

基于 Agent 和数据挖掘的分布式信息审计平台

王 强¹,皮德常¹,李伟奇²,吕 军²

(1.南京航空航天大学 信息科学与技术学院,江苏 南京 210016;

2.江苏南京金鹰国际软件系统有限公司,江苏 南京 210002)

摘 要:针对部门内部人员对部门信息实施的一系列不安全的操作行为,构建了一个基于 Agent 和数据挖掘的分布式终端行为审计平台。采用数据挖掘技术,实时地对系统所产生的报警信息和审计日志进行分析,并自动扩充规则库;同时,采用入侵检测技术帮助系统管理员做出实时的安全策略。考虑到多部门终端地理位置分散的分布式用户环境,系统引入 Agent。该系统为银行、证券、保险、政府和企业等涉密部门的信息系统提供了一个内部可信赖的安全审计环境。

关键词:信息安全;审计规则;Agent;数据挖掘

中图分类号:TP309

文献标识码:A

文章编号:1005-3751(2006)04-0141-03

Distributed Information Audit System Based on Agent and Data Mining

WANG Qiang¹,PI De-chang¹,LI Wei-qi²,LÜ Jun²

(1.Coll.of Info. Sci. and Techn.,Nanjing Univ.of Aeronautics & Astronautics,Nanjing 210016,China;

2.G E International Software-System Co.Ltd,Nanjing 210002,China)

Abstract:Considering employee's unauthorized operations and misuse of the information of department,this paper constructs a distributed terminal monitor system based on agent and data mining. In this paper,DM technology and intrusion detection technology are introduced to mine alarm information and system audit logs,make analysis and make real-time security strategy for system administrator. This paper also introduces multi-agents into the distributed user's environment,in order to adapt the multi-departments and multi-terminals' environment. This system provides an inside reliable security audit environment for bank,bond department,insurance,government,corporation and so on.

Key words:information security;audit rules;agent;data mining

0 引 言

近年来,由于网络的盛行,使得人们更多地关注来自部门外部入侵的信息安全问题,入侵检测技术得到了长足的发展,为此,人们研制了许多面向外部的入侵检测系统。然而,FBI和CSI对484家公司的信息安全调查发现:有超过85%的安全威胁来自企业内部;有16%未授权的存取来自内部;有14%专利信息被内部人员所窃取;有12%内部人员有财务欺骗;有11%资料或网络被内部人员破坏。由此可见,当前部门信息所面临的内部安全威胁要比外部的威胁更为突出。基于此,文中提出了一种面向信息部门内部的安全新方法。

1 信息安全技术

目前信息安全的研究已经不局限在传统意义上的人

侵检测、身份认证、密码学或防火墙之类的技术,由于信息领域一些新技术的出现,信息安全技术发展不仅更加深入,而且更加注重综合各种技术构造新的信息安全监控模型:

(1)Agent技术:Agent在分布式用户环境中得到了广泛的应用。Yoav Shoham给出了运用于计算机领域的Agent的描述性定义^[1]:“一个Agent是一个实体,它的状态由许多精神元件构成,如信念、能力、选择和义务等,这些以精神方式定义的元件能够表示意念的粗糙模糊的含义”。Agent作为一个智能化的软件独立体,可以自主地完成所分配的任务,适合分布式环境。

(2)数据挖掘技术:由于数据挖掘在大容量数据分析过程中以及文本挖掘中的优势,使得它在信息安全领域得到了广泛的应用,对审计数据的挖掘分析,发现和描述用户和系统的行为方式,给安全专家提供了可靠的安全策略;对审计规则库的挖掘分析,实现规则库的自扩充功能;对审计对象(文件内容、网页内容、邮件内容)进行文本挖掘,发现用户频繁访问模式,以免信息外泄。目前应用较多的挖掘方法有:分类,序列分析,聚类分析和关联规则等。这些数据挖掘方法在入侵检测系统中已经得到了应

收稿日期:2005-07-18

基金项目:科技部科技型企业技术创新基金项目(20023211053608)

