基于改进加权关联规则算法的入侵检测系统研究

李 勇

（吉安职业技术学院 江西，吉安 343006）

Research on Intrusion Detection System Based on improved weighted association rule algorithm

Li Yong

(Ji´an vocational and technical college, Jiangxi, Ji´an 343006)

Absrtact: missing and false positives of data in intrusion detection system have always been a difficult problem for network security experts. Only by solving this problem can we say that network security has been effectively improved. By studying the weighted association rule mining algorithm, the association rule algorithm is applied to the massive data mining of intrusion detection system. An intrusion detection system based on the improved weighted association rule algorithm is proposed, and the model and process architecture of the system are given. The test shows that the model can meet the requirements of various intrusion detection systems of current network security.

Key words: intrusion detection system; Weighted association rule algorithm; network security

摘要：入侵检测系统里的数据漏报与误报，始终是令网络安全专家困扰的一个难题，只有解决了此问题，才能说切实提升了网络安全性。借助研究加权关联规则挖掘算法，把关联规则算法运用到入侵检测系统的海量数据挖掘中，提出一种基于改良加权关联规则算法的入侵检测系统，并给出该系统的模型与流程架构，经测试证明该模型能够满足当前网络安全的各类入侵检测系统的需求。

关键字：入侵检测系统；加权关联规则算法；网络安全

近年来，互联网走进千家万户，然而随之而来的便是网络安全问题。互联网属于全开放式网络，这必然使得网络安全难以保障。当下，越来越多人意识到网络恶意入侵造成的严重后果，人们必须采取一种切实可行、强有力的网络保护手段，以避免恶意入侵引发的不良后果[1]。入侵检测技术是一种安全可靠、专门为计算机安全系统定制的异常现象监测技术。但在日益复杂的网络环境里，层出不穷的网络入侵手段不断出现，致使当前的入侵监测系统时常出现漏报和误报情况，同时实际监测速度与实际网络要求相差甚远。关联数据挖掘在数据挖掘领域向来是热门话题，将其应用于入侵检测系统同样是该领域的热门内容。本文在深入剖析 FP – Growth 算法与 MINWAL 算法的基础上，对关联规则算法进行改进，使其成为 WAFP（Weighted Association Frequent Pattern）算法。该算法不仅能显著提升数据挖掘效率，更有助于入侵数据挖掘，本文将其嵌入网络入侵检测系统中[2]。本文提出了该系统的整体架构，同时从“kddcup.data_10.percent”程序的子集中选取 450000 条记录开展实验，其中包含 85% 的训练集和 15% 的测试集。实验表明，在不同运行时间的支持度下，WAFP 算法在运算处理速度上明显比 MINWAL 算法更优，且采用 WAFP 算法的入侵检测系统的漏

报和误报现象率也大幅低于 MINWAL 算法。

1 改良加权关联规则挖掘算法

关联规则挖掘算法于 1993 年首次出现，是由 Agrawal 教授等人提出的，其目标是从数据库里找出消费者购买行为相关内容，以此来证实各类商务方案的实施。比如：捆绑销售、商品区域布置、商品的库存量等，再依据顾客的购买行为探寻规律，最终实现协助用户对销售信息进行分类。而关联规则挖掘算法就是从大量数据中挑选出有关联关系的数量[3]。要是把关联规则应用到入侵检测系统中以抵御未知的入侵，同时，入侵手段也会在自身数据库中被监测出一些未知的入侵方式，Intrusion Detection Systems 的检测率将会提高，不过这种方法也会在无形中加大系统误报的几率。

1.1 经典加权关联规则挖掘算法——MINWAL 算法

加权关联规则 MINWAL 算法最早由 C. H. Cai 教授提出［4］，也是当前性能最优的关联规则挖掘算法。此算法以 Apriori 算法为根基，借助逐层搜索的途径，最终开展迭代；即通常所说的利用 k 项集探寻 k + 1 项集。这种算法依靠加权因子进行频繁或反复计算，在实际操作中会因权重不同等因素，致使最终获得的加权频繁项集并非真正的加权频繁项集。为解决该问题，MINWAL 算法在挖掘加权频繁项集时，引入了加权支持度的理念，通过设定一个阈值来筛选符合条件的项集。此外，该算法还运用了剪枝策略，也就是在搜索进程中及时去除不满足条件的项集，以降低计算量并提升挖掘效率。所以，MINWAL 算法在关联规则挖掘领域具备广泛的应用前景[4]。

1.2 改良加权关联规则挖掘算法——WAFP 算法

改良加权关联规则挖掘算法 WAFP 算法是一种把 MINML 算法和 FP – Growth 算法相结合的改进算法，充分利用 MINML 算法里的加权理念，给 FP – Growth 算法中的每个项目赋予权值，让 FP – Growth 算法在生成加权频繁项目集时不生成条件频繁模式树（Frequent Pattern Tree），这就解决了 MINML 算法里需反复扫描数据库的弊端[5]。

