

基于云模式的媒体取证服务体系研究

惠雯^{1,2} 林闯² 赵海英³ 杨扬¹

(北京科技大学计算机与通信工程学院 北京 100083)¹ (清华大学计算机科学与技术系 北京 100084)²
(新疆师范大学计算机科学技术学院 乌鲁木齐 830054)³

摘要 根据现有媒体取证系统所面临的问题,结合云计算的思想,提出了基于云模式的媒体取证服务体系。针对大规模媒体取证服务所面临的异构性、可扩展性和效率问题等几大挑战,利用层次结构模型方法,创造性地提出了三层媒体取证服务体系架构模型,详细阐述了该模型所涉及的关键技术,即资源虚拟化、系统任务并行化和自适应能力,最后给出了应用实例及分析。

关键词 云计算,媒体取证,资源虚拟化

中图法分类号 TP391 **文献标识码** A

Cloud-based Architecture for Media Forensics

HUI Wen^{1,2} LIN Chuang² ZHAO Hai-ying³ YANG Yang¹

(School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)²

(College of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China)³

Abstract Current researches about media forensics face the following key challenges: heterogeneity, scalability, and efficiency. We proposed a new computing paradigm for media forensics, called "Media Forensics Cloud", which integrates the concept of cloud computing to handle large-scale media forensics service effectively. To address the above challenges, first, we presented a layered architecture of Media Forensics Cloud, which can provide efficient and scalable media forensics service. Then we presented the key technologies of Media Forensics Cloud, namely resource virtualization, forensic task parallelization, and service adaptation, by which users can efficiently access media forensics service from different terminals at anytime anywhere. Lastly, we presented a key study of Media Forensics Cloud to demonstrate the feasibility of our design.

Keywords Cloud computing, Media forensics, Resource virtualization

1 引言

随着数字媒体内容的广泛传播,其合法性逐渐成为一个迫切需要解决的问题。借助专业的媒体取证系统^[1-4]能够可靠而高效地识别盗版或违禁内容,提供有说服力的证据。然而,随着用户访问量的急剧增加、访问方式的不断变化,以及多媒体数据量的迅速扩充,现有媒体取证系统已无法满足日益复杂的取证需求。在设计大规模媒体取证系统时,会面临以下几方面挑战:

(1)异构性问题。异构性主要指网络异构性与设备异构性。3G和4G无线通信技术的快速发展,使得通过移动网络获取媒体取证服务的需求日益凸显。

(2)可扩展性问题。可扩展性是指,当用户达到一定数量后,系统的服务质量仍能满足用户需求,也就是系统能够有效

调度资源和协调用户请求,还能有效控制系统管理和维护的开销。

(3)效率问题。大规模媒体取证服务往往造成大量计算开销,这就要求系统能够实时分析和检测媒体内容。现有系统大多通过较高的硬件配置或高效的数据结构来提高系统整体效率^[5-8],但受服务器处理能力的限制,随着数据量的不断增加,服务实时性仍很难保证。

为解决上述问题,本文提出了一种基于云模式的媒体取证服务体系——媒体取证云,如图1所示。云计算作为一种全新的计算模式,为大规模分布式应用提供了很好的框架,它通过将各种互联的计算资源进行有效整合并实现多层次的虚拟化与抽象,可有效将大规模计算资源以可靠服务的形式提供给用户,从而将用户从复杂的底层操作中解放出来^[9,10]。借助云模式可以为实现大规模媒体取证服务提供一个很好的

到稿日期:2011-01-20 返修日期:2011-03-21 本文受国家自然科学基金(60873254,61003226),973国家重点基础研究发展计划(2011CB302600),教育部新世纪优秀人才支持计划项目资助。

惠雯(1982—),女,博士生,主要研究方向为多媒体通信,E-mail:hw97@126.com;林闯(1948—),男,教授,博士生导师,主要研究方向为计算机网络与系统性能评价;赵海英(1969—),女,副教授,主要研究方向为模式识别与图像处理;杨扬(1955—),男,教授,博士生导师,主要研究方向为多媒体通信、网络计算、无线通信、图像处理与模式识别等。

研究思路,使用户能够在任何地点、以任何方式部署服务,或使用服务检测媒体内容的合法性。图1中,媒体取证云中存有预先定义的敏感媒体特征,通过对待检测的媒体数据进行分析,将其特征与敏感媒体特征进行比对,即可判断其合法性。

