Computer Science and Application 计算机科学与应用, 2017, 7(4), 281-290
Published Online April 2017 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2017.74035

Extraction and Analysis of Location-Based Information in Android Forensics

Yi Zhang, Zelin Zhang, Yan Chen, Rui Mao, Lei Xu

Jiangsu Police Institute, Nanjing Jiangsu
Email: 464604236@qq.com

Received: Mar. 28th, 2017; accepted: Apr. 11th, 2017; published: Apr. 18th, 2017

Abstract

Android smartphones generate a lot of location-based information. In mobile forensics, location information, which effectively reflects users' behavior trace, shows very high investigative value and evidence value. This paper builds a forensics process model to extract, analyze and display location-based information in order to provide new ideas for case investigation.

Keywords

Android Forensics, Location-Based Information, Data Analysis, Data Visualization

Android取证中地理位置信息提取与分析研究

张　祎，张泽林，陈　岩，冒　睿，徐　雷

江苏警官学院，江苏　南京
Email: 464604236@qq.com

收稿日期：2017年3月28日；录用日期：2017年4月11日；发布日期：2017年4月18日

摘　要

Android手机在日常使用中产生了大量的地理位置信息。在手机取证中，这些信息能够有效反映用户的行为轨迹，具有重要的证据价值，同时也能为案件侦查提供线索。构建Android取证中地理位置信息提取分析的实现模型，探讨对Android取证中地理位置信息的提取、分析及证据可视化展示，为案件侦查取证提供新的思路。

文章引用: 张祎, 张泽林, 陈岩, 冒睿, 徐雷. Android 取证中地理位置信息提取与分析研究[J]. 计算机科学与应用, 2017, 7(4): 281-290. https://doi.org/10.12677/csa.2017.74035

关键词

Android取证，地理位置信息，数据分析，数据可视化

1. 引言

截止 2016 年第二季度，Android 手机的市场占比已经达到了 86.2% [1]。伴随着移动通讯技术的快速发展，可以预见手机取证技术，尤其是针对高普及率的 Android 系统的取证技术，将成为打击各类犯罪的重要技术手段。

Android 手机用户在日常使用中产生了大量的地理位置信息。在手机取证调查中，这些地理位置信息能够有效反映用户的行为轨迹，具有重要的证据价值，同时也为案件侦破提供重要线索。目前，公安机关在案件的侦破和审理中主要针对涉案人员移动终端设备中的通讯录、通话记录、聊天记录、照片等用户信息数据进行取证调查，对上述大量的地理位置信息并没有广泛利用。

本文围绕 Android 平台的地理位置信息，探讨在 Android 取证中的数据提取与分析，并辅助后续侦查和证据固定，以试图构建电子调查取证中地理位置信息提取分析的实现模型，为调查取证提供新的思路。全文分为四个部分：第一部分概述相关概念与实现模型，第二部分介绍各类地理位置信息及其提取，第三部分概述对获取信息的数据分析，第四部分介绍地理位置信息可视化展示的实现途径，第五部分总结已有成果并对未来研究提出展望。

2. 相关概念

2.1. 地理位置信息

本文所称地理位置信息，是指以各种形式表现且能够反映特定对象空间位置(通常还附带特定时刻或时间范围)的数据信息，不局限于 GPS 等定位系统产生的定位数据，还包括诸如表述地点位置信息的文本数据等其他各类形式。

2.2. 数据可视化

数据可视化(Data visualization)是关于数据之视觉表现形式的研究。数据可视化旨在借助图形化手段，清晰有效地传达与沟通信息[2]。目前可视化技术在商业金融、科学计算、气象海事、网络安全、犯罪学等领域得到了广泛而有效的运用，已成为各领域揭示数据集中数据之间的关系和背后隐匿信息的基本工具[3]。数据可视化虽然手段形式多样但实质都是将数据多维属性及其相互关系以二维或三维的图形图像进行展示的过程。

2.3. 数据取证分析的流程

根据取证实践将数据取证分析建立图 1 所示的流程模型。

第一部分是数据获取。在电子取证中，涉案电子数据主要来源于对涉案电子设备中数据的搜索和恢复。另外，在特定案件中，根据国家法律法规，调查人员有权依法从第三方机构调取涉案电子数据。