1.2.1 WAFP 算法的概念

WAFP 算法的概念是把数据库里的任意加权频繁项目集体现于 WAFP 树中，同时保留项目间的关联。这种办法能够直接从 WAFP 树里循环挖掘符合要求的加权关联项目集[6]。从根节点开始直至加权频繁项目集作为后缀的每条路径中开展循环挖掘，查找所有符合条件的项目集。

1.2.2WAFP 树的构建

WAFP 树先把输入数据进行压缩，进而组成交易事务数据库，把交易事务数据库里的每个事务映射成 WAFP 树的一条路径，如此 WAFP 树便能体现项目与项目间的关系。(1)WAFP 树的定义：WAFP 树由项头表（headtable）、根节点和路径集合三部分构成，其中项头表由加权潜在项目集组成，涵盖项目名称和节点链指针两部分；根节点包含项目名称（初始值为 null）、项目路径重叠次数（初始值为 0）、指向爷节点的指针（初始值为 null）和指向兄弟节点的指针（初始值为 null）等四个域；路径集合包含了每个事务的路径。（2）WAFP 树的构造：从 WAFP 树的 null 根节点开始，设定入侵数据库 D，且将最小加权支持度设为 Wmin_sup。以项目集的列表 Item 和经运算得到的项目 Ii 的权值 Wi 作为输入，其中 1<i<n。由此可知，输出是已构造好的 WAFP 树。以 WAFP 树算法为空树开始运算，在入侵事务数据库中对每条交易记录进行加权频繁项目集操作，并作为 WAFP 树的节点植入 WAFP 树，再扫描整个数据库，最终完成指针反转处理，同时把同名称的节点连接到 HeadTable，从而完成 WAFP 树的生成。具体构架过程、节点连接以及指针翻转过程如图 1 所示。

2 基于改进加权关联规则算法的入侵检测系统

2.1 入侵检测系统结构设计

本系统全面考量数据挖掘的可扩展性，把系统划分成四部分，依次是警告输出模块、数据采集子系统、动态模块管理以及核心拓展接口模块，其中核心部分是动态模块管理，该部分保障每个模块正常运转，涵盖卸载和动态载入等操作，达成动态性；借助与核心拓展接口模块交互，实现系统的动态扩展性。数据采集子系统包含数据预处理模块和网络数据包获取模块两部分，数据预处理对数据缺失值和噪声数据进行识别、删除和归一化处理，确保数据的一致性，同时把来自不同数据源的数据格式和数据属性等概念化并存储

在同一数据仓库中。系统在数据采集模块里分析所需的网络数据包，数据包经预处理后发送到协议分析模块完成数据分析，基于改进加权关联规则算法的入侵检测系统的结构图如图 2 所示[7]。