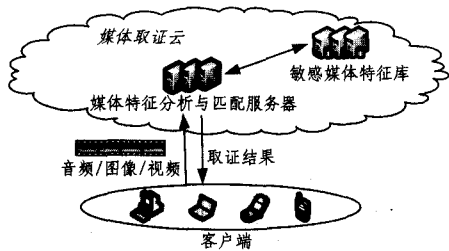


图1 媒体取证云

2 基于云模式的媒体取证服务体系架构

媒体取证服务是以用户为中心的,对用户来讲,需要在不同区域、不同网络 and 不同设备的情况下,可以连续、高效地获取服务。为了能够向大规模用户提供媒体取证服务,并支持服务的定制和扩展,本文利用层次结构模型方法,将基于云模式的媒体取证服务体系架构划分为3个层次:系统服务层、系统管理层和资源管理层,如图2所示。

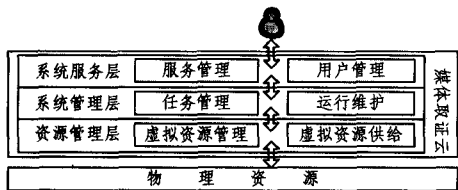


图2 媒体取证云体系架构

系统服务层的主要功能是媒体取证服务的定制和用户管理,它向上提供统一接口,供用户灵活定制自己的服务。在这一层中,每个服务请求都转化为一个服务任务,服务管理模块负责服务任务的管理与创建,用户管理模块则负责客户端环境的配置,以及按需计费辅助功能。为满足不同服务请求的需要,每个服务请求最终都是以独立的服务任务形式运行,系统管理层主要负责管理服务任务的执行和运行维护。在这一层中,每个服务任务都被分割为若干子任务,任务管理模块负责任务调度和资源分配,以及将任务部署到下层虚拟节点上运行,运行维护模块的功能是实时监控任务运行情况和虚拟资源的使用情况,同时对故障进行检测,并在故障导致任务执行中断时采取措施恢复任务的运行。最下层是资源管理层,主要功能是对底层物理资源进行虚拟化,屏蔽资源异构性对上层任务的影响。其中,虚拟资源管理模块用于为上层任务构建虚拟执行环境,并将虚拟资源最终映射到底层物理资源中;虚拟资源供给模块用于管理虚拟资源调度,为虚拟节点供给物理资源,并通过虚拟机的整合和隔离等机制,提高灵活性,降低成本。

3 基于云模式的媒体取证服务体系关键技术

3.1 资源虚拟化

资源虚拟化主要涉及两个问题,即虚拟资源管理和虚拟资源供给。

虚拟资源管理主要完成虚拟资源的映射过程,根据上层

任务的特征和对计算、存储等资源的需求,快速、有效地将虚拟资源映射到底层物理资源中。虚拟资源的映射原则是:选择具有足够CPU、磁盘、内存及其他硬件能力的物理节点,同时最小化整体资源使用代价。映射方法有两类:一是利用经济型目标函数选取资源,引入节点压力指标来量化物理资源使用,选择节点压力最小化作为目标函数^[11];二是将收益最大化作为目标函数^[12]。

虚拟资源供给主要完成虚拟资源调度,为虚拟节点供给物理资源。相关技术分为两类:静态供给技术和动态供给技术。静态供给技术在生命周期内不允许改变资源分配^[13],动态供给技术则允许基于动态变化的用户需求自适应地改变资源分配^[14]。

本文由资源管理层来实现资源虚拟化,所采用的方案如图3所示。虚拟资源管理模块主要实现虚拟资源的管理,包括虚拟执行环境的动态构建和虚拟资源的动态映射等,它主要由信息收集器、管理策略引擎和迁移执行器组成。虚拟机节点上的监控代理采集每个虚拟节点的虚拟资源使用信息,传输给信息收集器。信息收集器将这些信息存入信息表。管理策略引擎将从信息表中获取的信息作为依据,制定虚拟资源的动态映射和虚拟机节点的迁移策略,并将这些策略存入管理策略表。之后,管理策略引擎会触发迁移执行器发送命令给相关虚拟机管理器执行,以完成虚拟资源的动态映射和虚拟机节点的迁移。紧急情况下(如有故障发生时),信息收集器也可直接向管理策略引擎发出陷入信号。

