云环境下APT攻击的防御方法综述

张 浩[1] 王丽娜[2] 谈 诚[2] 刘维杰[2]
(国家数字化学习工程技术研究中心(华中师范大学) 武汉 430072)[1]
(武汉大学计算机学院 武汉 430072)[2]

摘 要 云计算以其快速部署、弹性配置等特性吸引了大量的组织和机构使用,然而近期出现的高级可持续性威胁(Advanced Persistent Threat,APT)相比传统的网络攻击具有攻击持续性、高隐蔽性、长期潜伏等特性,为实现云平台的信息资产的安全与隐私保护带来了极大的冲击和挑战。因此,如何有效地防护APT对云平台的攻击成为云安全领域亟待解决的问题。在阐述APT攻击的基本概念、攻击流程与攻击方法的基础之上,分析了APT新特性带来的多重安全挑战,并介绍了国内外在APT防护方面的研究进展。随后针对APT的安全挑战,提出了云平台下APT防护的建议框架,该框架融入了事前和事中防御策略,同时利用大数据挖掘综合分析可能存在的APT攻击以及用于事中的威胁定位与追踪。最后,介绍了安全框架中的关键技术的研究进展,分析了现有技术的优势与不足之处,并探讨了未来的研究方向。

关键词 云计算,高级可持续性威胁,大数据挖掘,威胁定位
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.001

Review of Defense Methods Against Advanced Persistent Threat in Cloud Environment

ZHANG Hao[1] WANG Li-na[2] TAN Cheng[2] LIU Wei-jie[2]
(National Engineering Research Center for E-Learning(Central China Normal University),Wuhan 430072,China)[1]
(Computer School,Wuhan University,Wuhan 430072,China)[2]

Abstract A large number of organizations and institutions have been attracted to use the cloud platform for its features,such as rapid deployment,flexible configurations. However,compared to traditional network attack persistent,the emerging attack mode advanced persistent threat(APT for short) is more persistent,high hidden and long-term buried,which makes the protection to protect security and privacy challenging. Therefore,how to protect the cloud platform from APT effectively becomes an urgent problem. The basic concepts,attack procedures and attack methods of APT were introduced ,and then we analyzed the multiple security challenges brought by APT new features,and introduced the research progress in APT protection aspects. To address the security challenges,we presented a proposal framework to protect cloud platform from APT,which includes the strategies before attack and during attack,and takes advantage of the data mining of big data to analyze the potential APT attack comprehensively and to position and track the threats. Finally,the research progress of some key technologies in our framework was introduced,the advantages and disadvantages were pointed out respectively,and some future research directions were given at the end.

Keywords Cloud computing,Advanced persistent threat,Data mining of big data,Positioning threat

1 引言

云计算[1]作为一种全新服务模式,以其弹性配置、按需购买、易于维护等优势在学术界和工业界得到了空前的发展。大量的国家基础设施及相关的应用服务逐渐转移到云计算平台中。然而在带来便利的同时,云计算也给服务的安全带来巨大的挑战。近年来出现了一种高杀伤力的攻击手段——APT攻击,它是攻击者为了获取某个组织甚至是国家的重要信息而有针对性地进行的一系列攻击行为。因此,国家及企业的机密信息所在的云平台就成了APT攻击的焦点,包括能源、电力、金融、国防等关系到国计民生或者国家核心利益的机密信息。例如,2011年9月发现的Duqu病毒被用来从工业控制系统制造商收集情报信息[2]。因此如何有效地防御云计算环境下的APT攻击,是安全领域中亟待解决的关键问题。

由于强大的攻击力、破坏力,APT的防护受到了各国政

到稿日期:2015-03-17 返修日期:2015-06-03 本文受国家自然科学基金项目(61373169,61103219,61303213),国家发改委重大专项(发改办高技[2013]1309),教育部博士点基金优先发展领域基金项目(20110141130006),华中师范大学中央高校基本科研业务费项目(CCNU15GF001,CCNU15A05010)资助。

张 浩(1986—),男,讲师,CCF会员,主要研究领域为云计算安全;王丽娜(1964—),女,教授,CCF会员,主要研究领域为云计算安全;谈 诚(1988—),男,博士生,CCF会员,主要研究领域为入侵检测;刘维杰(1991—),男,博士生,主要研究领域为网络安全。

府及国际组织结构的重视，如 FireEye、趋势科技等著名 IT 公司都研发了相关的 APT 防护产品。本文根据 APT 的攻击特点，分析了 APT 防御面临的挑战；分析了防御的安全需求以及影响，提出了云环境下 APT 防护建议性框架及重要的科研方向，该框架从攻击的各个防御环节出发，包括事前防御及事后追踪、分析等，以期为我国云计算下 APT 防护的科研、产业发展做出有益的探索。

2 APT 攻击的介绍及分析

2.1 攻击特点

APT 攻击的主要目标是核心利益所在的基础设施，可能是公司、研究所、银行，甚至是国家的能源、电力、金融和国防等重要部门。虽然这些部门通常拥有纵深的防御措施，但是传统的防护方法无法阻止 APT 攻击。APT 具有如下几种典型的特征：

1)攻击组织程度高。攻击多由黑客组织或者网络战部队发起，工具开发、攻击实施、数据分析的队伍分工明确，形成完整的技术链条或者产业链条。

2)长期的经营与策划。APT 攻击目的明确，大多以窃取核心资料为主要目的。攻击者通过各种渠道收集攻击对象的信息，根据需要采用不同的攻击技术和方法。而传统攻击对象随机，攻击方法单一。

3)技术水平高。综合使用各种攻击手段，采用技术手段和非技术手段(社工/骗术)进行渗透攻击，常使用攻击控制平台(Command & Control Platform)、僵尸网络(Botnet)等集中化、自动化工具进行网络攻击或侵袭，掌握并利用 0day 漏洞和第三方应用漏洞，甚至硬件漏洞。攻击方法和技术会不断进化。

4)成本高(资源多)。攻击者可以调动大量的计算和社会资源来为攻击服务，可以完成一些个人难以完成的任务，如爆破密码或者分布式集体取款。

5)攻击持续性。如果一种攻击方式不奏效，攻击者不会放弃，而是换另一种攻击方式继续尝试。在成功渗透进系统以后，会以攻陷的主机为跳板继续渗透系统的剩余部分。

6)高度的隐蔽性。在没有发布攻击任务(如敏感数据获取或者大规模提款等)时，攻击者会一直潜伏并收集情报。此时系统虽然已经被渗透或者完全控制，但是在表面上看不出任何异常。攻击者的通信和上传数据都使用加密传输协议，并且伪装成正常网络流量，传统的 IDS 难以察觉。