作者简介:王 强(1981-),男,山东临沂人,硕士研究生,研究方向为数据挖掘、信息安全审计;皮德常,副教授,博士,硕士生导师,研究方向为数据库系统、数据挖掘。

用^[2],文中将此技术引入到信息审计平台的构建。

(3)入侵检测技术:入侵检测^[3]作为一种外部防范信息安全的手段,其中有些技术对构建终端内控系统提供了支持。主要有:

a. 进程的资源使用率:记录因违规操作,导致系统资源使用异常,构造异常使用模式。

b. 文件的访问权限:在 Bell-LaPadula (BLP)^[4]模型中,Bell 和 LaPadula 提出了文件存取权限涉及的 3 个要素:主体,客体和存取属性。

c. 基于进程的系统调用序列:当用户执行某程序时产生一个系统调用序列,与预先建立的正常序列模式进行匹配,当不匹配率达到阈值时,就认为是操作失误或违规,并拒绝用户执行操作,屏蔽执行结果,起到信息保护的作用。

2 基于 Agent 和数据挖掘的分布式信息审计平台

2.1 APA 监控审计安全机制

应用过程审计(Application Process Audit, APA)是面向各种领域(行业)所设计的一个应用系统,它针对各种安全隐患和业务风险,将领域应用的业务处理特性与安全审计规则相结合,采用构件构架技术、信息审计和检测预警技术,建立领域应用审计知识管理库,通过对业务系统运行过程中各种内部、外部行为过程在信息系统各个层面上(应用系统、数据库、操作系统、网络)所遗留的痕迹信息进行实时和准实时扫描分析处理,以跟踪监测系统的运行,达到安全防护的目的,为内审部门提供了一套平台化、可描述、可配置的事中、事后审计工具。APA 系统的监控审计安全机制如图 1 所示。

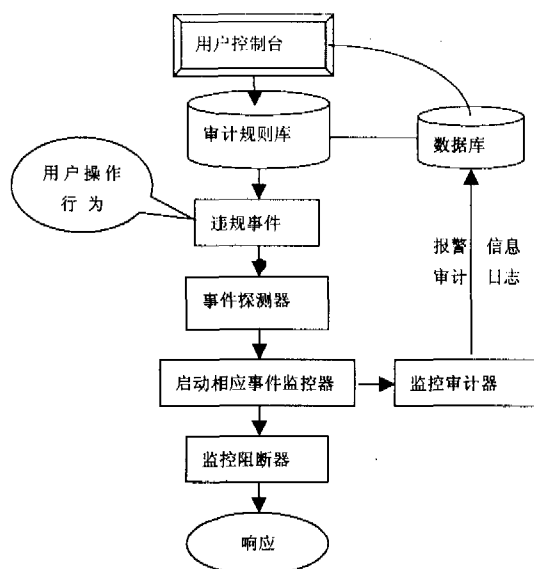


图 1 APA 审计安全机制

2.2 APA 的体系结构

基于 APA 的监控审计安全机制,提出了基于 Agent 和数据挖掘的终端内控系统体系结构。系统有如下几个 Agent 协同实现整个系统的监控工作:

(1)数据采集 Agent:采集终端运行信息和网络环境

信息,提交事件探测 Agent。

(2)事件探测 Agent:根据终端运行信息,基于预定义的审计规则和事件库,启动相应事件 Agent 对终端进行监控。

(3)事件 Agent:对终端采样信息和审计规则特征值进行匹配,判定事件是否违规,并做出裁决。其中在文件安全 Agent 中,根据 Bell-LaPadula(BLP)模型提出的三要素,给出了保密工作区概念。控制文档合法使用者的访问权限,包括打印、复制粘贴、保存和截屏,防止二次传播文档内容,保密工作区中文件只能进不能出,即使允许对文件进行存取操作,也要记录操作文件的新旧副本,记录用户对文件内容的更改;同时采用国际标准 AES-128 位加密算法对文件加密;提供保密工作区之间的访问管理;动态集中权限管理,提供访问权限的部署和回收;详细的文件访问日志记录,提供水印管理,以备事后审计。

(4)监控阻断 Agent:对各种类型的事件采取阻断处理,拒绝用户操作。

(5)监控审计 Agent:审核整个用户操作过程,记录审计日志和报警信息,以备事后审计和提交数据挖掘 Agent 作实时处理。

(6)数据挖掘 Agent:对审计规则库的分类和聚类分析,解决审计规则之间的定义冲突和部署冲突;对审计规则库的关联规则进行分析,实现审计规则库的自扩充;对审计日志和报警信息的实时挖掘,实时发现重大审计结果,自动调整规则部署,设置合理报警级别,并提交系统管理员;考虑到存取文件内容和在网络上传输信息对信息外泄的威胁,该系统提出了对文件监控和网络监控对象进行内容审计,记录操作副本入库,同时对审计副本库进行文本挖掘,发现用户对文件内容和网络传输信息内容隐含的操作模式,目前文本挖掘多采用分类、聚类算法,APA 系统采用了基于自组织特征映射神经网络(Self-Organizing Map, SOM)的聚类算法,它是一种无监督的自组织和自学习网络。该算法的工作主要有两步构成,首先是训练 SOM 网络,然后采用该网络对文档进行聚类。算法简要描述如下:

