云计算支持下的协作式数字取证技术①

路 红[1], 廖龙龙[2]

[1](南京理工大学 紫金学院, 南京 210046)
[2](南京体育学院 信息科学与技术教研室, 南京 210014)

摘 要: 为解决网络环境下电子证据分散、取证分析效率低、协作难度大等问题, 在分析计算机犯罪特点以及当前数字取证所面临的相关问题基础上, 针对数字取证与分析的协同需求, 设计了一种具有正循环反馈机制的云计算支持下的协作式数字取证模型, 并详细论述了其设计思想和体系架构. 最后, 研究了模型的系统实现方法、电子证据云存储调度策略、基于封锁机制的并发分析任务调度. 实验表明, 协作式数字取证技术可有效提高数字取证工作效率和分析结果的准确性.

关键词: 数字取证; 云计算; 电子证据; 计算机犯罪; 协同取证

Collaborative Digital Forensics System Based on Cloud Computing

LU Hong[1], LIAO Long-long[2]

[1](Zijin College, Nanjing University of Science and Technology, Nanjing 210046, China)
[2](Laboratory of Information Technology, Nanjing Sport Institute, Nanjing 210014, China)

Abstract: In order to solve the problems about decentralized electronic evidence, low efficiency of forensic analysis, cooperative forensics difficultly, etc. A collaborative digital forensics model with positive feedback loop mechanism based on cloud computing are designed according to the needs of digital discovery and analyze cooperatively, which is based on the characteristics of the computer crime and related problems of the current digital forensics. Then, design idea and architecture of the model are discussed in detail. Finally, the realization method and key techniques of the model are analyzed, such as scheduling policy for cloud storage and task scheduling for electronic evidence based on locking mechanism. Experiments show that collaborative digital forensics could effectively improve the efficiency of digital forensics and the accuracy of analysis results.

Key words: computer forensics; cloud computing; electronic evidence; computer crime; collaborative forensics

1 引言

自 1962 年计算机犯罪(Computer Crime)的概念被提出之后, 身份盗用、网络欺诈、E-Mail 威胁、垃圾信息骚扰、黑客攻击、网络跟踪等形式的计算机犯罪活动日益频繁. 网络通信技术以及数字化终端设备的普及和低成本化, 尤其是云计算技术、智能感知技术、社会网络服务、反侦查技术的发展, 使计算机犯罪呈现网络化、隐密化、跨区域合作等发展趋势, 导致数字取证面临取证困难、电子证据异构性、协作效率低等问题.

为解决数字取证所面临的问题, 文献[1]通过对证据图推理机制(Evidence Graphs)的研究, 提出了一种利用证据图推理方法进行数字取证分析的技术. 文献[2]基于模糊决策树算法设计了一种数字取证分析系统, 以帮助数字取证人员在互联网环境下对计算机犯罪案件进行取证分析. 文献[3]基于远程控制技术提供了三种动态获取电子证据的方法, 并设计了一种支持电子证据动态获取的计算机取证系统. 文献[4]提出了一种对多源日志数据文件进行格式化处理的方法, 采用逻辑公式对所构建的日志树结构模型进行形式化描述, 较好地解决了多源数据的处理问题. 文献[5]根据网络攻击时间与空间的相关性, 通过对数据流和数据包特

① 收稿时间:2014-04-17;收到修改稿时间:2014-05-07

征的融合，实现了对网络攻击过程的分析和攻击源的定位. 此外，工业界研发的一些专业数字取证分析工具，如计算机取证工具 Forensic Toolkit、分布式取证系统 ForNet、电子数据关联分析工具 Nuix Investigator、专业电子证据采集工具 X-Ways Capture 等，也被广泛应用在数字取证实践中.

现有的相关研究和开发主要针对某一个独立的数字终端设备或较小的计算机网络进行电子证据的采集和分析，无法满足数字取证环境动态变化、电子证据来源多样化、取证分析高效性等要求.

云计算通过网络将计算资源与存储资源进行有效整合，通过强大的云端计算和海量存储空间，为用户提供可伸缩的高性能计算服务和高可靠的网络数据存储服务[6]. 利用云计算技术能够实现数字取证的实时网络协作、电子证据安全存储与共享、海量电子证据智能分析等，构建满足协作式数字取证需求的服务系统. 为此，本文针对网络环境下计算机犯罪的新特点，基于云计算技术研究一种协作式数字取证技术，并分析其模型架构与系统实现方法.