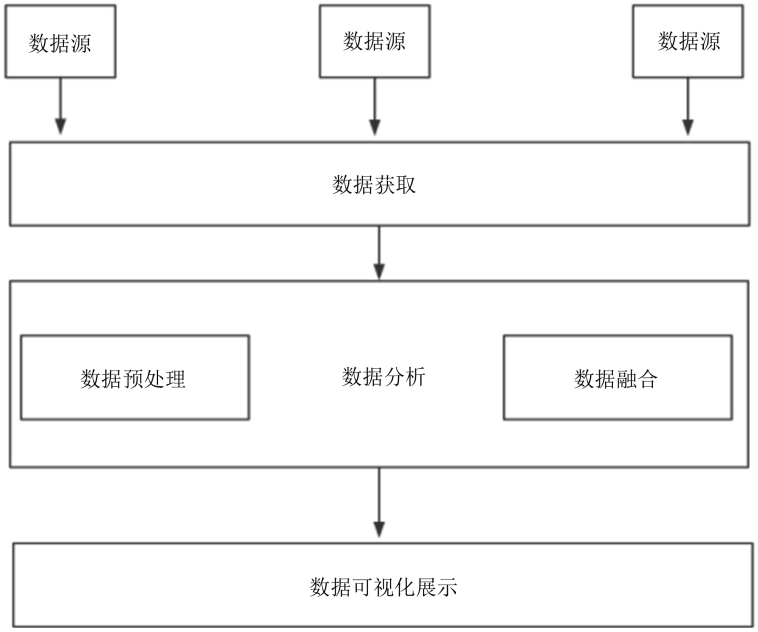

Figure 1. Procedure of forensics and data analysis
图 1. 数据取证分析流程

第二部分是数据分析。对于获取的原始数据往往并不能完全满足数据分析的要求，需要通过预处理对原始数据进行规整并集成。对于经过预处理的数据分析方法众多，本文主要介绍数据融合达到数据分析的目的。

第三部分是数据可视化。数据可视化要求根据数据特点和具体需求，综合利用多种手段工具，将数据分析结果进行图形化展示，以达到固定案件证据和挖掘涉案线索的目的。

3. Android 取证中地理位置信息及其提取

Android 手机用户在日常使用中产生了大量的地理位置信息，主要包括移动通信的基站信息，移动操作系统记录的位置信息，照片视频元数据中的附加信息，各类地图、打车、旅游、社交等移动应用的定位服务信息，WiFi 连接的位置信息等。

这些地理位置信息大部分都是由系统程序向用户声明且经用户允许下采集，存储于本地介质或者上传服务器云端。因此对地理位置信息的取证技术同时包含手机取证技术和网络取证技术。

Android 系统版本更新速度快，同时各个手机厂商对 Android 原生系统进行了定制，因此 Android 手机的系统平台具有多样性。本文以三星 S5 (SM-G9008V)作为实验平台，其系统环境为 Android 5.0 并已取得 ROOT 权限。数据的提取方法基于实际的数据提取，其方法对于基于镜像的提取也同样适用。

3.1. WiFi 信息

大量移动终端通过私人或者公共 WiFi 网络来访问互联网。Android 手机会自动存储 WiFi 网络的连接记录和网络的基本属性(例如 SSID 和 MAC 地址)。同时，WiFi 网络的无线接入点地理位置一般固定，根据 WiFi 接入信息可以得知关联的地理位置信息。

Android 系统使用一个修改版 wpa_supplicant 作为 daemon 来控制 WiFi，其配置文件/wpa_supplicant.conf 存储位置在/data/misc/Wi-Fi/目录下。该配置文件中定义的部分信息数据结构如表 1 所示。

3.2. 照片元数据

可交换图像文件格式(EXIF)通常附加于 JPEG、TIFF、RIFF 等文件之中，用于记录照片的属性信息和拍摄数据。其所记录的参数通常被称作照片元数据。照片元数据内容丰富，其中就包括 GPS 定位数据形式的地理位置信息。Android 智能手机在拍摄时，如果打开定位服务及相关选项会自动在 EXIF 中加入 GPS 定位数据。这一类地理位置信息配合拍摄时间较精确地反映了照片的拍摄时空信息。

