基于电子取证技术的持续审计模型研究

景　波, 刘　莹, 陈　耿

(南京审计学院 信息科学学院, 江苏 南京　211815)

摘　要: 缺乏系统的持续审计模型构造方法已成为制约当前计算机审计发展的瓶颈。在此背景下, 以持续审计的工作原理为基础, 本文提出一种基于电子取证技术的持续审计模型构造方法, 它可以使持续审计更具有实时性和连续性。该模型通过实时监控异常行为的发生, 一方面可以进行实时取证, 对异常行为做详细记录, 另一方面可以触发响应策略对不同强度的异常行为实施相应的响应。

关键词: 电子取证; 持续审计模型; 计算机审计; IT 审计; 信息系统审计

中图分类号: F239.1; TP399　**文献标识码**: A　**文章编号**: 1672－8750(2011)04－0058－05　**收稿日期**: 2011－03－25

作者简介: 景波(1975—　), 男, 江苏南京人, 南京审计学院信息科学学院副教授, 主要研究方向为 IT 审计、数据挖掘; 刘莹(1977—　), 女, 江苏南京人, 南京审计学院信息科学学院讲师, 主要研究方向为数据挖掘、分布式计算技术; 陈耿(1965—　), 男, 江苏南京人, 南京审计学院信息科学学院副院长, 教授, 博士, 主要研究方向为数据挖掘、审计信息化。

基金项目: 国家自然科学基金(70971067); 江苏省高校自然科学重大基础研究项目(08KJA520001); 江苏省"六大人才高峰"项目(2007148)

一、引言

电子取证和持续审计都是近年来的研究热点。目前电子取证和持续审计正处于起步阶段, 还没有形成系统的理论体系。笔者在审计实践中发现, 诸多审计人员在获取被审单位电子数据时较为被动, 无法直接将电子数据取证。笔者认为, 把电子取证和持续审计结合起来可能是一个很好的解决办法, 它将会使审计人员从被动获取被审单位数据转变为主动提取电子证据。

IT 审计已在我国开展了十几年, 但目前的审计工作仍主要关注"事后审计"和"周期审计", 极少能满足持续审计的实时性和全面性要求[1], 在国内几家大型国有企业开展的持续审计试点工作也只停留在数据层面的实时交换上。理论体系和技术手段的缺乏已影响到我国对复杂的商业环境开展持续审计工作的顺利进行, 因此我们有必要结合我国的审计信息化现状, 广泛借鉴其他领域对网络数据的监管经验, 展开对持续审计理论体系和技术手段的研究, 为促进我国 IT 审计发展提供参考。

电子取证是指对存储在计算机系统或网络设备中的潜在电子证据进行识别、收集、保护、检查、分析以及法庭出示的过程[2-3]。电子取证现在主要针对两个方面: 对事发后的计算机系统的静态证据收集分析; 结合入侵检测、防火墙、网络侦听等网络安全工具对实时发生的行为进行动态取证。2005 年 4 月我国成立了中国电子学会计算机取证专家委员会, 同年 6 月首届中国计算机取证技术峰会在北京召开。近年的科研项目也加大了对电子取证的研究力度, 如国家 863 计划之子课题"电子物证保护与分析技术"、国家博士基金项目"计算机动态取证技术研究"、国家"十五"重点项目"计算机侦查技术研究"等。目前电子取证的成熟产品主要有北京大学计算机研究所开发的"计算机犯罪证据固定与保全工具箱", 北京中网安达信息安全科技有限公司的"网络神捕"综合取证系统, 北京天宇宏远开发的硬盘拷贝机, 中

软公司开发的网络信息监控分析与取证系统，厦门美亚柏科开发的计算机犯罪取证勘察箱，上海金诺网安开发的介质取证系统 DiskForen，上海盘石数码开发的计算机现场取证系统，等等。

开展持续审计，一般应具备两方面的技术条件：一方面是被审计系统完全自动化，另一方面是相关事件或结果能够即时被审计人员提取[4]。持续审计完全可以采用电子取证技术来满足对数据连续采集和分析的要求，同时经过司法认证的电子取证手段也能使审计人员获取的被审单位数据具有法律效力。

