支持向量机在入侵检测系统中的应用

凌永发[1],解季萍[2]
(1. 西安交通大学 理学院,陕西 西安 710049;
2. 云南广播电视大学,云南 昆明 650223)

摘要:目前的入侵检测系统(IDS)存在着在先验知识较少的情况下推广能力差的问题。简述了 IDS 的基本原理,从本质上讲,入侵检测实际上是一个分类问题,就是通过检测把正常数据和异常数据分开。给出了入侵检测模型,论述了支持向量机(SVM)是在小样本学习的基础上发展起来的分类器设计方法,专门用于小样本数据,而且对数据维数不敏感。提出了基于 SVM 的通用入侵检测系统模型,它主要由审计数据预处理器、支持向量机分类器和决策系统 3 部分组成。说明了 SVM 系统模型的可行性、模型、工作过程、实现 4 方面的内容。

关键词:支持向量机;入侵检测系统;网络安全;统计学理论
中图分类号:TP 393.08 **文献标识码**:A **文章编号**:1006-6047(2005)08-0059-04

0 引言

随着互联网技术的飞速发展,网络的结构变得越来越复杂,网络安全也变得日益重要和复杂,信息技术的安全性越来越引起人们的普遍关注。但是系统脆弱性的客观存在以及各种各样入侵行为的存在,使得信息系统的安全保护难度大为提高。一个健全的网络信息系统安全方案应该包括安全效用检验、安全审计、安全技术、安全教育与培训、安全机构与程序和安全规则等内容,是一个复杂的系统工程。安全技术是其中的一个重要环节,目前使用的安全技术及手段主要有安全路由器、VPN(Virtual Private Network)设备、网络和系统安全性分析系统、防火墙、防病毒软件、用户认证、加密、入侵检测技术等。

近年来随着各种网络安全事件的发生,以及各种黑客技术在 Internet 上出现并散布开来,使得人们越来越清醒地认识到仅仅依靠防火墙维护系统安全是远远不够的。在这种情况下,人们把目光投向了各种基于审计分析和监测预警技术的入侵检测系统(IDS),将它作为整个网络安全体系中的一个重要组成部分。

入侵检测是一种主动的网络安全防御措施,它不仅可以通过监测网络实现对内部攻击、外部攻击和误操作的实时保护,有效地弥补防火墙的不足,而且还能结合其他网络安全产品,对网络安全进行全方位的保护,具有主动性和实时性的特点,是防火墙重要的和有益的补充。

收稿日期:2005-03-03
基金项目:国家自然科学基金(10371097);云南省计算机应用技术重点实验室开放基金资助项目

1 网络入侵检测系统

“入侵”是个广义概念[1],不仅包括发起攻击的人取得超出合法范围的系统控制权,也包括收集漏洞信息,造成拒绝服务访问等对计算机造成危害的行为。从入侵策略的角度可将入侵检测的内容分为:试图闯入、成功闯入、冒充其他用户、违反安全策略、合法用户的泄露、独占资源及恶意使用。而入侵检测是对入侵行为的发觉,它通过从计算机网络或计算机系统的关键点收集信息并进行分析,从中发现网络或系统中是否有违反安全策略的行为和被攻击的迹象。进行入侵检测的软件与硬件的组合便是入侵检测系统 IDS。

Study of sine-like signal measuring in electric power equipment

HUANG Tian-shu, ZHANG Kui, REN Qing-zhen
(School of Electronics and Information, Wuhan University, Wuhan 430072, China)

Abstract: For power network parameter measuring, when the frequency departure of sine-like signal or high-order harmonic disturbance occur, fast Fourier transform is used to control the initial sampling angle for maximum precision. The proposed method improves the system precision without increasing sampling speed, which has been verified by simulation with Matlab, mathematical proving and test in a data acquisition system with MAX 197 as its A/D converter.

Key words: frequency departure; fast Fourier transform; sampling angle

图1给出了入侵检测模型[2],该模型主要由主体、对象、审计记录、活动简档、异常记录和活动规则6部分构成。但是由于缺乏相应的通用标准,不同系统之间缺乏互操作性和互用性,大大阻碍了入侵检测系统的发展。

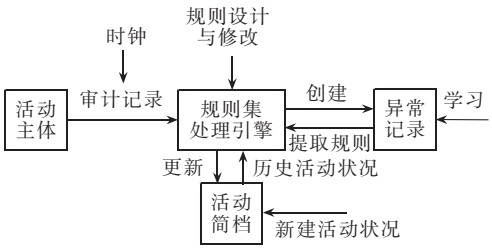

图1 入侵检测通用模型

Fig.1 The general model of intrusion detection

为了解决不同IDS之间的互操作和共存问题,制定了通用入侵检测框架CIDF(Common Intrusion Detection Framework)标准[3],试图提供一个允许入侵检测、分析和响应系统及部件共享分布式协作攻击信息的基础结构,而建立入侵检测数据交换格式IDEF(Intrusion Detection system Exchange Format)标准,并提供支持该标准的工具,则可更高效率地开发IDS系统。

