入侵检测中贝叶斯分类器改进的研究

高志森,张 铮,李 俊
(南京航空航天大学 信息科学与技术学院,江苏 南京 210016)

摘 要:介绍了一个改进的贝叶斯分类器,其中利用了滑动窗口技术改善入侵检测的实时性能和可控制性能。同时在入侵检测的结构中引入一个性能调节器,它可以动态调整系统参数,提高系统的运行性能,使系统成为一个自动的、有意识的安全系统。

关键词:入侵检测;贝叶斯;分类器;滑动窗口;性能调节器
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2006)11-0154-02

Improved Bayesian Classifier of Intrusion Detection

GAO Zhi-sen, ZHANG Zheng, LI Jun
(College of Information Sci. & Techn., Nanjing Univ. of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Introduces an improved Bayesian classifier, which uses the skipping window technology to reform the reaction time and facilitate the control of the intrusion detection system. Furthermore, adopt a performance adjuster in the IDS, which can dynamically adjust the system parameters to reform the running performance of the IDS, that make the IDS to be an automatic, conscious security system.

Key words: intrusion detection; Bayesian; classifier; skipping window; performance adjuster

0 引 言

在入侵检测方法研究过程中,为了降低误报率、提高对新攻击的检测能力,学者们引入了其他学科的各种技术,如机器学习中的神经网络、遗传算法,数据挖掘中的分类、聚类、序列分析等[1]。其中,贝叶斯(Bayesian)分类方法[2]也被应用于入侵检测中。由于它的简便性、有效性,基于贝叶斯方法的入侵检测系统[3,4]可以建立良好的用户模型,正确区分正常用户和异常用户,提高了IDS的检测率,减少误报。

文中所述的入侵检测系统也采用贝叶斯方法构造入侵检测模块-贝叶斯分类器。但提出了一种改进的贝叶斯分类器,它在实现、设计中引入了一种滑动窗口技术,使得分类器具有实时检测能力和可控性能。在入侵检测系统的结构设计中,提出增加一个性能调节器,它可以在系统运行时,动态地调节运行参数,大大地改善了检测系统的性能,使得入侵检测系统富有主动性、自适应性。

1 贝叶斯分类器

贝叶斯分类器,它是用于分类工作的贝叶斯网络。它把检测对象的行为进行分类,抽取每个对象中反映入侵特征的变量,组成一个矢量,根据这些矢量进行分类,对象分类成正常、一般异常、严重异常、警告等。

一般是把一个待分类的对象事例 O 用一个属性向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 表示,X_a 表示 O 的第 a 个属性,V_a 表示事例 O 在 X_a 上的属性值,$P(x)$ 表示 x 的概率,$P(y|x)$ 表示 x 条件下 y 的概率,n 表示属性的个数,C_i 表示第 i 个类,则事例 x 属于 C_i 类的概率为公式(1)[5]。计算出 x 属于各个类的概率后,再将 x 归入概率最高的类。

$$p(c_i|x)=\frac{p(x|c_i)p(c_i)}{p(x)}=\frac{p(c_i)\prod_{k=1}^{n}p(x_k=v_k|c_i)}{p(x)} \quad (1)$$

在该入侵检测系统,利用贝叶斯分类器作为入侵检测部件,初步设置3个类别,C_1="正常",C_2="异常",C_3="可疑"。为了反映入侵特征而选择的一组特征变量[6] $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 是SYN错误连接百分比、有REJ错误的连接所占百分比、同一端口服务连接所占百分比、不同服务所占百分比等。所以系统中一共有3个判别函数,它们实时提取统计参数,构成特征向量 $x=(x_1,x_2,\cdots,x_n)$,根据贝叶斯公式(1)计算出类别的后验概率值,再从中选择对应于判别函数为最大值的类作为决策结果。图1[3]表示了贝叶斯分类器的判别过程。

