学校编码：10384　　分类号_____密级_____

学号：X2012231197　　UDC_____

厦门大学

工　程　硕　士　学　位　论　文

基于数据挖掘技术的电子痕迹综合应用平台设计与实现

Design and Implementation of Electronic Evidence Integrated Application Platform Based on Data Mining Technology

杨智文

指　导　教　师：　姚　俊　峰　教　授

专　业　名　称：　软　件　工　程

论文提交日期：　2016　年　9　月

论文答辩日期：　2016　年　11　月

学位授予日期：　2016　年　12　月

指　导　教　师：_____

答辩委员会主席：_____

2016 年 9 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

2016 年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ √ ）1.经厦门大学保密委员会审查核定的保密学位论文，于年　月　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

2016 年　月　日

摘 要

随着计算机和网络技术的快速发展，通过计算机和互联网进行日常沟通已成为人们生活中不可缺少的一部分，很多日常行为的证据或线索不再是以传统物证的方式存在，而是以数字形式存储在计算机或网络中。另外，随着手机、计算机、平板电脑、移动电话、各种数码等通信终端的多样化，以及论坛、微博、电子邮件、即时消息等多种通信产品的出现，每个行为的主体之间的通联方式、沟通媒介多种多样，造成人们的通信行为轨迹分散，信息以片段的形式存在，进入所谓的“大数据时代”。这对电子数据取证提出了更大的挑战。随着各类电子介质的普及，电子数据的存储容量越来越大，并且利用各种计算机、手机、银行卡等载体进行的违法犯罪活动也呈蔓延的趋势。犯罪活动呈现出跨区域、团伙化的趋势。如何整合有限的取证工作人员、取证设备及各种孤立的社会数据资源，实现多个人员并行处理同一个案件、单个人员同时处理多个案件以及多个人员同时处理多个案件等工作模式，通过高效的分布式并行分析处理能力，快速进行案件线索扩展、情报研判和身份确认等工作，成为公安机关能否应对当前工作形势的关键，真正实现科技强警。

为此，我们研制的“电子痕迹综合应用平台”，能实现海量数据关联分析和碰撞，提炼和固化有价值的线索，为行业提供综合取证的解决方案。

关键词：电子痕迹；综合应用平台；数据挖掘

Abstract

With the rapid development of computer and network technology, through the computer and the network of information communication has become a part of social life is the indispensable, a lot of evidence or clues is no longer exist in the traditional evidence, but in digital form and stored in a computer or a network of computers. Also with the diversification of mobile phone, computer, tablet computers, telephones, digital and other communication terminal, and forums, twitter, email, instant messages and other communications products emerge in large numbers, each subject behavior between the communication mode, a variety of media diversity, resulting in line as the main communication behavior trajectory dispersion information exists in the form of fragment into the so-called "big data era". The electronic data forensics presents a greater challenge. The popularity of various types of electronic media, electronic data storage more and more big, illegal and criminal activities and use all kinds of computer, mobile phone, bank cards are also spread, cross regional, gangs and criminal activities showed the trend, how to integrate the limited evidence, forensic equipment and various isolated social data resources, to achieve multi a person with a case of parallel processing, a single staff to handle multiple cases and a number of staff to handle multiple case work mode, through efficient distributed parallel processing ability of case clues expansion, intelligence analysis and identification work quickly, become the key to the public security organs can deal with the current situation. The real implementation of strengthening police with science and technology.

Therefore, the development of "electronic trail integrated application platform, to achieve massive data association analysis and collision, refining and curing valuable clues, solutions for the industry to provide comprehensive evidence.

Key words: Electronic Evidence; Integrated Application Platform; Data Mining

目 录

Contents

第一章 绪 论

1.1 研究目的及意义

我国国家统计局数据表明，以2015年年底为界限，我国的移动用户群体人数总量接近12.8亿，其中，有接近1.15亿使用移动网络。目前存在的一个不争事实是，一般案件现场所涉及的嫌疑对象可能没有使用计算机，但几乎是人手必备手机的。在高通公司推出的千元智能机解决方案后，国产山寨智能机已逐渐取代了非智能机的市场，越来越多的人选择智能机。同时，随着智能机的功能逐步强大，现在的手机除了储存机主自身的身份信息和联系信息外，更多的承担原本由PC机来完成的并应用于互联网，用于沟通和交流的角色。手机作为一个私密的设备，往往记录了大量的涉案信息。如：被害人亲友、目标团伙中的其他个人信息等等。这些信息的获取可直接或间接的为案件提供线索及破案思路。但公安机关由于缺乏行之有效的手段，还无法通过读取手机中的各种信息，从而快速判断出对方身份，导致很多涉案人员在缺乏有效证据的情况下得以逃脱，或因重要线索缺失而导致案件进展缓慢等问题[1]。