IDS所检测的数据源一般分为2大类:即来自操作系统的审计数据和网络中流经的数据包。IDS正是通过处理和分析这些数据,检查是否有入侵或攻击行为,以保障系统和网络安全。

IDS的目的是检测发生在计算机网络或系统中的非法入侵行为,它研究以下问题:数据采集、数据聚类、行为的判断和分类、对入侵事件报警及响应等。然而,无论一个IDS如何设计,它的核心在于[4]:

a. 根据已有的攻击知识,推测当前的事件是否是可疑的攻击行为;

b. 根据已有的系统正常活动知识,推测当前的事件是否属于系统正常活动的范畴。

这些知识是IDS作出任何判断所依赖的基础。但是,在现实的计算机网络或系统中,这些知识是以大量实例的形式存在的,如系统日志信息、入侵攻击的手段描述等。

2 入侵检测方法

2.1 入侵检测

入侵检测可看作是一个分类问题,即对给定的审计数据进行分类:什么样的数据是正常的,什么样的数据是异常的。其中异常检测在入侵检测研究领域中得到人们很大的关注。ISA-IDS系统[5]采用了统计模型,在样本集中对每1个特征进行统计,最后找出1个中心值,然后再选择1个偏离门限。只要发生的事件超过这个门限,就被认为是入侵。这种方法设计简单,适用于具有简单分布的事件集合。也正在将人工智能方面的神经网络技术应用在异常检测方法中[6,7]。它将样本集中的入侵事件和正常事件作标志,再用它们训练1个神经网络,训练好的神经网络就可对未知事件进行检测。不过由于神经网络所固有的问题(难以设计最佳的神经结构),使得训练后的模型不大容易具有较好的推广能力。

基于机器学习的入侵检测,力图建立精确的用户模型,正确区分正常用户和异常用户[8~10],提高IDS的检测率,降低误报率。但是传统的各种机器学习算法多是基于样本数目趋于无穷大假设的,并且对数据的规律性要求较高。针对入侵检测领域碰到的多变小样本数据,可以引入机器学习领域中专门研究小样本学习的统计学习理论(SLT)[11],采用统计学习理论中最成熟的支持向量学习方法解决这类问题。支持向量机(SVM)和核学习方法主要用于解决有限样本学习问题,而且对数据的维数和多变性不敏感,具有较好的分类精度和泛化能力。

2.2 支持向量机

机器学习是人工智能应用的重要研究领域,它研究如何从观测数据中寻找规律,并利用这些规律对未来数据或无法观测的数据进行预测。统计学理论(SLT)为机器学习提供了坚实的理论基础[11,12],在此基础上发展了一种新的通用机器学习方法——支持向量机(SVM)。

SVM的基本思想是对于一个给定的具有有限数量训练样本的学习任务,如何在准确性(对于给定训练集)和机器容量(机器可无错误地学习任意训练集的能力)两个方面进行折衷,以得到最佳的推广性能。

3 基于支持向量机的入侵检测系统模型

目前的入侵检测系统存在着在先验知识较少的情况下推广能力差的问题。将支持向量机应用到入侵检测中,可以保证在先验知识不足的情况下,支持向量机分类器仍有较好的分类正确率,从而使得整个入侵检测系统具有较好的检测性能。

3.1 可行性分析

广义上讲,入侵检测属于模式识别和分类的范畴;从方法上讲有2种入侵检测方法:滥用检测和异常检测。滥用检测通过在被监控数据中发现已知的攻击签名判定是否存在攻击,主要采用模式匹配的方法;异常检测试图建立系统正常行为模式,然后通过发现与正常行为模式的偏离检测攻击。滥用检测的核心是黑客攻击模式的正确表达和快速识别;异常检测的核心是用户正常行为模式的建立及正常模式和黑客行为模式的识别和分类,这些都是典型的模式识别问题。因此,广义上讲入侵检测是属于模式识别和分类的范畴。支持向量机方法作为模式识别中的一种新的学习和分类方法,必然可以在入侵检测领域中找到其应用领域。从本质上讲,入侵检测实际上是一个分类问题,就是要通过检测把正常数据和异常数据分开。但是IDS中需要分类的数据更

加复杂,常常体现为高维、小样本和不可分性。SVM 是在小样本学习的基础上发展起来的分类器设计方法,专门用于小样本数据,而且对数据维数不敏感,SVM 方法还可以用于密度估计和孤立点发现[6],即不均衡数据集中无监督的异常检测问题。因此,SVM 方法适合于入侵检测领域高维异构不均衡数据集中的分类器设计和异常发现,将其应用于入侵检测领域是可行的。