2 改进的贝叶斯分类器

2.1 实时检测和滞后检测

一般,在用贝叶斯网络对数据进行分类时,需要进行一段时间 T 的数据积累,当内存中的数据足量时,才能在

收稿日期:2006-02-26
基金项目:国防科工委国防基础科研项目(S0500B003)
作者简介:高志森(1982-),男,浙江泰顺人,硕士研究生,主要研究方向为计算机网络、网络安全;李 俊,教授,硕士生导师,主要研究方向为计算机网络、网络安全。

积累的数据集上进行统计,再经过统计、属性值映射、特征变量获得、贝叶斯公式计算,最后由判决器判决类别。因此,贝叶斯网络分类器是一个滞后的检测器而不是一个实时的检测器。

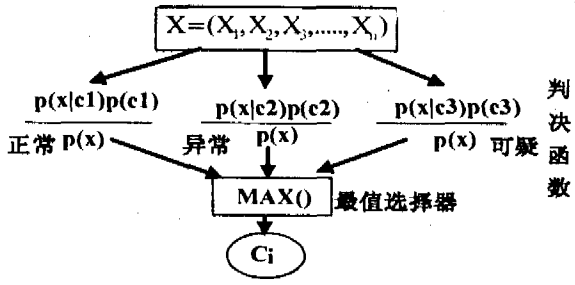

图 1　贝叶斯分类器

由于用户对 IDS 在实时检测方面有较高要求,所以需要对贝叶斯分类器进行改善,使它可以满足实时检测的要求。在系统中,采用了网络数据报传输中用来提高传输速度的滑动窗口方法,把数据收集器捕获的数据包看作是网络中需要传送的数据报,按照捕获的时间先后排成一个队列 list = ($d_1, d_2, \cdots$);之后,选择一个合适的窗口大小 n,窗口在队列上由前向后滑动,滑动的步宽为 t;每次在窗口中的 n 个数据作为一个贝叶斯分类器的测试集,对该测试集上进行贝叶斯分类的一系列操作,得到最后的判别类别。滑动窗口数据集如图 2 所示。

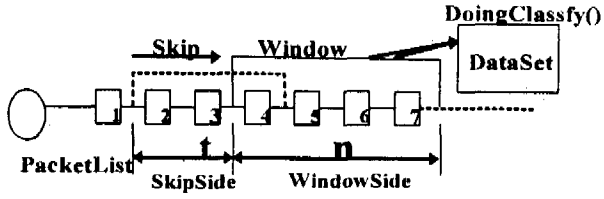

图 2　滑动窗口数据集

采用这种方法,分析的数据不再是滞后 T 时间的数据,而是可以主动控制的。当滞后时间太长了,只要适当地减少步长 t 就可以了。当 $t = 1$ 时,系统实时检测每一个数据包,没有滞后性。而且 t 和 n 可以根据分类器的运行状态进行调整。当系统负担过重、IDS 丢包率增大、IDS 反应迟缓时,可以适当增大步长 t(即适当降低实时性要求)和减少窗口 n(即减少检测数据),减轻系统负担,从而改善分类器的运行性能。

2.2　自适应性

一般的入侵检测系统需要设置一些参数,然而这些参数一般都是固定设置,这使得 IDS 缺乏自适应性和主动性,缺乏根据环境改变而自动调整的能力。我们的贝叶斯分类器中,也有几个参数需要设置,包括窗口大小 n、滑动步宽 t,这两个参数的设置对分类器的准确性和效率有重大的影响,然而它们又和网络的流量、丢包率、CPU 占用率等有关。因此,为了增强本 IDS 的自适应性,抵抗 DOS 攻击能力,在分类器中设置了一个性能调节器,它可以使 IDS 根据网络环境的变化和各部件的状态自动调整参数。