因此，相较于普通的恶意行为攻击，APT 攻击的目的性更加明确；APT 攻击普遍采用 0day 漏洞获取权限，其效率和能力远高于普通的攻击；APT 的攻击方式更加多元化，更加隐蔽。

2.2 攻击过程及手段

典型的 APT 攻击过程如图 1 所示。第一步是信息收集。收集的信息包括技术信息和社会信息。根据收集的信息，找出攻击者感兴趣的内容作为攻击目标，分析目标的可能位置，拟定攻击计划。第二步是外部尝试渗透。通过社会工程学手段发起试探，通过可信的关系介绍、IM、邮件等方式让内部员工相信自己可信。第三步是获得内部员工的控制权。攻击者利用 0day 漏洞进行针对性攻击，精确投放钓鱼邮件来控制员工的主机。由于云计算技术的兴起，现有一种新的攻击技术——云服务渗透。若某个员工使用了云存储同步盘，攻击者会利用网盘存在云服务的漏洞，向他的所有电脑同步木马，以渗透内部网络。第四步是逐步渗透到核心资产。渗透到组织内网以后，攻击者以其为跳板，从内部渗透到网络的其他部分，绘制网络结构图，确认攻击目标的位置，并得到访问权限。通过安置后门程序，或者通过专用的 SSL VPN 链接接入内部网络，使用攻击控制平台对系统进行全方位的控制。第五步是价值获取。攻击者在合适的时机利用可信用户的加密通道执行攻击动作，如上传敏感信息、修改账户余额或者进行破坏等。第六步是持续渗透。攻击结束后，攻击者可以考虑结束对目标系统的控制，清除痕迹或者长期获取，等待时机，持续收集信息。

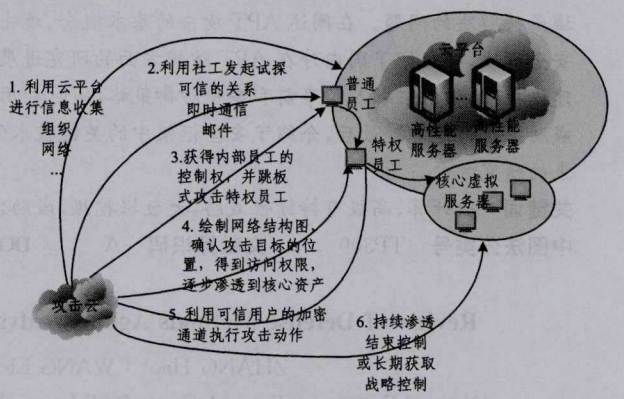

图 1 云环境下 APT 攻击流程

APT 并没有特定的攻击方法或者技术，实际上 APT 是一种攻击形式。为了实现目标，APT 攻击者会综合使用各种现有技术。其中，各种渗透技术则是 APT 攻击的关键，有些攻击方式因成功率高或者隐蔽性好，经常被 APT 攻击所使用。典型的攻击手段有如下几种：(1)通过 SQL 注入、任意文件下载等攻击手段正面突破门户网站 Web Server。(2)以被入侵的门户 Web 作跳板，对 DMZ 区及内网的其他服务器或终端进行渗透。(3)社会工程学，针对人的薄弱环节与信任体系，攻击人的终端，利用人的权限获取。例如，冒用高层主管邮件，发送带有恶意程序的附件，诱骗员工点击并获得内网终端权。(4)通过植入特种木马，回传大量的敏感文件(WORD、PPT、PDF、CAD 文件等)。(5)通过基于 CSRF 的 XSS 邮件进行攻击，目标邮箱自动发送邮件副本给攻击者，掌握目标近况。(6)攻击目标最近联系人、热点关注网站，再实施定点挂马以获得目标终端主机权限。(7)0day 漏洞利用，了解对方使用的软件和环境，有针对性地寻找只有攻击者知道的漏洞，绕过现有的保护体系实现漏洞利用。(8)通过虚拟机逃逸、虚拟机隐蔽信道等云平台特有的攻击方式来获取云平台的管理特权，进而掌控整个网络。

3 云平台下 APT 防护挑战及现状

3.1 云平台下 APT 防护面临的挑战

当前，APT 对云平台的攻击是多层次、多角度的，如网络层次的攻击、内核层次的攻击、应用层的攻击等，每个层次单一的攻击都有相应的技术来防护，然而其防护效果不尽如人意。因此，需要重点分析和研究 APT 的强渗透性、高隐蔽性、攻击的多样性特性对云平台防护带来的防护挑战，提升防御效果。云平台 APT 防护面临的挑战如下。

1)APT 攻击的多样性所引发的安全问题。由于 APT 的攻击目标明确,可能采取多种攻击方式(包括钓鱼网站、0day 漏洞等攻击方式),使得传统的面向已知攻击特征的防护如防火墙、杀毒软件等都无法有效防护 APT 攻击手段。传统的防护方式缺乏对未知威胁的感知能力。

2)APT 隐蔽性所引发的安全问题。由于 APT 在侵入到系统内部后可能通过密文的方式将内部的核心资料传递到外部,使得传统的网络检测无法有效识别。

3)APT 高渗透性所引发的安全问题。由于 APT 攻击需要获取机构的核心机密,所采用的过程可能是从最边缘的人员渗透到核心的特权人员,这种组织内部的渗透降低了组织内特权人员的警惕性和警觉度,大大提高了核心内容被窃取的可能性。

4)APT 长期的潜伏特性所带来的安全问题。由于 APT 攻击具有很强的耐心,在侵入系统后可能进行长达几年的潜伏,且处于信息收集的状态,使得传统的监控和检测机制无法跨越长时间的监控日志和检测日志来有效识别侵入的攻击。

3.2 云平台下 APT 防护现状

为了应对云计算下 APT 攻击,各国政府陆续制定和出台了一系列相关政策来确保国家安全。美国政府大力支持 APT 研究相关的公司(FireEye)。我国也成立了国家安全委员会,APT 的攻击防护将会成为国家安全的一个重要部分。同时,国内外针对 APT 攻击的相关产品也不断涌现,例如 FireEye 公司推出的 APT 防护产品,FireEye 利用沙箱技术和静态分析防止 0day 漏洞、未知型攻击、木马程序;Fortinet 公司推出的 APT 系列产品利用沙盒技术、网络流量分析等技术能有效地防止 0day、Rootkit、木马程序、DDoS 等攻击;趋势科技的 APT 产品 Deep Discovery 利用沙盒技术、关联规则等技术,能有效防止含文件漏洞攻击附件的电子邮件、0day 漏洞、僵尸程序、蠕虫等攻击;Bit9、Hbgary 等公司着力于终端的防护,通过例如白名单的方式来运行应用程序,进而防护 APT 攻击。

