场、酒店都会提供 Wi-Fi 网络以吸引顾客。如果能利用遍布在城市中的这张无形的 Wi-Fi 网络,探测并标记目标人或设备,将降低历史轨迹探测的成本,提高历史轨迹探测的精确度。本文提出了一种

基于 Wi-Fi 协议的历史轨迹探测方式,并给出具体的系统设计与功能实现。

2 Wi-Fi 协议利用原理

Wi-Fi 连接的过程可以描述成扫描、认证、关联和连接^[1] 4个步骤,详细流程如图1所示。

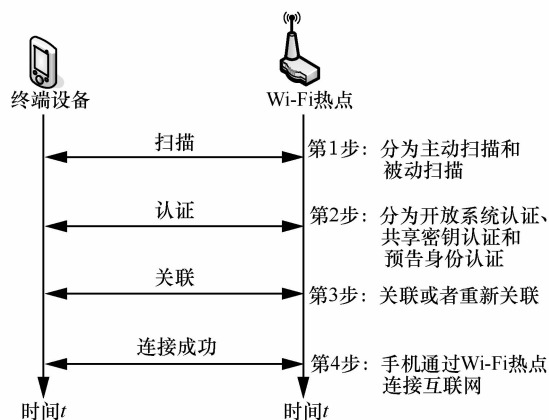


图1 Wi-Fi 连接过程示意

终端设备准备连接 Wi-Fi 接入点时,会首先扫描自身周围的 Wi-Fi 接入点,并根据扫描结果自动(已认证过)或手工输入密码完成身份认证,当双方认证成功,终端设备向 Wi-Fi 接入点发起关联以便获取网络的完全访问权限,当成功获得后,终端设备可以通过 Wi-Fi 接入点访问网络。

终端设备在探测 Wi-Fi 网络时存在被动接收和主动探测两种方式^[2]。

被动接收方式是终端设备定时将它的广播频率调整到一个固定的扫描频道,静默等待接收 Wi-Fi 接入点发布的信标(beacon)以完成 Wi-Fi 接入点的发现,过程如图2所示。

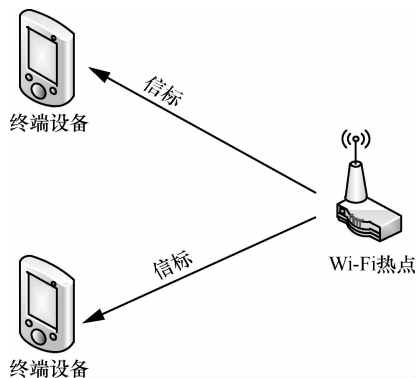


图2 被动接收方式示意

主动探测方式由终端设备主动向 Wi-Fi 接入点发送探针请求, Wi-Fi 接入点响应探针(probe)请求完成 Wi-Fi 接入点的发现,其中,探针又分为直接探针和广播探针。使用广播探针的状态下,终端设备广播一个无 SSID 的探针请求,所有的 AP 都会收到这个请求,并返回它们支持的 SSID,详细过程如图3所示。

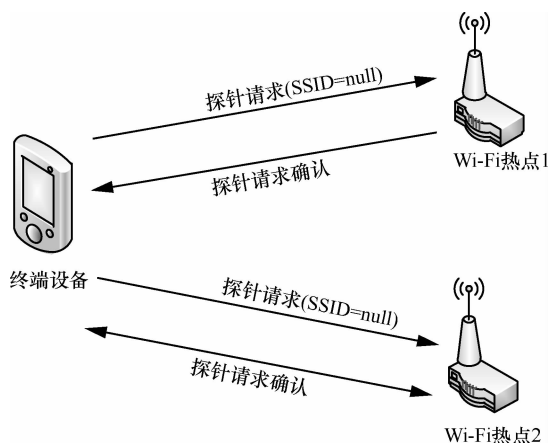


图3 广播探针探测示意

当终端设备使用直接探针时,终端设备直接发送特定名字的 SSID 探针请求,只有拥有请求的 SSID 的 AP 才能回应,过程如图4所示。

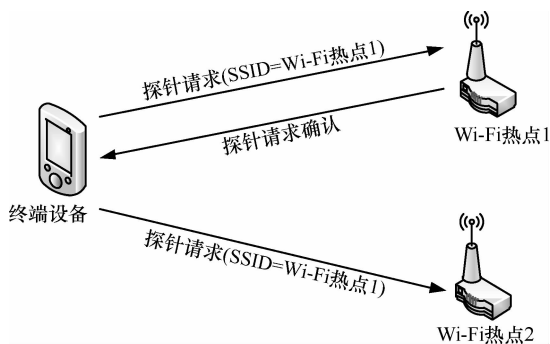


图4 直接探针探测示意

基于 Wi-Fi 协议的轨迹探测系统正是利用 Wi-Fi 协议中主动扫描的特性,自动捕获终端设备发出的直接探针请求,从中发现终端设备访问过的 Wi-Fi SSID^[3],进而描绘出终端设备的历史轨迹。