调节器由 3 个部分组成:信息收集模块(Message Collector)、性能调整模块(Performance Adjuster)、参数设置模块(Parameter Setter)。信息收集模块是一个文件收集服务器,它等待 IDS 部件的文件传输请求,接受部件发送的 XML 文件,而 XML 文件记录了该部件的标号和部件的运行性能信息。性能调整模块可以根据部件标号选择不同的性能调整算法,算法中实现了调整策略;用调整算法根据 XML 文件中的运行信息重新计算部件的参数。参数设置模块读取性能调整算法的结果,生成相应的参数设置 XML 文件,XML 文件中标明各种部件参数的新的值,重新传给各部件。贝叶斯分类器的结构图如图 3 所示。

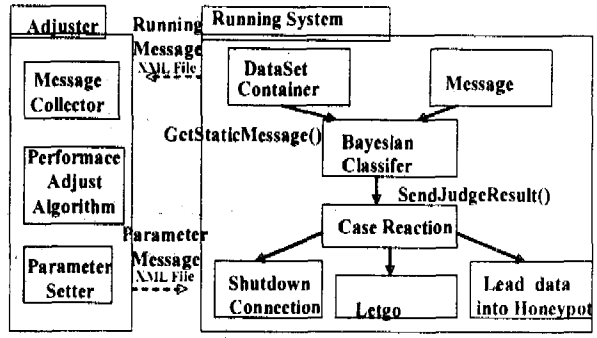

图 3　拥有调节器的贝叶斯分类器结构图

IDS 许多部件的参数都需要不断地调整才能适应环境的改变,保持良好的检测性能。该性能调节器可以为这些部件提供参数调节服务,只要提供实现了调整策略的算法和运行信息。在添加了性能调节器后的贝叶斯分类器,更具扩展性、自适应性、互动性,它成为一个有意识的检测器。

3　实　验

设计了一个简单的贝叶斯分类器,用它来对网络的攻击进行检测。分类器采用的特征属性有 10 个:端口、服务类型、会话持续时间、网络流量、TCP 堆栈的状态、REJ 标志、SYN 错误连接百分比、口令失败次数、root shell 的频度、su 命令次数。统计了每种攻击的平均检测率,实验表明,试验中的贝叶斯分类器检测效果可达到:扫描攻击为 60%,拒绝服务攻击为 20%,未授权访问为 55%,综合攻击为 60%,但没达到非常满意的程度。在事后的分析中,笔者认为是由于使用的特征属性不足,不能充分体现攻击的特征,应该适当地增加特征属性数目。

4　结束语

如何构造入侵检测的分类器,仍然是目前网络安全领域研究的一个重要方向。利用滑动窗口技术,设计改进的贝叶斯分类器,使得入侵检测更具有实时性和可控制性。同时在系统中增加了一个动态性能调节器,使得入侵检测系统可以主动地适应环境的变化,有效地避免了人工的参与。试验表明入侵检测系统在检测的实时性、系统的运行性能方面都有明显的改善和提高。鉴于其他数据挖掘算

(下转第 178 页)

一种特殊的形式或一套表达方式，如关联规则、分类规则或分类树、回归结构和聚类集等。

(3)数据挖掘、结果分析表述和挖掘应用。

此阶段运用使用兴趣度度量，并与数据挖掘模块交互，以便将搜索聚焦在有趣的模式上。它可能使用兴趣度阈值过滤发现的模式。运用统计学和关联规则等方法，把挖掘分析的结果放入一个个性化数据库，当学习者下次进入系统时，系统就可根据个性化数据库提供给其符合学习需求的页面。

4 结束语

网络教学平台的关键是针对用户的个性特征信息，通过系统的分析和判断，给予不同的学习环境和学习内容的呈现，通过运用数据挖掘技术可以从用户数据库及用户学习行为记录中挖掘出用户对知识点的理解程度，从而实现在学习过程中对用户的学习进行记录、指导、反馈，对用户选择的学习策略给予支持，大幅提高《大学物理》网络教学平台的教学效果，使个性化教学真正得以实现。

参考文献：

[1] 林君芬，余胜泉. 关于我国网络课程现状与问题的思考[J]. 教育技术通讯，2001(1)：55－59.

[2] 梁林梅，焦建利. 我国网络课程现状的调查分析与反思[J]. 开放教育研究，2002(6)：13－16.