二、电子取证的方法和特点

（一）电子取证程序的合法性

随着当前计算机技术的发展，审计人员从被审单位获得的书证、物证等证据信息大多以电子数据的形式出现，这些数据信息在证据法上就是电子证据。由于电子证据使用的磁性介质，其记录保存的内容经常被改动，而且即使被改动或添加也不易留下痕迹。另外由于人为原因或非人为原因（如环境、技术），电子证据的安全性和真实性受到威胁，一旦发生争议，这种电子证据在诉讼或仲裁中能否被采纳为证据就成为法律上的难题。因此笔者认为，电子数据的提供、收集、调查和保全应当符合法定程序，以电子取证的技术手段来进行依法审计，即通过窃录方式、非核证程序、非法软件得来的电子证据，不予采用[5]。

（二）电子取证的常用技术

电子取证可被视为“从计算机中收集和发现证据的技术和工具”[6]。以计算机证据的来源为标准，电子取证技术可分为单机取证技术、相关设备取证技术和网络取证技术。单机和相关设备取证包括数据恢复技术、加解密技术和系统口令获取、信息搜索与过滤技术、存储映像拷贝技术、源码反向工程技术等；网络取证则侧重在网络上跟踪犯罪痕迹或通过数据通信的信息资料获取证据的技术，常包括 IP 地址和物理地址的获取和识别技术、身份确认技术、电子邮件的鉴定和取证技术、网络监听和监视技术、数据包过滤技术、系统漏洞扫描技术等。

（三）证据分析技术

电子证据取证和鉴定最重要的工作是对需要鉴定的电子数据进行分析，并提交鉴定分析报告。电子数据分析包括对文件残存空间、未分配空间和自由空间中所含信息的发掘，对交换文件、缓存文件、临时文件中所含信息的复原，对计算机在某一特定时刻运行内存中数据的搜集与分析等，同时，它还要在已经获取的数据流或信息流中寻找、匹配关键词或关键短语进行数据分析。它的具体工作是：分析文件属性、文件的数字摘要和日志；发现目标系统中的所有文件，包括现存的正常文件、已经被删除但仍存在于磁盘上（即还没有被新文件覆盖）的文件、隐藏文件；尽可能地恢复已发现的被删除文件；最大程度地显示操作系统或应用程序使用的隐藏文件、临时文件和交换文件的内容；分析在磁盘的特殊区域中发现的所有相关数据。

三、基于取证技术的持续审计模型

持续审计的典型特征是利用技术优势来缩短内部审计周期、改善风险和控制安全系数。王刚认为，联网审计是对被审计单位财政财务收支的真实、合法、效益进行实时、远程检查监督的持续行为，是一种“全新的审计理念与审计模式”[7]。毕秀玲分析了持续审计的特征及产生和发展的动因并界定了持续审计的效用[8]。阚京华比较分析了几种常见的持续审计技术实现模型[9]。目前，常见的持续审计框架如图 1 所示。

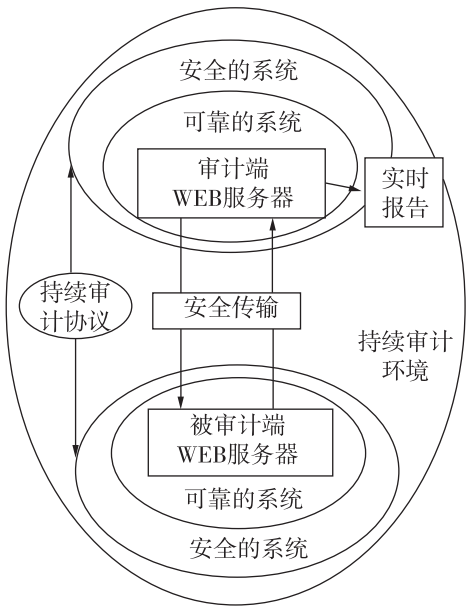

图 1　基于网络的持续审计概念模型

持续审计框架的主要组成部分包括互连的网

络服务器、持续审计协议、安全可靠的系统、实时更新的报告[10]。审计端一旦发现与审计所定义规则不一致的情况，将以例外报告或警告的方式通过网络通知审计人员，但是，传输数据的不确定性和电子数据的易失性、可篡改性会给所采集数据的可鉴证性带来较大风险。