3 系统设计与实现

基于 Wi-Fi 协议的轨迹探测系统包括 Wi-Fi 接入点采集系统、Wi-Fi 信息汇总资源库和 Wi-Fi 轨迹探测系统 3 个组成部分。系统总体结构如图5所示。

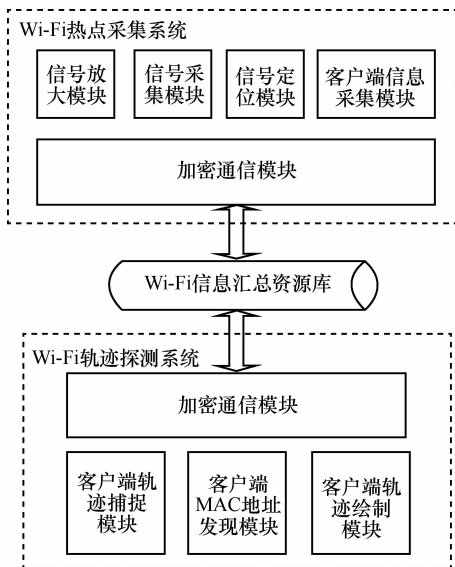


图5 探测系统结构

系统功能介绍如下。

(1) Wi-Fi 接入点采集系统

用来搜集 Wi-Fi 接入点的地理信息以及该接入点下连接的用户信息,包括 ESSID、BSSID、加密方式、信道、最大传输速率等^[4]。Wi-Fi 接入点采集系统由信号放大模块、信号采集模块、信号定位模块、客户端信息采集模块和加密通信模块构成。信号放大模块用于放大远端 Wi-Fi 信号,中继远端 Wi-Fi 接入点的数据分组以供 Wi-Fi 接入点采集系统分析采集使用;信号采集模块用于采集 Wi-Fi 接入点通信传输的数据分组,从中过滤并捕获 Wi-Fi 接入点采集系统所需的关键数据;信号定位模块将捕获的 Wi-Fi 接入点信息通过信号增益定位、公开 API 查询等方式,确定 Wi-Fi 接入点的地理信息;客户端信息采集模块用来采集每个 Wi-Fi 接入点下的客户端信息;加密通信模块用来传输 Wi-Fi 接入点采集系统与 Wi-Fi 信息汇总资源库之间的数据。

(2) Wi-Fi 信息汇总资源库

用来存储所有 Wi-Fi 接入点采集系统在外捕获的 Wi-Fi 接入点信息和客户端信息,同时向 Wi-Fi 轨迹探测系统提供查询接口,为绘制客户端轨迹提供数据支撑。Wi-Fi 信息汇总资源库采用关系型数据库构建,保证 Wi-Fi 信息汇总资源库具有较高的可扩展性,对高并发的查询请求具有良好的支撑能力,对于其中的数据,可以实现高可靠的容灾备份机制。

(3) Wi-Fi 轨迹探测系统

用于发现终端设备曾经连接过的 Wi-Fi 接入点,并通

过在 Wi-Fi 信息汇总资源库中的查询比对,绘制出该终端设备曾经出现过的历史轨迹。为了保证绘制轨迹的精准有效,Wi-Fi 轨迹探测系统使用客户端 MAC 地址发现模块搭建蜜罐 Wi-Fi,诱捕终端设备真实 MAC 地址。加密通信模块用来传输 Wi-Fi 轨迹探测系统与 Wi-Fi 信息汇总资源库之间的数据。

由于探针请求中不包含 Wi-Fi 接入点的 BSSID 及 MAC 地址,若仅使用 ESSID 作为标识,会因为重名而导致检索结果不唯一,所以在采集 Wi-Fi 接入点信息时,要补充采集 Wi-Fi 接入点下接入的客户端 MAC 地址,从而双向查询对比出终端设备接入过的 Wi-Fi 接入点,从而探测出该终端设备出现过的地理位置(即历史轨迹)。

