文章编号:1005-8451(2005)03-0039-03

基于协议分析的 SVM 入侵检测系统

时 瑞
(铁道第三勘察设计院 电算所,天津 300142)

摘 要:随着网络技术日益发展,计算机必须面对来自全球网的各种入侵。因而需要有效的入侵检测系统保护计算机远离这些未经允许的或恶意的行为,将模式识别的方法应用到入侵检测系统中,解决了传统检测方法的缺陷。提出一种基于协议分析的 SVM 入侵检测方法,并使用 KDD'99 中的数据对该方法进行了评估,证明该方法是有效的。

关键词:支持向量机;协议分析;入侵检测系统;检测方法
中图分类号:TP301 **文献标识码**:A

SVM Intrusion Detection System based on protocol analysis

SHIRui
(Computer Application Department of The 3rd Railway Survey & Design Institute, Tianjin 300142, China)

Abstract: As the internet spreads to each corner of the world, computers were exposed to miscellaneous intrusion from the World Wide Web. The effective Intrusion Detection System was needed to protect our computers from these unauthorized or malicious actions. With pattern recognition method used in Intrusion Detection System, the limitation in traditional detection methods was cancelled. A new intrusion detection approach based on protocol analysis was proposed. Evaluation has been done over dataset in KDD-99. The result of simulations showed that this method was an effective method.

Key words: SVM; protocol analysis; Intrusion Detection System; detection method

在网络安全问题日益突出的今天,如何迅速有效地发现各类新的攻击行为,对保证系统和网络资源的安全十分重要。入侵检测就是用来发现网络或主机日志文件中已知的或潜在的攻击行为。目前常用的入侵检测方法有两种:(1)误用检测,根据已知的系统和软件的弱点及其攻击模式进行编码,并通过与审计数据的匹配来检测入侵,误用检测具有较低的误警率,但不能检测出一些新出现的入侵行为,所以漏报率较高;(2)异常检测,通过对审计数据的训练学习,从中发现正常使用行为模式,以定量的统计方式描述可接受的行为特征,并由测试数据和正常行为模式的偏差捕获到异常,异常检测能够发现一些新的未知的入侵。本文提出一种基于协议分析的多分类器组合的入侵检测方法,并将该

收稿日期:2004-11-27
作者简介:时 瑞,高级工程师。

中形成了环路路由自环对网络的危害极大,不仅导致路由不可达,而且浪费了大量的网络的带宽。路由自环是所有路由协议必须解决的问题,也是衡量一个路由协议好坏的重要标志。OSPF 是一种基于链路状态算法的协议,其核心思想是:每一台路由器将自己周边的链路状态(包括接口的直接路由、相连的路由器等信息)描述出来,发送给网络中所有的路由器。每台路由器在收到其他所有路由器发送的链路状态信息后,运行 SPF 算法计算路由。

OSPF 计算出的路由不会有自环 OSPF 协议生成的,自治系统内部路由是无自环的,引入的自治系统外部路由则无法保证。

8 结束语

在介绍流量部署的时候,提到了流量的负载分担问题,本文给出的建议是将分支节点按照流量划分为两个组,每组主用一台上行设备,对于另外一条链路采取备用方式。在网络规划过程中,没有非常固定的对/错之分,只有适不适合,需要遵循的原则就是:在满足客户需求的基础上,用最简单、最有效的方式实现。当然,在网络规划过程中应该充分了解相关产品,将产品的局限性和缺陷在网络设计中规避掉,确保整个网络尽量的稳定。

方法应用于KDD Cup 1999 Data中，实验表明，该方法执行效果较好。

1 问题描述

入侵检测作为一个模式识别的任务，其基本流程可以用图1来描述。入侵检测的实质可以描述为对测试样本尽可能地进行正确分类，关键问题是特征提取、选择和模式识别方法选择。在此采用了多分类器组合的方法进行分类识别。

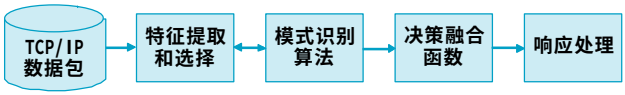

图1 入侵检测模式识别流程

1.1 多分类器组合

设计模式识别系统的目的就是尽可能正确地对测试样本进行分类。对某一分类问题，人们往往设计出不同的分类算法，并从中选择最优的算法作为最终的解决方案。然而在实验中人们发现，不同的分类器可以提供关于目标的互补信息，如果能充分利用这些互补信息的话，可以大大提高系统的性能。因而在入侵检测系统中，我们采用了多分类器组合的方法进行入侵的识别。目前，单个分类器的构造主要有两种形式：基于特征空间和基于样本空间。在本文中采用了基于样本空间的分类器构造方式，根据数据包的协议类型将样本空间划分成多个样本子空间，每个样本子空间对应一个分类器，多个分类器通过串行方式进行组合。

