基于决策树的安全审计策略自适应管理控制平台

张 良[1],何 华[2]

1. 中国航空工业集团公司沈阳发动机设计研究所,沈阳 110015
2. 北京启明星辰信息技术股份有限公司,北京 100193

摘要 目前国内众多信息安全企业针对电信运营商市场定制了4A(统一的账号、认证、授权、审计管理)解决方案。其中审计管理是展现4A整体效果,实施综合审计的最有力的功能模块之一,而审计策略的制定则是审计管理最核心的部分。现有审计策略定制方案主要为定制式,不具有通用性、可移植性等特点,且审计策略在制定的过程中,过多的人为因素带来的安全隐患往往是用户所不能接受的。本文所述基于决策树的安全审计策略自适应管理控制平台正是针对4A管理平台之审计管理子系统的业务需求和性能优化等方面存在的问题,结合用于处理海量数据的数据挖掘技术,实现了由系统自动生成审计策略,定期自适应优化审计策略等功能的审计策略通用平台。决策树优化的特殊方法使审计策略在应用过程中可以不断优化,从而满足不同业务系统的审计需求。

关键词 安全审计;数据挖掘;决策树;自适应

中图分类号 TM343 **文献标识码** A **文章编号** 1000-7857(2010)24-0067-04

Self-adapting Security Auditing Management Controller Platform Based on Decision Tree

ZHANG Liang[1], HE Hua[2]

1. *Shenyang Aero Engine Research Institute, Aviation Industry Corporation of China, Shenyang 110015, China*
2. *Beijing Venustech Inc., Beijing 100193, China*

Abstract Currently, quite a few domestic information security enterprises have customized the 4A solution (integrated Accounting, Authentication, Authorization, Auditing management) for the Telecom Operators in China. Auditing management for integrated auditing is one of the most powerful function model which reflects a global effect of the 4A solution. Moreover, the auditing strategy customization is the core of the auditing management. The existing auditing strategy customization scheme is mainly the customization mode, which lacks generality, transportability and other important features. During the process of working out an auditing strategy, many human factors would be involved in security threats, which are not acceptable by the enterprises or corporation. The Self-adapting Security Auditing Management controller platform based on decision tree is a general auditing strategy platform which implements automatically the generation of the audit policy by the system and a self-adapting optimizing auditing strategy periodically. The implementation includes a data mining technology to deal with huge amount of data. The business requirements and performance optimization are dealt with in the integrated auditing subsystem of the integrated security controller platform. The decision tree optimization method enables the auditing strategy being optimized continuously during its operation, to satisfy the auditing requirements of different business systems.

Keywords security audit; data mining; decision tree; self-adapting

收稿日期:2010-09-18;**修回日期**:2010-11-15

作者简介:张良,高级工程师,研究方向为计算机网络安全,电子信箱:dali@yeah.net;何华(通信作者),高级工程师,研究方向为计算机网络安全,电子信箱:hhvito@163.com

0 引言

国内外许多信息安全公司及机构在进行综合安全管理控制平台的研制时,需要实现真正有效的审计功能,由审计策略制定对功能的实现具有不言而喻的重要性。2002 年,中国审计署启动审计信息化系统建设项目(简称“金审工程”)时,许多学者提出要将数据挖掘技术应用到审计中,西南交通大学信息科学与技术学院刘胜国等[1]提出了基于审计与访问控制的授权策略研究,南京航空航天大学信息科学与技术学院王强等[2]提出基于 Agent 和数据挖掘的分布式信息审计平台,中国科学院软件研究所信息安全国家重点实验室何永忠等[3]提出了一种基于多级安全数据库管理系统的通用审计策略模型。

