弱关联规则下的联合数据库入侵检测方法研究

王世运
（琼州学院电子信息工程学院，海南 三亚 572022）

摘　要：联合数据库的入侵和普通入侵不同，其无显著的行为特征，入侵数据属性差异较大，很难形成统一的约束规范，导致传统的入侵检测方法，由于通过提取入侵行为特征进行入侵检测，无法有效且准确地完成联合数据库的入侵检测，提出一种弱关联规则下的联合数据库入侵检测方法，通过弱关联模式在联合数据库中支持程度与联合数据库记录总量的比求出弱关联模式的支持度，获取频繁弱关联模式集，采用改进的双置信度算法对频繁弱关联模式集的置信度进行计算，获取弱关联规则，依据弱关联规则，通过原始联合数据库对分类超平面进行计算，采用该超平面完成联合数据库的整体分类，采用主成分分析方法对联合数据库中的操作数据进行降维处理，通过差异分类方法，对联合数据库中的操作数据特征进行分类操作，实现弱关联规则下联合数据库的有效入侵检测。实验表明，所提方法具有很高的准确性及有效性。

关键词：弱关联规则；联合数据；入侵检测

中图分类号：TP122　　**文献标识码**：A　　**文章编号**：1001-7119(2015)03-0184-04

DOI:10.13774/j.cnki.kjtb.2015.03.045

Research of Federated Database Intrusion Detection Methods Based on Weak Association Rules

Wang Shiyun
(College of Electronics and Information Engineering, Qiongzhou University, Sanya Hainan 572022, China)

Abstract: Invasion of different joint database intrusion and ordinary, no significant behavioral characteristics of the intrusion data attributes are different, it is difficult to form a unified constraint specification, lead to the traditional intrusion detection methods, due to the intrusion detection by extract intrusion behavior characteristics, unable to effectively and accurately complete federated database intrusion detection, proposes a weak association rules under the joint database intrusion detection method, through the weak correlation patterns in the federated database support and joint database records of the total than the support of weak correlation patterns, frequent weak correlation patterns were obtained, with the improved algorithm of double degree of confidence the confidence level of frequent weak correlation patterns set calculation, get weak association rules, on the basis of weak association rules, using the original database to the classification hyperplane calculated, using the overall classification hyperplane complete joint database, USES the principal components analysis method for joint operations in the database data dimension reduction processing, through different classification method, classified characteristics of joint operation data in the database operation, implement united under weak association rule database intrusion detection effectively. Experiments show that the proposed method has high accuracy and effectiveness.

Keywords: weak association rules; joint data; intrusion detection

收稿日期：2014-01-22
基金项目：琼州学院校级青年科学基金项目:(QYQN201338)。
作者简介：王世运(1982-)，男，海南省定安县人，琼州学院实验师，主要研究方向:数字媒体开发、网站开发。

0　引言

入侵检测是指对恶意破坏联合数据库的行为进行检测并产生反应的过程,是当前网络安全研究领域的热点课题之一,受到越来越广泛的关注[1,2]。联合数据库的入侵和普通入侵不同,其无显著的行为特征,入侵数据属性差异较大,很难形成统一的约束规范。传统的入侵检测方法,通过提取入侵行为特征进行入侵检测,因此无法有效准确地完成联合数据库的入侵检测[3-5]。

本文提出了一种弱关联规则下的联合数据库入侵检测方法,通过弱关联模式在联合数据库中支持程度与联合数据库记录总量的比求出弱关联模式的支持度,获取频繁弱关联模式集,采用改进的双置信度算法对频繁弱关联模式集的置信度进行计算,获取弱关联规则。依据弱关联规则,通过原始联合数据库对分类超平面进行计算,采用该超平面完成联合数据库的整体分类,采用主成分分析方法对联合数据库中的操作数据进行降维处理,通过差异分类方法,对联合数据库中的操作数据特征进行分类操作,实现弱关联规则下联合数据库的有效入侵检测。实验表明,所提方法具有很高的准确性及有效性。