本文所讨论的持续审计模型利用电子取证的技术优势构造了高度自动化的审计过程并实时提供审计证据。电子取证技术的应用避免了目前审计中反复经被审单位二次认证的繁琐，并有效地保护了审计人员的审计意图。基于取证技术的持续审计模型是以网络活动为基础，以异常数据变化为核心，以完整、真实、有效记录计算机系统活动为目的的一种持续审计系统。系统模型如图 2 所示。

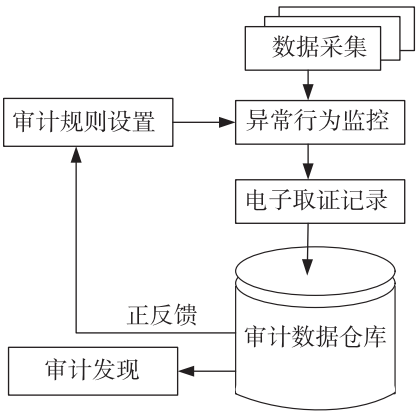

图 2　基于取证技术的持续审计模型

基于取证技术的持续审计模型由采集模块、数据分析模块、响应及其结果处理模块组成。采集模块使用目前技术成熟的经公安部认证的网络安全产品，并将采集的原始数据存入数据库。数据分析模块采用基于统计的抽样分析和基于关联的特征选择的方法，将数据按协议和特征进行统计分析，用数据挖掘方法建立数据仓库。响应及其结果处理模块用于处理可疑信息，除此以外，结果处理模块可以定期生成统计报表，同时反馈策略对不同强度的异常行为实施相应的审计规则响应。

（一）取证数据源

取证信息源主要分为防火墙信息、入侵检测信息、网络设备日志、数据库日志以及基于 NetFlow 技术的网络异常行为日志和主机数据。上述信息均来自经公安部认证的网络安全产品。主机数据来自于用户习惯信息、操作系统审计日志、系统日志和应用程序日志。其中，用户习惯信息是指因为用户使用习惯留下的痕迹；操作系统审计日志是由操作系统软件内部的专门审计子系统产生的，它记录当前系统的活动信息，并将这些信息按照时间顺序组织成为一个或多个审计文件；系统日志是指系统使用日志机制记录主机发生的事件；应用程序日志可以反映系统活动的较高层次信息，它是用户级别的系统活动抽象信息。

（二）数据分析模型

持续审计模型中事件分析模块的目的有两个：一是提供有关网络运行情况的统计及查询；二是对统计结果进行数据挖掘，作进一步的分析计算，以生成相应的审计发现模型。系统构成如图 3 所示。

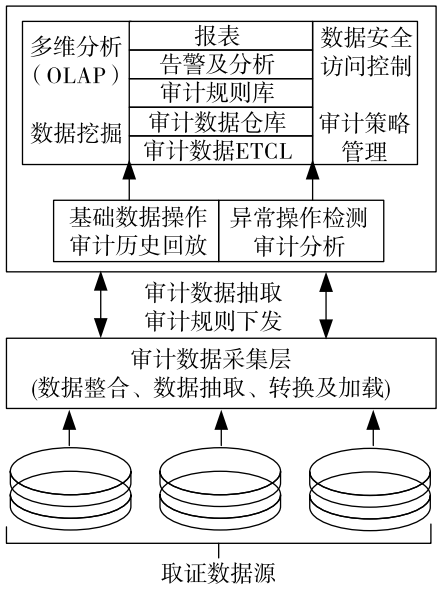

图 3　数据分析模型

网络运行情况的统计及查询采用基于统计的数据分析，并提供在线实时访问。系统将原始数据以小时为单位，按协议类型（TCP、UDP、ICMP 等）、源地址/端口、目的地址/端口等进行流量统计并记录统计结果，以便于数据挖掘分析和建立数据仓库。