4 云平台针对 APT 的防护框架建议

4.1 防护框架的需求

现有的 APT 防御产品和方案各有侧重,并不能有效地防护所有的 APT 攻击。为了应对云平台下 APT 攻击,APT 的防护框架需要能够有效地克服 APT 攻击特性所带来的安全挑战。本文通过分析云平台下 APT 的防御挑战,得出完善的 APT 防护框架具有以下的特性和功能:

1)为了应对 APT 攻击多样性带来的挑战,需要建立多维度的协同防御体系,从而能够从不同的角度防护可能的 APT 攻击。

2)为了应对 APT 攻击高度的隐蔽性,需要有效检测出网络系统的异常行为,包括通过密文传输的流量异常的文件窃取行为。

3)为了应对 APT 长期的潜伏能力,需要长期监视客户虚拟机的各个层次的状态。

4)为了应对 APT 攻击高超的技术水平,不仅需要检测已知威胁,还需检测未知威胁,尤其要能防护 0day 漏洞利用。

5)为了应对 APT 攻击强大的渗透性,需要根据安全级别对资源进行划分,并根据安全级别形成相应的安全域,防止 APT 攻击由外围向内部核心资源渗透;同时对机密的数据进行加密存储。

6)具有能够事中分析的能力。分析 APT 攻击发生的过程,重构攻击场景,挖掘攻击模式,理解攻击意图。综合应用各种分析技术方法,对 APT 攻击威胁的真实来源进行追踪溯源,对于保障安全具有重要的战略意义。

7)利用大数据分析技术对之前收集到的信息进行综合分析,将所有的模块互相关联,实现信息共享,在蛛丝马迹中寻找攻击的痕迹。同时能够利用关联分析对威胁进行综合研判。

4.2 防护体系架构

防护云平台中 APT 的首要任务是针对威胁建立云平台下多层次的 APT 防护框架,并积极开展各个防护层次的关键技术的研究。本节提出一个参考性的云平台下的 APT 防护框架,如图 2 所示。本文充分考虑了 APT 防护框架的需求,以及所有可能的攻击模式和防护方法,将本框架分为事前纵深防御架构和事中威胁分析、定位与追踪架构,通过意图理解和追踪定位来预测和诱导攻击,及时保护被攻击的对象。其中纵深防御架构包括基于虚拟化的多维细粒度监控模块,攻击检测、诊断及取证模块,数据安全防护模块。事中威胁分析、定位和追踪架构主要包括基于大数据的审计及意图理解模块、攻击场景重构模块、攻击溯源模块和综合研判模块。

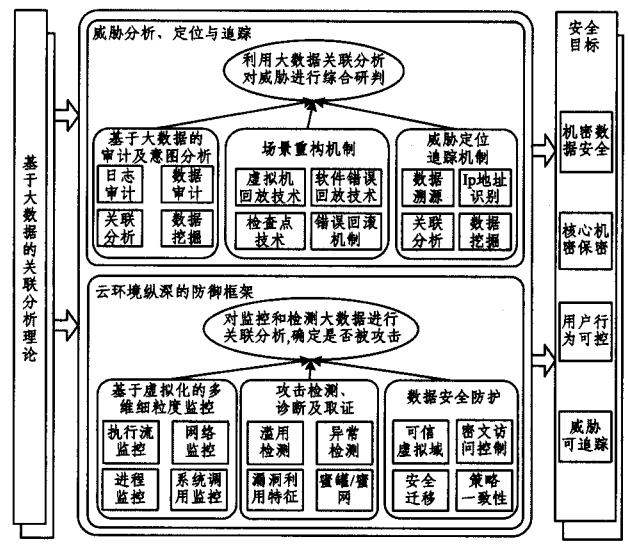

图 2 云环境下 APT 防御体系架构

基于虚拟化的多维细粒度监控模块主要监控网络和主机内部的各种状态。网络监控通过网络流量检测监控网络中存在的异常与威胁。该模块通过控监视系统内部的内存、文件操作、内核结构、进程和执行流等状态,发现微小的系统异常,最大限度地防止敏感信息的泄漏、被破坏和违规外传,并完整记录涉及敏感信息的操作日志,以便事后审计和追究泄密责任。

攻击检测、诊断、取证模块通过对核心服务器引入旁路诱骗机制(如蜜网/蜜罐技术),主动去诱捕可能存在的潜在攻击。虚拟蜜罐是防御方的情报收集系统,通过部署没有业务用途的安全资源,诱骗攻击者对其进行攻击,从而对攻击行为进行捕获和取证,了解攻击工具与方法。同时对虚拟机进行滥用检测、异常检测,以确保用户执行环境的动态完整性,使得用户的正常行为与预期保持一致。为了能够尽可能地发现

APT 攻击，需要结合监控数据和检测数据进行大数据关联分析，以发现任何可能的攻击。

数据安全防护模块根据机密数据的安全级别对资源进行安全划分，并根据用户的机密等级构建可信虚拟域，将高等级的资源进行安全隔离；同时将重要的资料进行加密存储，实行密文访问限制 APT 攻击由外向内的渗透。除此之外，还需要确保数据在动态中的安全，提供虚拟机的安全迁移和迁移过程中安全策略一致性机制。

基于大数据的审计及意图分析模块通过对攻击后的各种日志记录和数据记录进行审计发现攻击行为，包括系统日志、用户行为、资源使用、文件内容和网络流量等。对审计信息进行关联分析和数据挖掘，找出记录之间的相关性，发现新的特征。对挖掘出来的样本数据进行人工的分类或聚类，以机器学习的方法用分类结果构建特征库训练集，不断优化特征库的分类准确率。同时通过关联分析和数据挖掘发现攻击的意图，并为威胁追踪提供支撑。

攻击场景重构模块通过虚拟机的回放和软件回放等技术将状态及时恢复到软件正常的状态，同时利用回放技术再现攻击的过程，以便分析攻击的流程及特征。虚拟机在执行过程中发生的所有事件都可以被虚拟机监视器（Virtual Machine Monitor，VMM）捕获到，虚拟机回放技术是在虚拟化体系构架下，通过 VMM 来完整记录操作系统上的不确定性事件，从而对系统状态进行完全回放。

