电子证据的获取及可靠性关键技术研究*)

董晓梅 王大玲 于 戈 肖 珂 杨景辉
(东北大学信息科学与工程学院 沈阳110004)

摘 要 电子证据作为一种新的证据形式,逐渐成为新的诉讼证据之一。本文提出了一种结合法学、密码学、计算机网络安全技术、知识发现与人工智能技术来获取电子证据并保证其可靠性的研究方法,包括基于数据挖掘、入侵检测和入侵容忍技术的电子证据的高效获取技术,基于加密和网络公证技术的电子证据的可靠保全技术以及基于数据挖掘技术的电子证据的智能分析技术等。设计了一个智能取证系统 Intelligent-Forensic System(IFS),以解决电子证据的获取、传输、保存、分析、使用和可靠性保证问题。

关键词 电子证据,数据挖掘,入侵检测,入侵容忍

Study on Key Techniques for Acquiring and Reliability of Electronic Evidence

DONG Xiao-Mei WANG Da-Ling YU Ge XIAO Ke YANG Jing-Hui
(College of Information Science and Engineering, Northeast University, Shenyang 110004)

Abstract Being a new form of evidence, electronic evidence has been one of the new litigant evidence. In the paper, a research method for acquiring electronic evidence and ensuring its reliability incorporating law, cryptography, computer network security techniques, knowledge discovering and artificial intelligent techniques is proposed, which includes an effective acquiring technique of electronic evidence based on data mining, intrusion detection and intrusion tolerance, a reliable saving technique of electronic evidence based on encryption and network noatrization, and an intelligent analysis technique of electronic evidence based on data mining. An intelligent forensic system IFS is designed to solve the problems of acquiring, transmitting, storage, analysis, usage and reliability ensuring.

Keywords Electronic evidence, Data mining, Intrusion detection, Intrusion tolerance

随着计算机网络的广泛应用及以数字形式存在的信息的急剧增加,以计算机为犯罪目标和以计算机为犯罪手段的网络犯罪呈现出惊人的增长速度。电子证据作为一种新的证据形式,逐渐成为新的诉讼证据之一。广义的电子证据(Electronic Evidence),也称为计算机证据,是指借助电子技术或电子设备而形成的一切证据[1]。狭义的电子证据则是指数字证据(Digital Evidence),即能够证明发生了某一犯罪的证据,或者能够在犯罪行为与被害人之间、犯罪行为与犯罪人之间建立某种联系的数据[2]。这里我们使用狭义的电子证据概念,即数字证据。

目前,国外已经确认了电子证据的合法性。英、美、加拿大等国都制定了相关的法律,将电子证据作为合法的证据。在我国,电子证据还没有得到应有的重视,只在现行的《合同法》以及高法、高检、公安部的司法解释中,依稀可见电子证据法的极个别条款。但近年来,随着社会中出现的涉及电子证据的案件逐年增多,电子证据的重要性逐渐被人们所认识。大量的网络犯罪,每个案件的取证,都需要从计算机系统和网络中的数据中获取,安全地传输和保存,然后通过可靠的技术手段进行分析,最后提交为能说明事实真相的法律证据[3,4]。打击网络犯罪,获取可靠的电子证据已经成为亟待解决的问题,电子证据本身和取证过程的许多有别于传统物证和取证方法的特点,对司法和计算机科学领域都提出了新的研究课题[5]。作为计算机领域和法学领域的一项交叉研究课题——电子证据的获取和分析正逐渐成为人们研究与关注的焦点。

为此,本文提出了一种结合法学、密码学、计算机网络安全技术、知识发现与人工智能技术的研究方法,研究电子证据的获取及可靠性保证的关键技术,并在此基础上设计了一个智能取证系统 Intelligent-Forensic System(IFS)以解决电子证据的获取、传输、保存、分析、使用和可靠性保证问题。

1 计算机取证概述

计算机取证一般可分为3个阶段:证据获取、分析和法庭展示[6]。证据获取阶段保存系统的状态以备以后分析用。由于此时无法确定哪些数据可作为证据使用,因此通常的做法是使用工具将被怀疑设备上的所有数据复制到一个可信的设备上。分析阶段检查获取的数据,包括恢复被删除的文件和目录,从中找出证据,使用科学的方法根据证据做出结论。在法庭展示阶段,要出示结论和相应的证据。

