

基于动态网络行为可信度量的安全审计

李伟伟, 张 涛, 林为民, 邓 松, 时 坚, 汪 晨

(中国电力科学研究院, 江苏 南京 211100)

摘 要: 本文提出了一种基于动态网络可信度量的安全审计方法。该方法根据网络的实时日志信息, 周期性对其进行数据挖掘生成规则, 并将规则应用到数据流的过滤中。根据数据流与规则的匹配情况进行动态的可信度量, 使得系统对可信行为和危险行为形成不同的安全访问控制机制。较以前以固定规则应对变化的访问控制和过滤, 本文所提出的方法具有很好的适用性和灵活性。

关键词: 数据挖掘; 行为可信; 安全审计; 访问控制

0 引言

网络技术发展迅速, 要求网络安全系统除了具备传统的病毒防御和访问控制之外, 还要对用户行为进行监控、控制和预测, 使其行为和结果是具有可预期性、可知性和可控性即行为可信性^[1-2]。

安全审计是保障行为可信的重要手段。安全审计的主要内容是通过网络或者是主机的行为活动进行记录形成日志, 并对日志进行分析用于预防、调查、分析和事后追责从而确保用户和网络行为的可信。

现有对安全审计系统的研究工作中, 有对不同的领域和不同网络层次的研究^[3-7], 但是都是针对用户日志行为进行数据挖掘之后形成固定规则后直接用于网络过滤和控制。数据挖掘使用的训练数据很难囊括所有正常网络访问行为, 而且将固定的有局限性的规则应用于变化的网络访问过滤和控制, 使得其缺乏灵活性和适用性。

本文提出了一种基于动态网络行为可信度量 (Dynamic Amount of Network Behavior Trust, DANBT) 的安全审计机制, 可以根据系统网络访问的不断变化动态生成规则控制, 并且在规则控制的过程中对网络行为可信进行动态度量, 以此作为网络访问控制的依据。提高了访问控制的灵活性, 实现了动态的行为控制和监管体系, 达到了维护网络安全和事后追责的目的。

本文所做的主要工作主要有两点, 一是提出一种基于DANBT的安全审计方法, 并详细描述其原理。二是设计并实现了一个基于DANBT的安全审计系统。

本文的内容主要分为三个章节, 一是引言, 分析本文的研究背景, 介绍本文的研究内容以及研究成果, 介绍本文的组织结构。二是DANBT, 介绍了DANBT的形式化描述和原理流程。三是基于DANBT的安全审计, 设计并实现了基于DANBT的安全审计系统, 详细描述了各个功能组成模块的设计实现。

1 动态行为可信度量

1.1 形式化描述

DANBT实现的过程涉及到多个控制表的生成和转换。为了进一步描述, 在这里首先给出几个定义。

定义1: 规则表 (Rules Table, R-T), 是从日志服务器中通过数据挖掘分析出来的规则的集合, 即 $R = \{R_1, R_2, \dots, R_n\}$, 其中n为整数, 下同。每个规则条目 R_n 都描述了一种规则的特征。规则表是DANBT实现的基础和依据。

定义2: 临时表 (Cache Table, C-T), 是近期访问数据流信息的结合, 即 $C = \{C_1, C_2, \dots, C_n\}$, 其

中每个信息条目 C_n 都描述了数据流信息的特征和信任值。信任值是数据流与规则表匹配情况的度量体现。

定义3: 可信表(Behavior Trust Table, BT-T),是可以直接信任不需要通过规则匹配判断的数据流信息的集合, 即 $T = \{T_1, T_2, \dots, T_n\}$, 其中每个信息条目 T_n 都描述了数据流信息的特征和时间戳。时间戳是用于删除过期条目的依据。

定义4: 不可信表(Behavior Credible Table, BC-T),是可以直接禁止其通过不需要通过规则匹配判断的数据流信息的集合, 即 $B = \{B_1, B_2, \dots, B_n\}$, 其中每个信息条目 B_n 都描述了数据流信息的特征和时间戳。