攻击溯源模块主要由数据溯源技术、数据挖掘等实现。数据溯源技术是一种溯本追源的技术，通过追踪路径重现数据的历史状态和演变过程，实现数据历史档案的追溯。数据挖掘技术是从大量的、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、未知却潜在有用的信息的过程。

由于 APT 攻击的主要对象为网络空间，且网络空间本身具有数据类型和数据格式不一致及日志信息的行为、内容、结构化各异的特点，因此本架构还提供了大数据技术以对监控检测数据进行大数据关联分析，综合分析云平台是否存在被攻击的可能。同时还利用大数据分析技术对正在进行的攻击进行意图理解和反向追踪，实现对威胁的综合研判，从而及时采取相关的策略阻止攻击。

下面将详细分析本框架中的关键技术的研究进展。

5 云平台 APT 防护关键技术分析

5.1 基于虚拟化的监控技术

目前，国内外学术界围绕云计算环境的安全监控已开展了大量研究[3]，众多学者也提出了对软件的监控和支持运行监控的软件设计方法[4]。然而现有云平台操作系统不能提供较强的隔离性，绝大多数监控软件又都运行在目标操作系统中，这使得恶意程序能够感知监控软件的存在，一旦恶意程序渗透乃至控制操作系统，将会严重影响运行于其中的监控软件的正确性和完整性[5]。

近年来，国内外研究人员对系统监控技术展开了广泛的研究。从体系结构上来看，主要可分为两类：系统内部监控机制和系统外部监控机制。内部监控机制通过将监控功能驻留在目标系统内实现，典型的监控系统有 SIM[6] 和 Laries[7]。SIM 系统利用硬件虚拟化的内存管理机制在客户虚拟机中创建一块隔离的地址空间，并通过修改 VMM 的缺页处理程序实时保护该空间的完整性；Laries 系统将监控事件的感知器（Event Sensor）部署于被监控系统内部，而真正的安全监控模块仍驻留在被监控系统外部，并将受保护页面的影子页表设为只读属性，使得任何试图更改页面的操作都将引发缺页中断，进而陷入 VMM 的缺页处理函数中，而根据保护逻辑可以实现精确到机器字级别的内存保护。内部监控机制的优势在于可以较为精确地感知关键事件的发生，并通过直接读取被监控系统的内核代码和数据，省略了语义重构的过程，提高了监控系统的性能。但其不足之处在于需要在被监控系统中插入内核模块或建立独立的地址空间，且内存保护机制对于系统内的受保护模块不具有通用性，对于关键内核数据结构的监控难以确保其完整性，如 DKOM[8] 类的 Rootkit 专门用于破坏系统关键内核数据的完整性。目前，也有研究人员对通用的监控框架进行研究，如 XenAccess[9]、VMDriver[10] 等。XenAccess能够提供通用的函数库，为安全服务提供较好的扩展支持，但是它的功能有限，仅能对内存和磁盘读取进行监控，无法满足复杂的安全服务的需求。综上可知，已有的监控架构不具有很好的通用性，其具体的监控功能仍不够全面。

综上可知，研究人员利用虚拟机监视器 VMM 对上层虚拟机的完全控制权以及虚拟机间运行环境的隔离性，部署实施了良好的安全监控系统，但这种隔离性也为监控系统带来了语义的断层。因此，下一步需要探讨如何解决虚拟化技术的语义断层，并在此基础上实现内部监控优势和外部监控优势的无缝融合。

5.2 攻击检测、诊断、取证技术

系统内核是运行环境的核心，其攻击检测方面已有大量研究工作。为了解决非法代码的注入问题，相关研究[11,12]提出 W⊕X 方法，以保证内核的代码页不能同时具有可写和可执行权限。然而该方法无法确保代码、数据混合页面的完整性。Nickle[13]将客户虚拟机的内核代码拷贝至虚拟机监视器 VMM 的影子内存区域，并利用 VMM 的高特权级以及对客户虚拟机的完全控制权确保影子内存区域中的代码不被篡改。但 Nickle 系统中每次内存访问都将重定向到影子内存区域，从而导致严重的效率问题。卡内基梅隆大学设计了 SecVisor 系统[14]，它利用 Hypervisor 的高特权级保证只有用户允许的代码才能在内核中执行，同时制定了控制流在内核态和用户态切换时遵循的规则，最大程度上确保了内核代码的完整性。普渡大学[15]以软件方式模拟哈佛体系结构，将内存分为数据区和代码区，使得恶意代码的注入局限在数据区，从而无法影响代码区的完整性。Asrigo 构建了一个基于虚拟化技术的蜜罐，主要思路是利用旁路方式截取目标服务器的流量，通过 Hook 目标 OS 内核，并在 VMM 层部署事件处理器，当钩子函数被触发时，通过这些处理器进行相应的处理[47]。

动态信息流跟踪[16,17]可以有效地发现程序正常控制流的篡改以及不信任数据的非法使用。复旦大学提出了 SHIFT 动态信息流跟踪系统[18]和 BOSH 二进制混淆系统[19]，前者能够有效地检测不信任数据的使用情况并发出警告，有效地保护用户系统的安全；后者利用二进制混淆技术混淆程序的控制流，使得恶意程序无法提取程序的正常控制流，进而有效地保护程序控制流的完整性。

马里兰大学提出了 SBCFI 模型[20]，其通过强制执行控制

流完整性策略，周期性地检查内核中的控制数据（函数指针），确保其指向预先制定的可信区域，一旦出现指向未知区域情况，则表明系统的正常控制流受到破坏。但是该方法周期性地检查控制数据可能存在漏检现象，而且也没有提出控制流被破坏后的处理措施。北加州州立大学提出了一种抵御内核Root-kit的轻量级保护系统[21]，该系统增加了一个重定向层用于重定位函数的跳转，并利用虚拟化技术保护该重定向层。然而该方法只能检测和保护已知的数据，无法检测未知的控制数据。

目前数字取证软件在大规模、多任务、并行处理方面滞后于计算机硬件和系统结构的发展[22]。对此，解决的方案有两种：利用更高效的计算能力和更智能的分析方法。V. Roussev等提出利用云计算进行大规模的取证计算[23]，王丽娜等提出用虚拟机进行恶意代码的检测[24]，美亚公司开发了仿真取证环境以及多CPU平台的取证塔，AccessData开发了支持多核处理器的FTK软件；提高数字取证分析效率方法的研究包括分布式分析处理[22]、基于数据挖掘的搜索过程[25]、利用文件分类协助分析[26]等。研究人员还研究了基于网络的架构和虚拟基础设施来帮助大规模证据存储[27]、案件和数字资产管理系统[28]，以及地理上分布的协同分析等[29]。