1　弱关联规则的挖掘

弱关联规则的挖掘算法可利用下述两个步骤实现,首先求出支持度,从而获取频繁弱关联模式集,然后对频繁弱关联模式集的置信度进行计算,最终达到获取弱关联规则的目的。

1.1　支持度

支持度即给定模式在给定数据库中出现的频率。传统弱关联规则挖掘算法中的模式为数据集中属性的集合。而在对联合数据库中数据的弱关联规则进行挖掘的过程中,模式并非属性的集合,而是与属性相应的弱关联值的集合,即弱关联模式。弱关联模式的支持度可通过该模式在联合数据库中支持程度与联合数据库记录总量的比进行描述,如式(1)所示:

$$Support=\frac{sum\ of\ s_{ti}\ \ associated\ \ with\ one\ \ itemset}{number\ \ of\ \ records\ \ in\ \ T} \tag{1}$$

其中,s_{ti}用于描述联合数据库中所有记录对弱关联模式的支持程度。公式描述如式(2)所示:

$$s_{ti}=\min(f_1,f_2,\cdots,f_k) \tag{2}$$

其中,f_k用于描述弱关联模式中与第k个值相应的隶属度。

1.2　置信度

通过上述分析求出支持度后,即可获取频繁弱关联模式集。给出该集合中的一个频繁弱关联模式$\langle Z,C\rangle$,其中,Z用于描述弱关联变量的集合,C用于描述弱关联值的集合。通过该频繁弱关联模式,即可获取下述弱关联规则:*If X is A THEN Y is B*。其中,X与Y表示弱关联变量集合,$X\subset C,B=C-A$。*X is A*是规则前件,*Y is B*是规则后件。若上述规则的置信度超过给定的置信度阈值,则该规则即为所要挖掘的弱关联规则。下面,依据传统置信度算法,在分析传统算法的基础上,提出一种改进的双置信度算法,使得挖掘出的弱关联规则更加有效。

经典的弱关联规则挖掘算法通过式(3)对规则的置信度进行计算:

$$Confidence=\frac{Support\ \ of\ \langle Z,C\rangle}{Support\ \ of\ \langle X,A\rangle} \tag{3}$$

其中,$Support\ of\ \langle Z,C\rangle$用于描述与弱关联规则相应的频繁模式在整个联合数据库中的支持度,$Support\ of\ \langle X,A\rangle$用于描述规则前件在联合数据库中的支持度,其可利用前文所述的支持度算法获取计算结果。

分析式(3)可知,该算法只分析规则前件的支持度,未分析规则后件的支持度。采用上述规则一定会造成漏分现象。对于解决联合数据库入侵检测问题,一定会引起入侵检测的漏报率升高,因此,提出一种改进方法。

为了充分分析规则前件与规则后件的支持度,并且防止规则前件与规则后件在同一公式中对彼此造成干扰,本文通过两个公式分别对规则前件与规则后件的支持度进行计算,依据给定的两个置信度阈值对整个弱关联规则的有效性与完备性进行判断。规则前件与规则后件的置信度可分别通过式(4)、式(5)求出:

$$Confidence=\frac{Support\ \ of\ \langle Z,C\rangle}{Support\ \ of\ \langle X,A\rangle} \tag{4}$$

$$Confidence=\frac{Support\ \ of\ \langle Z,C\rangle}{Support\ \ of\ \langle Y,B\rangle} \tag{5}$$

通过式(4)求出的规则前件的置信度,被称作第一置信度。其主要负责弱关联规则有效性的判

断。如果一个弱关联规则的第一置信度超过该置信度阈值,则该规则被判断为有效;否则无效。通过式(5)求出的规则后件的置信度,被称作第二置信度。其主要负责弱关联规则完备性的判断。如果一个规则的第二置信度超过该置信度阈值,则该规则被判断为完备;否则不完备。