①初始化网络连接权值;

②将一个给定的输入向量加载到网络上;

③计算输入向量与权值向量的点积,并选择其中的最大者;

④更新所选结点及其临域结点的连接权值;

重复步骤②、③、④,直到满足终止条件。

(7)多 Agent 间的协作与通信:在系统设计之初,就考虑到要将系统复杂的监控审计任务细分为各个子任务,交给相应的 Agent 处理,为此,必须有一套统一的 Agent 协作语言。知识查询与操纵语言(Knowledge Query and Manipulation Language, KQML)^[5,6]是一种通用的 Agent 通信语言,同时也是一种交换知识和信息的描述性语言及协议,它定义了 Agent 间传递消息的格式和消息处理的协

议,通过提供一套标准的通信原语实现知识共享,对多智能系统中问题空间进行求解。APA 采用点对点方式的通信模式,由 Agent 本身确定消息的发送路径和方式。一个典型的 APA KQML 消息如下:

```
ask - one
:sender ProbeAgentID
:content (MonitorContent)
:receiver MonitorAgentID
:reply - with AuditContent
:language self - defined
:ontology Monitor - mode
```

其中, KQML 消息的行为原语是 ask - one; “: sender”是消息的发送者; “: receiver”是消息的接收者; “: content”代表消息内容; “: reply - with”表示下条消息对本条消息的响应; “: language”表示消息内容使用语言的名称; “: ontology”表示消息内容使用的实体集的名称; “: content”为内容层; “: sender, : receiver, : reply - with”构成了通信层; 行为原语“ask - one, : language, : ontology”为消息层。

在系统实现中,采用类似 Windows 消息处理流程的方法,自定义 KQML 消息:

```
KQMLMSG msg; //定义消息名
while (GetKQMLMessage (&msg, NULL, 0, 0))
```

```
TranslateKQMLMessage (&msg); //翻译消息
DispatchKQMLMessage (&msg); //分发消息
```

同时实现 SendKQMLMessage, ReplyKQMLMessage, 对 KQML 消息进行多 Agent 间通信处理。上述 Agent 之间的体系结构如图 2 所示。

3 系统实现及应用

APA 系统整体上分为用户控制台、控制服务器和 Agent 三个部分。用户控制台主要面向操作型的管理人员,考虑到操作员的需求,采用 Delphi 实现,其界面友好,操作方便;它主要负责审计规则和系统允许参数的定制和下发,以及审计信息查看和响应策略,如图 3 所示。Agent 运行在被监控的桌面终端机器上,从用户所登录的域控制服务器下载最新版本的 Agent 运行体到本地后,自动启动 Agent 并隐藏运行体,然后开始实时监控当前机器上的用户操作和系统运行情况,对用户的违规操作进行审计阻断,并实时上传信息。考虑到系统的效率,Agent 端代码采用 C++ 开发完成。控制服务器则负责审计规则、系统运行参数下发,审计信息上传等数据传输服务。

目前 APA 系统已经应用到银行、保险、证券、财税等许多部门,针对各种安全隐患和业务风险,基于行为过程进行信息审计。它独立于业务信息系统,对信息系统运行过程中遗留下的各种行为痕迹信息进行实时和准实时扫描处理,并依据合理性、合法性、真实性审计规则,进行实时分析和预警,以及时发现各种可疑、违规、违法事件。