5.3 攻击场景重构技术

计算系统局部还原与重放技术也是云平台取证与诊断技术中的一个重要研究内容[30]，是一种能够维持系统服务不间断的可用性恢复方法。已有的恢复方法的关注点在于恢复持久性数据和服务是否能够安全还原[31]。文献[32]提出了一项创新的安全技术RX，它可以从确定性和非确定性的软件错误中快速恢复程序，并且可以提供错误的诊断信息；文献[33]提出的Flashback是一项轻量级细粒度的回滚和重放技术，通过记录进程内存状态以及进程与系统的交互关系实现回滚和重放。系统可在发生攻击时有选择性地恢复合法的文件，通过自动解析规则克服无法准确确定污染操作及其影响的问题。ReVirt[34]、Rertace[37]以及ExecRecorder[38]中不需要记录系统中执行的每一条指令，仅仅记录影响进程执行的不确定事件以及这些事件发生的位置或者时间点，在回放时通过这些信息重构系统之前的运行状况；刘海坤结合以上系统，提出了一种基于虚拟机管理器Xen[39]的操作系统回放机制，该机制能够在不修改操作系统以及应用程序的条件下对系统状态进行完整快捷的回放。文献[35,36]提出了文件块级的语义分析方法和语义跟踪重放方法，其不仅可以帮助用户更好地理解文件系统行为，也可以实现文件系统的恢复。系统整体恢复技术会影响所有的进程，失去未感染进程实体的状态信息，导致未感染的服务进程失效，并失去用于保持业务连续性的信息。

现有恢复方法专注于进程级实体，不关注系统内核被破坏的情况，致使系统可能面临更大的风险；而且，通常监控记录的事件类型是系统恢复必需信息的一个子集，而系统恢复所需的数据量可能大于一般的监控日志。场景重构过程中最困难的是证据获取和分析，上面提到的研究中有的部分解决了获取问题，有的提高了搜索的速度，有的能让分析过程分布式进行，如何更有效地解决实时、大规模地获取和分析证据的问题还有待深入研究。

5.4 攻击溯源技术

基于云计算平台中多个层次的攻击分析，包括系统调用分析、关键内核结构的攻击诊断、文件和进程分析、网络流量分析等，建立统一的多层次攻击描述模型，并根据多层的攻击模型的海量日志信息，按照相关规则进行多攻击的关联分析，以发现云平台中可能存在的攻击，为云平台的攻击追溯提供可靠的依据。数据溯源（Data Provenance）技术在数据库领域得到广泛研究。其基本出发点是帮助人们确定数据仓库中各项数据的来源，例如了解它们是由哪些表中的哪些数据项运算而成，据此可以方便地验算结果的正确性，或者以极小的代价进行数据更新。数据溯源的基本方法是标记法，如在文献[40-42]中通过对数据进行标记来记录数据在数据仓库中的查询与传播历史。之后概念进一步细化为why和where两类[43]，分别侧重数据的计算方法以及数据的出处。除数据库以外，它还包括XML数据、流数据与不确定数据的溯源技术[46]。数据溯源技术也可用于文件的溯源与恢复。例如文献[44]通过扩展Linux内核与文件系统，创建了一个数据起源存储系统原型系统，以自动搜集起源数据。此外，也有其在云存储场景中的应用[45]。

当前，数据溯源面临的主要挑战有两个方面：1）怎样处理异构数据源，例如列数据库、文档数据库、Key/Value系统、XML数据库。保证数据挖掘系统能有效地处理不同类型的数据源是至关重要的。2）怎样处理海量数据。随着时间的推移以及互联网、物联网的应用，数据量呈指数级增长，为了有效地处理大数据，数据挖掘算法必须是高效率的、可伸缩的。

5.5 大数据挖掘技术

在云计算环境中，APT攻击采用多种途径进行攻击，孤立地进行恶意代码的检测和主机应用保护对防御APT攻击来说是很难奏效的。因此，APT的防护需要利用数据挖掘来关联分析各层面、各阶段的全方位信息数据，进而挖掘出可能存在的攻击和定位威胁来源。

基于数据挖掘的安全检测[48,49]是一类重要的安全检测手段，随着各类应用数据、业务数据和安全审计数据规模的急剧增加，传统的串行数据挖掘算法往往只能够处理一些小规模的数据，当迁移到大数据环境[54,55]下，它们的执行速度会降低甚至无法运行，这对目前的数据挖掘算法提出了严峻的挑战和考验。

并行计算是解决大数据挖掘任务的有效手段，它可以将计算任务分布在大量互连的计算机上，使各种应用系统能够根据需要获取计算资源、存储资源和其他服务资源。MapReduce[52]和Hadoop[50,51]等高扩展性、高性能的并行计算编程模型、分布式大数据处理框架以及相关关键技术使得大数据数据存储和分布式计算成为现实。基于分布式计算的大数据挖掘技术可以为更多、更复杂的基于数据挖掘的安全检测问题提供新的理论与支撑工具。目前大数据挖掘的安全检测技术主要研究基于分布式计算的新型大数据挖掘方法、各种数据挖掘算法的并行化策略，以及在MapReduce[53]上实现大规模的异常分析和检测等。

为了解决计算效能的问题，国内外学者相继提出了基于集群[56]、基于网格[57]、基于Agent[58]等的各种分布式数据挖掘平台，它们提高了数据挖掘系统的处理能力，但都有自身难以解决的问题。将分布式计算融入数据挖掘，通过大规模并行计算，可以解决大数据挖掘的效率问题。从近几年频繁项集挖掘算法[59]的研究趋势来看，高效的算法依然是研究热点

之一，为此出现了一系列的混合搜索策略和高效剪枝策略[60]。基于相似项发现的大数据挖掘技术主要解决文本的相似度和敏感话题发现等问题。其中的关键技术有最小哈希技术、局部敏感哈希、邻近搜索文档的相似度、协同过滤[61]等。另外，基于约束的频繁项集的挖掘、并行的挖掘算法、频繁项集的增量挖掘等都是目前的研究方向，有待进一步探讨。

大数据也是一把双刃剑[62]，利用大数据做关联分析有助于提高对未知攻击的识别，然而这些数据也可以被恶意攻击者用来分析系统的运行情况以及网络拓扑等情况。因此，在利用大数据的同时也要注意大数据的存储以及传输等方面的安全。