1.2 动态行为可信流程

DANBT即针对数据流条目近期的访问情况,对其可信值进行动态调整,对此条目的可信性进行度量,从而对此访问条目以后的访问进行预测和控制,达到保证网络访问安全的目的。

本文的 DANBT 主要通过以下三个方面来实现的:

(1) 周期性生成规则表: 以邻近时间间隔中大多为相似访问和系统中大多数访问为合法访问为依据,定时提取最新时间间隔的日志信息,通过数据挖掘方法生成最新的规则。通过规则下发进程下发到本机中,对本机的数据访问进行控制。其目的是防止以旧的网络规则约束新的网络行为。

(2) 动态生成可信和不可信表: 主要是通过规则表的过滤,对在临时表中条目的可信值进行动态度量来生成可信表和不可信表。以可信阈值为标准,当可信值达到可信阈值时,认为此条目可信加入可信表。对通过可信度量认为可信的网络访问,直接给予一定时间段的直接通行特权。对于不可信表是生成,同理。因为同时间段多为相似报文的访问,所以采用此方式在保证网络可信的情况下提高了访问控制的效率。

(3) 周期性清除过期的可信和不可信表信息: 主要是通过可在可信表和不可信表中增加时间戳,通过表更新进程扫描两个表,对超时过期的条目进行清除。其目的是将可信或者不可信处理给定一个时间段,超过之后需要重新度量,防止可信和不可信的终身制。

流程如图 1 所示。

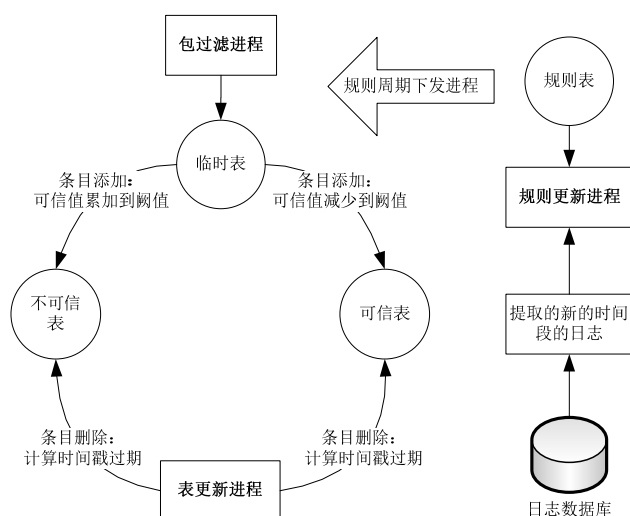


图1 表动态更新流程

2 基于动态行为可行度量的安全审计

2.1 体系结构

行为可信问题主要包括行为信任的评估、预测和控制。即通过对行为是否可信的判断，对未来的行为可信做预测并指导对用户行为进行控制。以此为目标，本文提出了通过数据挖掘方式实现的DANBT的安全审计系统模型。