3.2.1. EXIF 数据结构

对于 Android 智能手机默认存储照片格式 JPEG 文件，其以十六进制数值 FF D8 作为 SOI 标志(图像开始)，并以 FF D9 作为 EOI(图像结束)。文件头部有一系列“0xFF？？”格式的十六进制数值，称为 JPEG 标识，用来标记 JPEG 文件的信息段。EXIF 利用从“0xFFE0”-“0xFFD9”之间的应用标记(APPn)记录照片元数据[4]。EXIF 的官方文档关于元数据标签的定义较复杂，这里只列举常用的 GPSInfo 组件，见表 2。

3.2.2. 提取方法

针对 EXIF 的数据结构，我们使用 C#语言编写提取 JPEG 文件元数据的工具。以三星 SM-G9008V 手机拍摄的一张照片为例，读取其中 EXIF 数据。最后使用导出功能得到 html 格式的对应格式化摘要。摘要内容包括原照片浏览图、GPS 数据和 GPS 时间戳等，见图 2。

3.3. 即时通讯类应用

3.3.1. 地理位置信息存储形式

移动端常见即时通讯类应用有微信、手机 QQ、飞信、陌陌和米聊等，本文以较为广泛使用的微信为例。微信中的地理位置信息主要由用户聊天时地理位置分享产生，并保存在聊天记录数据中，Android 版聊天记录经加密存储在\data\data\com.tencent.mm\MicroMsg 路径下的 EnMicroMsg.db 数据库文件中，其中表 message 存储聊天内容，其表结构见表 3。

3.3.2. 提取方法

对于 Android 版微信，可以通过手机取证系统直接解密微信聊天记录，进一步获得地理位置信息。同时也可以借助免费的第三方软件实现微信聊天记录的找回操作。

Table 1. Data structure of Wi-Fi data
表 1. Wi-Fi 信息数据结构

字段名	ssid	Key_ngmt	autojoin	priority	psk
中文说明	接入点名称	认证密钥管理协议	自动加入	优先级	密钥

Table 2. Common GPSInfo components
表 2. 常见 GPSInfo 组件信息

字段名	标签名称	数据类型
GPSVersionID	GPS tag version	BYTE
GPSLatitudeRef	North or South Latitude	ASCII
GPSLatitude	Latitude	RATIONAL
GPSLongitudeRef	East or West Longitude	ASCII
GPSLongitude	Longitude	RATIONAL
GPSTimeStamp	GPS time	RATIONAL

Table 3. List structure of message (part)
表 3. message 表结构（部分）

字段	字段类型	说明
msgID	INTEGER	对话 ID/序号
msgSvrId	INTEGER	msg 服务号
type	INT	类型
status	INT	状态
isSend	INT	已发送
isShowTimer	INTEGER	计时器
createTime	INTEGER	创建时间
talker	TEXT	发起聊天者账号
content	TEXT	内容
imgPath	TEXT	图片路径
reserved	TEXT	存储位置
lvbuffer	BLOB	二进制标记
transContent	TEXT	传输内容
talkerId	INTEGER	发起聊天者 ID/序号