结束语 本文从云计算环境中的 APT 防护角度出发，对云环境中 APT 的监控、检测、攻击场景重构、追踪溯源等相关技术进行了阐述和分析，云计算中 APT 的防护虽然已经取得了一定的研究进展，但仍有巨大的研究空间。今后的研究工作可侧重以下几个方面。

1)基于虚拟化的 APT 监控和分析服务。在云计算环境中，虚拟化平台为 APT 的监控提供了便利，如用户为了确保监视器本身的安全性，将监视器置于特权域中，然而，这也使监视器失去了被监视系统的语义。透明性和语义鸿沟之间的平衡是一个挑战性问题。同时，由于云环境中物理资源的共享带来了一定的威胁，如虚拟边界威胁、恶意占用虚拟资源等，因此也需要针对虚拟化带来的安全问题进行有效的防护。

2)APT 攻击场景重构的信息完整性。现有场景重构专注于进程级实体，不关注系统内核被破坏的情况，致使系统可能面临更大的风险；而且，通常监控记录的事件类型是系统恢复必需信息的一个子集，而系统恢复所需的数据量可能大于一般的监控日志。

3)基于大数据的关联分析技术的研究。APT 采用多途径攻击，孤立地检测恶意代码和监控主机对防御 APT 攻击来说是很难奏效的。因此，需要通过日志、网络流量等大数据关联分析来应对 APT 攻击，这就需要大数据分析技术的进一步发展。同时，也要注意大数据的存储以及传输等方面的安全，防止攻击者获取云平台的大数据来进行挖掘分析，进而有针对地进行攻击。

参 考 文 献

[1] Toosi A N, Calheiros R N, Buyya R. Interconnected Cloud Computing Environments: Challenges, Taxonomy, and Survey[J]. ACM Computing Surveys (CSUR), 2014, 47(1): 1-47

[2] Bencsáth B, Pék G, Buttyán L, et al. Duqu: Analysis, detection, and lessons learned[C]// ACM European Workshop on System Security (EuroSec 2012). 2012

[3] Zeng Jin, Sun Hai-long, Liu Xu-dong, et al. Dynamic Evolution Mechanism for Trustworthy Software Based on Service Composition[J]. Journal of Software, 2010, 21(2): 261-276 (in Chinese)
曾晋，孙海龙，刘旭东，等. 基于服务组合的可信软件动态演化机制[J]. 软件学报，2010，21(2)：261-276

[4] Wen Jing, Wang Huai-min, Ying Shi, et al. Toward a Software Architectural Design Approach for Trusted Software Based on Monitoring[J]. Chinese Journal of Computers, 2010, 33(12): 2321-2334(in Chinese)
文静，王怀民，应时，等. 支持运行监控的可信软件体系结构设计方法[J]. 计算机学报，2010，33(12)：2321-2334

[5] Xiang Guo-fu, Jin Hai, Zou De-qing, et al. Virtualization-Based Security Monitoring[J]. Journal of Software, 2012, 23(8): 2173-2187(in Chinese)
项国富，金海，邹德清，等. 基于虚拟化的安全监控[J]. 软件学报，2012，23(8)：2173-2187

[6] Sharif M, Lee W, Cui W, et al. Secure In-VM Monitoring Using Hardware Virtualization[C]// Proceedings of the 16th ACM Conference on Computer and Communications Security. 2009: 477-487

[7] Payne B, Carbone M, Sharif M, et al. Lares: An architecture for secure active monitoring using virtualization[C]// Proceedings of the IEEE Symposium on Security and Privacy. 2008: 233-247

[8] Butler J. DKOM (Direct Kernel Object Manipulation) [EB/OL]. http://www.blackhat.com/presentations/win-usa-04/bh-win-04-butler.pdf

[9] Payne B D, Carbone M A, Lee W. Secure and flexible monitoring of virtual machines[C]// The 23rd Annual Computer Security Applications Conf. New York: ACM Press, 2007: 385-397

[10] Xiang G, Jin H, Zou D, et al. VMDriver: A driver-based monitoring mechanism for virtualization[C]// Proc. of the 29th Int'l Symp on Reliable Distributed Systems. Washington: IEEE Computer Society, 2010: 72-81

[11] Team P X. Documentation for the PaX project-overall description[EB/OL]. http://pax.Grsecurity.net/docs/pax.txt

[12] Microsoft. A detailed description of the Data Execution Prevention (DEP) feature in Windows XP Service Pack 2[OL]. http://support.microsoft.com/kb/875352

[13] Riley R, Jiang Xu-xian, Xu Dong-yan. Guest-Transparent Prevention of Kernel Rootkits with VMM-Based Memory Shadowing[C]// Proceedings of the 11th Symposium on Recent Advances in Intrusion Detection (RAID). 2008: 1-20

[14] Seshadri A, Luk M, Qu Ning, et al. SecVisor: A Tiny Hypervisor to Provide Lifetime Kernel Code Integrity for Commodity OSes [C]// Proceedings of 21st ACM SIGOPS Symposium on Operating Systems Principles. 2007: 335-350

[15] Riley R, Jiang Xu-xian, Xu Dong-yan. An Architectural Approach to Preventing Code Injection Attacks[C]// DSN. 2007: 30-40

[16] Crandau J, Chon F. Minos: Control Data Attack Prevention Orthogonal to Memory Model[C]// 37th International Sympcsium on Microarchitecture. 2004: 221-232

[17] Suh G, Lee J, Zhang D, et al. Secure Program Execution via Dynamic Information flow Tracking[C]// Proceeding of International Conference on Architectural Support for Programming Languages and Operating Systems. 2004: 85-96

[18] Chen H, Wu X, Yuan L, et al. From Speculation to Security: Practical and Efficient Information Flow Tracking Using Speculative Hardware[C]// Proceeding of the 35th International Symposium on Computer Architecture (ISCA'08). Washington DC, USA: IEEE Computer Society, 2008: 401-412

[19] Chen H, Wu X, Yuan L, et al. Binary Obfuscation Using Taint Tracking[C]// International Conference on Architectural Support on Programming Language and Operating System. 2008

[20] Petroni N L, Hicks M. Automated Detection of Persistent Ker-

nel Control-Flow Attacks[C]// Proceedings of the 14th ACM Conference on Computer and Communications Security. 2007: 103-115

[21] Wang Z, Jiang X, Cui W, et al. Countering Kernel Rootkits with Lightweight Hook Protection[C]// Proceedings of the 16th ACM Conference on Computer and Communications Security. 2009: 545-554

