基于DM的入侵检测系统结构方案*

詹瑾瑜** 熊光泽 孙 明

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】 介绍了入侵检测系统和数据挖掘技术的概念、特点和关键技术，分析了入侵检测系统中信息收集的主要数据来源，结合传统的入侵检测方案的缺点，提出了一种基于数据挖掘技术的具有自我学习、自我发展能力的入侵检测系统的体系结构模型，此模型针对不同的信息来源应用不同的数据挖掘方法进行识别。

关 键 词 入侵检测系统; 数据挖掘; 知识库; 防火墙

中图分类号 TP309

Framework of an Intrusion Detection System Based on DM

Zhan Jinyu Xiong Guangze Sun Ming

(College of Computer Science and Engineering , UEST of China Chendu 610054)

Abstract In this paper, the conception, the characteristic and key technique of Intrusion Detection System and Data Mining are introduced. And the main data source of information collection in the Intrusion Detection System is analyzed. This paper combines the shortcoming of the framework of traditional Intrusion Detection, and also specifies a framework of a self-training and self-development Intrusion Detection System based on Data Mining. The framework adopts different methods of Data Mining to identify according to different information.

Key words intrusion detection system; data mining; knowledge data base; fire wall

人们在得益于信息革命所带来的新的巨大机遇的同时，也不得不面对信息安全问题的严峻考验。随着网络技术的进步，网络攻防的战斗也愈演愈烈。防范网络攻击，最常用的对策是构建防火墙。防火墙是一种应用层网关，按照设定的规则对进入网络的IP分组进行过滤，同时也能针对各种网络应用提供相应的安全服务，但对于那些成功地绕过防火墙而进入系统的黑客来说，防火墙只能望“黑”兴叹了。入侵检测系统(Intrusion Detection System，IDS)是近年出现的新型网络安全技术，是一套软件和硬件的结合体。IDS能弥补防火墙的不足，为受保护的网络提供有效的入侵检测及采取相应的防护手段；入侵检测是一个全新的、迅速发展的领域，并且已成为网络安全中极为重要的一个课题；入侵检测的方法和产品也在不断的研究和开发中，已经在网络攻防实例中初步展现出其重要价值。

本文分析了传统的入侵检测方法，如概率审计法等，一般采用固定模式的入侵原型(或许可原型，将待检测的数据包通过模式匹配算法与系统中固有的入侵原型(或许可原型)进行匹配以检测入

2002年6月25日收稿
*总装备部预研基金资助项目
** 女 23岁 博士生.

侵，这样的入侵检测系统在一定程度上能检测到非法的入侵，但由于系统总的入侵原型(或许可原型)是固定的，只能通过人工的方法进行扩充，而且无法对大量的系统中没有入侵原型的入侵行为(或没有许可原型的正确行为)做出准确的判断而造成漏报(或误报)，使具有入侵行为动机的数据包顺利通过而造成严重的损失(或使正确行为数据包被拒绝)。基于传统方法中存在的不足，提出了一种具有自我学习、自我发展能力的入侵检测方案，实现系统在非人工的方式外丰富入侵原型(或许可原型)，更好的检测非法的入侵行为，该方案中引入数据挖掘(Data Mining，DM)技术，发挥其能从大量数据中有效地提取特征和规则的优势，以解决目前入侵检测系统的迫切需要的自我学习和自我发展能力问题。

1　入侵检测技术

入侵检测技术是对计算机网络或计算机系统中若干关键点的信息进行收集和分析[1,2]，从中检测到网络或系统中可能存在的各种非法攻击、恶意破坏、错误操作等违反安全策略的行为或迹象，并对此做出有效的防范和防卫行为。入侵检测系统不仅能够检测到网络上的非法入侵，还能对计算机内部的一些非法的越权处理、滥用职权和错误操作等行为做出准确的判断，是防火墙的合理补充，帮助系统对付网络攻击，扩展了系统管理员的安全管理能力(包括安全审计、监视、进攻识别和响应等)，提高了信息安全基础结构的完整性。入侵检测系统被认为是防火墙之后的第二道安全闸门，在不影响网络性能的情况下能对网络进行监测，从而为系统提供对内部攻击、外部攻击和误操作等的实时检测和保护。