国外的很多专家和研究机构在计算机取证领域进行了大量的研究工作:Lee Garber 认为,计算机取证是分析硬盘驱动器、光盘、软盘、Zip 和 Jazz 磁盘、内存缓冲以及其它形式的储存介质以发现犯罪证据的过程[7];计算机取证资深专家 Judd Robbins 给出了计算机取证的基本步骤[8];哥伦比亚大学法学院在一个专家系统的研究项目中论证了电子证据在法庭中的可信性[9];科罗拉多大学的 Roger King,Carolyn Stanley 则给出了一套增加电子证据在法庭上的可信性的方法[10]。在对黑客的关注上,全球最权威的网络安全论坛 FIRST(Forum of Incident Response and Security Teams)与美国警方在这个层面上高度一致[11,12]。早在1989年,美国国防部就存在着一个计算机取证实验室(CFLab)。到了1995年,美国一份名为"秘密服务"的调查报告就指出,至少有70%的法律部门拥有自己的计算机取证实验室。

*)本课题得到国家863计划 CIMS 主题(2003AA414210)、国家自然科学基金(60173051)和教育部优秀青年教师科研教学奖励计划资助。

目前，进行计算机取证工作，多是收集电子证据上传到取证服务器统一保护和分析，采用基于代理(Agent)的C/S结构。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[13~15]。

也有一些公司开发了计算机取证的软件工具，如Guidance Software公司的FastBloc数据采集器；该公司开发的Encase工具则提供了图形界面来查看和管理所有证据，软件还可以记录操作者的操作时间[16]；New Technologies公司开发的SafeBack是一个高级的证据保存工具，用于美国联邦执法部门处理计算机证据[17]。The Coroners Toolkit则利用文件的最近修改、访问时间进行分析，并根据数据节点的值提取出文件列表以进行恢复[18]。

在电子证据领域中不同以上的是关于公平反拒认协议的研究，它利用加密和数字签名的技术通过第三方保存流通数据，在防止发送方拒认和接受方拒认的同时保证保存传输数据的第三方不会了解数据内容[19,20]。

但是，大多数研究工作和商业工具都致力于帮助警察局办案人员进行事后的证据收集，而调查过程和工具则要求可疑的存储媒体和相关的硬件的物理映像的完整性，因而需要进行长达数月的分析，以恢复数字信息。而对于某些重要的信息系统，需要实时评估和分析遭到的攻击，不允许对受害计算机进行离线的检疫[21]。因此，目前的计算机取证方法和工具有很大的局限性。此外，虽然也有一些研究探索在取证分析中使用人工智能方法[22]，但目前对于电子证据的分析进行的深入研究仍然不多，多数工具只能做简单的分析，分析的自动化水平和智能性较低。同时，由于法律对电子证据的可采性的规定，在目前的技术条件下，电子证据的使用还有很多限制。如何使获取的证据符合法律的规定，仍是一个重要的研究课题。

从目前的研究状况来看，电子证据的获取和可靠性研究还处于起步阶段，还没有建立一套完整的电子证据获取规范和可靠性保证机制。而与之相关的法律方面的理论研究仍然显得很肤浅，并且存在大量的误区和混乱。以电子证据的获取和可靠性为目标的研究工作有很大的理论指导意义和实际应用价值，必将产生重大的社会效益。进行此项研究，既可以为相关法律立法提供技术上的支持，又有利于打击日益严重的网络犯罪。我国在这方面的研究处于起步阶段，学术界对此认识仍然不够，所做的工作非常有限，还有很多这方面的工作有待我们去完成。