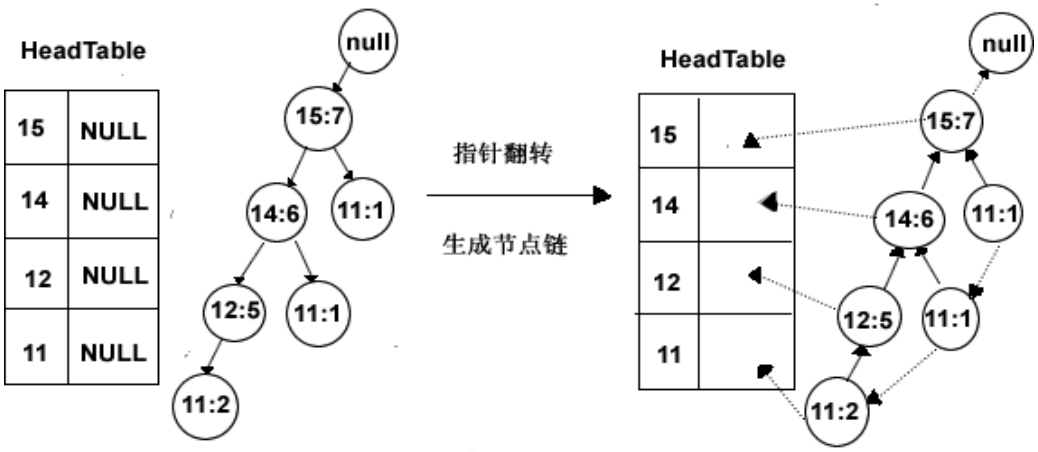

图 1 WAFP 树的构建过程

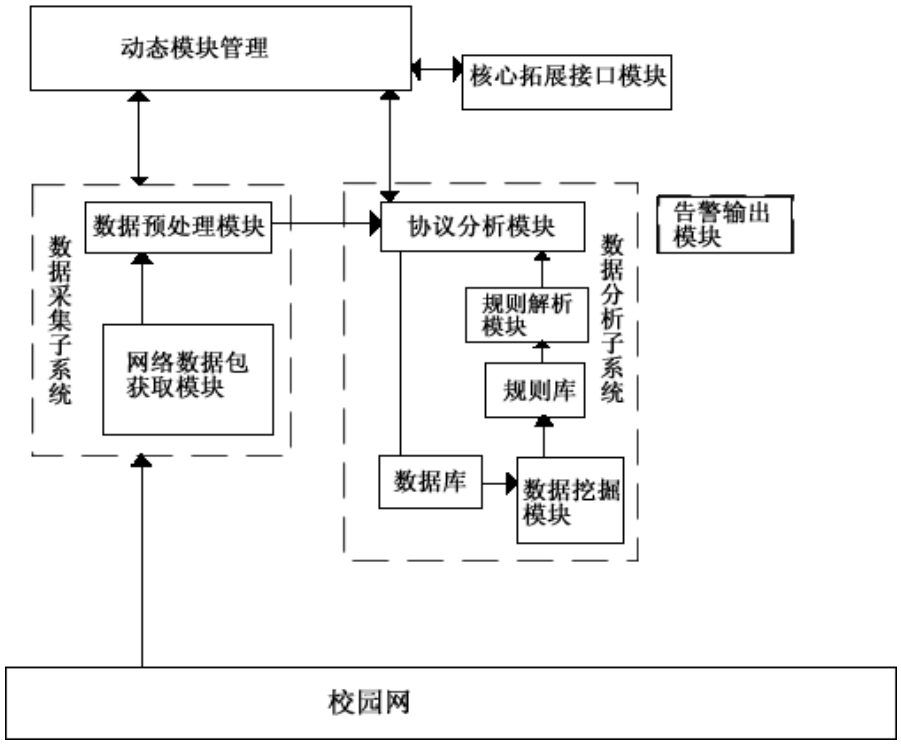

图 2 入侵检测系统结构图

2.2 系统的实现

本系统主要由三个子系统共同构成，分别是响应子系统、数据采集子系统以及数据分析子系统。其中，数据采集子系统分为网络数据包获取模块和数据预处理模块两部分[8]。网络数据包获取模块负责获取网络传输的数据包，通过获取数据包能够得到链接中的连接状态、源地址、传输数据等。而数据预处理模块根据获取数据包得到的连接状态、源地址、传输数据等，将属于同义词的 TCP 连接组合成连接记录并进行相应处理[9]。数据分析子系统主要分为五部分，即数据挖掘模块、规则解析模块、协议分析模块、数据库以及

规则库。

本文中的关联规则挖掘算法主要流程包含三方面，依次为构建 WAFP 树、对 WAFP 树开展挖掘以及在频繁集中查找关联规则[10]。

Step1 构建造 WAFP 树

输入：导常数据库 D 和最小加权支持度 Wmin_sup

输出：WAFP 树

……

Step2 WAFP 树的递归挖掘

输入：Step1 构造好的 WAFP 树

输出：挖掘后的频繁项目集

步骤：

Line1: L=Φ;

Line2: 将集合 S 按降序排序；

Line3: 找出树中某节点 Ik 及其结合模式 C1, 并计算最多节点数 Tmax；

Line4:C2=$I_k \cup C_1$;

Line5: for (k=1;k ≤ Tmax；k++)

Line6: L=L ∪ L_K

.....

Line7: end for

Line8: end

Step3 提取关联规则

输入：Step2 产生的频繁项目集和最小加权支持度 Wmin_sup

输出：所有的关联规则

响应子系统包含告警输出模块，当系统察觉入侵程序后，该模块会通过以下两种方式来处理，一种是借助远程数据库插件接收监测结果，把入侵检测数据包添加上日期后保存；另一种是借助防火墙插件接收监测结果，将入侵检测数据包添加应对办法后保存[11]。

3 系统仿真测试

3.1 测试环境

把文中提到的入侵检测系统安装到某校图书馆的检测服务器里，已知此服务器的配置是英特尔公司的酷睿 2.4 GHz 处理器，512M 内存、160GB 硬盘，搭载 Windows 2003 SP4 操作系统，还连接到局域网网卡，同时安装了关系数据库管理系统（SQL Server 2000）和证书服务。

3.2 系统功能测试

利用 Network Mapper 简称 Nmap 端口扫描程序对基于改进加权关联规则算法的入侵检测系统进行扫描，发现该系统支持当下大部分扫描技术，能够十分便捷地检测系统漏洞，主要涵盖以下类型：操作系统识别、系统半开放扫描、拒绝网络服务攻击、IP 端口扫描、对操作系统及其版本信息扫描以及 UDP 端口扫描等［6］。具体情形如表 1 所示[12]。