入侵检测系统主要分两大类：基于网络的入侵检测系统和基于主机的入侵检测系统[3~6]。基于网络的入侵检测系统保护整个网段，监听网络中的所有数据包，从中发现有攻击特征的包并对它做出防范或防卫行为。基于主机的入侵检测系统不管网络上的数据包，而是检测系统内部的数据、日志信息、应用程序等，从中发现被攻击的各种迹象并做出响应的缩小损失的处理。现在多数入侵检测系统都兼顾了以上两种入侵检测系统的功能，同时对网络上和系统内的非法入侵做出反应。

入侵检测系统通过入侵检测的模式匹配算法将当前检测的数据包与系统中的知识库，进行比较，从而判断当前的数据包是否为入侵行为，然后根据检测结果做出响应。其过程原理图如图1所示。

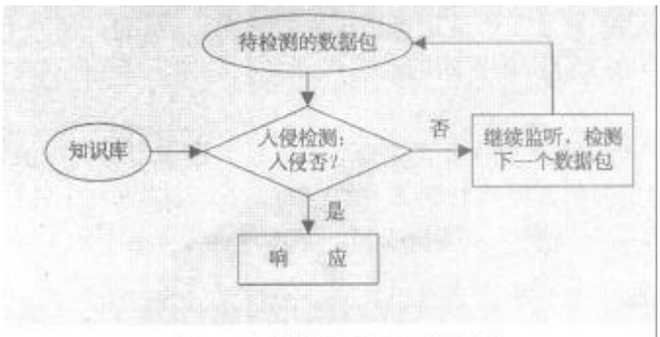

图1　入侵检测系统过程原理图

入侵检测系统所使用的知识库的来源主要是通过信息收集得到的数据，其内容包括系统、网络、数据及用户活动的状态和行为，重要的有以下面4个方面：

1) 系统和网络日志文件

黑客经常在系统日志文件中留下他们的踪迹，因此，充分利用系统和网络日志文件信息是检测入侵的必要条件。日志文件中记录了各种行为类型，每种类型又包含不同的信息，通过查看日志文件，能够发现成功的入侵或入侵企图，并很快地启动相应的应急响应程序。

2) 目录和文件中的不期望的改变

网络环境中包含重要信息的文件和私有数据文件经常是黑客修改或破坏的目标。目录和文件中的不期望的改变(包括修改、创建和删除)，特别是那些正常情况下限制访问的，很可能就是一种入侵产生的指示和信号。

3) 程序执行中的不期望行为

网络系统上的程序执行一般包括操作系统、网络服务、用户起动的程序和特定目的的应用，每个在系统上执行的程序由一到多个进程来实现。一个进程出现了不期望的行为可能表明黑客正在入侵你的系统.

4) 物理形式的入侵信息

这包括两个方面的内容：1) 未授权的对网络硬件连接；2) 对物理资源的未授权访问。黑客会想方设法去突破网络的周边防卫，如果他们能够在物理上访问内部网，就能安装他们自己的设备和软件。

2 数据挖掘技术

数据挖掘[7, 8](Data Mining, DM)也称数据库中的知识发现(Knowledge Discovery in Database, KDD)，狭义上是指从数据库提取知识，具体的说应是从数据库中对数据进行处理，从而获得隐含的、事先未知的、潜在的而又是非常有用的知识，提取出的这些知识一般表示为概念、规律、模式等形式。数据挖掘技术是一种决策支持过程，它能自动分析数据源，得出归纳性的推理，挖掘出潜在的模式。数据挖掘方法有多种，其中比较典型的有关联分析、序列模式分析、分类分析、聚类分析等.

