



基于 Hadoop 的高效分布式取证:原理与方法^{*}

吴松洋¹, 张熙哲¹, 王旭鹏¹, 李祥学²

(1. 公安部第三研究所 上海 201204; 2. 华东师范大学 上海 200241)

摘要:随着信息技术的发展以及各种智能设备的普及,设备的平台多样化使得现有电子数据勘查取证分析装备已不能满足网络和存储技术所需要的高速数据镜像存储和海量数据相关性分析等要求,并表现出操作复杂、效率低等缺陷。设计并实现了一种高效的基于 Hadoop 的分布式取证系统,它能够支持多介质并行取证的工作场景,并通过调度控制服务将不同的证据介质中的数据存储到不同的分布式数据存储服务器上,每个取证任务运行时都可以独占一个取证介质,从而实现多介质的并行取证分析。实验数据显示,搜索一个 2~4 GB 的文本数据的响应时间可以达到仅 0.1 s。

关键词: Hadoop; 分布式系统; 取证; 海量数据; 多介质

doi: 10.3969/j.issn.1000-0801.2014.01.005

An Efficient Distributed Forensic System Based on Hadoop: Principle and Method

Wu Songyang¹, Zhang Xizhe¹, Wang Xupeng¹, Li Xiangxue²

(1. The Third Research Institute of Ministry of Public Security, Shanghai 201204, China;

2. East China Normal University, Shanghai 200241, China)

Abstract: With the development and popularization of information technology and intelligence device, the diversity of different device making forensic analysis of existing equipment cannot meet today's networking and storage technology requirements, and exhibit complex operation, low efficiency, on high speed disk image storage and massive data correlation. An efficient distributed forensics system based on Hadoop technique, which can support multiple concurrent media scene forensics work, was designed and implemented, and through the dispatch control services would be evidence of different data storage media to a different distributed data storage server, each forensic task runtime could monopolize a forensic medium to achieve a parallel multiple media forensic analysis. Data show that responsible acknowledge duration will be 0.1 s for a 2~4 GB text file.

Key words: Hadoop, distributed system, forensic, massive data, multiple media

1 引言

随着社会信息化的快速推进,越来越多的数据以电子形式保存。统计数据显示,目前全世界所有信息数据采用电子存储媒介的占 92.22%, 并且未来采用电子方式作为

信息资源存储媒介的比例仍呈上升趋势。因此,存储于计算机及其他信息设备中的电子数据逐步成为针对和利用计算机网络犯罪案件的重要证据和诉讼依据。随着社会信息化进程的不断加快,对电子数据勘查取证分析的需求也越来越多,越来越复杂;信息技术和存储技术不断更新,工

^{*} 国家“十二五”科技支撑计划基金资助项目(No. 2012BAK07B01)



作量也越来越大。赌博案件和淫秽色情案件涉及的硬盘存储介质多,取证分析任务重,工作效率低。因此,电子数据勘查取证必须有相应的技术、工具和系统予以支撑,以解决上述问题。

2 分布式取证框架

分布式取证框架利用分布式系统的特点,采用并行执行机制,能够对大量数据进行分布式处理,涉及多介质证据接入管理、电子数据取证分析管理、分布式存储和运算以及分布式规模化电子数据存储与分析,支持多介质并行取证的工作场景,通过调度控制服务将不同证据介质中的数据存储到不同的分布式数据存储服务器上,每个取证任务运行时都可以独占一个取证介质,从而实现多介质的并行取证分析。

2.1 多介质证据接入管理

分布式取证是一种具备多介质证据^[1]接入的控制软件系统,能够对网内所有介质只读接口进行统一控制,支持多案件、多用户并行操作,支持将源存储介质的镜像^[2]存储于各种网络存储系统中,支持容错处理、边复制边分析等功能。

2.2 电子数据取证分析管理服务

电子数据取证分析^[3]管理服务子系统的功能包含:镜像文件导入功能,支持将镜像数据导入分布式存储或光纤存储中;为多用户提供访问和处理人机接口的功能;对案件进行综合管理、分析服务的功能;对证据数据的有效提取、导入、分析、报告导出、展示;对取证分析的整个过程进

行详细的跟踪记录,实现技术、工具、方法、过程的综合管理;支持多用户并行分析操作,包括多人同时对一个存储介质进行分析,或一人同时对多个存储介质或案件进行分析,如图1所示。

2.3 分布式存储和运算

存储原始的案件证据数据,为上层和分布式计算任务提供原始数据访问。分布式运算将承担系统中最耗时的应用分析逻辑,能够对电子数据进行并行化分析,大大提高工作效率,有效提高电子数据的应用和处理水平。