[22] Joyce R, et al. MEGA: A tool for Mac OS X operating system and application forensics[J]. Digital Investigation, 2008, 5(suppl): 83-90

[23] Roussev V, et al. A cloud computing platform for large-scale forensic computing[M]// Advances in Digital Forensics V. 2009: 201-214

[24] Wang Li-na, Gao Han-jun, et al. Detecting and Managing Hidden Processvia Hypervisor[J]. Journal of Computer Research and Development, 2011, 48(8): 1534-1541(in Chinese)
王丽娜,高汉军,等. 利用虚拟机监视器检测及管理隐藏进程[J]. 计算机研究及发展,2011,48(8):1534-1541

[25] Pollitt M, et al. Virtualization and digital forensics: A research and teaching agenda[J]. Journal of Digital Forensic Practice, 2008, 2(2): 66-73

[26] Beebe N, et al. Clark. Dealing with terabyte data sets in digital investigations[M]// Advances in Digital Forensics. 2006: 3-16

[27] Mell P, Grance T. Draft nist workingd efinition of cloud computing[EB/OL]. (2009-4-24)[2011-09-10]. www. csrc. nist. gov/groups/SNS/cloud-computing/index. html

[28] Solomon J, et al. User data persistence in physical memory[J]. Digital Investigation, 2007, 4(2): 199-211

[29] Dorn G, et al. Analyzing the impact of a virtual machine on a host machine[M]// Advances in Digital Forensics V. 2009: 69-82

[30] Lin Chuang, Su Wen-bo, Meng Kun, et al. Cloud Computing Security: Architecture, Mechanism and Modeling[J]. Chinese Journal of Computers, 2013, 36(9): 32-37(in Chinese)
林闯,苏文博,孟坤,等. 云计算安全:架构、机制与模型评价[J]. 计算机学报,2013,36(9):32-37

[31] Lu Dun. Modeling and Reasoning of the Software Component Based System Recovery Basedon Survivability Specfication[J]. Journal of Software, 2007, 18(12): 3031-3047(in Chinese)
卢暾. 基于可生存性规范的软件构建系统恢复的建模与推理[J]. 软件学报,2007,18(12):3031-3047

[32] Qin F, Tucek J, Sundaresan J, et al. Rx: treating bugs as allergies-A safe method to survive software failures[J]. ACM-SIGOPS, 2005, 39(5): 235-248

[33] Srinivasan S M, Kandula S, Andrews C R, et al. Flashback: A lightweight extension for rollback and deterministic replay for software debugging[D]. USA: University of Illinois at Urbana, Champaign, 2004

[34] Dunlap G W, King S T, Cinar S, et al. ReVirt: enabling intrusion analysis through virtual-machine logging and replay[J]. ACM Sigops Operating System Review, 2002, 36(S1): 211-224

[35] Prabhakaran V, Arpaci-Dusseau A C, Arpaci-Dusseau R H. Analysis and evolution of journaling file system[C]// USENIX. Anaheim, USA, 2005: 105-120

[36] Grizzard J B, Gardner R W. Analysis of Virtual Machine Record and Replay for Trustworthy Computing[J]. Johns Hopkins APL Technical Digest, 2013, 32(2): 528-535

[37] Xu M, Malyugin V, Sheldon J, et al. ReTrace: Collecting execution trace with virtual machine deterministic replay[C]// Proceedings of the Third Annual Workshop on Modeling, Benchmarking and Simulation(MoBS'07). California, USA, 2007

[38] de Oliveira D A S, Crandall J R, Wassermann G, et al. ExecRecorder: VM-based full-system replay for attack analysis and system recovery[C]// Proceedings of the First Workshopon Architectural and System Support for Improving Software Dependability(ASID'06). SanJose, Caiifornia, ACM Press, 2006: 66-71

[39] Barham P, Dragovic B, Fraser K, et al. Xen and the art of virtuaiization[C]// Proceedings of the 19th ACM Symposium on Operating Systems Principies (SOSP'03). LakeGeorge, New-York, USA: ACM Press, 2003: 164-177

[40] Cui Y, Widom J, Wiener J L. Tracing the lineage of view data in a warehousing environment[J]. ACM Transactions on Database System(TODS), 2000, 25(2): 179-227

[41] Cui Y, Widom J. Practical lineage tracing in data warehouses[C]// International Conference on Data Engineering(ICDE). SanDiego, USA, 2000: 367-378

[42] Cui Y, Widom J. Lineage tracing for general data warehouse transformation[J]. The International Journal on Very Large Data Bases, 2003, 12(1): 41-58

[43] Buneman P, Khanna S, Tan W C. Why and where: A characterization of data provenance[C]// 8th International Conference on Database Theory(ICDT). London, UK, 2001: 316-330

[44] Muniswamy-Reddy K K, Holland D A, Braun U, et al. Provenance-aware storage systems[C]// USENIX Annual Technical Conference. Boston, USA, 2006: 43-56

[45] Muniswamy-Reddy K K, Macko P, Seltzer M. Provenance for the cloud[C]// The 8th USENIX Conference on Fileand Storage Technologies. SanJose, USA, 2010: 15-28

[46] Gao Ming, Jin Che-qing, Wang Xiao-ling, et al. A Survey on Management of Data Provenance[J]. Chinese Journal of Computers, 2010, 33(3): 374-389(in Chinese)
高明,金澈清,王晓玲,等. 数据世系管理技术研究综述[J]. 计算机学报,2010,33(3):374-389

[47] Asrigo K, Litty L, Lie D. Using VMM-based sensors to monitor honeypots[C]// Proceedings of the 2nd International Conference on Virtual Execution Environments(VEE'06). 2006: 13-23

[48] Phua C, Lee V, Smith K, et al. A comprehensive survey of data mining-based fraud detection research[J]. arXivpreprintarXiv: 1009. 6119, 2010

[49] Schultz M G, Eskin E, Zadok E, et al. Data mining methods for detection of new malicious executables[C]// Proceedings of 2001 IEEE Symposium on Security and Privacy(S&P2001). IEEE, 2001: 38-49

[50] He Y, Lee R, Huai Y, et al. RCFile: A fast and space-efficient data placement structure in MapReduce-based warehouse systems[C]// 2011 IEEE 27th International Conference on Data Engineering(ICDE). IEEE, 2011: 1199-1208

[51] Floratou A, Patel J M, Shekita E J, et al. Column-oriented storage techniques for MapReduce[J]. Proceedings of the VLDB Endowment, 2011, 4(7): 419-429