本项目计划研发一套适合案件侦办人员使用的手机数据采集分析系统。该系统主要部署在派出所采集室中，且能够快速提取现场各类人员的手机通讯簿、通话记录、短信等逻辑数据和QQ账号。在后台建立的刑嫌人员的手机信息资源库，并可实现将其与GA现有的情报信息资源如在逃人员库、重点人口库等资源进行深度的数据挖掘、关联分析，进行人员身份的确认、线索扩展、线索摸排等工作，并以图表的形式向办案人员提供关联分析结果，这就可以便于办案人员进行决策数据的搜集与进行研判信息的确定。

1.2 国内外发展现状

随着计算机、网络技术和移动通信技术的飞速发展，使用计算机进行信息的处理和存储的方式已经被大众所认可和接受，广泛的电子数据信息采集的基础已经确立。人们使用手机、计算机结合有线和无线网络进行娱乐、沟通也逐步成为生活中不可或缺的一部分。在案件的配合的过程中，我们渐渐的发现，在多数个

体活动，大部分的线索、证据有两种存在方式：（1）传统物证；（2）数字存储形式，存储介质可以使手机，也可以是计算机、网络等。而多数情况下下，人们更倾向于后者。

随着移动通讯设备的不断发展和我国使用手机打电话、上网人群数量的急速增加，根据腾讯2014年第3季度的业绩报告显示，QQ同时在线人数已经达到1.85亿。而其中一半的数据来自手机QQ。随着智能机的普及，使用手机上网浏览网页、查询各种信息、发微博、聊QQ、收邮件已经越来越普遍。手机除了储存机主的身份和联系关系之外，还更多的承担了原先PC承担的在互联网进行沟通和交流的工具的角色。但目前由于缺乏有效手段，无法通过读取手机中的各种信息来迅速判断对方身份，从而可能导致很多在逃人员或与案件有重大关系的人员在无法提供更多证据的情况下只能被迫放走的情况[2]。

目前，虽然手机取证类得软件在国内已经有多家进行研发，也能够部分解决手机数据提取的问题。但由于最初设计思路是用于取证调查工作使用，在针对现场针对目标手机进行快速取证方面都存在一定的不足。同时，即使取到了手机中的数据。但由于缺乏其他相关数据的支持，无法即使为办案人员即使提供判断依据功能[2]。

本项目的建设就立足在解决此类问题，根据不同的使用场景，设计不同的采集前端用以采集刑嫌人员所持有的手机中的信息。智能终端采集数据分析系统的后台，从功能上主要划分为，汇聚采集数据，管理采集设备、将数据清洗过滤送“网综”平台以及在采集数据的基础上做技战法这四大类。

1.3 系统应用前景

近几年，我国政府高度重视信息安全技术及产业的发展，尤其是总理提出的“互联网+”的概念，各地政府都先后成立了信息安全产业基地。对于信息存储安全以及使用方面得到了国家各执法机关的高度重视，此外，这些机关也意识到数字化时代中信息技术也是越来越重要的，我们可以凭借信息技术有效打击或防范一系列危害社会安全事件以及各种犯罪活动。至今，国内的各大警院、政法大学

等高等院校也都设立了计算机取证、信息安全等相关专业，立足为国家培养急需的人才。

本项目就是基于目前公安机关的工作现状及外界大环境设立的，系统的功能是依据公安机关的工作需求来开发的，相信可以给最终用户提供真真切切的帮助。同时，我们的系统也会持续的改进，以满足公安机关不断变化的需求。

1.4 论文研究内容

本项目计划研发一套适合公安机关案件侦办人员使用的以手机数据采集分析为主的海量数据分析系统。该系统主要部署在派出所采集室中，且能够快速提取现场各类人员的手机通讯簿、通话记录、短信等逻辑数据和QQ账号。在后台建立的刑嫌人员的手机信息资源库，并可实现将其与公安机关现有的情报信息资源如在逃人员库、重点人口库等资源进行深度的数据挖掘、关联分析，进行人员身份的确认、线索扩展、线索摸排等工作，并以图表的形式向办案人员提供关联分析结果，这就可以便于办案人员进行决策数据的搜集与进行研判信息的确定。

1.5 论文组织结构

本论文在结构上分成了六章，行文安排为：

第一章绪论

确定研究对象为电子痕迹综合应用平台。主要概述介绍该应用平台的研究内容，分析了研究目、研究的现实意义，综述了当前该研究的国内外发展现状与趋势，最后，对客户的实际需求进行了梳理；