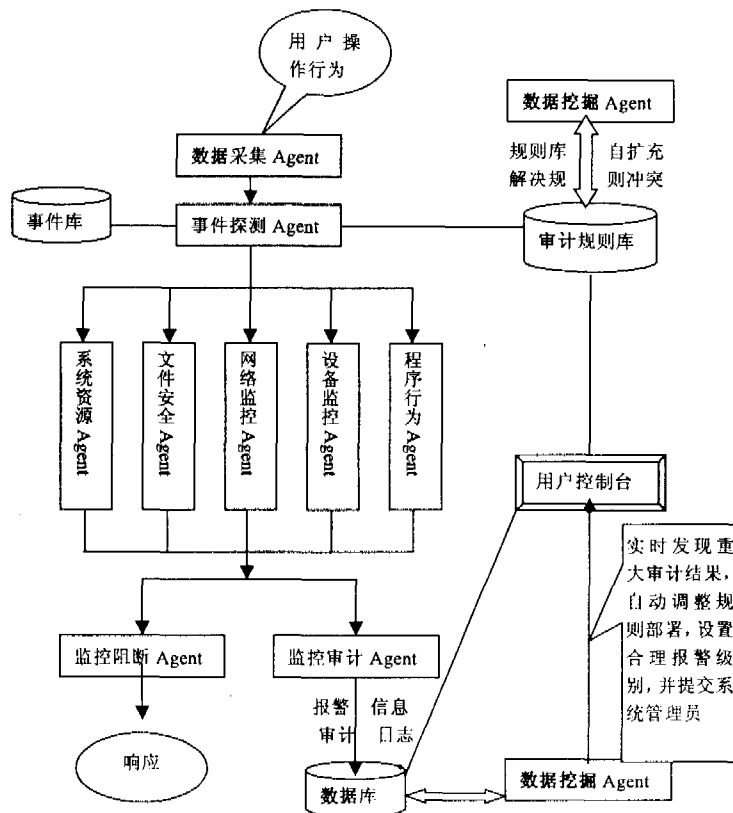


图2 APA分布式信息审计平台框架图

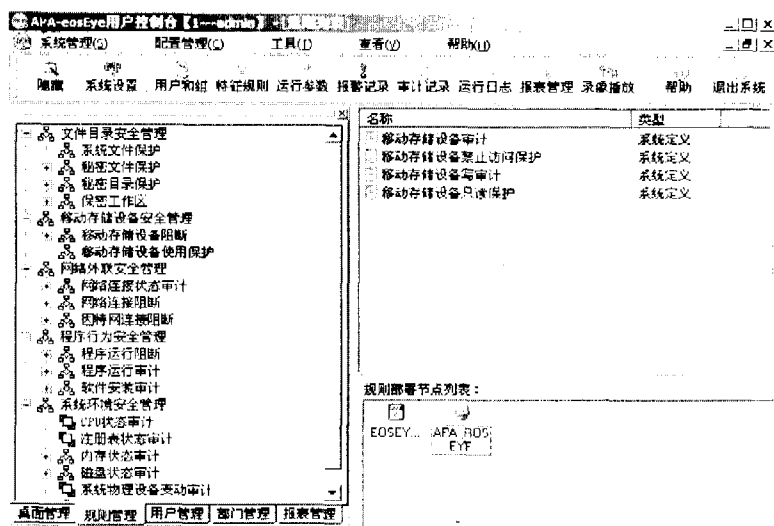


图3 用户控制台管理界面

4 结束语

针对部门信息所面临的内部安全威胁,文中提出了一种安全防范体系结构,采取适当的审计策略,及时纠正和

(下转第 146 页)

RegistrationControl (DeviceObject, WMIREG_ACTION_REGISTER)注册驱动程序成为一个 WMI 的提供者,然后驱动程序要提供一个 DispatchSystemControl 例程来处理 IRP_MJ_SYSTEM_CONTROL 请求。在一定的设备事件发生,而且这个某个注册的事件已经被 WMI 使能的情况下,驱动程序通过 WmiFireEvent 触发事件。对于应用程序来说,在连接到 WMI 之前,必须先初始化 COM,获得一个指向 IWbemLocator 接口的指针。在连接到 WMI 之后,就可以用 IWbemServices 类的 ExecNotificationQuery 方法进行查询,这个函数立即返回,并且返回的一个指向 IEnumWbemClassObject 类的指针,可以利用这个类的 next 方法进行查询,判断事件是否发生。

WMI 触发事件法通常是在标准的驱动程序把标准的事件发送给操作系统组件时使用,通常驱动开发者不会使用到这种方法。当然,用户也可以在驱动程序中使用自定义事件。虽然这种通知方法增加了应用程序代码的复杂性,但是它能够在触发事件的同时,把相应的数据发送给应用程序。