2.4 分布式规模化电子数据存储与分析

此系统能够支持高性能分布式电子数据取证分析系统核心网络,具备万兆网络的超高速传输、分布式分析处理功能,采用海量数据的分布存储技术,将案件数据分布式存储在各个存储节点,从而可以对案件数据进行分布计算和分析。案件数据在各个节点之间的存储考虑冗余和备份,以提高系统的顽健性和可靠性。具备失败转移功能,单个节点失败不影响取证的分析及其系统运行;当存储容量不足时,能够在不影响系统正常运行的情况下,方便地插入新的存储节点,实现方便扩展。分布式分析功能包括:散列、搜索、特征恢复、即时通信、邮件、上网日志、基本信息等分析功能。

3 基于 Hadoop 的取证系统设计理念

Hadoop^[4]是一个能够对大量数据进行分布式处理的软件框架^[5],假设计算元素和存储会失败,因此它维护多个工作数据副本,确保能够针对失败的节点重新进行分布处

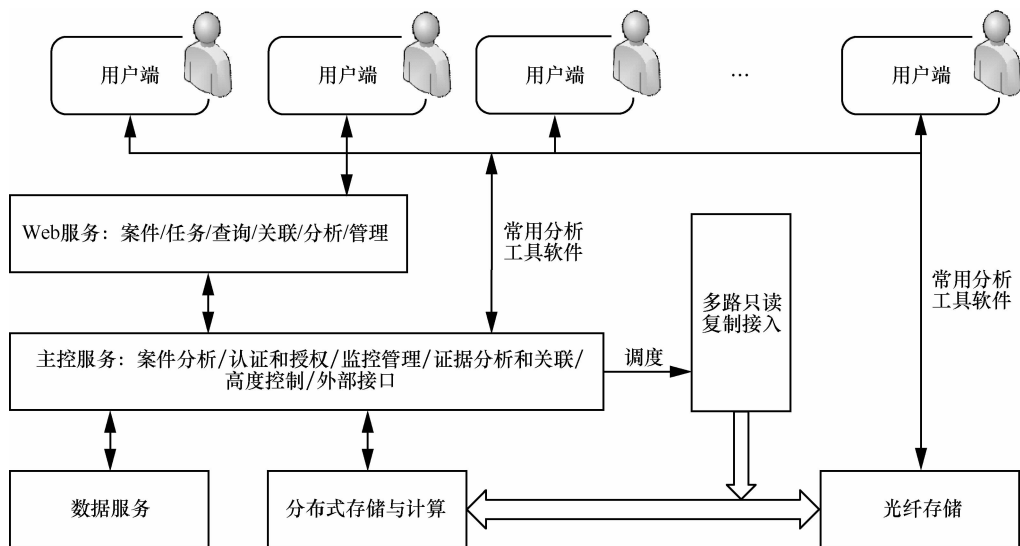


图1 分布式电子数据取证分析系统的业务流

理。它以并行的方式工作,通过并行处理加快处理速度,因此能够处理 PB 级数据。此外,Hadoop 可以依赖于低端服务器,因此成本比较低,并且具有高可靠性、易于扩展、高效性以及高容错性等优点。

3.1 基于 Hadoop 的分布式取证系统

基于 Hadoop 的分布式电子证据分析系统^[6,7]是一个统一的取证分析工作平台。案件数据和文件分布式存储在多台设备中,可自定义数据存储的数量,以支持案件数据的冗余和可靠性,排除存储系统的单点故障,为上层提供文件和镜像访问接口,可通过简单增加设备的方式对系统容量进行扩展。

本系统从整体上可以分为多介质证据接入平台、电子数据取证分析管理服务子系统、高性能核心网络、分布式数据存取与分析系统以及扩展接口等部分。分布式电子证据分析系统的整体架构如图 2 所示。

3.2 基于 Hadoop 的分布式管理系统

服务管理系统是将分布式电子取证系统所提供的各项服务进行抽象,形成统一的对外服务接口;然后再将这

些系统服务进行统一的管理,以便于系统不同组成部分能够通过服务管理模块进行相互调用。服务管理模块主要包括服务注册、服务查询、服务调度等子功能。服务管理模块的工作示意和所提供的 3 项主要功能如图 3 所示。

3.3 基于 Hadoop 的取证系统的特点

分布式存储结构借鉴 Hadoop 的 HDFS(分布式文件系统),具备透明性、并发控制、可伸缩性、容错性以及安全需求,具体介绍如下。

(1) 透明性

每台存储服务器作为一个文件节点,整体存储服务器组成文件节点群,文件以大数据块的方式存放在文件节点中。所有文件节点由元数据节点统一管理。用户统一通过元数据节点获取文件,而不必考虑文件具体存放在哪台服务器上。

(2) 并发控制

支持负载均衡,当某个文件节点存储的数据超过某个阈值时,元数据节点将新增的数据添加到其他文件节点,防止某个节点数据过大,造成吞吐量下降。数据访问采取 write-one-read-many 模式,一次性写入、多次读取,采用在文