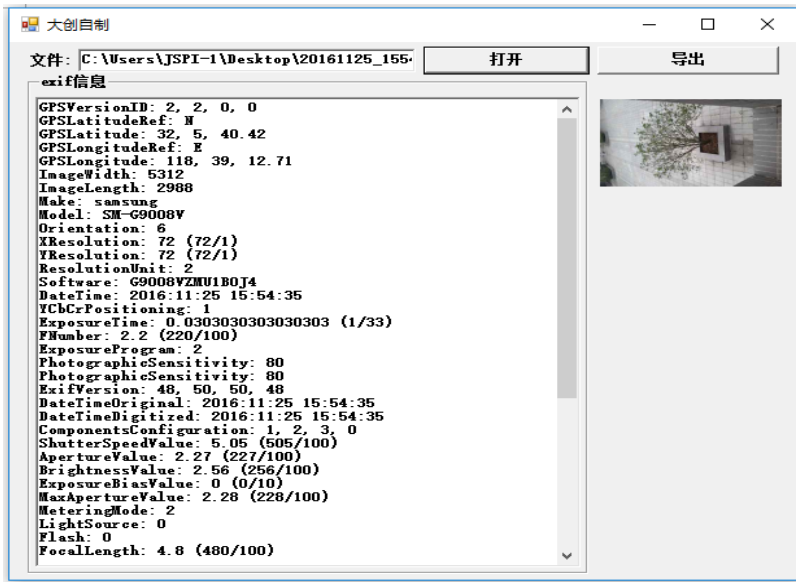

Figure 2. Program to read EXIF
图 2. EXIF 信息读取程序

如下是对微信(版本号：6.3.31) EnMicroMsg.db 数据库解密并提取位置分享信息的实验。首先将手机 IMEI(国际移动设备身份码)和微信 uin (微信用户信息识别码)字符串连接(见图 3)，其 MD5 值前七位即为数据库解密密码。利用工具 SQLCipher.exe 打开 EnMicroMsg.db 数据库文件，输入密码，即可成功打开文件。读取表“message”中记录的聊天记录，使用关键词“location” 执行 SQL 查询语句，从而获取所需的地理位置信息(见图 4)。

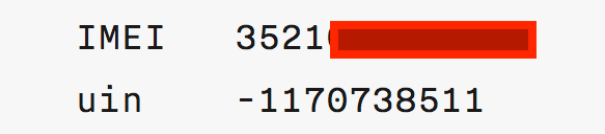

Figure 3. IMEI and uin
图 3. IMEI 与 uin

```
<?xml version="1.0"?>
<msg>
        <location x="32.092300" y="118.658653"
 scale="16" label="浦口区江苏警官学院(浦口校区)旁"
 maptype="0" poiname="[位置]" />
</msg>
```

Figure 4. Shared location data
图 4. 分享地理位置数据

3.4. 导航地图

经济社会高速发展的今天，复杂的路况已经成为户外出行的首要问题，由此百度地图、高德地图、腾讯地图等手机地图导航软件应运而生。本文以手机百度地图软件中地理位置信息的提取为例。

3.4.1. 相关信息存储形式

百度地图应用中地理位置信息主要包含用户定位记录、导航记录、步行记录等，这些数据存储在/data/data/com.baidu.BaiduMap/databases 路径下。在众多生成的数据文件中，名为“tracks.db”的数据库文件中各表详细记录使用者相应的地理位置信息，“tracks”数据库结构见表 4。

以存储用户定位记录的表 Track_Location 为例，其表结构如表 5 所示。

3.4.2. 提取方法

如图 5 用户定位记录所示，访问 data/data/com.baidu.BaiduMap/databases 目录，利用数据库查看软件 SQLite Expert Professional 即可查看数据库 tracks.db，并导出数据。

3.5. 网约车应用

网约车软件的诞生改变了传统打车市场格局。这里以滴滴出行为例进行相关地理位置信息的提取。

通常用户在使用应用时，会授权软件定位当前的地理位置，滴滴再根据司机的实时位置，向附近的车辆发送订单请求，一经接受即订单生效，以此来快速提供出行服务。这些位置信息包括常用地址信息(家和公司地址信息)，起点和终点信息。

由于滴滴出行应用将大量订单和用户信息数据存储于服务器，因此对此应用的数据提取需要可视化取证和介质取证相结合。滴滴出行应用的一部分订单信息和用户个人信息以数据库文件的形式存储在 data/data/com.sdu.didi.psnger/databases/路径下。

3.5.1. 数据库结构

在 datebases 中有许多数据库文件和 journal 文件，其中 im_database_281475098446424.db 中存储着有关订单信息，如图 6 所示。图中 Android_metadata 为 Android 数据库原文件数据，im_message_table 为会话信息，im_user_table 为用户相关的额表信息。

图 6 展示提取 im_user_table 表得到的用户与滴滴司机活动信息，然而具体位置信息并未包含于本地数据库中，而是存储在应用服务器上。因此需要使用可视化取证手段提取更多详细信息，见图 7。