1.5 使用 PNP 通知策略

WDM 驱动程序模型支持 PNP,用户程序可以注册标准的 PNP 事件(例如一个设备的接口被打开),使得特定的 PNP 事件发生时,能够得到通知,当特定的 PNP 事件发生时,这些事件会自动触发,无需驱动程序的干预。然而可以利用这种 PNP 通知策略,在驱动程序里触发自定义的事件。在特定的事件发生时,驱动程序调用 IoReportTargetDeviceChangeAsynchronous() 触发一个事件,用户应用程序需要调用 RegisterDeviceNotification() 注册 PNP 事件,使得当事件发生时,用户的窗口获得一个消息^[5]。

这种方法中,特定的事件发生之后,应用程序得到的是一个发送到窗口的消息,这使得应用程序的处理比较简单,但是要求应用程序至少要有个窗口。应用程序在得到消息的同时,还可以同时得到驱动程序传回来的数据。

(上接第 143 页)

补偿在应用过程中由于各种违规操作所带来的信息泄密等严重后果。在 APA 系统中采用了 Agent 技术和数据挖掘技术,从而使其很好地适应了分布式的用户环境,并实时对审计结果做出响应,减少了部门保密信息受到的持续威胁。目前工作重点是引入移动 Agent,增加 APA 系统的智能性和移动性,以更好适应信息安全审计。

目前该系统已经在中国电子科技集团十四所和成都十所、总参六十所等一些军工企业得到了应用,反应良好。

参考文献:

- [1] Shoham Y. Agent - Oriented Programming[J]. Artificial Intelligence, 1993(60): 51 - 92.
- [2] Lee W, Stolfo S, Mok K. Mining Audit Data to Build Intrusion

通常情况下开发 USB 设备类型的程序的时候使用这种方法比较多。

2 小 结

把设备事件通知给应用程序的方法是设备驱动程序设计中的关键和核心所在。在上述的几种方法中,每一种方法都有其特性和局限,在实际的应用中,要根据自己的要编写的驱动程序的类型和需要通知的事件的特性来选择使用。笔者在开发网络中间层驱动程序时,使用了命名事件法,中间层驱动程序截获网络上发送或者接收的数据,通过命名事件通知给应用程序,之后应用程序在共享的缓冲区里面读取发送或者接收的数据并处理;在开发 PCI 网络安全卡的驱动程序时,使用了挂起 IRP 法,另外因为程序的逻辑要求在注册一个事件之后,等待这个事件发生之前,还要进行发送数据的操作,所以程序中采用了挂起 IRP 中的异步法;在开发某 USB 多媒体键盘驱动程序时,由于要在应用程序中动态监测 USB 键盘的插拔,故采用了 PNP 通知策略法。

参考文献:

- [1] 张惠娟,周利华,翟鸿鸣. Windows 环境下的设备驱动程序设计[M]. 西安:西安电子科技大学出版社,2002. 33 - 37.
- [2] Richter J. Windows 核心编程[M]. 王建华,张焕生,侯丽坤,等译. 北京:机械工业出版社,2000. 190 - 227.
- [3] Oney W. Programming the Microsoft Windows Driver Model, 2nd Edition Ebook[M]. Washington, USA: Microsoft Press, 2003.
- [4] Cant C. Windows WDM 设备驱动程序开发指南[M]. 孙义译. 北京:机械工业出版社,2000. 204 - 219.
- [5] Microsoft Corporation. Microsoft Windows 2000 Driver Development Kit [EB/OL]. <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwinforms/html/wnf-TaskVision.asp>, 2004.

Detection Models[A]. The Fourth International Conference on Knowledge Discovery and Data Mining (KDD'98)[C]. New York, NY: [s. n.], 1998.

- [3] Sang - Jun Han and Sung - Bae Cho. Detecting intrusion with rule - based integration of multiple models[J]. Computers & Security, 2003, 22(7): 613 - 623.
- [4] Bell D E, LaPadula L J. Secure computer systems: Unified exposition and multics interpretation[R]. Mitre Technical Report ESD - TR - 75 - 306[s. l.]: Mitre Corporation, 1976.
- [5] Finin T, McKay D, Fritzson R. An Overview of KQML: A Knowledge Query and Manipulation Language[R]. Technical Report, KSL Laboratories, 1992.
- [6] Finin T, Weber T. Draft Specification of the KQML Agent - Communication Language[EB/OL]. <http://www.csee.um-bc.edu/kqml/kqmlspec/spec.html>, 1993 - 08.