对于下文分析的联合数据库的入侵检测,双置信度与检验入侵检测效果的两个重要指标相对应:检测率与误报率。在弱关联规则的两个置信度均较高的情况下,规则为有效且完备的,由此提出的联合数据挖掘算法才具有较高的检测率与较低的误报率。

2 联合数据库入侵检测方法

在上述分析的基础上,对联合数据库入侵进行检测。分类是一种在统计学的基础上对数据进行分析的方法,考虑到联合数据库的特征,可引入一种联合数据库分类的思想,完成联合数据库的入侵检测。通过原始联合数据库对分类超平面进行计算,采用该超平面完成联合数据库的整体分类。假设联合数据库中操作数据构成的数据集合用 z_j 进行描述,和其相对应的权值系数用 b_j 进行描述,则联合数据库中的操作数据需满足式(6)所示的要求:

$$\sum_{j=1}^{p} z_j b_j = 1,\ b_j > 0 \tag{6}$$

通过式(7)可求出联合数据库中操作数据特征的极大值:

$$\sum_{j=1}^{p} b_j - \frac{1}{2}\sum_{j,k=1}^{p} b_j b_k z_j z_k l(y,y_j) \tag{7}$$

式中,$l(y,y_j)$ 为核函数。

通过式(8)可完成联合数据库中操作数据的二次规划:

$$z(y) = \mathrm{sign}(\sum_{j=1}^{p} \beta_j z_j l(y_i,y) + c) \tag{8}$$

通过式(9)即可求出联合数据库中操作数据二次规划的对偶规划:

$$\begin{cases} \max \sum_{k=1}^{m} \beta_k - \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} z_j z_k \beta_j \beta_k (y_j \cdot y_k) \\ s.t.\ \sum_{k=1}^{m} z_k \beta_k = 0 \\ 0 \leqslant \beta_k \leqslant v(y_k) D,\ k = 1,2,\cdots,m \end{cases} \tag{9}$$

通过上述分析的方法,即可将最优超平面问题变成求解二次规划的对偶规划问题。该规划的最优解可通过式(10)求出:

$$\beta^* = (\beta_1^*, \beta_2^*, \cdots, \beta_m^*)^{\mathrm{T}} \tag{10}$$

式(11)描述的是最优分类函数:

$$g(y) = \mathrm{sgn}\{(x^* \cdot y) + c^*\} \tag{11}$$

式(11)需满足式(12)描述的约束条件:

$$\begin{aligned} x^* &= \sum_{k=1}^{m} \beta_k^* z_k y_k \\ b^* &= y_i - \sum_{j=1}^{l} y_j \alpha_j (x_j \cdot x_i) \\ j &\in \{ j \mid 0 < \beta_j^* < v(y_i) D \} \end{aligned} \tag{12}$$

通过式(13)即可得最优分类函数:

$$g(y) = \mathrm{sgn}\left\{\sum_{k=1}^{m} \beta_k^* z_k L(y,y_k) + c^*\right\} \tag{13}$$

其中,$c^* = z_j - \sum_{k=1}^{m} z_k \beta_k L(y_k, y_j)$,$j \in \{ j \mid 0 < \beta_j^* < v(y_i) D \}$。

通过上述分析的方法,依据主成分分析方法的基本思想,对联合数据库中的操作数据进行降维处理,使得冗余数据大大降低。引入差异分类方法,对联合数据库中的操作数据特征进行分类操作,从而有效实现联合数据库的入侵检测。

3 仿真结果

为了验证本文方法的有效性,需要进行相关的实验分析。实验在Visual C ++6.0环境下进行编程。入侵检测数据集共包含四种攻击类型:拒绝服务攻击(DoS),远程到本地攻击(R2L),普通用户到超级用户攻击(U2R)以及扫描攻击(Probe)。入侵检测的主要目的是对攻击类型进行分类标识,从而采取有效措施降低入侵造成的危害。