第二章 基本概念及相关技术介绍

基于前述的客户需求，本章节的核心内容为进行相关系统开发基本概念的介绍，并对相关的技术进行概述，为后续方案制定提供支持，为后续系统设计做铺垫；

第三章 对电子痕迹综合应用平台的需求分析

以公安机关内部用户为调研对象，分析他们在使用该系统时的切实需要用到

的技术功能进行需求分析与系统总体设计；

第四章 系统设计

系统主要涉及：（1）服务器端软件；（2）关键技术。本章的方案设计基于这两点展开；

第五章 系统实现

本章首先概述了相关的开发工具、开发及运行环境，建立符合一线公安机关工作人员需要的能够进行现场手机数据提取并上传数据到后端的工作机制与手机数据采集软件；构建基础手机信息资源库，为将来串并案分析提供决策数据。

第六章 系统测试

本章的核心内容是完成系统测试方案的制定，并对研发的系统执行压力测试来验证系统的功能；

第七章 总结与展望

本章总结了系统设计与实现的过程，并对后续系统的完善提出了几点建议。

第二章 基本概念及相关技术介绍

2.1 电子证据

信息化时代，热门越来越依赖与计算机和网络，这两者遍布于生活中的方方面面。例如，网络聊天、网络邮件、网络购物、网络办公、网络学习等，通过网络可以完成几乎所有的日常事务，即人们所说的“E 化生存方式”[3]。

电子数据取证主要是围绕电子证据来开展工作的，其目的就是将存储在计算机及相关设备中能够反映嫌疑对象的犯罪信息转化为有效的，能够当诉讼证据提供给法庭。

关于电子证据的定义和概念，目前没有一个统一的称谓，其实电子证据之所以叫电子证据，可能跟原来的纸面传统书证相对应，电子证据现在更多地被称为以计算机为基础的证据（Computer-based Evidence），传统书证被称为以纸面为基础的证据（Paper-based Evidence）。随着技术的发展，一种比较广泛的解释是，电子证据的存在方式是电子形式，可具体为某些证据材料或者与证据相关联的派生物，这种证据形成可借助于电子技术以及设备。通俗讲，就是利用计算机进行存储的对案件有证明作用的数据文件，可以是电话信息、电子聊天资料、电报、E-mail、传真资料、EDI、电子签名、电子数据交换等[3]。

2.2 电子取证规范

犯罪行为为电子数据取证的主要对象，包括不法分子对计算机的入侵操作、破坏操作、欺诈行为、攻击行为等，依照国家法律规范凭借计算机软硬件技术来实现数字证据的识别、保持和提交过程。找出入侵者以及入侵过程为取证的最终目的[4]。

电子数据取证的中心环节是要确保收集到的证据是有效的、真实的、有及时性，这个问题也是现今电子取证的首要考虑因素。但是电子证据比较容易受到破坏，所以，相关的法律法规就做了相关规定来保障电子证据的真实性界定、准确性界定、完整性界定以及正当性界定，其中，国际计算机证据组织就电子数据取证提出了以下的原则[4]：

2.2.1 不损害原则

主要是保障嫌疑人的合法权益，督促取证人员规范取证，相关要求是取证人员不能对嫌疑人的部分数据进行改动，主要是指计算等存储介质中的数据。这也是在电子数据取证分析时要必须使用只读锁连接到待分析的存储介质的必要性。

2.2.2 避免使用原始证据

原始证据不能作为取证人员的分析资料。如果情况特殊，那么相关的能力强的工作人员可以胜任对原始计算机等存储介质中数据的查询工作，此时操作人员因为对原始数据进行了访问则必须给出相应的解释以及理由。

2.2.3 记录所进行的操作

对调查期间出现的电子证据要进行记录等操作。第三方采用这些记录时可以获取一致的结果。

这也是对取证人员的取证过程进行监管的一个功能，有助于评判所取得的电子证据的有效性。

2.2.4 遵循相关的法律、法规

因为各个国家及地区都有相应的法律法规，取证人员在遵循技术原则的基础上，还必须遵循当地的法律法规来进行电子数据取证操作。

2.3 相关技术标准

《GAWA1006-2013 数据传输交换规范》

《GAWA2003-2014 智能终端数据快速取证设备》

《GAWA1001 网安数据元素集》

《GAWA1002 网安数据元素代码集》

《GAWA1003 网安源数据格式规范》

《GAWA1004 网安基础数据格式规范》

《GAWA1005 网安数据传输交换通用格式规范》

《GAWA1006 网安数据传输交换规范》

《GA-T 517-2004 常用证件代码》

《GB 02261.1-2003-T 个人基本信息分类与代码》

《GBT_02659-2000 世界各国和地区名称代码》

《GBT2260-2007 各地区编码》

《GBT-12403-1990》

2.4 系统架构

2.4.1 建设模式

电子痕迹综合应用平台建设在公安网里，数据按照集中存储、管理的思路，系统建议采取以下方式：

省、市两级建库，地市、区县、派出所应用的模式，在全省进行数据共享，并预留未来与其他系统进行对接的接口，进一步逐渐形成全省网内数据共享。

2.4.2 网络拓扑图

网络拓扑图如图 2-1 所示：