2　协作式数字取证模型设计

2.1 设计思想

数字取证源于已发现的计算机犯罪活动，或针对可疑计算机系统的入侵检测，其目标是形成具有法律效力的证据报告，以帮助侦查计算机犯罪案件. 协作式数字取证要确保电子证据的合法性和完整性、取证过程的机密性和高效性、犯罪事实的可重现性、数据存储的安全性和可共享性、协同取证的灵活性和高可靠性，其基本流程如图 1 所示. 因此，协作式计算机取证流程不是一个简单的线性过程，而是一种具有正反馈循环机制的协同取证分析过程，电子证据的网络共享与协同分析有助于帮助计算机取证人员更准确、更及时地定位和采集相关电子证据，提高计算机取证效率及其分析结果的准确性.

(1) 取证工作准备

取证工作准备直接影响后续的电子证据采集与分析，是保证数字取证合法性和有效性的基本保障，包括分析计算机犯罪现场、确定数字取证方案和保护取证对象三个方面，即根据计算机犯罪案件设计数字取证方案、选择相应的取证工具、调配相关取证工作人员，同时保护犯罪案件涉及的相关电子证据和攻击痕迹，为重现数字取证现场提供有利条件.

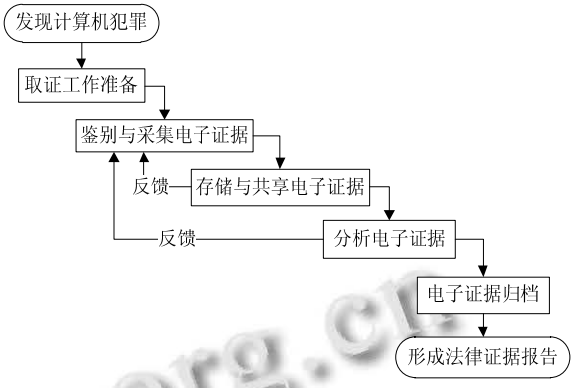

图 1　协作式数字取证流程

(2) 鉴别与采集电子证据

为提高数字取证工作效率，首先对电子证据来源进行预测，确保电子证据的采集具有针对性，以免错过获取电子证据的最佳时机. 对于采集到的电子证据，要对其来源、获取方法与时间、取证人员、有效性评价等进行注释，以便取证分析人员对电子证据的准确性和完整性进行鉴别.

(3) 存储与共享电子证据

针对同一宗计算机犯罪案件，不同的取证人员首先将采集的电子证据上传到数字取证系统. 一方面，有助于帮助取证人员挖掘新的电子证据来源或调整现有的电子证据采集方法，及时获得更全面的电子证据. 另一方面，有助于取证人员对电子证据进行关联性和有效性评估，通过协作提高电子证据的准确性.

(4) 分析电子证据

取证分析人员利用数据挖掘、关联分析、统计分析等技术对已获取的电子证据进行协作式取证分析，通过分析结果的比较归纳与统计分析，形成具有法律效力的证据报告.

(5) 电子证据归档

电子证据归档是根据电子证据归档的规范性要求，对法律证据相关的原始电子证据、取证方案、取证工具、取证时间、取证现场、取证分析过程等进行完整保存，以便法官在审理案件过程中随时调取相关电子证据.

2.2 模型框架

根据图 1 所示的协作式数字取证流程，基于云计算平台设计了如图 2 所示的数字取证模型，包括取证

接口层、证据获取层、证据分析层、数据存储层和协同工作层. 云计算平台通过虚拟化技术、资源调度技术、状态监控技术等为协作式数字取证提供虚拟存储空间和高性能计算服务, 满足协作式数字取证对电子证据采集、存储、共享与分析的需求.

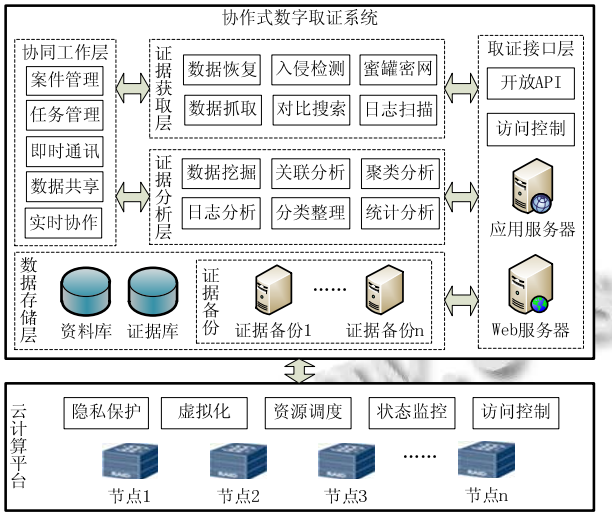

图 2 协作式数字取证模型架构

(1) 取证接口层

取证接口层包括开放 API、访问控制、Web 服务器和应用服务器, 是用户访问证据获取层、证据分析层和数据存储层的入口, 也是系统各个功能模块之间进行数据传输、功能协调、系统扩展与集成的重要组件. 此外, 取证接口层具有用户管理、身份认证、操作日志管理等功能, 是整个协作式数字取证系统的交互接口. 为使取证人员方便地查询所需的功能、快速完成相应操作, 取证接口层支持通过 Web 浏览器和应用程序客户端两个方式实现用户与服务器端的交互.