1.2 协议分析

特征模式匹配是第1代和第2代入侵检测系统在网络数据包中检查某个攻击特征存在的一种技术，但是这种技术有两个根本的缺陷：（1）计算负荷大，导致检测速度跟不上网络数据的传输速度，使得IDS丢包严重，漏报率高；（2）检测准确率低，会错过通过对原始攻击串做对攻击效果无影响的微小变形而衍生所得的攻击和未知的攻击。而协议分析技术则有效地解决了这些问题。 网络通信协议的核心是TCP/IP协议，TCP/IP协议是一组不同层次上多个协议的组合。在TCP/IP协议实现时，上层协议的一些细节可以在下层协议的实现时得到体现。比如在IP的首部有协议字段，可以确定是TCP协议还是UDP协议；而TCP首部有端口，可以确定上层应用协议的类型究竟是HTTP，还是SMTP，或者其他协议。这种网络协议严格分层的特点为协议分析提供了依据[2]。在建立分类器前，首先对数据包的协议进行分析，然后对每一类协议的特点分别进行特征提取和选择。

2 多分类器入侵检测模型

基本思想：从网络上截获数据包，通过数据预处理模块对数据包进行处理，同时根据数据包头文件中的协议类型对数据进行分类。然后对每一类协议的数据进行提取特征，选择出重要的特征，并建立分类器用于识别网络正常行为和各种入侵行为，最后通过一定的策略将多个分类器的输出结果组合到一起，得到最终的识别结果。图2描述了这一过程。

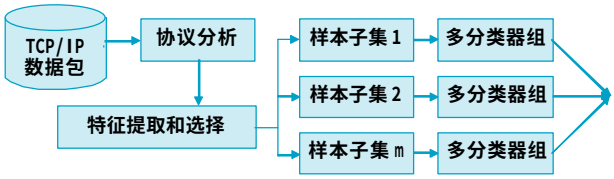

图2 多分类器入侵检测模块

图3对多分类器组的具体结构进行了详细的描述。在入侵检测系统中，不仅要实时地识别正常与攻击，还要能够识别是哪一类攻击。因此，在设计的模型中，用normal支持向量机作为第1层分类机，用以识别入侵事件和正常事件；对于入侵事件，将多输出转化为多个2输出的支持向量机分类器，再将入侵事件细分为各类不同的攻击类型，从而做出相应的反应。

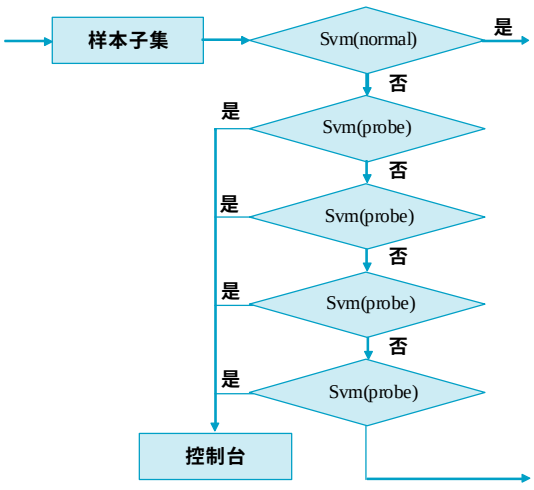

图3 多分类器组结构

本文所使用的实验数据是 DARPA（Defense Advanced Research Projects Agency）为1999年KDD (Knowledge Discovery and Data Mining) 竞赛所建立的基本数据。DARPA 资助了一系列有关入侵检测系统的研究。1998 年，美国麻省理工学院林肯实验室建立了一整套入侵检测的基准数据，该数据集提供了从一个模拟的局域网上采集来的 9 个星期的网络连接数据，其训练数据包含了 500 万个连接数据，测试数据集包含了 200 万个连接数据，并将数据记录划分为5类 Normal,Dos,U2R,Probing,R2L[1]。同时总结出了判断每一个正常与异常 TCP/IP 连接的 41 个特征，其中，与协议相关的 3 个特征是离散的特征，被用来做协议分析。如表 1 所示。

表 1　与协议相关的 3 个特征

协议类型	TCP UDP ICMP
服务网	Auth bgp courier csnet_ns ctf daytime discard domain domain_u echo eco_i ecr_i efs exec finger ftp ftp_data gopher hostnames http http_443 icmp imap4 IRC iso_tsap klogin kshell ldap link login mtp name netbios_dgm netbios_ns netbios_ssn netstat nnsp nntp ntp_u other pm_dump pop_2 pop_3 printer private remote_job rje shell smtp sql_net ssh sunrpc supdup systat telnet tftp_u tim_i time urp_i uucp uucp_path vmnet whois X11 Z39_50
标志	OTH REJ RSTO RSTOS0 RSTR S0 S1 S2 S3 SF SH