系统的主要架构如图2所示。

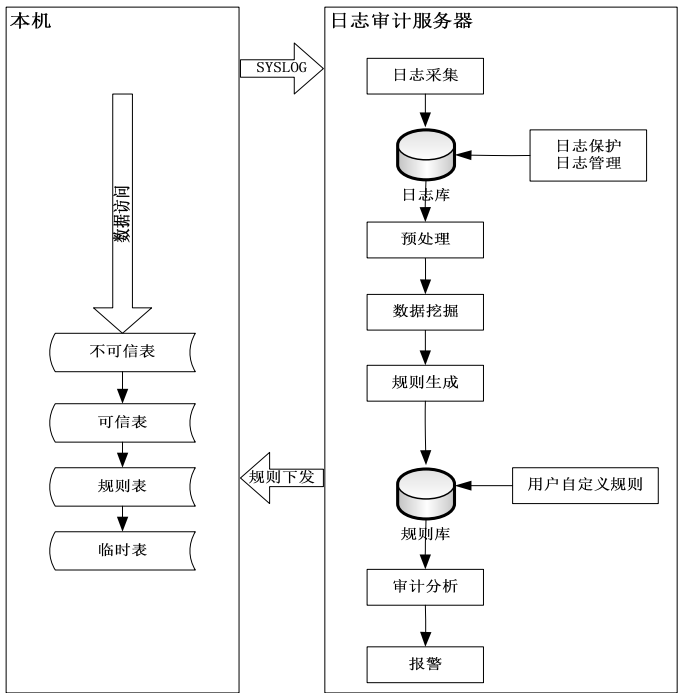


图2 基于DANBT的安全审计系统模型

主要系统流程如下：

Step1: 本机与日志审计服务器通过双方身份认证和安全通信信道建立之后，通过SYSLOG发送日志信息到日志审计服务器，确保日志和日志服务器的安全；

Step2: 服务器采集到日志之后，保存数据库并以大多数行为可信为前提，通过数据挖掘算法周期性从日志系统中发现和更新规则保存至规则库；

Step3: 通过本机和日志服务器建立的通信安全信道，将规则通过规则下发进程下发到本机；

Step4: 本机在网络访问控制和过滤过程中，通过对临时表中保存的数据缓存条目的信任值的动态评估形成以行为可信为依据的可信表和不可信表，从而对本机的网络访问数据流进行基于行为可信的安全访问控制。

2.2 功能组成

2.2.1 日志采集

日志信息就是操作系统或者是应用程序记录所发生的事件并保存这些事件发生的时间、对象、行为等属性，便于日后的分析和追责。日志信息如果单纯保存在本地很容易被篡改或删除。本系统采用了单独的日志数据库来存放日志信息，一方面方便日志的存储，另一方面也加强了日志的安全。日志服务器对日志的存储采用了周期性存储的方法，方便后续数据挖掘周期性生成的规则是基于最新的日志生成的最新的规则。

在本系统中主要采取了两个措施对日志采集的过程安全进行防护，确保日志的真实性：

(1) 对日志发送端和接收端之间的通信采用的是身份认证之后的隧道实时传输技术，实现了对输入

日志本机身份的安全和日志传输过程的安全。

(2) 采用了标准SYSLOG的机制，日志以安全日志传输协议发送到日志服务器中进行存放。规则表也是通过安全通道下发到本机中，提高了日志和规则的安全性和可靠性。

2.2.2 日志预处理

数据预处理^[8,9]是数据挖掘的一个重要过程。因为原始的日志信息中存在着无用的信息和噪声数据，需要在数据挖掘之前通过对日志库的日志做预处理，形成适合做数据挖掘的日志信息。合理高效的数据预处理可以提高数据挖掘的高效性和准确性。主要的预处理手段包括：数据清理，数据集成，数据变换，数据规约。数据清理主要是填写缺失值、消除冗余。数据集成是涉及集成多个数据库、数据立方或文件。数据变换是数据的规范化和聚集。数据规约是数据集的简化表示。本文主要用到了数据清理和数据变换。

网络行为日志属性内容比较简单，对它的处理主要是不完整和冗余数据的清理和IP地址的转换。

(1) 对于缺少字段的条目的预处理采用剔除的方式。在本系统中认为缺少字段的数据会影响挖掘的结果，造成挖掘过程陷入混乱，导致不可靠的输出。采用剔除的方式最方便和快捷的消除这种不可靠的冗余条目。

(2) 对于IP地址的预处理采用的是HASH的处理方式，将32位IP地址的后两个8位做异或得到此字段的值。采用这种方式可以提高后期查找等操作的处理速度，而且通过对这种HASH算法的测试发现，冲突率很低。