对统计结果进行数据挖掘时，系统主要分析数据的关联性和潜在的行为趋势。该阶段分析将决定系统对实际网络环境的适应程度和系统的自学习能力。它将提取出的行为特征反馈到审计规则中来修正数据采集层的活动。在数据分析层中，数据仓库是数据和模型的存储中心，它多层的存储结构能够有力支持复杂运算和海量存储，而联机分析处理（OLAP）可以对多维的数据进行有效的分析处理和趋势发现。

目前常用的分析方法有以下五种：

1. 基于统计模型的检测分析方法。统计模型中常用参数包括审计事件的数量、间隔时间、资源消耗情况等；可用于检测分析的典型统计模型有操作模型、均值和标准差模型、多元模型、马尔柯夫过程模型等。

2. 基于人工神经网络的检测分析方法。神经网络具有自适应、自组织和自学习的能力，可以处理一些环境信息十分复杂、背景知识不清楚的问题。在采用统计处理方法很难达到高效准确的检测要求时，系统可以构造智能化的基于神经网络的检测分析器。

3. 基于专家系统的检测分析方法。与运用统计方法与神经网络的方法不同，用专家系统对网络行为进行审计主要是针对有特征的违规行为。所谓的规则即是知识。专家系统的建立依赖于知识库的完备性，而知识库的完备性又取决于审计记录的完备性与实时性。违规行为的特征抽取与表达，是专家系统检测分析的关键。

4. 基于数据挖掘技术的检测分析方法。数据挖掘（DataMniing，DM）是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程。DM 的对象可以是数据库、文档以及其他结构化和半结构化的数据。DM 的目标是从数据库中发现有用的信息、知识、规律、规则等。数据挖掘持续审计模型及实现技术为当前审计和计算机交叉领域的研究热点之一[11]。

5. 基于模糊理论的检测分析方法。违规活动往往不是孤立的活动，而是带有目的性的一个有机的活动序列。违规者往往首先收集有关信息，然后试图获得某操作的访问权限，之后窃取或修改系统的某些信息。这种检测分析方法在一定程度上可以将违规行为检测分析转化为一个多模糊证据综合判决的问题。

（三）基于犯罪意图的证据链检测

尽管上述分析方法已经在很多领域有所应用，但对于特定的审计目标来说，它们仍存在很多不足之处——或者不具有通用性，或者实施成本太高，这样就不能较好地满足目前的需要。同时，模型中的审计规则库可以根据现行法律制度、企业业务逻辑关系、不同数据间的钩稽关系、审计人员的实际经验和审计专家的科学预测来设立，但其效率有待经一步提高。

在具体的审计工作中，很多案件是已从别处发现了违规或违法事实、需要通过电子数据完善其证据链的情况。笔者受此启发，模型中将取证犯罪事实提升到符合犯罪意图的证据收集上。在分析电子数据时，犯罪意图将有助于指导审计人员理清犯罪步骤，而步骤间的因果关系则是发现犯罪痕迹的线索。由于计算机犯罪大多数不是孤立的行为，而是经过多个单步异常行为后才完成的一个完整的犯罪过程，因此，笔者提出一种基于犯罪意图的证据链检测。该检测方法建立的依据是复合犯罪的单步行为间具有因果关系，前因即实施犯罪所必须具有的前提条件。算法的基本思想是：首先建立异常行为特征库，使用特征描述语言表征所有的行为。异常特征库中记录了每一种单步犯罪行为的前提条件和犯罪结果以及单步行为发生的环境条件。然后，设计一种自动规则生成器，系统根据自动生成的规则集对所采集信息进行匹配和检测[12-13]。该方法的优点是事先不知道整个作案过程，也不需要产生大量的关联规则，其自动规则生成器的构造对效率影响较大；其缺点是不能发现新的犯罪，并且 CPU 开销较大。