1) 关联分析

关联分析，即利用关联规则进行数据挖掘，而关联规则是描述事物之间同时出现的规律的知识模式，关联分析的目的是挖掘隐藏在数据间的相互关系。在数据挖掘研究领域，对关联分析的研究开展得比较深入。

2) 序列模式分析

序列模式分析和关联分析相似，它把数据之间的关联性与时间联系起来，为了发现序列模式，不仅需要知道事件是否发生，而且需要确定事件发生的时间。其目的也是为了挖掘数据之间的联系，但序列模式分析的侧重点在于分析数据间的前后序列关系。

3) 分类分析

分类分析就是通过分析示例数据库中的数据，为每个类别做出准确的描述建立分析模型或挖掘出分类规则，能够把数据集中的数据映射到某个给定的类上。目前已有多种分类分析模型得到应用，其中几种典型模型是线性回归模型、决策树模型、基本规则模型和神经网络模型。

4) 聚类分析

与分类分析不同，聚类分析输入的是一组未分类的数据，并且这些数据的分类情况事先未知，通过聚类分析后把数据划分到不同的组中，组之间的差别尽可能大，而组内的差别尽可能小。

3 应用数据挖掘技术实现入侵检测系统

3.1 理论分析

数据挖掘技术可以自动高效地分析处理大型数据库，并从中挖掘出潜在的规律、模式等知识。所以在入侵检测系统中采用数据挖掘技术从历史数据构成的数据源中提取有用的知识，建立知识库，可以使入侵检测系统具有自我学习、自我发展的能力，不需人为就可自主地丰富入侵检测系统中入侵原型(或许可原型)的数量，使入侵检测系统具有智能性。

前面已经介绍了数据挖掘技术的4种基本方法：1) 关联分析；2) 序列模式分析；3) 分类分析；4) 聚类分析，它们可以协同工作在入侵检测系统中进行知识的提取。

1) 利用关联分析方法来总结出某种操作和入侵行为或各种入侵行为之间的相互关系，即某种操作通常伴随着某种入侵行为，或某两种入侵行为通常相伴发生等知识。

2) 黑客进行入侵行为通常是有步骤的，所以可以利用序列分析方法对各种入侵行为和某些操作发生的先后关系做出归纳，比如黑客要实施入侵行为前，通常要扫描计算机的端口，如果某时刻入侵检测系统发现有人在扫描计算机的某个端口就应做好提前的准备，以防范非法的入侵。

3) 黑客的入侵行为有不同的目的、不同的危害，可以利用分类分析方法将不同目的、不同危害的入侵行为进行分类处理，同时还可以利用分类分析方法将用关联分析方法和序列模式分析方法归纳总结出的各种知识进行分类处理。

4) 在入侵检测系统中还可以采用聚类分析方法利用通过前面4种方法归纳总结出的知识对用户的行为进行划分，并用显式或隐式的方法对不同的划分进行描述，获得更多的入侵原型(或许可原型)。

3.2　构造模型

入侵检测的过程原理在前面已经介绍过，在此不赘述，而数据挖掘技术的过程通常可以分成数据准备、挖掘、表述以及评价4步，根据入侵检测系统和数据挖掘技术的特点，本文提出一种将此二者合理地融合的思路——一种基于数据挖掘技术的入侵检测系统，其体系结构如图2所示：

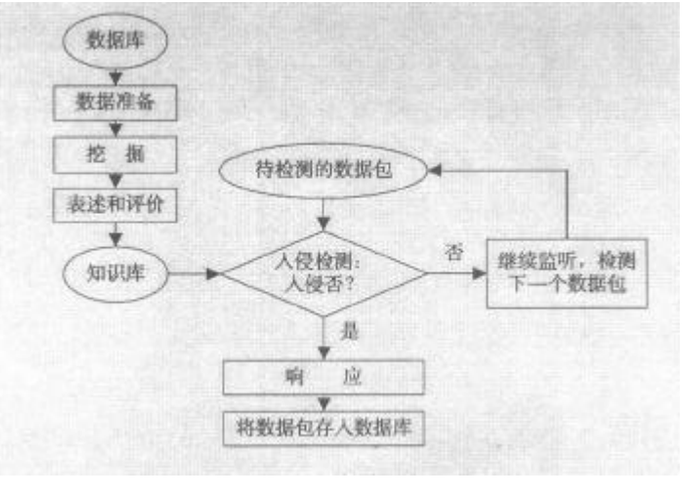

图2　基于数据挖掘技术的入侵检测系统体系结构图

1) 数据库：存放通过信息收集得到的各种数据包，由于信息收集不断的在进行，所以数据库中的数据包也就不断的丰富，这样可以向系统提供各种各样的数据，以挖掘出更多的有用知识。

