学校编码：10384　　分类号_____密级_____
学　　号：20051302484　　UDC _____

厦门大学

硕 士 学 位 论 文

基于关键词匹配的打印数据获取系统

A printing data acquisition system based on keyword matching

余 扬 武

指导教师姓名：吴顺祥 教授
专 业 名 称：系 统 工 程
论文提交日期：2008 年 4 月
论文答辩时间：2008 年 5 月
学位授予日期：2008 年 月

答辩委员会主席：_____
评 阅 人：_____

2008 年 5 月

厦门大学学位论文原创性声明

兹呈交的学位论文，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的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（　），在　　年解密后适用本授权书。

2、不保密（　）

（请在以上相应括号内打“√”）

作者签名：　　日期：　　年　　月　　日

导师签名：　　日期：　　年　　月　　日

摘　要

随着信息产业快速发展，打印机的使用非常广泛，为维护信息安全，维护单位利益，越来越多的具有前沿认识的管理者意识到对打印机的使用进行深层次的保护极为重要。毫无疑问，对打印机保护的最佳手段是要保证对打印内容的完全监控。

电子证据综合了文本、图形、图像、音频及视频等多种媒体信息。电子证据的来源主要有系统日志，IDS、防火墙、ftp、www 和反病毒软件日志，系统的审计记录(Audit trails)，网络监控流量(Network monitor traffic)，E-mail，Windows 操作系统和数据库的临时文件或隐藏文件，数据库的操作记录，硬盘驱动的交换(swap)分区、slack 区和空闲区，软件设置，完成特定功能的脚本文件，Web 浏览器数据缓冲，书签、历史记录或会话日志、实时聊天记录等等。

如今打印机打印的内容也可以作为电子证据，如犯罪嫌疑人打印过的报表、文档、图片等。目前，针对此类电子证据的取证办法不多，本系统较好地解决了打印数据的获取问题。

本文针对目前的计算机取证中的电子证据问题，设计并实现了一种基于关键词匹配的打印数据获取系统。简单介绍了计算机取证的相关知识以及打印系统原理，对获取打印内容过程中涉及到的几个关键技术进行了详细阐述。同时为实现更高效率的打印数据获取，提出了相应的关键词匹配算法。最后进行了性能测试，测试结果验证了该系统的效果。

关键词：电子证据；数据获取；关键词匹配

Abstract

As the fast development of information industry and the widespread application of printer, more and more managers with advanced insight have been aware of that it is very significant to secure the application of printer in order to protect the information and the interest of users. Undoubtedly, the best way to do that is to keep the content of printer under surveillance.

Electronic evidence is included many aspects like text, graphic, image, audio and video. The source of Electronic evidence mainly comes from system log, IDS, firewall, ftp, www, anti-virus log, system's audit trails, Network monitor traffic, E-mail, temporary files or hidden files in Windows and database, operation log of database, swap section of hard disk driver, slack space, idle space, software configuration, script to finish specific function, Web explorer cache, bookmark, history, dialogue and chatting log.

Today, the content of printer can be used as electrical evidence, such as the report forms, documents and pictures printed by suspects. But, it is not easy to get such evidence. This paper presents a practical solution about acquiring the content of printer.

In order to solve the problems of the electronic evidence about computer forensics, a printing data acquisition system was designed and implemented based on keyword matching. In this paper, the author gives a brief introduction about computer forensics and the principle of printing. In addition, the author expatiate some key technologies in the process of acquiring printing data. Moreover, a key word matching algorithm is proposed in order to acquire the printing data in higher efficiency. The testing results prove that this system has a good performance.