[3] 刘 莉. 远程学习者研究现状及发展趋势——远程教育专家访谈录[J]. 中国远程教育，2003(5)：7－12.

[4] 黄 萍. 高校学生网络自主学习行为的调查研究[J]. 开放教育研究，2004(6)：77－80.

[5] 舒 蓓，申瑞民，王加俊. 个性化的远程学习模型[J]. 计算机工程与应用，2001(9)：90－92.

[6] 康晓东. 基于数据仓库的数据挖掘技术[M]. 北京：机械工业出版社，2004.

(上接第 155 页)

法的特有功能，今后可以继续把新的数据挖掘算法引入到入侵检测中，改善检测的准确性、可靠性。

参考文献：

[1] Lee Wenke, Stolfo S J, Mok K W. A Data Mining Framework for Building Intrusion Detection Models[C]//Proceedings of the 1999 IEEE Symposium on Security and Privacy. Los Alamos, CA: IEEE Computer Society Press, 1999: 120－132.

[2] Wong M L, Leung K S. An Efficient Data Mining Method for Learning Bayesian Networks Using an Evolutionary Algorithm－Based Hybrid Approach[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(4): 378－404.

[3] 张 琨，徐永红，王 珩，等. 用于入侵检测的贝叶斯网络[J]. 小型微型计算机系统，2003，24(5)：913－915.

[4] Kruegel C, Mutz D, Robertson W, et al. Bayesian Event Classification for Intrusion Detection[C]//Proceedings 19th Annual Computer Security Applications Conference. Los Alamos, CA: IEEE Computer Society Press, 2003: 14－23.

[5] 白耀辉，陈 明，王举群. 利用朴素贝叶斯方法实现异常检测[J]. 计算机工程与应用，2005(34)：131－132.

[6] 牛建强，曹元大，阎 惠. 基于数据挖掘的 CIDF 协同交换[J]. 计算机工程，2003，29(14)：35－36.

(上接第 174 页)

代的地位与影响。信息技术极大地推动了图书馆的现代化进程，同时也带来了信息的爆炸式增长。在知识经济时代，解决好海量信息的存储、检索、开发与利用，是关系到图书馆未来的生存与发展的重大问题。因此必须制定国家数字图书馆发展战略，做好整体建设规划。同时还要制定和完善数字图书馆相关的保密、版权等方面的法律。

7 结束语

数据挖掘技术及其应用是目前国际上的一个研究热点，在数字图书馆领域，面对大量的信息，用数据挖掘技术找出数字图书馆用户感兴趣的信息加以组织利用，加强客户关系的管理，提高满意度。基于数字图书馆建设与应用数据挖掘的有利时期已经到来，优质的网络信息服务事业前景不可限量。综合应用数据挖掘技术和人工智能技术，获取用户知识、文献知识等各类知识，将是实现知识检索和知识管理发展的必经之路。

参考文献：

[1] Poe V. Building a Data Warehouse for Decision Support[M]. [s.l.]: Prentice PTR, Prentice－Hall Inc, 1996.

[2] 赵洗尘. 数字图书馆及其建设[J]. 现代图书情报技术，1999(1)：28－31.

[3] 刘 霞. 关于数字图书馆建设的几个问题[J]. 图书情报知识，1998(1)：30－32.

[4] 王 珊. 数据仓库技术和联机分析处理[M]. 北京：科学出版社，1998.

[5] 刘海虹，刘伯萤. 数据挖掘技术[J]. 丹东纺专学报，2001(1)：15－18.

[6] 郝先臣. 数据挖掘工具和应用中的问题[J]. 东北大学学报：自然科学版，2001(2)：183－187.

[7] 卢增祥. Bookmark－智能化网络信息服务系统[J]. 高技术通讯，1999(6)：30－32.

[8] 郑巧英，杨宗英. 图书馆自动化新论——信息管理自动化[M]. 上海：上海交通大学出版社，1998.