2) 数据准备：从数据库中获取数据包，并进行选择和预先处理，比如去除数据包中存在的二义问题，以供挖掘之用。

3) 挖掘：利用前面所描述的各种数据挖掘方法对预处理过的数据包进行分析，从中提取特征、规则等有用的知识。

4) 表述及评价：通过对数据的挖掘提取出了有用的特征和规则等知识，再根据这些知识表述出各种入侵原型和许可原型，并对这些原型进行评价。

5) 知识库：存放了系统所挖掘出的各种入侵原型和许可原型，由于每加入新的数据包进数据库，再进行数据挖掘后得到的入侵原型和许可原型可能会增加，所以这个知识库是一个不断扩展的知识库。

6) 入侵检测：根据模式匹配算法将待检测的数据包与知识库中的知识进行比较，做出判断并给出相应的响应，如果是入侵行为，则应向系统报警并采取一定的防卫措施；如果是许可行为，则继续监听，等待下一个数据包的到来。

4 结 束 语

本文介绍了传统方法的缺点，描述了入侵检测系统和数据挖掘技术的概念、特点以及方法，并提出了一种将二者相融合、具有智能性的入侵检测系统的体系结构，按照这种思路设计出的入侵检测系统能通过不断的信息收集丰富数据库，使系统在非人为的情况下挖掘出更多的知识和原型丰富知识库，实现该入侵检测系统自我学习、自我发展的能力，在检测入侵方面的准确性也较传统的固定原型的入侵检测系统有了很大的提高。

参 考 文 献

1 Heady R, Luger G, Maccabe A, *et al*. The architecture of a network level intrusion detection system. Department of Computer Science, University of New Mexico, 1990

2 Denning D E. An intrusion-detection model. IEEE Transactions on Software Engineering, 1987, 13(2): 222-232

3 Hofmeyr S A. An immunological model of distributed detection and its application to computer security. University of New Mexico, 1999

4 蒋建春, 马恒太, 任党恩, 等. 网络安全入侵检测:研究综述. 软件学报, 2000, 11(11): 1460-1466

5 盛思源, 战守义, 石耀斌. 自适应入侵检测系统. 北京理工大学学报, 2002, 22(1): 72-75

6 陈 硕, 安常青, 李学农. 分布式入侵检测系统及其认知能力. 软件学报, 2001, 12(2): 225-232

7 胡 侃, 夏绍玮. 基于大型数据仓库的数据采掘. 软件学报, 1998, 9(1): 53-63

8 边肇祺, 阎平凡, 杨存荣. 模式识别. 北京: 清华大学出版社, 1998

· 科研成果介绍 ·

大面积单/双 YBCO 高温超导薄膜的研制

主研人员：李言荣 刘兴钊 陶伯万 姜 斌 罗 安 徐 进 何世明

为提高大面积 YBCO 高温超导双面薄膜性能的两面一致性和面内均匀性，在研究了工艺参数对 YBCO 双面薄膜性能影响的基础上，采用独特的基片旋转方法，取得重大突破。同时将“自外延技术”应用于研制大面积 YBCO 高温超导双面薄膜，提高了5 mm(2英寸)双面 YBCO 高温超导薄膜的性能。在 $LaAlo_3$单晶基片上研制的5 mm YBCO 高温超导双面薄膜性能优异，T_c=91.8 K, ΔT_c=0.4, J_c(77 K)＞1×10^6 A/cm^2, R_s(77 K, 10 GHz)=0.1～0.31 mΩ，所研制的5 mm 以内的各种尺寸 YBCO 高温超导双面薄膜已小批量生产，应用到有关产品中，为高温超导器件的研制奠定了基础。

·渠 涌·