(2) 证据获取层

根据犯罪现场分析和电子证据来源预测, 利用数据恢复、入侵检测、蜜罐密网、网络数据抓取、对比搜索、日志扫描等技术, 分布式获取与案件相关的电子证据, 为证据分析层提供犯罪案件相关的完整、准确的电子证据.

(3) 证据分析层

证据分析层是协作式数字取证模型的核心层, 主要利用关联分析、聚类分析、日志分析等技术对已获取的电子证据进行智能分析, 发现电子证据之间的时间先后关系、逻辑关系、与犯罪案件的相关度等, 通过分类整理形成可以呈交法庭的法律证据报告. 一方面, 多名取证分析人员针对犯罪案件相关的同一组电子证据进行协作式分析, 有助于提高取证分析的工作效率和分析结果的准确性. 另一方面, 针对同一组电子证据进行分析结果的比较分析, 有助于激发新的取证思路, 发现新的电子证据来源, 保证电子证据获取的全面性.

(4) 数据存储层

数据存储层为用户提供数据存储与管理服务, 包括资料库、证据库、电子证据备份系统. 资料库存储系统配置信息、系统用户资料、系统操作日志等, 通过实时监控数字取证人员的系统操作日志, 规范数字取证过程, 防范计算机取证人员的违规或违法取证操作. 证据库存储采集到的原始电子证据和取证分析结果, 能够对新增的电子证据自动进行多副本数据备份.

(5) 协同工作层

协同工作层为证据获取层和证据分析层提供协作管理机制和协作工具, 实现基于取证任务分配的协作工作流程控制、电子证据审核与共享、在线协作式取证分析等. 案件管理始于计算机犯罪案件立案调查, 结束于该犯罪案件完成取证分析或终止调查, 是数字取证与分析的服务对象. 任务管理是协作式数字取证过程管理的基础, 可实现取证人员调配、取证任务管理、取证过程协同控制, 保证协作式数字取证工作高效地进行. 即时通讯是为取证分析人员提供实时消息通信、视音频通信、高清视频会议、取证经验与工具共享、实时协作等, 是实现协作式数字取证与分析的重要基础.

3 模型实现的关键技术

3.1 实现方法

云计算平台是协作式数字取证模型实现的基础, 为其提供电子证据网络存储和高性能取证分析服务. 开源 Apache Hadoop 包括分布式文件系统 HDFS、并行处理模型 MapReduce、非关系型数据库系统 HBase 等核心组件, 能够为用户提供底层细节透明的云计算基础架构[7]. 其中, MapReduce 在分布式文件系统 HDFS 基础上实现了分布式计算和并行任务处理, 分布式文件系统 HDFS 利用 MapReduce 实现可靠的文件存储与管理, HBase 数据库管理系统利用 MapReduce 实现海量数据的高效管理与高并发访问[8]. 我们利用

Apache Hadoop 2.1.0-beta 构建了一个包含 1 个 Master 节点和 3 个 Salve 节点的云计算实验平台作为协作式数字取证系统部署的基础云计算平台. 由于协作式取证分析涉及海量电子证据的关联分析、聚类分析、偏差分析等复杂计算, 采用 MapReduce 并行处理模型可提高电子证据分析效率.

TestDisk 是一款功能强大的开源磁盘数据恢复软件, 具有修复分区表、恢复被删除的分区、支持从 FAT/exFAT/NTFS/ext2 等文件系统中恢复被删除的文件, 可用于对已删除或已篡改电子证据的获取. NetworkMiner 是一款支持 Windows、Linux、Mac OS X、FreeBSD 等系统平台的开源网络取证分析工具, 能够主动抓取以太网和 WiFi 网络上的数据包, 支持对离线 PCAP 文件的解析和网络流量的分析, 适用于动态电子证据的获取和电子证据的解析. 为实现基于 Web 的电子证据共享和协同分析处理, 使数字取证人员无需安装和更新系统客户端, 便可方便地进行协作式数字取证与分析, 我们采用基于 Ajax 动态网页开发框架实现 B/S 架构的 Web 客户端. 实时消息通信协议 XMPP 支持即时消息通讯、文件传输、视音频通信等, 利用 XMPP 应用服务器 ejabberd、开源 XMPP Web 客户端开发包 jaxmpp2 实现数字取证人员在线状态管理、Web 即时通讯、数据传输与共享、视频会议、在线协同工作等功能.