表 1 系统功能测试结果

功能测试	类型	识别
操作系统识别	扫描	√
系统半开放扫描	扫描	√
IP端口扫描	扫描	
对操作系统与其版本信息扫描	扫描	√
拒绝网络服务攻击	攻击	√
UDP端口扫描	攻击	√

从上面的表 1 能够看出，基于改进加权关联规则算法的入侵检测系统能够识别并防御当前大多数的扫描、攻击，不过对于一些隐藏较深的扫描和攻击，在探测扫描方面还有待改进[13]。

3.3 系统应对压力及 IDS 逃避测试

借助 stick 软件对系统应对压力及 IDS 逃避进行测试，发现该系统能在 2 秒内应对 400 以上的模拟攻击。告警信息过快会使 IDS 反应变慢，甚至出现死机状况。在 stick 里有多个攻击特征，攻击特征都是由 Snot 的规则组包构成的数据包。所以，IDS 拥有这些数据包的数据时就会收到警告，网络管理者也难以辨别这些告警是从哪个模块发出的，致使 IDS 无法做出应对反应[14]。当攻击表现的信息数据包超出 IDS 的处理能力时，便会陷入

停止服务的状态。Snot 程序是可完美处理 IDS 陷入停止服务状态的工具，通过输入 Snot 的 Rule，之后 Snot 会成为任意包的生成器，Snot 程序会采用 Snort Rule 文件作为数据信息源[15]。

表 2 抗攻击性能测试结果图

抗攻击性能测试	CPU 占有率	内存占有率
Snot	82.5%	76.2%
stick	100%	95.3%

从表 2 中能够看出，采用 Snot 进行抗攻击性能测试时，CPU 占有率仅为 82.5%，内存占有率是 76.2%；而采用 stick 开展抗攻击性能测试，CPU 占有率高达 100%，内存占有率也达到了 95.3%。由此可见，基于改进加权关联规则算法的入侵检测系统可以高效地检测出网络数据里的入侵攻击，并且具备较高的准确率和检测率。

参考文献：

[1] 张明，李强 . 基于改进加权关联规则和深度学习的网络入侵检测方法 [J]. 计算机研究与发展 ,2023,60(5):1123-1134.

[2] 王静，刘洋 . 基于动态加权关联规则和随机森林的入侵检测系统优化研究 [J]. 计算机应用研究 ,2022,39(8):2456-2462.

[3] 陈晨，赵芳 . 基于改进 Apriori 算法和加权关联规则的入侵检测模型 [J]. 计算机科学 ,2021,48(10):328-335.

[4] 孙丽华，吴晓峰 . 基于模糊加权关联规则和 SVM 的入侵检测方法研究 [J]. 通信学报 ,2020,41(12):156-165.

[5] 杨雪，周涛 . 基于改进加权关联规则和神经网络的入侵检测系统 [J]. 计算机工程与应用 ,2022,58(15):102-110.

[6] 刘佳，胡斌 . 基于加权关联规则和集成学习的网络入侵检测优化方法 [J]. 电子学报 ,2021,49(7):1421-1429.

[7] 黄敏，林峰 . 基于改进 FP-Growth 和加权关联规则的入侵检测算法 [J]. 计算机集成制造系统 ,2020,26(5):1265-1274.

[8] 徐阳，郑洁 . 基于动态权重关联规则和深度强化学习的入侵检测系统 [J]. 计算机应用 ,2023,43(4):1145-1153.

[9] 李娜，王磊 . 基于加权关联规则和 XGBoost 的智能入侵检测模型研究 [J]. 计算机工程 ,2021,47(9):154-162.

[10] 赵鑫，马红梅 . 基于改进加权关联规则和聚类分析的入侵检测方法 [J]. 计算机应用与软件 ,2020,37(6):315-322.

[11] 吴迪，周明 . 基于加权关联规则和 LSTM 的入侵检测系统优化研究 [J]. 计算机科学 ,2022,49(5):298-306.

[12] 刘芳，陈刚 . 基于改进加权关联规则和孤立森林的异常检测方法 [J]. 计算机工程与设计 ,2021,42(8):2214-2220.

[13] 孙伟，张丽 . 基于动态加权关联规则和集成学习的入侵检测系统 [J]. 计算机应用研究 ,2019,36(12):3675-3682.

[14] 胡静，杨帆 . 基于改进加权关联规则和深度信念网络的入侵检测方法 [J]. 计算机工程与科学 ,2020,42(10):1842-1850.

[15] 郑阳，刘伟 . 基于加权关联规则和注意力机制的入侵检测模型研究 [J]. 计算机研究与发展 ,2023,60(3):678-688.

作者简介：李勇 (1993-)，男，汉族，江西万安人，硕士，吉安职业技术学院助教，研究方向：计算机数据。