在算法改进方面，我们可以根据特定犯罪行为会在不同的日志中留有不同痕迹的事实，对算法进行扩展，增加日志事件、证据支持关系和证据支持度，将特定行为与特定的日志事件关联起来，采用证据支持度来评价重构后的犯罪场景中攻击步骤的正确性[14-15]，同时根据异常日志事件来推断可能的破坏行为，合并被漏报分离的攻击场景。对于改进算法的应用，由于篇幅所限，笔者将另文讨论。

四、结论

电子取证和持续审计都是近年来的研究热点，正处于起步阶段，还没有形成系统的理论体系。在此背景下，以持续审计的工作原理为基础，笔者提出了一种基于电子取证技术的持续审计模型构造方法，它可以使持续审计更具有实时性和连续性。该模型通过实时监控异常行为的发生可进行实时取证，对异常行为做详细记录；同时它可以触发响应策略对不同强度的异常行为实施相应的响应。该模型为及时发现审计线索和破获重大犯罪事件提供了详细的电子证据和良好的分析平

台。笔者希望本文能为我国实施面向重大事件的持续审计提供理论和实践上的指导。

参考文献:

[1]何家弘. 电子证据法研究[M]. 北京:法律出版社,2002.

[2]蒋平,黄淑华,杨莉莉. 数字取证[M]. 北京:清华大学出版社,中国人民公安大学出版社,2007.

[3]王玲,钱华林. 计算机取证技术及其发展趋势[J]. 软件学报,2003(9):1635-1644.

[4]Kogan A, Sudit F E, Vasarhelyi A M. Continuous online auditing: a program of research[J]. Journal of Information Systems,1999,13:87-103.

[5]Alles M G,Kogan A,Vasarhelyi A M. Feasibility and economics of continuous assurance, auditing[J]. Practice and Theory,2002,21:125-138.

[6]Cheung S, Lindqvist U, Fong W M. Modeling multistep cyber attacks for scenario recognition[C]. The Third DARPA Information Survivability Conference and Exposition (DISCEX III). Washington: IEEE computer Society Press,2003:284-292.

[7]王刚. 关于联网审计的几个基本问题[J]. 审计月刊,2005(5):49-50.

[8]毕秀玲. 持续审计基本问题研究[J]. 审计研究,2008(4):16-20.

[9]阚京华. 信息技术环境下连续审计技术实现模型分析比较[J]. 科技管理研究,2009(10):285-287.

[10]Woodroof J, Searcy D. Continuous audit model development and implementation within a debt covenant compliance domain[J]. International Journal of Accounting Information Systems,2001,2:169-191.

[11]谷瑞军,陈圣磊. 数据流挖掘及其在持续审计中的可用性研究[J]. 南京审计学院学报,2011(1):36-40.

[12]Dain O, Cuningham R. Building scenarios from a heterogeneous alert stream[EB/OL]. [2011-02-03]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.5277&rep=rep1&type=pdf.

[13]Kim J Sun, Kim M, Noth B N. A fuzzy expert system for network forensics[C]. The 2004 International Conference on Computational Science and Its Applications (ICCSA 2004), Perugia,2004:117-129.

[14]Ning Peng, Xu Dingbang, Healey G C, etc. Building attack scenarios through integration of complementary alert correlation methods[C]. Proc. Of the 11th Annual Network and Distributed System Security Symposium. NDSS,2004:233-255.

[15]Reith M, Carr C. An examination of digital forensics models[J]. Int'l Journal of Digital Evidence,2002,3:1-12.

(**责任编辑:**黄　燕)

On Continuous Auditing Model Based on Electronic Forensics Technology

JING Bo, LIU Ying, CHEN Geng

Abstract: The lack of systematic guidelines for the construction of continuous auditing models has been a bottleneck hampering the development of computer-aided audit. Under this background and the principles of continuous auditing model, it proposes a construction method based on the electronic forensics technology. The model makes continuous auditing more real-time and continuous. Through real-time monitoring abnormal occurrence, on the one hand it may carry on the real-time synchronized forensics and make the detailed records of abnormal behavior; On the other hand it may activate the response system to implement corresponding response to the abnormal behavior of different intensity.

Key words: electronic forensics; continuous auditing model; computer-aided audit; IT audit; information system audit