(下转第 43 页)

[2] Hsu C M. Forecasting Stock/futures Prices by Using Neural Networks with Feature Selection[C]//Proceedings of 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference(ITAIC 2011). 2011:1-7

[3] Ma Li-li, Li Quan. Chinese Investors' Risk Preferences in China [J]. Statistical Research, 2011, 28(8):63-72(in Chinese)
马莉莉,李泉. 中国投资者的风险偏好[J]. 统计研究,2011,28(8):63-72

[4] Duan Jing, Liu Yong-fang, He Qi. The Effects of Decision Makers' Roles and Related Variables on Risk Preferences[J]. Acta Psychologica Sinica, 2012, 44(3):369-376(in Chinese)
段婧,刘永芳,何琪. 决策者角色及相关变量对风险偏好的影响[J]. 心理学报,2012,44(3):369-376

[5] Chen Xu-hong. Extension and Application of Markowitz Mean-Variance Model[D]. Shanghai: Shanghai Jiaotong University, 2009(in Chinese)
陈许红. Markowitz 均值-方差模型的推广及应用[D]. 上海:上海交通大学,2009

[6] Xu Xu-song, Chen Yan-bin. Portfolio choice model based on mean absolute deviation and it's simulated annealing algorithm[J]. Journal of Management Sciences in China, 2002, 5(3):79-85(in Chinese)
徐绪松,陈彦斌. 绝对离差证券组合投资模型及其模拟退火算法[J]. 管理科学学报,2002,5(3):79-85

[7] Huang Fu-kai. An Empirical Study of Downside Risk and Value Premium of A-Share Market: Based on Semi-Variance Analysis [J]. South China finance, 2014(3):57-62(in Chinese)
黄赋凯. 我国 A 股价值溢价实证研究-基于半方差方法的分析[J]. 南方金融,2014(3):57-62

[8] Hu Xiao-wen, Hui Jun. Weighted semi-variance risk measuring model for portfolio investment[J]. Journal of Hefei University of Technology(Natural Science), 2007, 30(4):518-520(in Chinese)
胡小文,惠军. 加权半方差风险度量模型[J]. 合肥工业大学学报(自然科学版),2007,30(4):518-520

[9] Chen Ju-hua, Sui Shan-shan, Wang Jian-jiang. Compensation Regulation, Risk Preference and Enterprise Investment Efficiency: An Analysis Based on Meeting Effect Theory[C]//Proceedings of the Fourth Conference on Accounting on Both Sides of the Taiwan Straits. 2012(in Chinese)
陈菊花,隋姗姗,王建将. 薪酬管制、风险偏好与企业投资效率:基于迎合效应的分析[C]//第四届海峡两岸会计学术研讨会论文集. 2012

[10] Zhao Yan. Research and Application on BP Neural Network Algorithm[C]//Proceedings of 2015 International Industrial Informatics and Computer Engineering Conference (IIICEC 2015). 2015:1444-1447

[11] Wang Shuang-cheng. Bayesian Network Learning, Reasoning and Application[M]. Shanghai: Lixin accounting press, 2009(in Chinese)
王双成. 贝叶斯网络学习、推理与应用[M]. 上海:立信会计出版社,2009

[12] Dong Hong-hui, Jia Li-min, Tian Yin, et al. Quantitatively Analysis of Train Communication Network Based on Multi-attribute Utility Function[C]//Proceedings of 17th International IEEE Conference on Intelligent Transportation Systems I. 2014:2330-2335

[13] Chen Xiao-wei, Park G K. Expected Uncertain Utility Function and Its Risk Averse Analysis[C]//Proceedings of the Thirteenth International Conference on Information and Management Sciences. 2014:305-309

[14] Wu Jheng-long, Su Chen-chi, et al. Stock Price Predication using Combinational Features from Sentimental Analysis of Stock News and Technical Analysis of Trading Information[C]//Proceedings of International Conference on Economics, Business and Management(ICEBM2012). 2012:39-43

(上接第 7 页)

[52] Li Bo-duo, Edward M, Diao Yan-lei, et al. A platformf or scalable one-pass analytics using MapReduce[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data(SIGMOD'11). Athens, Greece, 2011:985-996

[53] Blanas S, Jignesh P, Ercegovac V, et al. A comparison of join algorithms for log processing in MaPreduce[C]//Proceedings of the ACMSIGMOD International Conferenceon Management of Data(SIGMOD'10). Indianapolis, Indiana, USA, 2010:975-986

[54] Qin Xiong-pai, Wang Hui-ju, Du Xiao-yong, et al. Big Data Analysis-Competition and Symbiosis of RDBMS and MapReduce [J]. Journal of Software, 2012, 23(1):32-45(in Chinese)
覃雄派,王会举,杜小勇,等. 大数据分析-RDBMS 与 MapReduce 的竞争与共生[J]. 软件学报,2012,23(1):32-45

[55] Meng Xiao-feng, Ci Xiang. Big Data Management: Concepts, Techniques and Challenges[J]. Journal of Computer Research and Development, 2013, 50(1):146-169(in Chinese)
孟小峰,慈祥. 大数据管理:概念,技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169

[56] Guo J, Li Y, Du L, et al. Research on Distributed Data Mining System Based on Hadoop Platform[C]//Proceedings of International Conference on Computer Science and Information Technology. SpringerIndia, 2014:629-636

[57] Skillicorn D, Talia D. Mining large data sets on grids: Issues and prospects[J]. Computing and Informatics, 2012, 21(4):347-362

[58] Sakaeda T, Kadoyama K, Yabuuchi H, et al. Data mining of the public version of the FDA adverse event reporting system[J]. International Journal of Medicalsciences, 2013, 10(7):796-803

[59] Tong Y, Chen L, Yu P S. UFIMT: an uncertain frequent itemset mining toolbox[C]//Proceedings of the 18th ACMSIGKDD International Conferenceon Knowledge Discovery and Datamining. ACM, 2012:1508-1511

[60] Weisrock D W, Smith S D, Chan L M, et al. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics[J]. Molecular Biology and Evolution, 2012, 29(6):1615-1630

[61] Koren Y. Collaborative filtering with temporaldynamics [J]. Communications of the ACM, 2010, 53(4):89-97

[62] Feng Deng-guo, Zhang Min, Li Hao, et al. Big Data Security and Privacy Protection[J]. Chinese Journal of Computers, 2014, 37(1):246-258(in Chinese)
冯登国,张敏,李昊,等. 大数据安全与隐私保护[J]. 计算机学报,2014,37(1):246-258