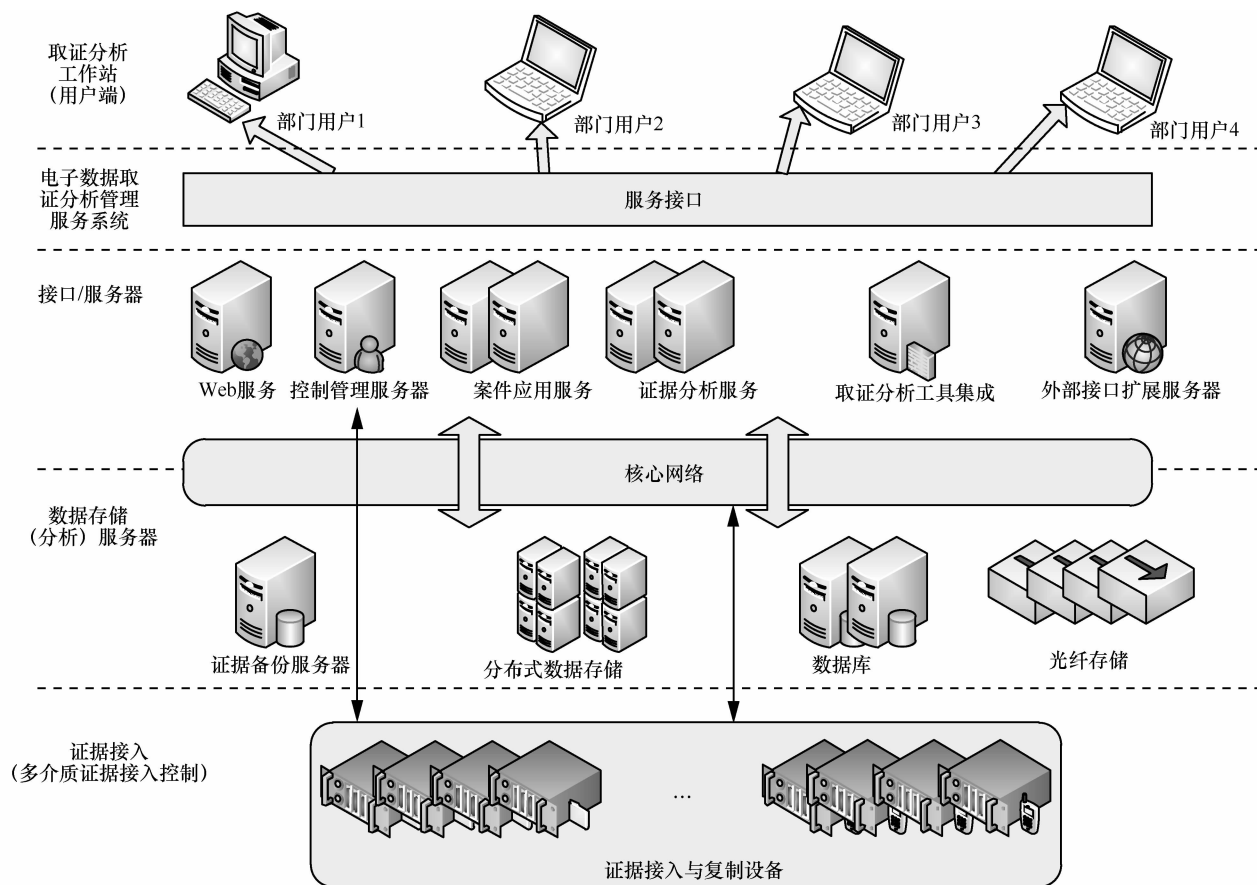


图 2 分布式电子证据分析系统的总体架构

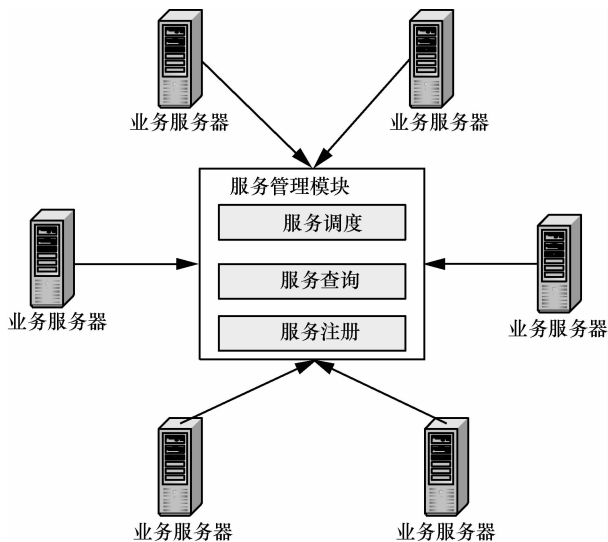


图 3 服务管理模块的工作示意和所提供的主要功能

件末尾添加文件和文件合并的方式修改数据,以降低磁盘随机输入/输出。

(3)可伸缩性

支持存储设备的横向扩展,当增加服务器时,只需在元数据节点中添加新的文件节点,对文件服务器的配置文件进行同步即可。

(4)数据可靠性

数据在复制到文件节点时支持冗余复制,文件节点和元数据节点之间通过心跳分组检测。当文件节点出现问题时,元数据节点支持将该文件节点管理的数据分拆到其他节点,不影响正常访问,元数据节点采用双备份机制,防止元数据丢失。