3.2 电子证据的云存储与共享

云存储系统采用存储冗余数据的方式自动保存数据的多个副本, 利用低廉的计算机设备即可构建用于电子证据存储与共享的私有云存储系统, 在实现电子证据数据安全存储和电子证据归档的同时, 降低电子证据存储与共享的成本. 电子证据归档通常要求存储多个副本, 以便通过数据完整性校验判断归档的电子证据是否被篡改或损坏, 即使某一份归档的电子证据被损坏或丢失, 也可以根据其他归档电子证据副本快速进行恢复.

当并发用户数较多、数据传输量较大时, 传统网络数据存储系统往往受到并发数和数据传输速度的限制, 而云存储系统中的每份数据采用多份副本进行备份, 数据文件以数据块的形式分散存储在不同集群的 DataNode 节点中, 当用户访问云存储系统中的某个数据文件时, 可以同时从多个 DataNode 节点中读取包含该文件的数据块, 解决并发数据访问和数据传输速率较低的问题.

Hadoop 中的一个作业被划分成若干个任务并行处理, 但由于异构网络环境下节点性能差异或个别节点故障导致部分任务进展缓慢甚至启用备份任务机制, 为提高云存储系统的吞吐率、减少请求响应时间, 本文采用 LATE 算法进行云存储任务的调度[9]. 设 *SCP* 表示同一时刻备份任务执行数, *STT* 表示当前任务是否需要进行备份, *SNT* 表示节点处理任务快慢的阀值, *PGS* 表示进程速率, *PRS* 表示任务按照分片策略得到的进程分, t 表示任务已执行时间, T 表示任务执行的剩余时间, 则:

$$PGS = PRS / t \tag{1}$$

$$T = (1 - PRS) / PGS \tag{2}$$

LATE 算法的基本过程为: 若有一个节点请求新任务, 且整个 Hadoop 系统中备份任务执行数小于 *SCP*, 则执行以下步骤:

1) 如果节点速率小于 *SNT*, 则忽略这个请求, 结束执行, 否则继续执行;

2) 对正在执行的任务进行排序, 按照式(1)和式(2)计算任务完成所需的时间, 从低到高对任务进行排序;

3) 重新备份执行速率低于 *SNT* 且排在第二位的任务.

3.3 协作式电子证据分析

协作式电子证据分析由处在不同空间的取证分析人员在不确定的时间内进行, 除需进行电子证据共享和实时互动之外, 还需对电子证据的协同分析过程进行冲突检测和并发控制. 多人针对同一电子证据的分析可能产生不同甚至相互矛盾的分析结论, 为避免这种情况以及前面提交的分析结果被后面提交的分析结果覆盖, 系统规定在同一时刻仅允许一名取证分析人员对同一个电子证据进行在线分析处理.

为解决取证分析人员协作过程中并发操作之间的冲突问题, 本文采用封锁机制对并发取证分析任务调度处理, 只有获得电子证据封锁权限的用户才能对其进行分析. 基于封锁机制的并发取证分析调度算法如下: 设电子证据为 Q, 多个并发取证分析指令分别为, 共有 X 锁和 S 锁两种类型的锁. 在指令对 Q 进行分析之前, 必须首先申请获得 Q 上的 X 锁; 在释放 Q 上的 X 锁之前, 其他操作指令仅能获得 Q 上的 S 锁. 如果指令获得了 Q 上的 S 锁, 则仅能查看电子证据 Q 的分析结果, 而不能对其进行分析处理. 这样, 如果所有的并发取证分析指令都遵守上述封锁机制, 则可以证明针对这些并发取证分析指令的任意调度是冲突可串行化的.

本系统中的并发控制模块采用 Java 的 Socket 通信

技术和 Vector 类提供的对象串行化机制实现, 分为服务器端和客户端. 客户端负责向服务器端发送处理指令和预处理对象, 服务器端运行的并发控制服务自动处理来自不同客户端的并发取证分析请求, 通过给预处理对象施加相应的 X 锁或 S 锁保证并发分析操作的有序执行. 这样, 一份电子证据在同一时刻只响应一个取证分析人员的处理指令, 其他取证分析人员需等待该电子证据的 X 封锁被释放后, 才能申请获得对其进行 X 封锁和分析处理; 在等待获取分析操作权限的过程中, 取证分析人员可在线查阅他人的分析方法与结果, 实时调整自己的取证分析计划, 避免针对同一电子证据的重复分析和并发冲突, 保证取证分析结果的一致性, 缩短并发用户的平均响应时间.