现今,多数审计策略解决方案仍停留在策略人工制定、手动配置的阶段,对于需要处理海量安全事件的用户而言,审计策略的制定非常繁杂。即使最终经过很长的时间,花费很多精力制定出审计策略,也难免出差错;即使策略正确无误,在将其加载到综合安全管理控制平台,使之运转的过程中,也存在规则录入量大、审计逻辑复杂不易实现等问题,并且整个过程过多地依赖于人,势必存在很大安全隐患。针对审计策略人工制定繁杂、工作量大,审计策略录入复杂逻辑不易实现,运行效率较低等问题,本文结合基于机器学习的人工智能决策支持系统(IDSS)框架,设计出综合安全管理控制平台综合审计解决方案——审计策略自适应系统。

1 综合安全管理控制概述

随着网络技术的发展,信息安全越来越受到网络用户的重视。随着各行业业务的发展和 IT 技术的应用,行业内部业务系统数量持续增加,网络规模迅速扩大,安全问题不断出现。每个业务系统分别维护一套用户信息数据,管理本系统内的账号和口令,孤立地以日志形式审计操作者在系统内的操作行为,这些已远不能满足行业客户安全保障需求,因此提出综合安全管理控制解决方案(图 1)。综合安全管理控制解决方案包括统一用户账号管理、统一认证管理、统一授权管理和统一安全审计 4 要素。

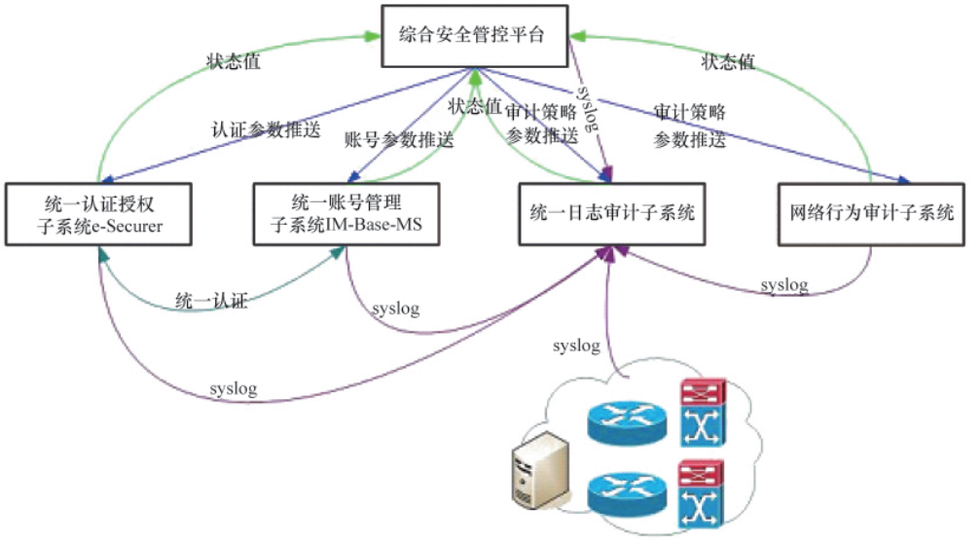

图 1 综合安全管理控制解决方案框图

Fig. 1 Framework diagram of integrated security controller solution

统一安全审计,即综合审计是体现综合安全管理控制平台综合管理价值的重要方面。其主要功能为:① 安全日志采集;② 安全日志多维分析;③ 安全日志实时展现;④ 报表分析;⑤ 审计策略配置;⑥ 数据存储。其对象包括:① 综合安全管理控制平台的参数变更及各告警值通过 syslog 日志的方式传递给综合审计子系统;② 统一认证授权子系统对用户进行统一接入认证后,产生的认证记录通过 syslog 日志的方式发送给综合审计子系统;③ 统一账号管理子系统对用户账号进行维护的操作通过 syslog 日志的方式发送给综合审计子系统;④ 网络审计引擎部件将采集到的日志信息通过 syslog 日志方式发送给综合审计子系统。

如上所述,综合审计子系统的审计对象是来自认证、授权、网络审计及综合管理控制平台的 syslog 日志,其数据量至少可达 3000~4000 条/s,对于这样海量数据的处理,需要引入数据挖掘技术。由于审计的重点是访问控制类审计,最终要实现的是对日志消息进行分类——是否为合法用户的登录或访问,是否为授权用户的合法操作等。