Key words: Electronic evidence; Data acquisition; Keyword matching

目 录

Contents

第一章 绪 论

1.1 打印数据获取的研究背景和意义

1.1.1 研究背景

随着信息技术与网络技术的飞速发展，越来越多的计算机联成网络，提供信息共享服务，给人们带来了工作的高效率和生活的高质量。与此同时，以计算机信息系统为犯罪对象和以计算机信息系统为犯罪工具的各类新型犯罪活动越来越多，比如非法截获信息、传播计算机病毒、利用计算机技术伪造篡改信息、进行诈骗及其他非法活动。打击利用计算机为工具实施的犯罪，确保信息安全对于国家的经济发展和社会稳定具有重大现实意义，计算机取证便是其中的一个重要手段。

计算机取证包括物理证据获取和信息发现两个阶段。物理证据获取是指调查人员来到计算机犯罪或入侵的现场，寻找并扣留相关的计算机硬件；信息发现是指从原始数据(包括文件，日志等)中寻找可以用来证明或者反驳什么的证据[1]。与其他证据一样，电子证据必须是真实、可靠、完整和符合法律规定的[2]。

犯罪的证据可能存在于系统日志、数据文件、寄存器、交换区、隐藏文件、空闲的磁盘空间、打印机缓存、网络数据区和记数器、用户进程存储区、堆栈、文件缓冲区、文件系统本身等不同的位置。其中，对打印数据的再现，作为电子证据获取的重要技术和内容有着越来越重要的意义，已经成为计算机取证领域的一个新兴和重要的研究方向。

1.1.2 打印数据获取的重要意义

如今打印机的使用已经非常广泛，打印机打印的内容也可以作为电子证据，如犯罪嫌疑人打印过的报表、文档、图片等。目前，针对此类电子证据的取证办法不多，本系统较好地解决了打印数据的获取问题。这些打印数据通过信息发现就可以作为电子证据。

本系统通过对进行打印过的硬盘进行关键词搜索的方式，完整且真实的还原了打印信息，让计算机取证调查人员对打印数据进行有效的分析和研究，结合打印数据的恢复，达到独特的打印内容监控和审核功能。并通过将犯罪嫌疑人打印过的报表、文档、图片等文件的再现，对打印电子证据的获取提供现实意义。

1.2 计算机取证

目前,国内许多法学界学者将计算机证据定义为:在计算机或计算机系统运行过程中产生的以其记录的内容来证明案件事实的电磁记录物，也称为电子证据 [3]。

电子证据的来源有很多，主要有操作系统日志， IDS、防火墙等安全设备的日志，网络上采集的数据流，传输的数据等[4]。

“计算机取证”首先由 International Association of Computer Specialists (IACIS) 在 1991 年举行的第一次年会中正式提出。计算机取证专家 JuddRobbins 的定义是：计算机取证不过是将计算机调查和分析技术应用于对潜在的有法律效力的证据的确定与获取。计算机紧急事件响应和取证咨询公司 New Technologies 进一步拓展了该定义:计算机取证包括了对以磁介质编码信息方式存储的计算机证据的保护、确认、提取、归档[5]。

因此，计算机取证是指对能够为法庭接受的、足够可靠和有说服性的、存在于计算机和相关外设中的电子证据的确认、保护、提取和归档的过程，也称为计算机法医学[6]。

1.2.1 计算机取证的概念

计算机取证是指对计算机入侵、破坏、欺诈、攻击等犯罪行为，利用计算机软硬件技术，按照符合法律规范的方式， 对能够为法庭接受的、足够可靠和有说服性的、存在于计算机、相关外设和网络中的电子证据的识别、获取、传输、保存、分析和提交认证的过程。

计算机取证学是计算机科学、法学和刑事侦查学的交叉学科。取证的目的是找出入侵者（或入侵的机器），并解释入侵的过程。取证的实质是一个详细扫描计算机系统以及重建入侵事件的过程。

1.2.2 计算机取证的步骤

(1) 现场勘查：保护现场，对现场进行勘查，获取物理证据；

(2) 识别证据：识别可获取的信息的类型，以及获取的方法；

(3) 传输数据：将获取的信息安全完整地传输到取证分析的机器上；

(4) 保存数据：确保与原数据一致，不对原数据更改和破坏；

(5) 分析证据：以可见的方式显示，结果要有确定性；

(6) 提交证据：以证据的形式按照合法的程序向司法机关提交证据。