2 电子证据的获取及可靠性关键技术

2.1 数字证据的高效、准确识别与获取技术

在海量的各种数据中以合法的手段高效、及时、准确地获取电子证据。包括静态数据的获取和动态数据的捕捉。目前已有一些针对静态证据的取证方法和取证工具，可以获取计算机文件，对磁盘做映像备份、恢复、重建交换文件、缓存文件、临时文件和已删除文件中信息等。但是，对于网络中的动态证据，如系统日志、网络数据包等，仅有以上的方法是远远不够的。而对于存储于磁盘上和流动于网络之间的大量数据，也需要进行有针对性的识别，从中获取有价值的证据。此外，入侵者对网络系统发起攻击后，往往会试图删除系统的日志文件，毁灭其犯罪的证据。因此，只有及时地检测到这类攻击行为，采取相应的措施，才能够维持系统的正常运转，并及时地记录下入侵行为的证据。因而，仅靠传统的计算机取证手段和现有的工具是远远不够的。

我们将传统的手段与数据挖掘、网络入侵检测和入侵容忍技术相结合，提出一种能够高效、及时、准确地获取电子证据的取证方法。包括：

(1)基于数据挖掘技术的静态证据的高效识别与获取方法。使用数据挖掘技术，可以针对案件特点和法律规定，通过对数据进行预处理、关联规则和分类规则挖掘，总结出其中的规律，在大量的数据中识别和发现与案件事实相关的信息。

(2)基于入侵检测和入侵容忍技术的动态证据识别与获取方法。要想获得罪犯入侵的充分证据，就必须能够及时地检测到入侵行为，并尽量减小其对系统的破坏程度，及时地记录其入侵的证据。我们将入侵检测和入侵容忍技术应用于取证系统中，一旦检测到入侵行为，立即采取相应措施，保护系统，保存证据。入侵检测系统可以检测到各种入侵和可疑行为，而入侵容忍系统可以对入侵和可疑行为进行隔离，保障系统的连续运转。

2.1.1 *高效的静态证据识别与获取* 借鉴和改进已有的证据收集方法和工具，基于我们研究的数据挖掘算法，提出了一种高效的静态证据识别与获取方法。

(1)将Web挖掘中数据预处理和Web挖掘数据源重构技术融于入侵证据的获取过程中，以便更加及时、充分地获得所需要的证据；

(2)将法律中的相应规定作为领域知识或本体知识，对获取的证据进行裁减和优化处理，提高证据的准确性和有效性；

(3)将关联规则、分类规则挖掘技术应用于证据获取过程中，通过建立各种现象和证据之间的分类模型，获得通过现象获取证据的分类规则。

2.1.2 *动态证据的识别与及时获取* 基于我们设计的分布式协作入侵检测与响应机制，结合并改进已有的计算机取证方法，实现动态证据的识别与及时获取。

(1)基于我们研究的分布式入侵检测及容忍结构，构造动态证据识别与获取机制；

(2)综合使用人工神经网络、数据挖掘和计算机免疫学等方法，结合误用入侵检测和异常入侵检测方法，对发现的入侵行为和可疑行为进行记录，对可疑行为进行隔离监视，对入侵行为采取及时的处理措施；

(3)使用我们提出的基于XML消息交换的分布式入侵检测与响应协作机制，有效地发现一些复杂的入侵行为。

2.2 电子证据的可靠保全技术

收集到的电子证据的传输基于安全的传输协议，数据在安全传送到取证系统的机器上后，就要对数据进行保全，以防止内部或外部非法人员的篡改和删除，保证电子证据的连续性。在司法实践中，证据的保全有常规保全和公证保全等方法。同传统证据相比，电子证据的保全有一定的特殊性，体现为一种虚拟空间的保全，因而传统的保全方法，如对证人证言和当事人陈述制作询问笔录和录制资料，对物证和视听资料进行勘验、扣押、封存，对书证的复制和存档，对电子笔录进行打印、核正、签字盖章封存等方法具有很大的局限性，难以适应对电子证据的保全要求。网络公证保全作为一种新型的电子证据保全方法，不仅适合于对电子证据的保全，而且还可以

增强证据的证明力。我们的思路是结合密码学技术，基于数据加密和数字签名算法，设计电子证据的保密传输和网络公证协议，由网络公证机构对电子证据进行网络公证，保证电子证据的可靠性和真实性，增强电子证据的证明力。