2 决策树分类方法

分类问题数据挖掘的解决方法有很多,如:基于统计学的算法(回归、贝叶斯分类等)、基于距离的算法(简单方法、K 最近邻等)、基于决策树的算法、基于神经网络的算法、支持向量机算法和基于规则的算法等[4]。众多数据挖掘分类方法中,决策树分类方法具有以下特点[4-5]:① 易理解:用决策树的形式表达审计策略,形式简单直观,易被人理解,每一条路径代表一条规则,每一片叶子代表最终的分类结果;② 无参数:

决策树生成算法均不需要对训练集中的数据分布做任何假设或是需要先验知识等,另外,很适合探索型的知识发现;③ 速度快:决策树的建立相比于其他分类方法有明显的性能优势;④ 准确率高:决策树生成分类规则的准确率一般情况下优于其他分类方法。鉴于决策树分类方法的以上特点,适用于审计策略自适应系统。

决策树分类方法是将搜索空间划分为一些矩形区域,然后根据元组落入的区域对元组进行分类。如何划分搜索空间(由测试属性及其值决定),即如何选择属性是区别各算法的关键,值得一提的是,审计策略自适应系统所述的解决方案为用户提供多个候选算法[7],使得用户可以根据不同的实际数据和用户需求选择适当的算法,建立准确的审计策略。

3 审计策略自适应系统

在综合审计子系统顺利采集完整、安全的日志前提下,提出审计策略自适应系统,其主要功能为:① 自主生成审计策略;② 策略生成多算法选择;③ 自适应优化策略。

系统初始化时,根据用户导入的审计资料(如实体权限配置表、文件列表等)自动生成审计策略,并将其应用至综合安全管理控制平台; 定期对审计后的日志进行机器学习,实现审计策略的自适应优化,后推送至综合安全管理控制平台进行日志审计。系统框图如图 2 所示,各系统功能描述如下:

1) 人机交互系统,即综合安全管理控制平台审计策略配置模块,主要功能实现在系统初始化时,将用户的审计数据和用户所期望的审计策略形式(决策树型、规则型)导入。

2) 机器学习系统[8],即决策生成系统,包含以下子系统:① 数据预处理子系统: 对用户导入的需审计数据进行预处理,填充属性值、理顺权限、操作、账号的对应关系等,并将清洗过的数据存储到数据库;② 决策树生成子系统:对已经清洗过的数据进行数据挖掘,通过计算信息增益 \ 增益率 \GINI 系数等确定根节点属性及其各子节点测试属性。③ 问题求解子系统:即算法选择–决策显示模块,主要功能是根据不同审计数据具有的特色到方法库中选择不同的决策生成算法,主要对应的操作人员是本系统的配置人员。

3)审计逻辑,即审计策略应用模块,在整个系统的使用过程中起枢纽作用:① 对由学习系统传过来的审计策略——决策树进行存储、加载及应用;② 对用户修正过的审计策略进行存储、加载、应用,并将所有审计规则按产生时间顺序,添加版本号入知识库。