2.2.3 关联规则

如何从庞杂的用户历史行为数据中发现用户的不可信问题，是研究用户行为可信的关键。在本系统根据网络中的大多数包为可信正常访问的原则，对经过了预处理的日志采用关联分析数据挖掘^[10-12]生成规则。生成规则的置信度接近百分之百就是我们期望的规则，放入规则库并下发本机。

关联分析^[13-16]就是从给定的数据集中发现频繁出现的项集模式知识（即关联规则），它反映一个事件和其他事件之间依赖或关联的知识。如果两项或多项属性之间存在关联，那么其中一项的属性值就可以依据其他属性值进行预测。Apriori算法是关联分析的常用算法。

Apriori的主要思想是使用一种称作逐层搜索的迭代方法，k项集用于探索(k+1)项集。首先，通过扫描数据库，累积每个项的计数，并收集满足最小支持度的项，找出频繁1项集的集合。该集合记作L1。然后，L1用于找频繁2项集的集合L2，L2用于找L3，如此下去，直到不能再找到频繁k项集。找每个Lk需要一次数据库的全扫描。

2.2.4 规则表生成

规则表是通过对日志信息进行预处理和数据挖掘之后生成的对数据报文进行过滤的规则集，也是生成可信表和不可信表的重要依据。规则表主要包含两个来源，通过这两种规则的来源方式，这样可以使得规则库更加灵活和完善。

(1) 一种方式是通过数据挖掘方式发现的系统隐藏规则。这中规则是通过周期生成下发到本机。

(2) 一种方式是用户随时将发现的问题归纳、总结形成新的规则。这种规则不进行更新将永久有效，直到用户进行删除。

2.2.5 审计表

审计表包含临时表，可信表，不可信表。这三个表是系统实现DANBT安全审计的主要方式。通过规则表对经过的数据报文的过滤、统计、度量、转化和更新，形成基于可信的访问控制。

采用这种方式一方面可以根据最新的日志情况来度量报文的可信度。另一方面在一定时间段会有大量相同地址的报文经过，通过这种可信表和不可信表的转化可以加快报文的处理速度。

2.2.6 用户可视化的界面

用户的可视化界面主要是方便用户对规则库的查看和配置。主要功能点：

(1) 可以通过管理员自身的经验，进行规则库的补充或者删除，但是对可以配置规则的管理员的身

份有严格的认证。

(2) 可以通过数据库的查询，查看日志信息以及规则表的内容。通过对日志信息的查询可以方便用户行为跟踪和事后的追责。

2.3 审计流程

网络访问报文的审计，主要是对表的匹配、对信任值的修改、条目转化的过程。表的查询和修改会对系统的包处理效率有很大的影响。为了提高效率，在本系统的审计流程中对表的操作，我们采用了哈希算法。将数据元素的关键字 K 作为自变量，通过一定的函数关系（称为哈希函数），计算出的值，即为该元素的存储地址。表示为： $Addr = H(key)$ 。通过这种方式可以大大加快表的检索和插入删除操作的速度。