联合数据库入侵检测正确率可通过下式进行描述:

$$\text{检测准确率} = \frac{\text{准确检测出的异常数据量}}{\text{全部异常数据量}} \tag{14}$$

表1描述的是采用本文方法和传统方法在2种测试数据集上的检测准确率结果。分析表1可知,本文方法的检测准确率明显高于传统方法,这是因为本文方法采用主成分分析法对操作数

据进行降维后,有效避免了高维数据造成的“维数灾难”问题,增强了本文方法的入侵检测准确率及计算效率,验证了本文方法的准确性。

表1 两种方法的入侵检测正确率比较

Table 1 The intrusion detection accuracy compared two methods

攻击类型	本文方法/%	传统方法/%
DoS	99.98	97.32
R2L	87.54	59.87
U2R	75.39	62.59
Probe	99.42	72.73

分别采用本文方法和传统方法对联合数据库的检测率、漏报率以及假报率进行统计与比较。结果如图1所示。

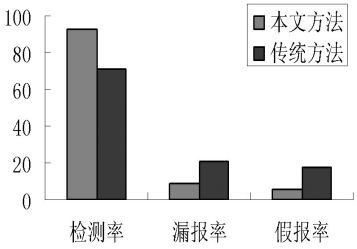

图1 两种方法几项评价指标比较结果

Fig.1 Several indexes comparison results of the two methods

分析图1可以看出,采用本文方法获取的检测率明显高于传统方法,而漏报率与假报率明显低于传统方法,验证了本文方法的有效性。

4 结论

本文提出了一种弱关联规则下的联合数据库入侵检测方法,通过弱关联模式在联合数据库中支持程度与联合数据库记录总量的比求出弱关联模式的支持度,获取频繁弱关联模式集,采用改进的双置信度算法对频繁弱关联模式集的置信度进行计算,获取弱关联规则,依据弱关联规则,通过原始联合数据库对分类超平面进行计算,采用该超平面完成联合数据库的整体分类,采用主成分分析方法对联合数据库中的操作数据进行降维处理,通过差异分类方法,对联合数据库中的操作数据特征进行分类操作,实现弱关联规则下联合数据库的有效入侵检测。实验表明,所提方法具有很高的准确性及有效性。

参考文献:

[1] 郭虎升,王文剑.基于神经网络的支持向量类和支持向量机的入侵检测研究[J].计算机仿真,2008,25:130-132.

[2] 张新有,曾华燊,贾磊.入侵检测数据集KDD CUP99研究[J].计算机工程与设计,2010,31(22):4809-4816.

[3] 陈莉,焦李成.基于关系代数的关联规则挖掘算法[J].西北大学学报(自然科学版),2005,35(6):692-694.

[4] 祝万涛,欧阳为民,辛洪亮.基于数据挖掘的自适应异常分析[J].计算机工程与设计,2007,28(2):264-266.

[5] 陶树平.关联规则和分类规则挖掘算法的改进与实现[J].计算机工程,2003,29:186-187.

(上接第173页)

[3] 任丰原,黄海宁,林闯. 无线传感器网络[J].软件学报,2003,14(7):1282-1291.

[4] 梁宗保,李鹏.基于Zigbee技术的无线传感器网络网关设计与实现[J]. 计算机与现代化,2013,6:133-138.

[5] 杨诚,聂章龙. Zigbee网络层协议的分析与设计[J]. 计算机应用与软件,2009,26(12):203-206.

[6] 张朋,陈明,陈亚萍. 无线传感器网络操作系统关键技术[J]. 计算机应用研究,2007,10:24-25.

[7] 曹莉,曾黄麟,乐英高.基于Zigbee和MSP430无线温度控制系统设计[J]. 四川理工学院学报:自然科学版,2012,25(1):52-55.