4) 自适应模块,不同系统的审计重点不同,为了使审计策略更加贴近业务系统需求,引入自适应模块,定期更新训练集,优化决策树。

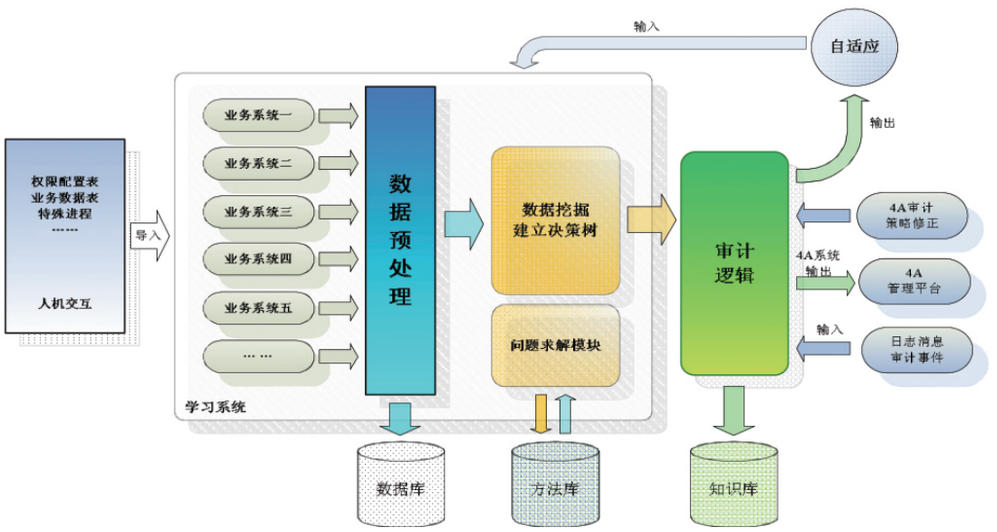

图 2 系统流程框图

Fig. 2 System flowchart

审计策略自适应系统具有以下优势:① 审计策略自主生成。在一定程度上极大地减少了审计策略制定人员的工作量,使综合安全管理控制平台更具人性化、智能化。② 无需规则录入。减少由人为因素而带来的安全隐患。③ 算法多选择。本系统提供多种决策树建立算法,针对不同的训练集选用较为合适的分类算法。④ 系统自适应。审计策略可以通过定期对审计后日志信息进行机器学习而达到审计策略优化,使不同的业务需求其审计策略不同。⑤ 高性能。由于审计策略是以树的结构进行存储,每次进行审计时,按深度或广度优先方式遍历树比逐条遍历以规则集形式存储的策略效率要高很多,以二叉树为例,最坏情况复杂度为 $O(n)$。⑥ 支持人为修正策略。在实际综合安全管理控制应用中,本系统建议采用以机器生成策略为主,人工修正策略为辅。决策树生成的策略很大程度上依赖于训练集的好坏,现实中得不到完美训练集,审计策略一定存在不合理之处,人工修正相当于为审计策略增加了保险。

4 应用

用户先通过综合安全管理控制平台的审计策略配置模块导入需要审计的业务数据、权限配置文件等,后台决策自适应系统通过机器学习自动生成相应的业务审计策略,并将其进行优化后(对生成的决策树进行修剪),输出到审计逻辑,供综合安全管理控制平台进行实际审计应用。

由审计策略自适应系统产生的审计规则在审计策略模块展现给用户,用户可根据实际业务系统的需求对其修正。修正是通过审计策略配置界面,以规则集的形式将需要修正的策略输入系统,加载应用,具体行为模型图如图 3 所示。