具体实现中,Wi-Fi 接入点采集系统分为固定基站式和随身携带式两种设备形态,固定基站式设备用于大范围覆盖、长期采集 Wi-Fi 接入点信息,是整个 Wi-Fi 接入点的移动终端定位系统的基础支撑。拟采用多网卡多天线的结构覆盖全向方位,天线采用多台高增益定向天线拼接实现全向覆盖,随身携带式设备用于对重点地区的详细侦测采集,意图发现嫌疑接入点信息的精确位置。Wi-Fi 接入点采集系统采用 C++ 开发,运行在 Linux 平台上,使用 airodump-ng 库抓取无线数据,覆盖 2.4 GHz、3.6 GHz、4.9 GHz 和 5.8 GHz 4 个频段,通信加密采用 AES 算法结合对称加密的方式。

Wi-Fi 信息汇总资源库采用 Java 开发,运行在 Hadoop 集群上,数据存储在 HBase 分布式数据库中,使用 Impala 检索模块、MapReduce 并行编程技术以及基于 Lucene 的高性能索引模块,实现海量数据中的高速查询。

Wi-Fi 轨迹探测系统中 Wi-Fi 探测模块采用 C++ 开发,运行在 Linux 平台上,使用 airodump-ng 库抓取无线数据;前台采用 JavaEE + Oracle + Tomcat 开发实现,运行在 Windows 平台上;控制层通过 Spring MVC 框架实现;持久层使用 Hibernate 实现;业务逻辑层采用 Spring 框架实现;表示层采用 JSP + HTML + CSS + AJAX 实现;通信加密采用 AES 算法结合对称加密的方式。

具体部署结构如图 6 所示。

Wi-Fi 接入点采集系统采集数据分组包括 Wi-Fi 接入点的 BSSID(无线路由器的硬件地址)、MB(无线路由器支持的最大速率)、ENC(加密协议)、AUTH(密钥模式)、ESSID(无线路由器的广播序列号)和接入终端的 station MAC(接入端的硬件地址)、PWR(信号强度)、Rate(接入端支持的最大速率)以及经纬度信息,数据结构如图 7 所示。

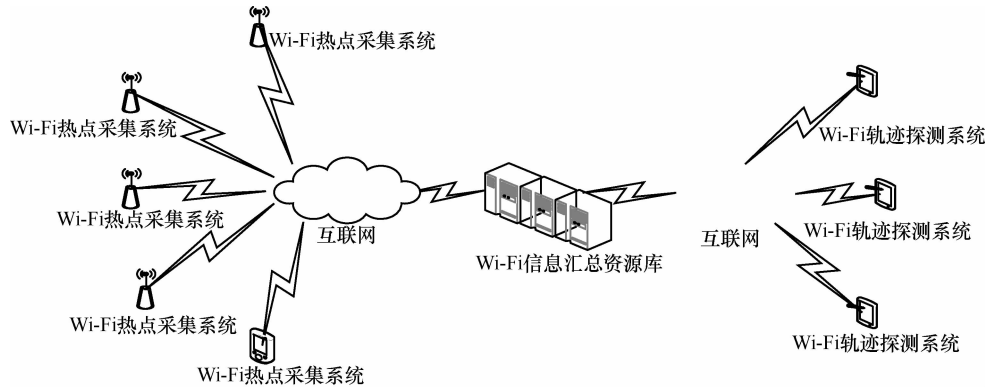


图 6 系统部署结构

BSSID	PWR	RXQ	Beacons	#Data, #/s	CH	MB	ENC	CIPHER	AUTH	ESSID
E4:D3:32:9A:44:56	-50	51	212	10014 389	11	54e	WPA2	CCMP	PSK	TP-LINK_1B1B0A
BSSID	STATION	PWR	Rate	Lost	Frames	Probe				
E4:D3:32:9A:44:56	9C:4E:36:9D:9C:40	-44	48e-36e	497	9964					
E4:D3:32:9A:44:56	D4:F4:6F:51:48:D9	-45	48e-24	4087	103					
E4:D3:32:9A:44:56	C8:B5:B7:59:02:D1	-58	0 -54e	0	28					

图 7 采集数据结构

Wi-Fi 信息汇总资源库包括 3 张数据表，分别是 Wi-Fi 接入点信息表、终端设备信息表和接入点/终端设备关系表，表结构分别如图 8、图 9、图 10 所示。