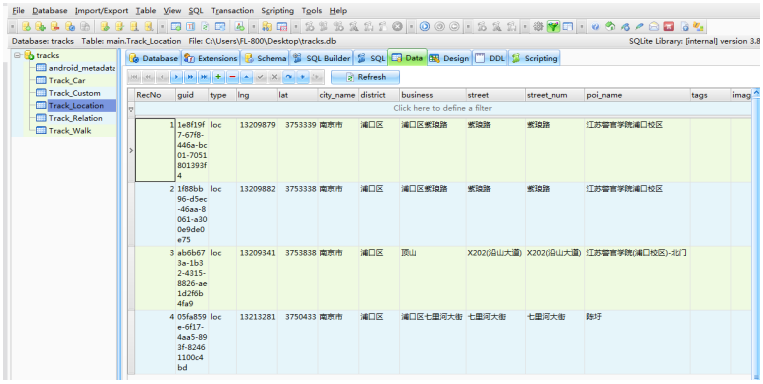

Figure 5. Record of user's tracks
图 5. 用户定位记录

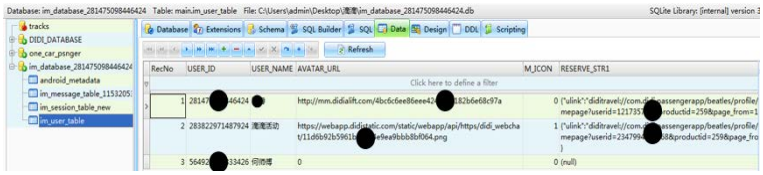

Figure 6. Orders of Didi Taxi
图 6. "滴滴出行"订单信息

Figure 7. Activities of users and drivers
图 7. 机主与司机进行的滴滴活动

Table 4. Database structure of tracks
表 4. tracks 数据库结构

表名称	说明
android_metadata	Android 元数据
Track_Car	用户导航数据
Track_Custum	用户常用地址
Track_Location	用户定位记录
Track_Relation	用户线路查询记录

Table 5. List structure of Track_Location
表 5. Track_Location 详细表结构

字段	字段类型	说明
guid	TEXT	数据库唯一标识符
type	TEXT	表的类型
lng	REAL	经度
lat	REAL	纬度
city_name	TEXT NOT NULL	城市名称
district	TEXT	行政区名称

从图 7 中可读出常用地址信息与订单的起终点信息，其中起终点信息以图针的形式在地图上标记出来，能更好地观察出行车路线与方位信息。

4. 数据处理

4.1. 数据预处理

地理位置信息数据来源途径较多，并且不同途径所获取的数据形式与结构不同，因此需要通过预处理对原始数据进行规整并集成统一。本研究提出可以使用基于 Javascript 对象表示法的地理空间信息数据交换格式——GeoJSON 格式进行统一编码。以 GeoJSON 数据结构为模板将各类信息调整为规范的数据结构，以规范地表示诸如点、线、面、多点、多线、多面等多种几何类型的地理位置信息，同时添加自定义属性。将地理位置信息规范为 GeoJSON 格式存储后，即可使用 Python 和 Javascript 等编程语言对其快捷读取与操作。

4.2. 数据融合

将来源不同的地理位置信息进行数据规整与集成，利于统一操作，从而进行数据融合和合成分析。数据融合的目的是合并或者综合多个数据集，成为一个完整的性能更好的数据集[5]。时间信息和空间信息是地理位置信息的两大基本要素。从这两要素出发，可以将地理位置信息的数据融合分为基于时间的数据融合和基于空间的数据融合，前者能体现信息对象某段时间的移动轨迹，后者能够反映某个位置或某个位置范围信息对象的惯性行为。

5. 数据可视化

5.1. 电子取证中的数据可视化

在电子数据取证中，数据可视化的实现目前主要有两种。第一种是使用商业软件产品，如 IBM i2 Analyst’s Notebook、美亚柏科可视化数据智能分析系统、智器云火眼金睛等。此类软件直接面向警务工作开发，操作简单，但价格不菲。第二种是利用开源工具和脚本语言实现，如 R 语言和 Python 语言及相关包，针对实际工作需求，开发扩展性较强的可视化工具，节约了购买商业工具授权的成本，然而具体实现的理论和技术基础要求高。可以说数据可视化的实现途径各有优势，在具体取证工作中应该根据实际需求，综合利用。