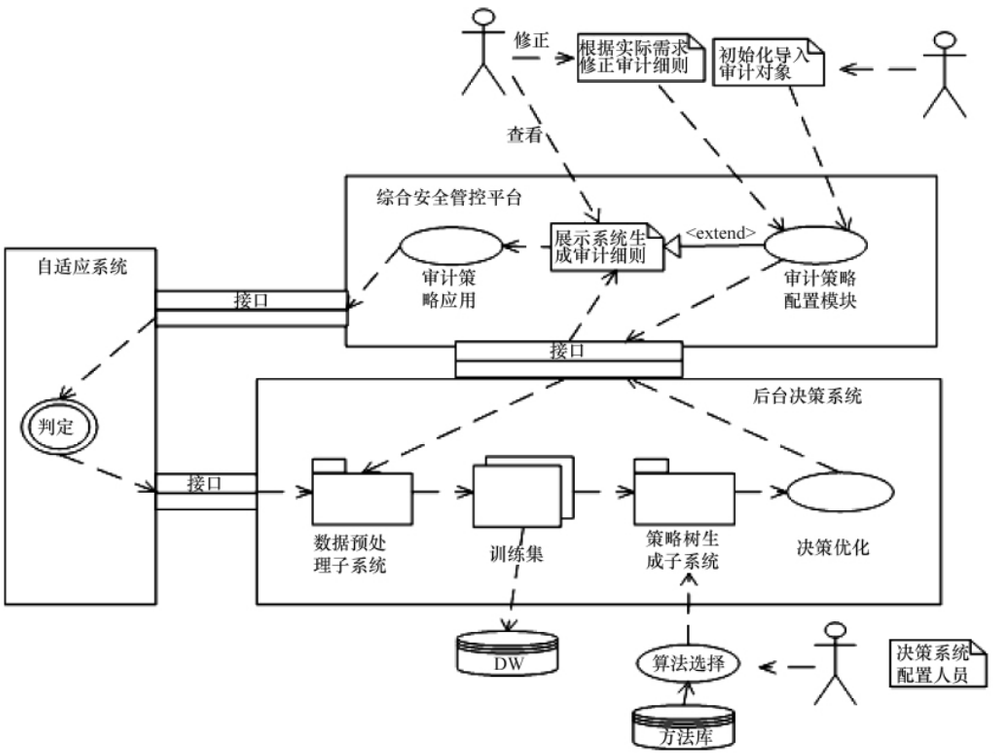

图 3 行为模型图

Fig. 3 Behavior model chart

5 结论

针对数据挖掘技术的应用,在安全审计需求的基础上,设计实现了一种基于决策树的安全审计策略自适应系统。该系统主要为了解决审计模型通用性和审计策略的复杂逻辑实现两方面的难题,并且使得系统自适应机器学习后,制定出的审计策略更贴近用户的需求,大大减少了由用户自行生成审计规则的工作量,使系统更智能化、更具竞争力。

参考文献(References)

[1] 刘胜国, 徐志根, 刘雁林, 等. 基于审计与访问控制的授权策略研究[J]. 计算机工程与设计, 2006, 27(22): 4268-4270.
Liu Shengguo, Xu Zhigen, Liu Yanlin, *et al*. *Computer Engineering and Design*, 2006, 27(22): 4268-4270.

[2] 王强, 皮德常, 李伟奇, 等. 基于 Agent 和数据挖掘的分布式信息审计平台[J]. 计算机技术与发展, 2006, 16(4): 141-143, 146.
Wang Qiang, Pi Dechang, Li Weiqi, *et al*. *Computer Technology and Development*, 2006, 16(4): 141-143, 146.

[3] 何永忠, 李斓, 冯登国. 多级安全 DBMS 的通用审计策略模型[J]. 软件学报, 2005, 16(10): 1774-1783.
He Yongzhong, Li Lan, Feng Dengguo. *Journal of Software*, 2005, 16(10): 1774-1783.

[4] Brodley C E, Utgoff P E. Multivariate decision tree[J]. *Machine Learning*, 1995, 19(1): 45-77.

[5] Breiman L, Friedman J, Olshen R, *et al*. Classification and Regression Trees[M]. London: Chapman & Hall, 1984.

[6] Witten H I, Frank E. Data mining: Practical machine learning tools and techniques[M]. 2nd ed. Amsterdam: Elsevier, 2005.

[7] Dunham M H. 数据挖掘教程[M]. 郭崇慧, 田凤占, 靳晓明, 等译. 北京: 清华大学出版社, 2005.
Dunham M H. Data mining introductory and advanced topics [M]. Guo Chonghui, Tian Fengzhan, Ji Xiaoming, *et al* trans. Bejing: Tsinghua University Press, 2005.

[8] 杨善林, 倪志伟. 机器学习与智能决策支持系统 [M]. 北京: 科学出版社, 2006.
Yang Shanlin, Ni Zhiwei. Machine Learning and Intelligent Decision Support System[M]. Beijing: Science Press, 2006

(责任编辑 刘志远)