(5)数据安全性

取证数据是刑侦和调查工作的基础,防篡改和防泄密是最核心的安全需求。基于文件过滤驱动技术,实现了文件防篡改模块和文件透明加密。在取证数据文件产生时,防篡改模块使用防篡改密钥生成指纹文件,并且将数据和指纹同时压缩备份,当受保护的数据文件被非法篡改后,防篡改模块能通过对指纹文件发现非法篡改行为并实时恢复正确的数据文件。透明加密模块保证了只有在工作区域内数据文件才能被正常使用,即使数据被窃取,得到的也只是加密后的文件。

4 基于 Hadoop 的取证系统架构

Hadoop 取证系统包含了多介质证据接入平台、电子数据取证分析管理服务子系统、高性能核心网络、分布式

数据存取与分析系统以及扩展接口等部分。系统整体上采用 B/S 架构,提供取证 Web 服务,通过镜像散列、即时通信等分析功能对数据进行分析,对分析得到的结果进行归档和存储,并且支持文本搜索和高级查找。

4.1 取证 Web 服务

系统整体采用 B/S 架构^[89],后台主要为用户提供分布式取证系统功能操作和管理功能,包括案件分析业务逻辑处理、案件信息的存档管理、对调查员提供案件任务的调度数据管理等功能。后端采用 RESTEasy+EJB3+DAO,包括请求响应、业务逻辑处理和数据库访问标准 3 层架构,并结合实际业务需求选择不同数据库的开发架构。遵循 CRUD 原则,前端通过 RESTEasy JavaScript-API 请求和操作资源,服务器端返回统一的 JSON 格式的数据。Web 服务整体架构如图 4 所示。

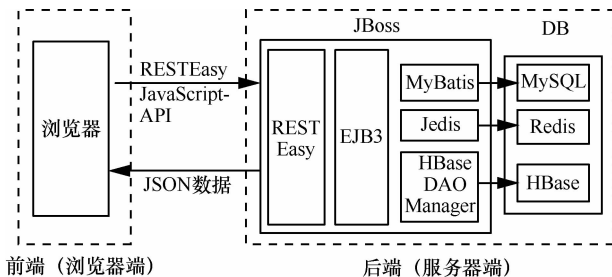


图 4 Web 服务整体架构

4.2 证据分析和存储

证据分析和存储模块提供对证据数据的分析和分析结果的存储功能,包括镜像散列、即时通信、邮件、上网日志、基本信息等分析功能,对分析结果进行归档和存储,支持文本搜索和高级查找。

(1)证据分析功能

证据分析功能主要包括对证据数据进行镜像,同时生成镜像散列值,支持 MD5、SHA-1、SHA-2 等多种散列值,保证镜像数据不被修改,支持第三方取证工具接入,针对即时通信、邮件、上网日志、基本信息等数据进行分析^[90];取证分析工具集成模块对外提供服务接口,方便第三方取证工具接入。整体架构采用 SOA(service-oriented architecture,面向服务的体系结构),根据需求通过网络对松散耦合的粗粒度应用组件进行分布式部署、组合和使用。服务接口实现采用标准 Web Services,基础组件包括 SOAP、WSDL 和 UDDI。

WSDL 用来描述服务;UDDI 用来注册和查找服务;SOAP 作为传输层,用来在消费者和服务提供者之间传送消息。XML(extensible markup language,可扩展的标记语

言) 作为 Web Services 平台中表示数据的基本格式, XSD 作为数据类型系统。第三方取证工具接入取证分析工具集成服务, 通过 UDDI 注册(registry)表查找服务, 取得服务的 WSDL 描述, 然后通过 SOAP 调用接口。取证分析工具集成模块整体架构如图 5 所示。

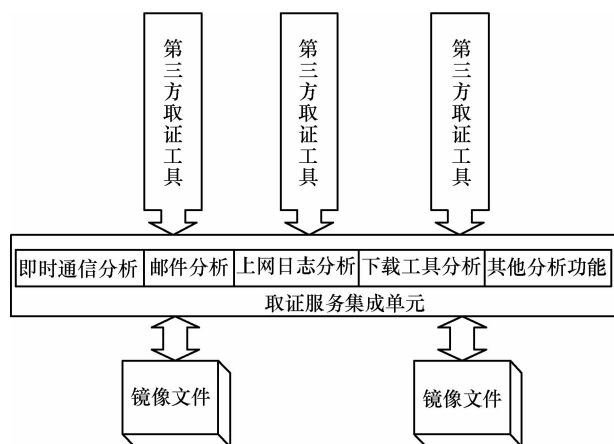


图 5 取证分析工具集成模块整体架构

(2) 证据存储功能

证据分析结果采用关系数据库 MySQL 存储, 主从数据库分布式架构, 支持读写分离。主数据库主要负责数据写入, 从数据库复制数据读取, 通过二进制日志文件进行数据同步。InnoDB 作为存储引擎, 具有事务(transaction)、回滚(rollback)和崩溃修复能力, 是多版本并发控制(multi-versioned concurrency control)的事务安全(transaction-safe)型表。在数据访问时提供了行级锁(locking on row level), 以减少数据冲突, 加快写入速度。