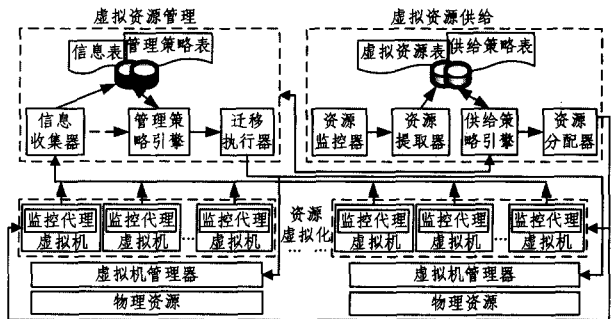


图3 虚拟资源管理和虚拟资源供给

虚拟资源供给模块主要实现虚拟资源的供给,它由资源监控器、资源提取器、供给策略引擎和资源分配器组成。资源监控器监控互联网上的物理资源信息及其实时运行状况,将这些信息发送给资源提取器。资源提取器将现有互联网上的可用物理资源进行抽象,记录到虚拟资源表中。供给策略引擎根据虚拟资源表中的信息,与虚拟资源管理模块交互,了解虚拟节点的资源需求情况,制定资源供给策略,并将这些策略存入供给策略表。最后,供给策略引擎触发资源分配器,为不同虚拟机动态分配资源。

3.2 系统任务并行化

系统任务并行化要解决的问题是,如何实现上层服务任务的大规模并行处理,以保障资源的按需分配。在并行处理模型中,一个任务通常被划分为若干子任务同时执行。传统方法是采用主从模型^[15],但该方法只有当任务被划分为相同的子任务时,才能真正实现负载均衡,这在网络状况比较复杂或需要处理的任務和数据量比较大的情况下,具有一定的局限性。MapReduce^[16]是Google提出的一种软件架构,用于处理海量数据的并行运算。MapReduce模式由若干个Map(映

射)操作和若干个 Reduce(化简)操作组成。一个 Map 操作是对原始数据进行划分,并分配给并行 Worker 处理。尽量保证不同 Map 操作之间的独立性,以实现任务处理的充分并行化。一个 Reduce 操作是对每个 Map 所产生的一部分中间结果进行合并的操作,每个 Reduce 所处理的 Map 中间结果是互不交叉的,所有 Reduce 所处理的 Map 中间结果经过简单连接就可形成完整的结果集。该模式是目前大规模任务分布式处理技术发展的主流。

本文由系统管理层来实现系统任务并行化,具体方案如图 4 所示。任务管理模块接受服务任务后,首先由任务引擎对服务任务进行分析,将其分割为可并行处理的子任务,尽量消除子任务之间的相关性。然后,任务引擎在虚拟节点中指定 Master 和 Worker 节点,与虚拟资源管理模块交互构建虚拟执行环境,并在 Worker 节点中指定执行 Map 和 Reduce 操作的节点,将操作分发到指定节点上。访问控制器在整个过程中对资源访问进行控制,通过分布式锁来控制冲突访问时的一致性等问题,通过访问控制列表的形式来保障安全性。任务引擎最终将节点任务分配信息保存于任务分配表中。同时,任务索引维护管理器利用任务分配信息,建立全局索引,以便 MapReduce 产生的中间结果和数据资源能够被快速定位和访问,任务索引信息保存在任务索引表中。

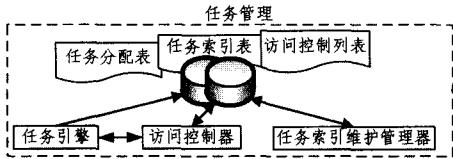


图 4 系统任务管理

3.3 自适应能力

除了通过资源虚拟化来屏蔽底层物理资源的异构性以外,还要解决不同接入网络和接入设备的异构性问题,以更高效地支持媒体取证服务^[17]。本文中,虚拟适配层用于解决这个问题,如图 5 所示。虚拟适配层分为内容适配层和网络适配层两部分。内容适配层根据客户端接入设备的特点,如处理能力、屏幕分辨率、待机时间等,对传输的内容进行处理。网络适配层通过获取当前可用的接入网络方式和不同接入方式的不同传输性能和传输质量测量信息,为不同的数据选择最优传输方式与网络承载方式。