4 协作式取证实验

某公司的一份商业机密文件泄漏, 公安机关接到报案后, 分派 *A*、*B*、*C*、*D* 共 4 名数字取证人员针对该公司 16 台可疑联网计算机进行数字取证分析, 以重现该机密文件外泄过程. 该案件的数字取证与分析过程如下: 首先, 分别为 4 名数字取证人员分配相应的取证任务, 如每人负责 4 台计算机中近 10 天内已删除文件的数据恢复、来往电子邮件内容获取、Web 浏览器历史数据抽取、HTTP 数据解析等工作. 其次, 数字取证人员对所获取的数据进行分析和过滤, 并将与案件相关的电子证据上传到服务器端的云存储空间进行共享. 最后, 取证数字取证人员对云存储空间的相关电子证据进行协作式分析, 生成具有法律效力的证据报告.

表 1 实验结果

项目	人员 *A*	人员 *B*	人员 *C*	人员 *D*	总计
恢复文件数	345	445	145	315	1250
数据恢复时间(分)	23	35	32	41	131
Web 历史记录数(条)	36	135	235	86	492
发送 E-Mail 数(份)	21	4	7	18	50
接收 E-Mail 数(份)	12	2	6	78	98
发现线索数	465	665	165	68	1363
取证时间(小时)	34	29	45	39	147

应用本系统进行数字取证的过程, 将成为一个以取证任务分配和并发分析调度为基础的协作式电子证据采集与分析过程, 取证人员的注意力将集中在电子证据的采集、共享和协同分析方面. 由表 1 的实验结果可以看出, 多名数字取证人员同时针对一个计算机犯罪案件进行电子证据的协作式获取和分析, 可明显缩短取证分析时间, 提高数字取证的工作效率.

5 结语

云计算支持下的协作式数字取证系统, 具有磁盘数据恢复与云存储、Web 浏览器历史数据抽取与分析、HTTP 数据抓取与解析、E-Mail 内容获取与显示、多媒体通信数据采集与分析、网络打印文档还原等功能. 系统设计遵循了分层原则、模块化原则和跨平台原则, 它具有以下三个方面的特点: 第一, 面对来自不同地理位置的分布式网络攻击, 协作式数字取证系统能够快速鉴别和定位攻击源, 通过在线协作实现电子证据的及时获取; 第二, 电子证据的安全云存储可防止电子证据在存储或传输过程中被篡改或泄漏; 第三, 基于 B/S 架构的系统实现方法提高了数字取证系统的可用性和维护的便捷性, 便于数字取证人员通过多种终端设备进行在线协作与取证分析.

基于该系统的数字取证流程与分析过程符合计算机犯罪案件调查取证的国际规范与要求, 能够为计算机犯罪案件的侦查与法律证据的形成提供技术支持. 同时, 该系统的应用有助于实现电子证据的云安全存储与备份, 促进数字取证流程的规范化和工作效率的提升, 解决电子证据分布式获取与协同取证分析问题.

参考文献

1 Wang W, Daniels TE. A graph based approach toward network forensics analysis. ACM Trans. on Information and System Security, 2008, 12(1): 1–33.

2 刘在强,林东岱,冯登国.一种用于网络取证分析的模糊决策树推理方法.软件学报,2007,18(10):2635–2644.

3 史伟奇,张波云,谢冬青.基于远程控制技术的动态取证软件系统.计算机工程,2007,33(16):117–119.

4 戴江山,肖军模,陈波,郑君杰,刘晶.基于代理的主动型网络取证系统.解放军理工大学学报(自然科学版),2006,7(l):27–31.

5 Arasteh AR, Debbabi M, Sakha A, Saleh M. Analyzing multiple logs for forensic evidence. Digital Investigation Journal, 2007, 4(1): 82–91.

6 罗军舟,金嘉晖,宋爱波,东方.云计算:体系架构与关键技术.通信学报,2011,32(7):3–21.

7 吴胜艳,许力,林昌露.基于门限属性加密的安全分布式云存储模型.计算机应用,2013,33(7):1880–1884,1902.

8 郭东,杜勇,胡亮.基于 HDFS 的云数据备份系统.吉林大学学报:理学版,2012,50(1):101–105.

9 李丽英,唐卓,李仁发.基于 LATE 的 Hadoop 数据局部性改进调度算法.计算机科学,2011,38(11):67–70.