3.2　基于 SVM 的入侵检测模型

基于支持向量机的入侵检测系统主要由审计数据预处理器、支持向量机分类器和决策系统 3 部分组成,如图 2 所示。

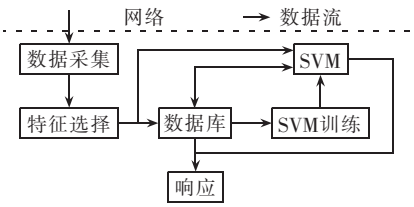

图 2 基于 SVM 的入侵检测模型

Fig.2 The intrusion detection model based on SVM

审计数据预处理器用来对大量的系统审计纪录进行处理或变换。由于支持向量机的分类器只能对维数相同的数字向量进行分类,但系统审计数据中的数据不但长度不尽相同,而且很有可能不是数字类型,所以必须将原始数据转换成支持向量机能够识别的数字向量。支持向量机分类器对这些数字向量进行分类,产生判决结果。当然,这些判决结果可以直接作为整个入侵检测系统的输出,但为了进一步提高整个系统的正确率,可以设定一些判决准则,例如发生数目、百分比等进行最终的判定,这个过程是由决策系统完成的。

3.3　工作过程

支持向量机进行入侵检测分为训练和检测 2 个阶段。在训练阶段,由先验信息得到训练样本,包括正常样本和异常样本。选择合适的支持向量机的核函数,并调整其参数找出最优参数使得支持向量机对于训练样本的分类性能达到最优,该阶段的流程如图 3 所示。

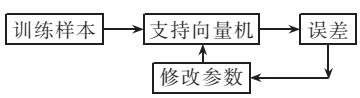

图 3 训练过程

Fig.3 Training process

检测阶段由预处理、用支持向量机进行分类和判决模块组成,见图 4。

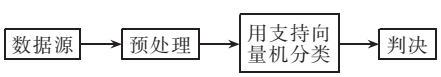

图 4 检测过程

Fig.4 Detection process

入侵检测使用的数据源主要来自主机数据和网络数据。这些数据的类型是多样的,有字符类型,用户的命令,还有数字型,如占用 CPU 的时间等。而支持向量机要求的输入必须是同维的数字,需对收集的数据进行处理,转换成支持向量机能够处理的同维数字矢量。然后用支持向量机对预处理后得到的数字矢量分类,并将分类的结果提交给决策系统。决策系统根据一定的判决准则(如发生频率、百分比等)对分类器的结果判决,并输出最终的判决结果。

3.4　实现

标准的 SVM 主要处理 2 类数据的分类问题,并且具有相当好的结果[13]。但是,根据入侵检测数据的特点,正常数据与异常数据相比,其数量要大得多,所以并不适于 2 类分类。并且由于更为关心异常类(入侵行为所对应的一类),因此与传统的 SVM 所处理的 2 类分类问题有所不同。针对当前大量 Web 入侵的事实,分析当前国内外入侵检测系统及采用技术的基础上,提出了一种基于 SVM 技术的入侵检测系统模型。这种模型在统计分析大量已有的网络数据的基础上生成入侵事件的 SVM 分类器,利用生成的分类器函数判断访问以及连接是否为攻击事件并对其分类,以供网络管理员分析。

模型将入侵行为以及被怀疑为入侵行为或存在入侵倾向者所对应的数据规定为负样本,而正常数据规定为正样本。根据 SVM 理论,2 类 SVM 分类器的分类超平面由靠近分类超平面的支持向量决定。而在实际操作中,只需要选用比较靠近分类面的样本点即可。但是,当负样本的数量不足时会严重影响分类器的泛化性能。并且,由于用于训练的负样本数量不足,将直接导致分类器的错误率过高。因此,支持向量机至今仍没有真正应用到入侵检测这一重要的领域。模型讨论用一种基于 1 类 SVM 算法的方法,针对计算机系统的网络数据的特点,构造一种可应用于异常检测的 SVM 方法。其基本思想是,在选定了 1 个核函数后,把空间中的坐标原点视为另外一类中唯一的点。并且引入松弛变量,进而可以应用传统的 2 类支持向量机。该算法可以归纳为:存在一个适当的核函数将数据映射到一个特征空间,然后用最大间隔将这些映像点与原点分割开。

整个系统的工作过程分为训练和检测 2 个阶段。在训练阶段,根据已知的正常审计数据和异常审计数据训练支持向量机,并得到支持向量和相应的参数。在检测阶段,预处理器先将未知状态的审计数据处理成数字向量的形式,然后通过支持向量机分类器,根据判决函数对这些数字向量进行分类,并将分类结果提交给决策系统作出最后判断。