5.2. 地理位置信息可视化的实现途径

5.2.1. 基于百度地图 API 的展示

百度地图 URI API 是为开发者提供直接调用百度地图产品以满足特定业务场景下应用需求的程序接

口。按照接口规范构造一条标准的 URI，例如“http://api.map.baidu.com/geocoder?location=39.990912172420714,116.32715863448607&coord_type=gcj02&output=html&src=test”(参数说明见表 6)，将前期提取的地理位置信息传递给百度地图-URI API 反地址解析接口，经过逆地理编码后，定位信息即可以标注形式展示出来，如图 8 所示。

5.2.2. 基于 Leaflet.js 的展示

Leaflet.js 是一个用于创建交互式地图的开源 Javascript 库，其允许载入 GeoJSON 格式的地理位置信息并创建带有特定标识图案的地图，再结合 HTML 便能以网页的形式进行可视化展示。Leaflet.js 实现的可视化图案要素有图钉，折线轨迹，圆圈和弹窗等。基于 Leaflet.js 能够有效地表达位置痕迹和移动轨迹。为了方便高效地完成大量地理位置信息输出，使用 Python folium 库读取 GeoJSON 规范数据并输出 HTML 文件完成位置信息可视化展示。图 9 所示为以简化实验数据为例的实现代码和输出效果。

6. 结语

在案件侦查中，地理位置信息是重要的电子证据之一，能够为案件分析、确定相关人员的行动轨迹

Figure 8. Visual display via Baidu Map
图 8. 百度地图可视化展示

```
#coding=utf-8
import folium
import json

map = folium.Map(location=[32.093,118.652],zoom_start=12)
marker_cluster = folium.MarkerCluster().add_to(map)

folium.CircleMarker([32.093,118.652],      #绘制已知地点范围
                    radius=500,
                    popup='江苏警官学院',
                    color='#3186cc',
                    fill_color='#3186cc',
                   ).add_to(map)

folium.GeoJson(open('test.json'),   #读取GeoJOSN文件
               name='json'
              ).add_to(map)


map.save('./test.html')
```

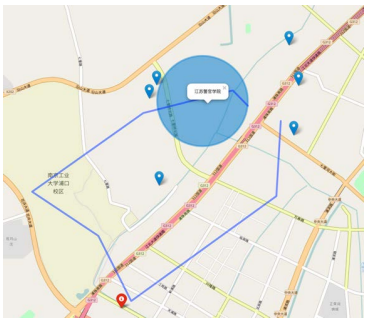

Figure 9. The code and examples of visual display
图 9. 实现代码与可视化效果示例

Table 6. Parameters of Baidu Map API
表 6. 百度地图 API 参数说明

参数名称	参数说明
location	lat<纬度>, lng<经度>
output	表示输出类型，web 上必须指定为 html 才能展现地图产品结果
coord_type	坐标类型，可选参数
zoom	展现地图的级别，默认为视觉最优级别
src	App 名称

提供帮助。本文构建 Android 取证中地理位置信息提取与分析的实现模型，通过数据可视化技术进行证据展示，为 Android 平台电子数据取证提供新的方法与思路。

另一方面，面对实际工作中海量复杂的数据，传统的数据分析手段已经捉襟见肘。因此，将机器学习和数据挖掘等技术运用于电子数据取证，实现对电子数据进一步有效和深度的分析，通过实战运用探讨手机取证获得的地理位置信息的应用价值，是未来研究和实践的方向。

参考文献 (References)

[1] Statista: 2011 年-2016 年全球智能手机出货量对比. http://www.199it.com/archives/509159.html

[2] Dunleavy, D. (2015) Data Visualization and Infographics. *Visual Communication Quarterly*, **22**, 68-68. https://doi.org/10.1080/15551393.2015.1029070

[3] Feng, Y.D., Wang, G.P. and Dong, S.H. (2001) Information Visualization. *Journal of Engineering Graphics*, 324-329.

[4] CIPA. DC-008-2016. Exchangeable Image File Format for Digital Still Cameras: EXIF Version 2.3.1[S]. CIPA, 2016.

[5] 魏彦飞, 等. 一种新的数据融合方法——广义融合法[J]. 天文研究与技术, 2016(3): 318-325.