为了建立单个分类器，首先将选择的数据集按照协议类型划分成不同的子数据集，由于 DARPA 提供的十分之一数据集中只包含了 TCP、UDP 和 ICMP 这 3 种协议的数据，因而在实验时仅考虑了这 3 种协议，鉴于 SVM（支持向量机）在速度及对于大数据集、复杂的分类性能方面比神经网络更好，因而采用 SVM（支持向量机）作为分类器，把特征中与协议相关的 3 个离散的特征按照 KDD 所给样本出现的可能进行排列，共列举出了 189 种可能，分别对每一种可能构造一个分类器。SVM（支持向量机）的核函数采用径向基函数。

3　实验结果与分析

使用 DARPA 提供的 311 029 条的数据集对分类器分别进行了训练和测试，在表 2 和表 3 中分别给出了 2 种协议数据的实验结果。通过实验结果可看出本文提出的方法执行效果较好。由于篇幅所限，只列出了对总识别率影响较大的 100 条以上数据集的分类器识别率。

表 2　使用协议分类的识别结果

协议类型（服务网）,标志	Count 计数	Normal 识别率	u2r 识别率	r2l 识别率	Dos 识别率	Probe 识别率
icmpecr_iSF	158875	99.62%	100.00%	100.00%	99.95%	100.00%
TcphttpSF	38834	99.98%	100.00%	99.99%	99.63%	100.00%
TcpprivateREJ	36629	99.99%	100.00%	100.00%	97.75%	99.87%
UdpprivateSF	22572	99.87%	100.00%	100.00%	99.91%	99.88%
tcpprivateS0	16337	100.00%	100.00%	100.00%	99.95%	99.95%
TcpsmtpSF	7889	99.87%	100.00%	99.90%	99.97%	100.00%
tcppop_3SF	3651	99.62%	100.00%	100.00%	100.00%	100.00%
udpdomain_uSF	2991	100.00%	100.00%	100.00%	100.00%	100.00%
tcpftp_dataSF	2058	97.47%	99.22%	100.00%	100.00%	100.00%
TcpotherREJ	1631	100.00%	100.00%	100.00%	100.00%	100.00%
TcptelnetSF	1247	92.78%	97.75%	98.07%	100.00%	95.02%
TcpftpSF	650	87.69%	98.46%	97.85%	100.00%	100.00%
TcphttpRSTR	540	100.00%	100.00%	100.00%	99.63%	99.63%
icmpeco_iSF	535	97.38%	100.00%	100.00%	100.00%	97.38%
tcphttpS0	311	100.00%	100.00%	100.00%	100.00%	100.00%
UdpotherSF	283	81.56%	100.00%	93.62%	100.00%	81.56%
tcptelnetS3	278	100.00%	100.00%	100.00%	100.00%	100.00%
tcpprivateRSTO	246	100.00%	100.00%	100.00%	100.00%	100.00%
tcpprivateRSTR	196	96.94%	100.00%	100.00%	100.00%	96.94%
tcptelnetS0	181	100.00%	100.00%	100.00%	97.78%	97.78%
tcptelnetRSTO	181	100.00%	100.00%	94.44%	88.89%	83.33%
TcpfingerSF	176	98.85%	100.00%	100.00%	100.00%	98.85%
tcppop_3RSTO	167	100.00%	100.00%	100.00%	91.57%	91.57%
tcpsunrpcREJ	136	100.00%	100.00%	100.00%	100.00%	100.00%
tcpimap4RSTO	136	100.00%	100.00%	100.00%	100.00%	100.00%
tcpimap4REJ	95	100.00%	100.00%	100.00%	89.36%	89.36%

表 3　识别正确率对比表

识别率	识别正确率（不使用协议分析）	识别正确率（使用协议分析）
NORMAL	96.06%	99.88%
DOS	98.86%	99.88%
R2L	97.32%	99.96%
U2R	99.94%	99.99%
PROBE	99.29%	99.40%

4　结束语

本文提出了一种基于协议分析的多分类器组合的入侵检测方法，将样本空间按照协议类型进行划分，为每一种协议数据建立一个分类器，然后将多个分类器进行组合，组合后的结果作为分类识别结果。根据实验结果可知，基于协议分析的多分类器组合的入侵检测方法比采用全部特征建立的分类器具有更好的检测效果，因而在实际的入侵检测系统中是完全可行的。

参考文献：

[1] Tarun Ambwani. Multi Class Support Vector Machine Implementation to Intrusion detection[C]. 2003 International Joint Conference on Neural Network,pp.2300－2305.

[2] 李晓莺，曾启铭.利用协议分析提高入侵检测的效率[J]. 计算机工程与应用，2003，6）:169－170.