Wi-Fi 轨迹探测系统捕获数据格式如图 11 所示。
Wi-Fi 轨迹探测系统将捕获的探针数据中的 station 和 probe 字段向 Wi-Fi 信息汇总资源库发起查询请求得到该终端设备的经纬度，之后通过 google map API 将终端设备出现的地理信息描绘到地图上，探测结果如图 12

所示。

4 结束语

终端设备历史轨迹的探测在多个应用领域都有着迫切的需求和重要的意义^[5]。通过对 Wi-Fi 协议连接过程的分析，论证了基于 Wi-Fi 协议探测终端设备历史轨迹的可行性，同时设计实现了基于 Wi-Fi 协议的轨迹探测系统的功能原型，并给出了系统探测的结果。

ap id	essid	bssid	encrypt	type	channel	mb	longitude	latitude
1 1	Reader	00:69:9e:85:47:35	wpa2	psk	6	54e	39.9414572757	116.3182917981
2 4	NewCenturyHotel	8ca9:82:3d:47:db	wpa2	psk	1	54e	39.9375490499	116.3276885408
3 3	Tenda_397E18	c8:3a:35:39:7e:18	wpa2	psk	6	54e	39.9364080499	116.3245715408
4 2	TP-LINK_1B1B0A	e4:d3:32:9a:44:56	wpa2	psk	11	54e	39.9380610499	116.3243735408

图 8 Wi-Fi 接入点信息表结构

station id	station mac	pwr	rate
1 1	9c:4e:36:9d:9c:40	-34	0-1
2 2	f0:24:75:d0:3b:81	-37	0-1

图 9 终端设备信息表结构

id	ap_id	station_id	time
1 1	1	1	2015年 09月 06日 星期日 8:18:42 CST
2 2	2	1	2015年 09月 06日 星期日 8:34:52 CST
3 3	3	1	2015年 09月 06日 星期日 9:07:01 CST
4 4	4	1	2015年 09月 06日 星期日 9:21:33 CST

图 10 接入点/终端设备关系表结构



BSSID	STATION	PWR	Rate	Lost	Frames	Probe
(not associated)	80:19:34:78:CD:FA	-40	0 - 0	8	20	xsyq
(not associated)	38:48:4C:3D:AF:E0	-35	0 - 1	0	1	
(not associated)	58:1F:28:4D:30:4F	-35	0 - 0	0	4	Lizhiming
(not associated)	08:10:78:31:DE:2C	-35	0 - 1	0	8	
(not associated)	9C:4E:36:9D:9C:40	-37	48e-36e	0	25	TP-LINK_1B1B0A, TP-LINK_C887D0
(not associated)	D0:22:BE:E7:0C:DC	-38	0 - 0	0	5	

图 11 探针数据捕获



图 12 历史轨迹探测结果展示

参考文献:

- [1] 周天彤, 李强, 郑东. 对无线网络移动认证协议的分析与改进[J]. 计算机应用与软件, 2012, 29(3): 19-21.
ZHOU T T, LI Q, ZHENG D. The analysis and improvement of wireless mobile network authentication protocol[J]. Computer Applications and Software, 2012, 29(3): 19-21.
- [2] BHARDWAJ J. What is your phone saying behind your back [EB/OL]. (2012-10-02)[2015-09-20]. <http://nakedsecurity.sophos.com/2012/10/02/what-is-your-phone-saying-behind-your-back/>, 2012-10.
- [3] SIDIROPOULOS N, MIODUSZEWSKI M, SCHAAP E. Open Wi-Fi SSID broadcast vulnerability[EB/OL]. (2012-12-24)[2015-09-20]. https://www.OS3.nl/_media/2012-2013/Courses.
- [4] PISCATAWAY N. Wireless LAN medium access control (MAC) and physical layer(PHY) specifications [J]. IEEE Computer Society, 2012: C1-1184.
- [5] 李婷, 裴韬, 袁焯城, 等. 人类活动轨迹分类、模式和应用研究综述[J]. 地理科学进展, 2014, 33(7): 938-948.
LI T, PEI T, YUAN Y C, et al. Research of human activity trajectory classification, patterns and applications [J]. Progress in Geography, 2014, 33(7): 938-948.

[作者简介]



王奕钧(1983-),男,公安部第一研究所工程师,主要研究方向为网络信息安全。