针对各种条件查询采用聚集索引和非聚集索引相结合的方式, 索引和查询过的数据会保留在内存中, 加快查询速度。MySQL 对于传统的文本查找采用模糊匹配实现, 无法利用索引, 需要遍历全部记录, 查询速度非常慢。针对文本查找添加 Sphinx 搜索引擎, 可以快速建立索引, 支持实时搜索。证据分析结果存储架构如图 6 所示。

4.3 案件应用服务

案件应用服务模块主要连接 Web 前端界面和分布式存储系统, 通过案件数据服务为多个用户提供统一的案件访问、管理和同步分析服务。案件应用服务模块可以将服务请求方和后台服务处理方解耦, 提供透明的数据访问和分析; 添加 RabbitMQ 服务中间件, 支持多用户分布式访问, 支持分析服务的负载均衡, 支持服务访问请求优先级调度, 支持服务访问请求数据分组缓存。

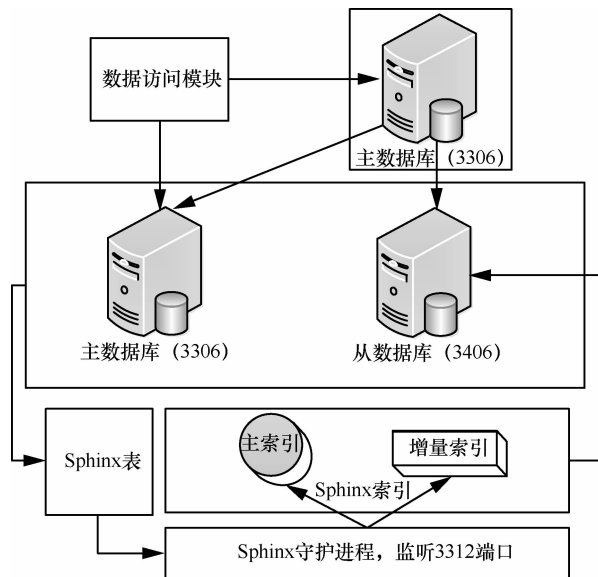


图 6 证据分析结果存储架构

RabbitMQ 是基于高级消息队列的 AMQP 的技术实现。用户服务请求先发送到队列, 服务处理模块从队列中取出请求处理。整体结构包括发送者、接收者、转换器和队列。请求发送方和接收方采用生产者和消费者模式; 转换器提供路由功能, 可以接入指定队列; 队列可以提供数据缓存, 并能够对两端服务进行解耦。案件服务中间件整体结构如图 7 所示。

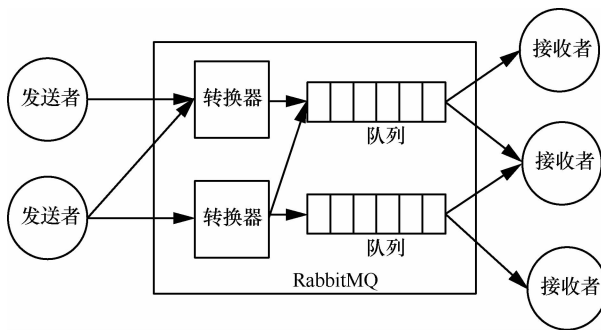


图 7 案件服务中间件整体结构

用户业务请求可以根据配置发送到指定队列, 不同队列配置不同的优先级, 高优先级队列的请求在数据处理模块空闲时被优先处理, 服务请求支持添加、删除和队列切换。数据处理模块支持横向扩展, 每个处理模块根据配置挂载指定队列的监听器, 监听器可以从队列中获取数据, 请求发送和处理采用异步的方式。