4　结语

入侵检测方法对于大量先验知识的依赖,增加了入侵检测系统收集数据、建立知识库的难度和时间。如果在先验知识少的条件下,入侵检测方法的检测性能仍能满足要求,这样能换取其他方面性能

的提高。由于统计学习理论和支持向量机建立了一套较好的有限样本下机器学习的理论框架和通用方法,既有严格的理论基础,又能较好地解决小样本、非线性、高维数和局部极小点等实际问题,因此成为20世纪90年代末发展最快的研究方向之一,它在网络安全中的应用将会越来越广泛。

参考文献:

[1] ANDERSON J P. Computer security threat monitoring and surveillance[R]. Fort Washington, Pennsylvania: James P Anderson Co., 1980.

[2] DOROTHY E D. An intrusion-detection model[J]. **IEEE Transactions on Software Engineering**, 1987, 13(2): 222-232.

[3] MUKHERJEE B, HEBERLEIN L T, LEVITT K N. Network intrusion detection[J]. **IEEE Network**, 1994, 13(2): 26-41.

[4] BALAJINATH B, RAGHAVAN S V. Intrusion detection through learning behavior model[J]. **Computer Communications**, 2001, 24(2): 1202-1212.

[5] YE N, EMRAN S, LI X. Statistical process control for computer intrusion detection[A]. **DARRA Information Survivability Conference & Exposition Ⅱ, 2001. DISCEX'01 Proceedings**[C]. [s.l.]: [s.n.], 2001.3-14.

[6] LEE S, HEINBUCH D. Training a neural-network based intrusion detector to recognize novel attacks[J]. **IEEE Transactions on Systems, Man and Cybernetics, Part A**, 2001, 31(4): 294-299.

[7] BONIFACIO J, CAMSIAN A. Neural networks applied in intrusion detection systems[A]. **The 1998 IEEE International Joint Conference on Neural Networks, Proceedings**[C]. [s.l.]: IEEE, 1998. 205-210.

[8] FORREST S, PERRELASON A S. Self-on-self discrimination in a computer[A]. **Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy**[C]. California: IEEE Computer Society, 1997. 202-212.

[9] GHOSH A K, MICHAEL C, SCHATZ M. A real-time intrusion detection system based on learning program behavior[A]. **Recent Advances in Intrusion Detection (RAID)**[C]. Toulouse: Spinger-Verlag, 2000. 120-132.

[10] LEE W, XIANG D. Information-theoretic measures for anomaly detection[A]. **The 2001 IEEE Symposium on Security and Privacy**[C]. Oakland: IEEE Computer Society, 2001. 130-143.

[11] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995.

[12] VAPNIK V, LEMER A. Pattern recognition using generalized portrait[J]. **Automation and Remote Control**, 1963, 24(3): 6-13.

[13] BURGES C J C. A tutorial on support vector machines for pattern recognition[J]. **Data Mining and Knowledge Discovery**, 1998, 2(2): 121-167.

[14] JOACHIMS T. Text categorization with support vector machines: Learning with many relevant features[A]. **The European Conf. on Machine Learning (ECML'98)**[C]. German: [s.n.], 1998. 16-21.

[15] OSUNA E, FREUND R, GIROSI F. Training support vector machines: An application to face detection [A]. **CVPR'97**[C]. Puerto Rico: [s.n.], 1997. 78-81.

[16] ELEAZAR E, ANDREW A. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data[DB/OL]. http://www.cs.columbia.edu. 2002.

(责任编辑:汪仪珍)

作者简介:

凌永发(1973-),男,江西上犹人,副教授,副院长,博士后,主要研究方向为网络算法和网络控制(**E-mail**: yfling73@163.com);

解季萍(1964-),女,云南昆明人,副教授,主要研究方向为计算机网络技术。

Application of support vector machine in intrusion detection system

LING Yong-fa[1], XIE Ji-ping[2]

(1. Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, China;
2. Yunnan Radio & TV University, Kunming 650223, China)

Abstract: The actual IDS(Intrusion Detection System) has poor expansion ability when there is less knowledge. The principle of IDS is introduced briefly. As an assortment in nature, IDS detaches the normal data from exceptional data by detection. The intrusion detection model is presented. The SVM(Support Vector Machine) is an assortment machine, which is specially design for small sample data and insensitive to data dimension. The general IDS model based on SVM is brought forward, which comprises three parts: audit data pretreatment processor, SVM assortment machine and decision-making system. Four aspects are focused on: feasibility, model structure, working process and implementation.

This project is supported by the National Natural Science Foundation of China(10371097) and Yunnan Province Computer Application Technology Pivot Laboratories Opening Foundation Imburse.

Key words: support vector machine; intrusion detection system; networks security; statistical learning theory