我们基于密码学技术，结合数据备份、错误恢复和纠错编码等技术，设计了电子证据的安全传输和保全机制。

(1)基于我们研究的混沌序列密码算法和网络安全协议，结合数据备份、错误恢复和纠错编码等技术，设计电子证据的保密传输协议；

(2)基于我们提出的基于 RSA 和离散对数的门限签名算法，设计网络公证协议，提出电子证据的可靠保全方法，保障证据的可靠性和连续性。

2.3 数据的智能分析和处理技术

考虑到收集来的数据之间的关联性，我们采用数据挖掘、智能推理机制等方法，对收集的数据进行处理，分析出犯罪嫌疑人与犯罪事实之间的关系，以此可以得到司法部门认可的电子证据。具体做法如下：

(1)将犯罪证据和相应的法律条文、证据规则进行形式化描述，表示成为数据挖掘和人工智能技术能够分析和应用的数据集；

(2)应用聚类分析技术，将多种不同的证据根据其相似性归类，以便进一步的分类、预测和相关性分析；

(3)以证据作为数据项(item)，应用关联规则挖掘技术，进行不同类别的证据之间、相同类别证据之间的相关性分析，得到同时出现的证据之间的关联规则；

(4)以犯罪证据为测试属性、以犯罪事实为分类属性，应用分类规则挖掘技术，获得犯罪证据与犯罪事实的分类模型，以便根据证据对犯罪行为进行推断；

(5)应用时间序列模式挖掘技术，从相当一段时间内的犯罪证据和犯罪事实的发展规律中发现规律；

(6)建立分析评价模型，通过对上述各种规则挖掘结果的分析和评价，获得挖掘过程中所需的各种参数的设定范围，保证数据挖掘的准确性。

2.4 取证过程和方法的合法性

证据的获取、分析和使用的过程应具有合法性，不能违反法律的规定，不能侵犯他人的合法权利。对于电子证据，凡是生成、取得等环节不合法，其不合法程度足以影响证据真实性的，或者足以影响某一重大权益的，则可能会被排除。如，通过窃录方式获得的电子证据；通过非法搜查、扣押等方式获得的电子证据，情节严重的；通过非核证程序(未经审核合格的计算机程序)得来的电子证据等。要借鉴西方发达国家的证据法中的各种证据规则，结合我国诉讼法和其他相关法律、法规的原则，研究取证方法和过程的合法性问题。同时，分析得到的证据，还必须具有关联性和客观性。如果是直接证据，则必须有说服力；如果是间接证据，各种证据还必须要形成证据链条，要能够彼此互相印证，以证明案件事实，不能相互矛盾。对于证据的分析和认定，包括真实性和证明力(可靠性、关联性、完整性)的认定，也需要制定相应的法律规范。电子证据法律的发展和完善，是以相关的科学技术的发展为基础的；同时，研究电子证据的获取、分析和正确使用的技术，最终目的是要服务于立法和司法工作。我们将在技术研究成果的基础上，制定电子证据的获取、分析和使用规范，将我们的研究成果应用于司法实践中，并可作为我国的立法工作的参考。

3 智能计算机取证系统 IFS

在电子证据的获取和可靠性保证技术的研究基础上，我们设计了一个智能计算机取证系统 IFS。IFS 系统由入侵检测器、数据挖掘与入侵响应部件、网络公证、加密传输以及智能推理与数据挖掘部件组成。系统总体结构如图1所示。

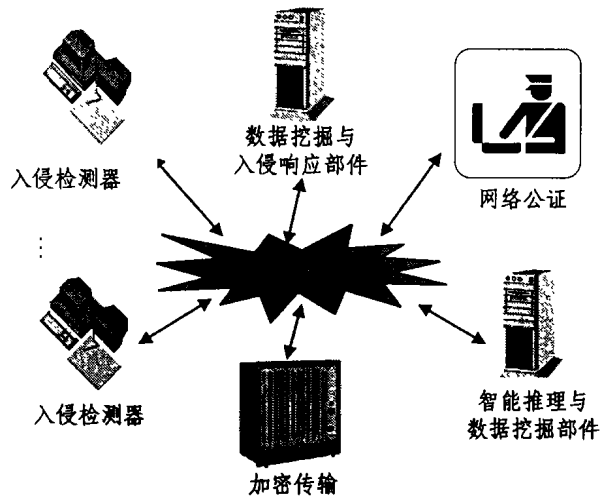

图1 系统总体结构

系统中在每台需要保护的计算机上安装入侵检测器，由证据获取主机上的入侵检测响应部件和数据挖掘部件完成电子证据的识别与获取；证据经过网络公证和加密后传输至分析主机；分析主机上的智能推理和数据挖掘部件实现对证据的分析。各部件之间的关系如图2所示。