4.4 数据防篡改模块

采用文件底层过滤驱动防篡改方法对取证数据文件进行实时监控。生成数据文件时, 采用密钥生成指纹

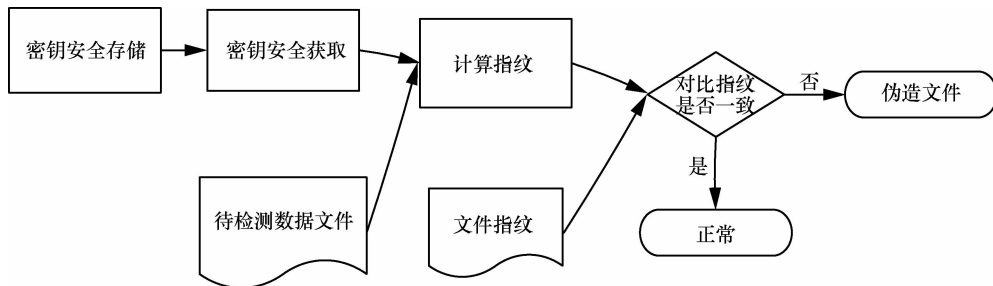


图 8 取证数据防篡改实时检测

和备份;之后在内核驱动层实时监控对数据文件的更新操作,当发现文件被篡改(新的文件无法匹配对应的指纹)时,立即从备份文件中恢复正确的文件,如图 8 所示。

4.5 数据透明加密模块

透明加密^[11]无需用户改变使用习惯,有效地保障了取证数据安全。安装该模块以后,就形成了一个可信区域,该区域内的任何文件操作、交流都不受影响。一旦离开该区域,用户只能看到加密后的数据文件,从而达到防复制、防泄密的目的。透明加密模块包括 3 个部分:应用层服务、内核层加/解密和密钥安全存储,如图 9 所示。应用层服务是与内核驱动交互的接口,负责驱动的启动、停止,保护策略的设定,密钥的读写(需要移动安全设备,如 USB 智能卡等);内核层加/解密驱动负责文件的透明加/解密操作;密钥安全存储包括移动安全设备,负责密钥存储、密钥管理和读写的接口。

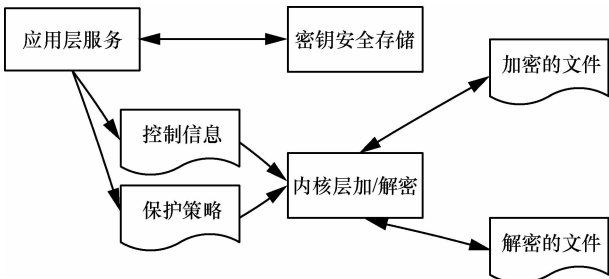


图 9 实时透明加密模块

密钥是透明加密模块的关键,必然也是攻击者企图窃取的重点对象。系统采用公私钥加密体制实现用户身份的强认证和对称密钥的保护。在为每个用户初始化对称密钥存储介质时,就采用该用户的公钥进行加密,这样,只有掌握用户私钥才能解密对称密钥。相应地,透明加密模块也需要集成其他系统内核层的一系列自我保护机制,如程序加壳、防静态分析、防动态调试、防止非法钩子、防止进程内存代码篡改等。

5 基于 Hadoop 的取证系统性能

Hadoop 取证系统将数据存储于分布式系统中,能够提供流数据模式访问,处理超大文件,可以运行于大量分布式服务器集群中,主要包括元数据节点和数据节点,用来管理文件系统的命名空间以及存储数据。

5.1 分布式数据存储

案件信息主要存储在分布式系统中,分布式数据存储模块底层数据存储基于 HDFS^[12],数据存储于分布式文件系统中。分布式数据存储集群采用主从结构,其中包括两类服务器节点,分别为元数据节点和数据节点。