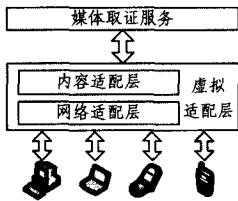


图 5 服务自适应

4 典型应用场景

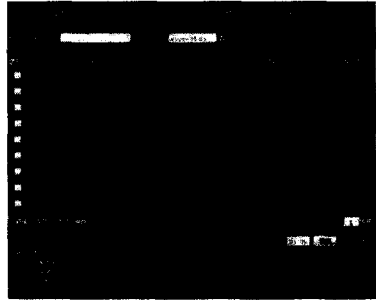
目前,作者研究开发的媒体取证云原型系统已在数字版权保护领域得到初步应用。下面以该系统为例,通过一个典型应用场景来分析媒体取证云的工作流程。

媒体取证服务提供商利用媒体取证云部署其服务,使用户可以对视频共享网站进行监控。当版权管理机构(用户)向媒体取证云提交需要监控的视频网站后,一旦有网民上传非

法盗版视频,媒体取证云便收集非法信息及相关证据,反馈给版权管理机构,并阻止该视频出现在网站中,如图 6 所示。



(a) 网站监控



(b) 审核结果

图 6 媒体取证云原型系统

以上场景的工作流程如图 7 所示。

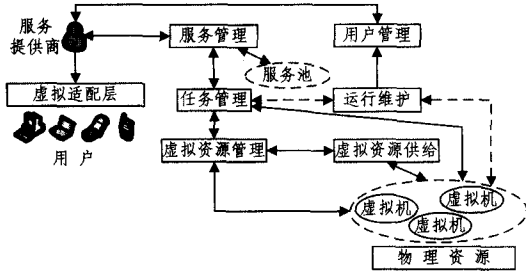


图 7 典型应用场景

第 1 步 当用户请求媒体取证服务时,首先需要经过系统服务层的用户管理模块的认证和授权。用户通过系统服务层提供的统一接口定制自己的服务,服务管理模块根据用户的需求,从服务池中查找相应的服务工具和组件,创建服务任务,然后交由系统管理层的任务管理模块分配执行。同时,用户管理模块进一步负责配置客户端环境,并提供安全性保证。

第 2 步 任务管理模块对服务任务进行分割和调度,并与下层虚拟资源管理模块交互,分配下层资源以支持任务的完成。在这一过程中,运行维护模块会实时监控任务的运行情况和虚拟资源的使用情况,对异常进行处理,保障系统的可靠性和可用性,同时还可配合用户管理模块对用户进行按需计费。

第 3 步 资源管理层根据资源情况和任务需求对下层物理资源进行虚拟化和抽象,形成虚拟执行环境,保障上层任务的执行。

第 4 步 所有任务完成后,任务管理模块会通过服务管理模块将最终结果反馈给用户。

结束语 本文针对现有媒体取证系统所面临的问题,结合云计算的思想,提出了基于云模式的媒体取证服务体系。该体系模型重点可解决大规模媒体取证服务中所面临的三大挑战,即异构性、可扩展性和效率问题。进一步的研究工作主要从以下几方面着手:

(1) 虚拟资源映射方案的复杂度高、开销大,缺乏具体资源选择方法,虚拟资源的大规模映射的思路、方法和机理都有

(下转第 42 页)

测网络中有数据包篡改、丢包、谎报等行为的恶意节点来评估节点的可信度。实验结果证明,该 TCTR 算法能减小能耗、平衡负载,从而延长网络生存时间,同时能快速有效地识别与隔离恶意节点,从而提高路径和网络的安全度。

参 考 文 献

- [1] 荆琦,唐礼勇,陈钟. 无线传感器网络中的信任管理[J]. 软件学报,2008,19(7):1716-1730
- [2] 陈钟,刘鹏,刘欣. 可信计算概论[J]. 信息安全与通信保密,2003(11)
- [3] Dai Hong-jun, Jia Zhi-ping, Dong Xiao-na. An Entropy-based Trust Modeling and Evaluation for Wireless Sensor Networks [C]//Proceedings of the 2008 International Conference on Embedded Software and Systems, Washington, DC, USA, 2008: 27-34
- [4] Xiong Fei, Xu Qi-jian. Active Trust Transmission Mechanism for Wireless Sensor Network [C]//Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application, Washington, DC, USA, 2008: 626-632

(上接第 30 页)

待进一步研究。虚拟资源供给算法现在大都基于某种假设,还需要对更多约束和限制条件进行考虑,以及从全局角度进行优化设计。

(2)在进行大规模任务的调度和分配时,仍有很多问题亟待解决,如任务依赖关系的处理、节点的动态选择等,需要根据具体任务特性和范围进行优化设计。另外,利用虚拟资源基本元素间的相互关系,需要形成更能均衡成本、服务质量和效率等因素的资源按需调度方法。