网络访问行为可信的安全审计数据包流程如图3所示。

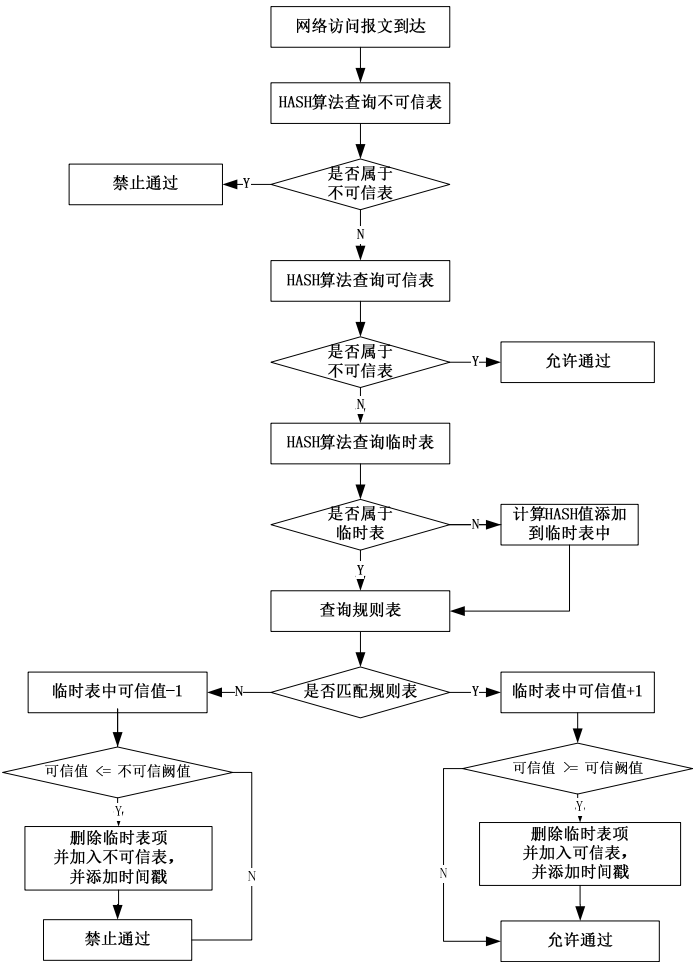


图3 DANBT审计数据包流程

3 结束语

本文通过研究一种基于数据挖掘的安全审计，实现了对行为可信的动态度量模型。通过对面向可信网络的用户行为信任的评估、预测与控制，并通过多种用户行为可信管理机制的有效组合，达到控制的静态与动态、历史与实时的有效统一。

参考文献：

[1] 林闯,田立勤,王元卓.可信网络中用户行为可信的研究[J].计算机研究与发展,2008,45(12):2033-2043.

-
- [2] 江泓,何恩. 行为分析技术及其在可信网络中的应用前景[J].信息安全与通信保密,2009(2):67-69.
- [3] 张蕾,电力企业信息系统的数据安全审计[J].电子信息化,2005,3(9):42-43.
- [4] 杨超峰,刘庆云,刘利军.多层次网络安全强审计模型BrosaAudit[J].计算机工程,2006,32(10):159-160.
- [5] 曾德胜,彭灿明,陈源,等. 基于数据挖掘的审计系统研究[J]. 长春工程学院学报(自然科学版),2011,12(1):124-127.
- [6] 徐诚. 基于日志的网络安全审计系统设计[J]. 软件导刊,2010,9(9):132-133.
- [7] 赵平,汪海航,谭成翔. 基于防火墙日志的网络隔离安全审计系统设计与实现[J].计算机应用研究,2007,24(7):114-116.
- [8] Jiawei Han,Micheline Kamber. 数据挖掘概念与技术[M].范明,孟小峰,译. 北京:机械工业出版社.30-51.
- [9] 王锐,马德涛,陈晨. 数据挖掘技术及其应用现状探析[J]. 电脑应用技术,2007,69:20-23.
- [10] 李丹丹.数据挖掘技术及其发展趋势[J].电脑技术应用,2007(69):38-40.
- [11] 黄力.数据挖掘理论在安全审计分析中的应用[J].微计算机信息,2007,23(93):199-200.
- [12] 王刚,黄丽华,张成洪,等.数据挖掘分类算法研究综述[J].科技导报,2006,24(12):73-77.
- [13] 武建华,沈均毅,王元元. 一种改进的关联分类算法[J]. 计算机工程,2009,35(9):63-66.
- [14] 邓景毅.关联规则数据挖掘综述[J].电脑学习,2006(3):4-5.
- [15] 王义,贾宇波,东兴. 基于关联规则的数据挖掘研究[J]. 工业控制计算机,2011,24(3):86-87.
- [16] 崔国华,侯澄志,洪帆. 审计日志的关联规则挖掘[J].华中科技大学学报(自然科学版),2002,30(9):28-30.