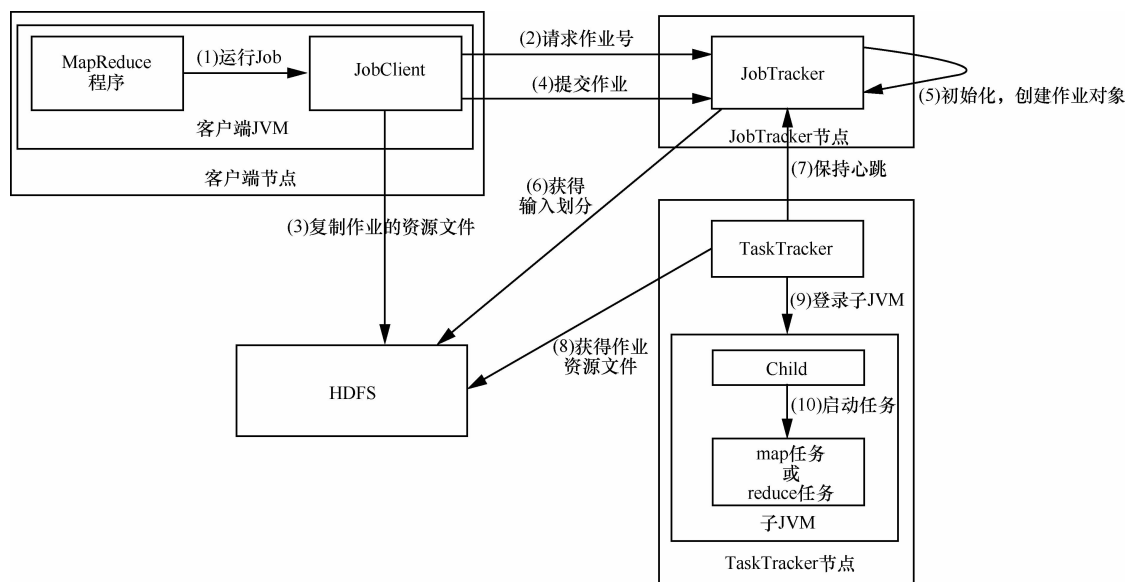
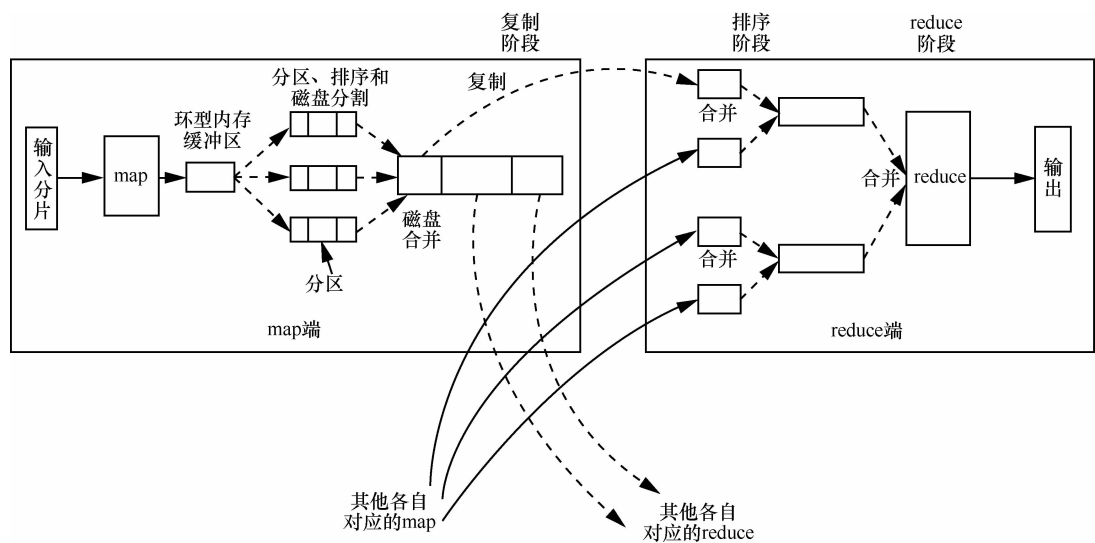
元数据节点用来管理文件系统的命名空间,其将所有文件和文件夹的元数据保存在一个文件系统树中。这些信息也会在硬盘上被保存成以下文件:命名空间镜像(namespace image)及修改日志(edit log),同时保存了一个文件包含的数据块以及分布的数据节点。然而这些信息并不存储在硬盘上,而是在系统启动时从数据节点收集而成,数据节点是文件系统中真正存储数据的地方。客户端通过元数据节点找到文件保存的数据节点,向数据节点请求写入或者读出数据块。

5.2 分布式数据

分布式数据分析建立在分布式数据存储系统之上,基于 MapReduce 编程模型^[13],案件分析任务可以并行分割成多个 map 任务,多个任务并行执行,由统一的任务调度器进行管理。map 任务结束后,map 任务的输出作为 reduce 任务的输入,reduce 任务会对输入结果进行汇总处理,包括数据合并、过滤、排序等,产生最终结果。MapReduce 任务处理流程如图 10 所示^[14]。

用户可以通过 Web 管理界面查看每个任务的运行进度,查看各个数据节点的负载信息,同时可以通过日志文件查看任务失败原因。分布式数据分析流程如图 11 所示。

分布式数据分析模块从分布式数据存储模块中获取数据,数据分析结果经过汇总过滤后存储到关系型数据库



中,并对相关数据建立索引,用户可以通过 Web 服务直接查看分析结果。搭建了 Hadoop-V 虚拟机系统和 Hadoop-A 实体机 2 个实验系统,并与 non-Hadoop 结构的实体机系统进行对比,结果见表 1。

进一步测试了 Hadoop 集群与 non-Hadoop 集群在文

本搜索性能方面的差异。每个集群中各自包含 10 台服务器。集群网络环境,核心网络 10 Gbit/s;单台服务器硬件配置,CPU 为 Intel Xeon E7-8837,8 核;内存为金士顿 16 GB×3;硬盘为希捷 2 TB×2,10 000 转/s。选取不同大小的磁盘镜像,进行同样取证分析功能,功能主要是进行文本搜索,通过镜像匹配大小长度为 10 B 的字符串,实验结果见表 2。

实验表明,在同样的网络和存储条件下,基于 Hadoop 的分布式取证系统能够高效地存储 200~300 TB 的海量数据,搜索一个 2~4 GB 的文本数据的响应时间可以达到仅 0.1 s,并且单个 CPU 最大支持文本处理的水平达到