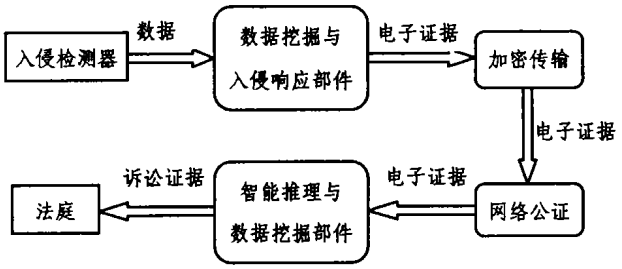

图2 系统各部件之间的关系

结束语 本文提出的关于电子证据的获取和可靠性关键技术的研究方法，涉及网络安全、数据处理、人工智能、行为科学、密码学、法学等多种学科，对其进行深入研究具有很高的理论研究价值和广阔的应用前景。目前，该领域的研究国内外都处于探索阶段。我们提出的研究方法和 IFS 系统在以下几方面有所创新：(1)基于数据挖掘、入侵检测和入侵容忍技术的电子证据的高效获取技术。(2)基于加密和网络公证技术的电子证据的可靠保全技术。(3)基于数据挖掘技术的电子证据的智能分析技术。(4)结合计算机技术和法学原理的电子证据获取、分析的合法性问题研究。(5)诉讼活动中电子证据的获取、分析和使用规范。

参 考 文 献

1 何家弘主编. 电子证据法研究. 北京：法律出版社，2002

2 Casey E. Digital Evidence and Computer Crime：Forensic Science，Computers，and the Internet. Academic Press，2000

(下转第148页)

名的构造不是很简捷，因为协议的初始化要求仔细地选择安全参数。下面我们先介绍一下这个算法(图4)。其中，k 是安全参数，k' 是消息泄漏参数。我们定义 $R=2^{k+k'}S$ 和 $M=2^{k+2k'}S$，其中 $S\geqslant 2\cdot Ord(g)$ 用来限定秘钥的范围。

定理7 这个协议是统计盲签名协议。

定理8 在随机 oracle 模型中，对这个盲签名协议采用平行攻击时，构造一个一次以上的伪造(one-more forgery)比对 N 进行因子分解要困难。

4 安全和效率

从使用的角度来看，选择一个1024位长的模 N 和一个160位长的非对称基 g 是合适的。消息泄漏参数可以固定为64位，安全参数则根据情况从24位变化到128位。所以，它的安全性依赖于1024位模的因子分解。从证明者的观点来看，这些协议非常有效。确实，我们只需要做少量必要的实时计算，包括一次乘法和一次加法，而且所使用的数字都非常小。

由于一些数据可以预先计算，因此证明者在执行证明的过程中只要进行一次乘法和一次加法。对于盲签名，银行只要用一个168位的数去乘一个128位的数，并把结果加到一个360位的数上。这样，银行可以以很小的存储代价每秒盲签成百万条的消息。

下面，我们对算法的效率能进行一个直观分析(表1)。

结论 我们列出了许多基于复合离散对数问题的算法。从认证算法到盲签名算法，我们已经证明了有效的算法安全性至少和因子分解问题一样。而且，我们提出了一种新的盲签名算法，其计算负载非常小，所以该算法非常适合成千上万个用户的大范围应用。