(3)在虚拟适配层还可以考虑更多的内容适配,如基于语义的内容适配等。

参 考 文 献

- [1] Douze M, Gaidon A, Jegou H, et al. INRIA-LEAR's video copy detection system [C]//Proc. the TRECVID Workshop. 2008
- [2] Xu Peng, Xie Le-xing, Chang S F, et al. Algorithms and system for segmentation and structure analysis in soccer video [C]//Proc. the IEEE ICME. 2001: 928-931
- [3] Gauch J, Shivadas A. Identification of new commercials using repeated video sequence detection [C]//Proc. the IEEE International Conference on Image Processing. 2006
- [4] Shen Heng-tao, Zhou Xiao-fang, Huang Zi, et al. Statistical summarization of content features for fast near-duplicate video detection [C]//Proc. the 15th International Conference on Multimedia. 2007: 164-165
- [5] Zhao Wan-lei, Tan Song, Ngo C W. Large-scale near-duplicate web video search: Challenge and opportunity [C]//Proc. the IEEE International Conference on Multimedia and Expo. 2009: 1624-1627
- [6] Shang Li-feng, Yang Lin-jun, Wang Fei, et al. Real-time large scale near-duplicate web video retrieval [C]//Proc. the International Conference on Multimedia. 2010: 531-540
- [7] Hoad T, Zobel J. Fast video matching with signature alignment

- [5] Ganerwal S, Srivastava M B. Reputation-based Framework for High Integrity Sensor Networks [C]//SASN '04. Washington, USA, 2004
- [6] Tanchaiwatt S, Dave P, Bhindwale R, et al. Secure Locations: Routing on Trust and Isolating Compromised Sensors in Location-aware Sensor Networks [C]//SenSys '03. California, USA, 2003: 324-325
- [7] Li Xiao-qi, Liu M R, Liu Jiang-chuan. A TrustModel-based Routing Protocol for Secure Ad Hoc Networks [EB /OL]. http://www.cse.cuhk.edu.hk/~lyu/paper_pdf/Aero04_TAODV.pdf
- [8] Kannan R, Ray L, Iyengar S S, et al. Maxmin length-energy-constrained routing in wireless sensor networks [C]//Proceedings of 1st European Conference Workshop on Wireless Sensor Networks, Berlin, Germany, January 2004: 234-249
- [9] Kannan R, Iyengar S S. Game-theoretic models for reliable path-length and energy-constrained routing with data aggregation in wireless sensor networks [J]. IEEE Trans. on Selected Areas of Communications, 2004, 22(6): 1141-1150

[C]//Proc. the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval. 2003: 262-269

- [8] Lejsek H, Tormöesdöttir H, ásmundsson F, et al. Videntifier Forensic: large-scale video identification in practice [C]//Proc. the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence. 2010: 1-6
- [9] Vouk M A. Cloud computing-Issues, research and implementations [J]. Journal of Computing and Information Technology, 2008, 16(4): 235-246
- [10] Armbrust M, Fox A, Griffith R, et al. Above the clouds: A Berkeley view of cloud computing [R]. UCB/EECS-2009-28. EECS Department, University of California, Berkeley, 2009
- [11] Zhu Yong, Ammar M. Algorithms for assigning substrate network resources to virtual network components [C]//Proc. the IEEE INFOCOM. 2006
- [12] Yu Min-lan, Yi Y, Rexford J, et al. Rethinking virtual network embedding: Substrate support for path splitting and migration [J]. ACM SIGCOMM Computer Communication Review, 2008, 38(2): 17-29
- [13] Lu Jing, Turner J. Efficient mapping of virtual networks onto a shared substrate [R]. WUCSE-2006. Department of Computer Science and Engineering, Washington University in St. Louis. 2006
- [14] Courconbetis C, Weber R R. Economic issues in shared infrastructures [C]//Proc. the VISA. 2009: 89-95
- [15] Heymann E, Senar M, Luque E, et al. Evaluation of an adaptive scheduling strategy for master-worker applications on clusters of workstations [J]. High Performance Computing, 2000: 310-319
- [16] Dean J, Ghemawat S. Mapreduce: Simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107-113
- [17] Motahari Nezhad H R, Benattallah B, Martens A, et al. Semi-automated adaptation of service interactions [C]//Proc. the International Conference World Wide Web. 2007: 993-1002