表 1 Hadoop 系统性能对比指标				
	核心网 络/Gbit/s	数据存 储/TB	搜索(2~4 GB 文本数据)	可用性(单个 CPU 最大支持 文本)/GB
non-Hadoop	10	200	>0.6 s	10
Hadoop-V	8	200	~0.3 s	50
Hadoop-A	10	300	~0.1 s	100



表2 Hadoop 系统文本搜索对比数据

取证分析	Hadoop 集群耗时/s	non-Hadoop 集群耗时/s
50 GB 镜像(完整)	296	260
100 GB 镜像(完整)	515	495
200 GB 镜像(完整)	980	972
500 GB 镜像(完整)	2 300	2 277
500 GB 镜像(分割成 10 份)	362	2 287
1 TB 镜像(完整)	4 534	4 494
1 TB 镜像(分割成 10 份)	585	4 615
2 TB 镜像(完整)	9 729	9 589
2 TB 镜像(分割成 10 份)	1 070	9 709

了 100 GB。同时,随着数据量的增多(大于 1 TB),Hadoop 集群更能发挥分布式系统协同并行工作的优势,迅速降低数据搜索的时间开销。

6 结束语

根据分布式系统原理和 Hadoop 原理,设计实现了一个统一取证分析工作平台的分布式取证软件框架,包括分布式数据存储和分布式电子数据取证分析,支持海量数据的分布式存储和并行分析计算,包括散列、搜索、特征恢复、即时通信、邮件、上网日志、基本信息等分析功能,并且涉及高性能核心网络^[5]的需求,包括硬件和软件系统的传输、处理性能不低于 10 Gbit/s,能够全面提升海量数据的处理能力,提高取证分析工作效率。数据存储系统采用分布式存储和光纤存储配合的形式,支持分布式存储不低于 200 TB,传输速度不低于 10 Gbit/s,并可以根据需要进行无缝扩展。案件数据和文件分布式存储在多台设备中,可自定义数据存储的数量,以支持案件数据的冗余和可靠性,排除存储系统的单点故障,为上层提供文件和镜像访问接口,可通过简单增加设备的方式对系统容量进行扩展。

参考文献

- Wang X Q. How to discover the truth in data. *China Information Security*, 2009(11):23~24
- Intel Corporation. Understanding the Flash Translation Layer (FTL) Specification, December 1998
- King C, Vidas T. Empirical analysis of solid state disk data retention when used with contemporary operating systems. *Digital Investigation*, 2011(8):111~117
- Wikipedia Apache Hadoop. <http://en.wikipedia.org/wiki/Hadoop>, 2013
- Analyzing big data with Hadoop, decorated a new history. <http://www.bloter.net/archives/68650>, 2011
- Zhou W C, Tao T, Boon T L, *et al.* Declarative secure distributed information systems. *Computer Languages, Systems and Structures*, 2013,39(1):1~24
- Jens D, Jorge-Arnulfo. Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, 2012,5(12):2014~2015
- Prashant S, Kamalakara K. A multi-agent simulation framework on small Hadoop cluster. *Engineering Applications of Artificial Intelligence*, 2011,24(7):1120~1127
- Wang S G, Su W, Zhu X L, *et al.* A Hadoop-based approach for efficient web service management. *International Journal of Web and Grid Services*, 2013,9(1):18~34
- Wu T Y, Chen C Y, Kuo L S, *et al.* Cloud-based image processing system with priority-based data distribution mechanism, 2012,35(15):1809~1818
- Zhao M, Mao R, Jiang R. Transparent encryption file system model based on filter driver. *Computer Engineering*, 2009(1): 51
- Kathleen E, Shrideep P. On the performance of high dimensional data clustering and classification algorithms. *Future Generation Computer Systems*, 2013,29(4):1024~1034
- Satish N S, Pelle J, Eero V. Adapting scientific computing problems to clouds using MapReduce. *Future Generation Computer Systems*, 2012,28(10):184~192
- Jiang D W, Deng C O, Shi L, *et al.* The performance of MapReduce: an in-depth study. *Proceedings of the VLDB Endowment*, 2010,3(1/2):472~483
- Erik R. The Odd couple: hardware and software. *IEEE Micro*, 2012,32(4):2~2
- Neal L. Bringing big analytics to the masses. *Computer*, 2013,46(1):20~23

[作者简介]



吴松洋,男,博士,公安部第三研究所信息网络安全技术研发中心副主任、助理研究员,主要研究方向为网络安全和云计算。

张熙哲,男,博士,公安部第三研究所信息网络安全技术研发中心助理研究员,主要研究方向为虚拟机安全体系。

王旭鹏,男,硕士,公安部第三研究所信息网络安全技术研发中心实习研究员,主要研究方向为海量数据处理。

李祥学,男,博士,华东师范大学副教授,主要研究方向为密码学。

(收稿日期:2013-07-30)