表1 算法的效率分析

算法	认证	签名	盲签名
模	|N|=1024位；|p|=|q|=512位		
Ord(g)	160位		
安全参数	k=24	K=128	
消息泄漏参数	K'=64		
|S|(>|Ord(g)|)	168位		
|R|(=|S|+k+k')	256位	360位	
|M| (=|S|+k+2k')			424位
在线代价(证明者)	Mult(24,168) +Add(256,192)	Mult(128,168) +Add(360,296)	
通信	360位(45字节)		
签名大小		488位(61字节)	552位(69字节)

参考文献

1 Schnorr, Guillou. Quistquater Composite discrete logarithm and secure authentication. In: Third Intl. Workshop on Practice and Theory in Public Key Cryptosystems, PKC, 2000

2 Brickell E F, McCurley K S. An interactive identification scheme based on discrete logarithm and factoring …, Advances in Cryptology-Eurocrypt'90, LNCS 473, Springer-Verlag, pp. 481～486

3 Girault M. An identity-based identification scheme based on discrete logarithms modulo. a composite number Advances in Cryptology-Eurocrypt'90, LNCS 473, Springer-Verlag, pp. 581～586

4 Poupard. Practical multi-candidate election system. PODC, 2001. 274～283

5 Feige U, Shamir A. Zero knowledge proofs of identity. In: Proc. the Nineteenth Annual ACM Symposium on Theory of Computing, New York City, 1987. 210～217

6 Pointcheval. Provably Secure Blind Signature Schemes (1996) David. ASIACRYPT: Advances in Cryptology--International Conference on the Theory and Application of Cryptology

7 Girault M, Stern J. On the Length of Cryptographic Hash-Values used in Identification Schemes. Identification schemes. In: Proc. of Crypto 94, Lecture Notes in Computer Science 839, 202～215

(上接第145页)

3 梁锦华，蒋建春，戴飞雁，卿斯汉. 计算机取证技术研究. 计算机工程，2002，8：12～14

4 钱桂琼，杨泽明，许榕生. 计算机取证的研究与设计. 计算机工程，2002，6：56

5 David WJ, Stringer-Calvert. Digital evidence. Communications of the ACM, 2002, 45(4): 128

6 Carrier B. Open Source Digital Forensics Tools: The Legal Argument: [Research report]. http://www.atstake.com/.

7 Garber L. Computer forensics: high-tech law enforcement. IEEE Computer, 2001, 34(1): 22～27

8 Judd R. An Explanation of Computer Forensics. http://www.computerforensics.net/forensics.htm

9 MacCrimmon M T. Expert Systems in Case-Based Law: The Hearsay Rule Advisor. In: Proc. of the second intl. on conf. on Artificial intelligence and law. Vancouver, British Columbia, Canada, 1989. 68～73

10 King R, Stanley C. Ensuring court admissibility of computer-generated records. ACM Transactions on Information Systems (TOIS), 1985, 3(4): 398～412

11 Farmer D, Venema W. Forensic Discovery. FIRST. 2002 Hawaii, USA, 2002

12 Schaffer G P. The Role of Computer Forensics in the Investigation of Network Intrusion Activity. FIRST. 2002 Hawaii, USA, Jun 2002

13 Lunn D A. Computer Forensics: An Overview. http://www.sans.org/rr/incident/forensics.php.

14 Carrier B. Defining digital forensic examination and analysis tools Using Abstraction Layers. International Journal of Digital Evidence. 2003, Volume 1, Issue 4, http://www.ijde.org/.

15 Schneier B, Kelsey J. Secure audit logs to support computer forensics. ACM Transactions on Information and System Security (TISSEC), 1999, 2(2): 159～176

16 http://www.guidancesoftware.com/encase/frame_encase.html

17 http://www.forensics-intl.com/safeback.html

18 http://fish.com/tct/

19 蒋晓宁，叶澄清. 一种新的公平反拒认协议. 计算机工程与应用，2000，5：40～42

20 蒋晓宁，叶澄清. 电子证据与反拒认协议. 通信学报，2000，7：76～81

21 Giordano J, Maciag C. Cyber Forensics: A Military Operations Perspective. International Journal of Digital Evidence. 2002, Volume 1, Issue 2. http://www.ijde.org/.

22 Mukkamala S, Sung A H. Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques. International Journal of Digital Evidence. 2003, Volume 1, Issue 4. http://www.ijde.org/