基于数据挖掘的异常入侵检测系统研究*)

吕志军 袁卫忠 仲海骏 黄 皓 曾庆凯 谢 立
(南京大学计算机软件新技术国家重点实验室 南京210093)
(南京大学计算机科学与技术系 南京210093)

摘 要 网络上不断出现新的攻击方法,要求入侵检测系统具有能检测新的未知攻击的异常检测能力。本文提出了一个基于数据挖掘的异常入侵检测系统ADESDM。ADESDM系统提出了同时从网络数据的协议特征,端口号和应用层数据中挖掘可疑行为的方法。在挖掘过程中,不但采用了基于强规则的关联规则挖掘方法,还针对强规则挖掘方法的缺点,提出了基于弱规则的关联规则挖掘方法,来检测那些异常操作少,分布时间长等不易检测的的网络攻击。同时利用网络通信的时间、方向、端口号、主机地址等属性之间的影响,建立以各属性为节点的贝叶斯网络作为异常判别器,进一步判别关联规则挖掘中发现的可疑行为,提高了系统检测的准确率。

关键词 异常入侵检测,数据挖掘,强规则,弱规则,贝叶斯网络

Research of Anomaly Detection System Based on Data Mining

LV Zhi-Jun YUAN Wei-Zhong ZHONG Hai-Jun HUANG Hao ZENG Qing-Kai XIE Li
(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)
(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract Intrusion detection system(IDS)must be capable of detecting new and unknown attacks. In this paper, we propose an Anomaly Detection System based on Data Mining(ADESDM). Firstly, ADESDM mine suspicious behaviors in the protocol header, ports and application data with strong association rules and weak association rules; then, it sends the suspicious behaviors to the Deciding Module based on Bayesian Belief Net(DMBBN). In real network communications, the attributes, such as time, direction, ports and IP addresses, are influencing each other. The DMBBN illustrates the conditional probabilities and relationship among the above attributes, and uses them to determine whether the suspicious behaviors are normal ones or attacks. Thus, system can reduce the false alarm rate.

Keywords Anomaly intrusion detection, Data mine, Strong rule, Weak rule, Bayes net

1 引言

随着网络技术和网络应用的不断发展,针对网络和计算机系统的入侵越来越多,入侵检测系统(intrusion detection system, IDS)正成为网络安全体系中的重要组成部分,成为网络安全技术研究的重点。根据检测方法,IDS可分为异常检测(Anomaly Detection)系统和误用检测(Misuse Detection)系统两大类。异常检测的优点是同时可以发现新的未知的入侵行为,有一定的学习能力,但难点是不容易建立正常行为的模式,且具有一定的误报率。误用检测的特点是对入侵行为检测的准确性高,但只能发现已知的入侵行为,且难以更新。

网络上不断出现新的攻击方法,要求IDS能及时检测这些新出现的攻击方法。由于有关新的攻击方法的知识比较少,基于误用检测的IDS难以及时检测到这些未知攻击。但由于攻击行为的特殊性,它和正常行为有一定的区别[10],如果能很好地建立正常行为模型,则基于异常检测的IDS能及时检测这些新的攻击,关键是如何更好地建立正常行为模型。目前,进行异常检测的方法很多,有神经网络技术[4],统计技术[4],模糊集合和数据挖掘技术[18]等。和其它几种技术相比,数据挖掘方法善于从大量的数据集合中确定最重要的特征,在异常检测中能获得正常模式更简洁的定义,而不是像传统的方法那样简单列举出所有模式。已有的数据挖掘检测方法通常采用强规则方法检测网络数据包头的特征来检测入侵行为[18]。这种方法可以检测网络活动频繁,且包头有特征的入侵行为;但对异常操作少,分布时间长的网络攻击(慢扫描攻击等),以及通过应用层数据进行攻击的入侵行为检测率很低,且数据挖掘检测方法的误报率比较高。

针对目前数据挖掘检测方法中存在的问题,本文提出了一个基于数据挖掘的异常入侵检测系统(Anomaly Detection System based on Data Mining, ADESDM):该系统不但采用了基于强规则的检测方法,还提出了基于弱规则的检测方法来检测那些异常操作少,分布时间长的网络攻击,并且同时检测数据包特征和应用层数据,提高了系统的检测率。另外,针对数据挖掘检测方法的误报率高的问题,建立以网络连接各属性为节点的贝叶斯网络作为异常判别器,进一步判别网络攻击的真实性,降低了误报率。

2 系统的结构和功能

ADSBDM系统由数据预处理过程,数据挖掘过程,异常行为判别过程三部分组成,分别工作在训练阶段和检测阶段

*)本文研究得到国家863计划2001AA142010(智能入侵检测与预警系统)、2002AA141090(安全服务器)资助。吕志军 博士研究生,研究方向:信息安全;袁卫忠 博士生,研究方向:网络与信息安全;仲海骏 硕士生,研究方向:网络安全;黄 皓 教授,博导,研究方向:网络安全、分布式计算;曾庆凯 教授,主要研究方向为计算机网络安全、分布式计算;谢立 教授,博导,研究方向:操作系统,信息安全。

(见图1、2)。

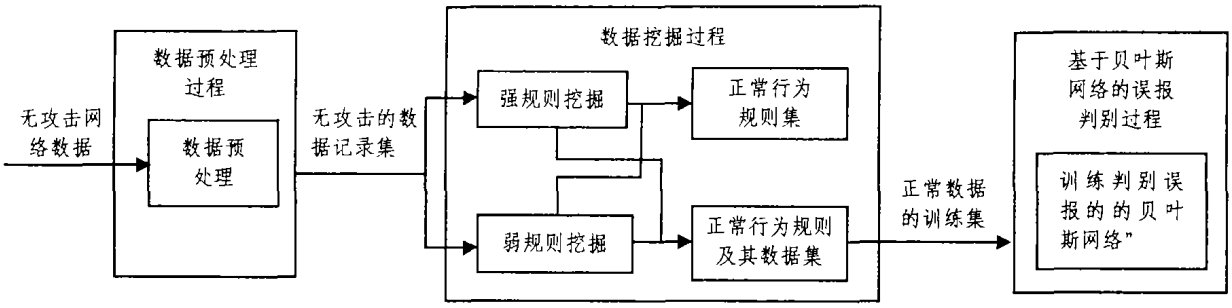

图1 ADESDM 系统的训练阶段

在训练阶段,数据预处理过程根据数据挖掘的要求,将原始的无攻击的网络训练数据预处理成适合数据挖掘的数据记录。数据挖掘过程,从预处理过的网络数据记录中挖掘出所有的规则集,并根据先验知识选出满足强规则和弱规则的正常行为规则集,并将正常行为规则集及其支持数据一起送给异常行为判别过程。异常行为判别过程根据先验知识,事先建立网络通信的主要属性之间的依赖关系,将这些属性作为节点,建立贝叶斯网络(Bayes Net, BN);然后使用数据挖掘过程输出的正常行为规则集及其支持数据训练贝叶斯网络,确定各属性节点间的依赖程度,即建立各属性节点的条件概率表(Conditional Probability Table, CPT)。

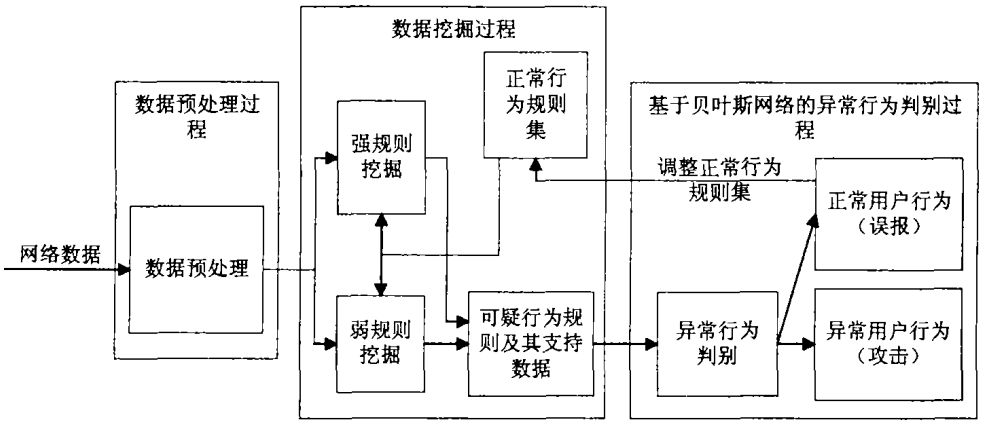

图2 ADESDM 系统检测阶段

在检测阶段,数据预处理过程将实际网络数据预处理成适合数据挖掘的数据记录。数据挖掘过程根据正常行为规则集,分别采用强规则和弱规则挖掘方法,从预处理过的网络数据记录中检测出不符合正常行为规则的可疑行为规则集,并将可疑行为规则集及其支持数据送给异常行为判别器进行进一步的判断,以确定其是异常行为的概率,这样就可以降低异常入侵检测的误报率。异常行为判别过程通过贝叶斯网络中各属性的依赖关系和依赖程度,使用支持可疑行为的数据确定待判别的可疑行为是异常行为的概率。当该概率大于一定的阈值时,判定其为正常行为,否则为异常行为。当判定可疑行为是正常行为时,将其支持数据回馈给挖掘过程,重新调整挖掘过程中的规则集中各规则的阈值。

3 异常检测过程和关键技术

3.1 数据预处理

数据预处理模块处理原始网络数据包,去除错误数据包,重组分片的 IP 包,抽取对应的特征组成网络数据记录集,提供给数据挖掘过程。

网络连接属性和数据内容的选择是关联规则挖掘中的重点;因此,预处理模块从原始的网络数据中提取能用于入侵检测的属性,分别生成具有流量属性的数据记录集 *FR* 和具有应用层数据内容的数据记录集 *DR*。*FR* 包含属性(*Time*, *SIP*, *Sport*, *DIP*, *Dport*, *Dir*, *IPFlag*, *SYNFlag*, *FINFlag*, *otherTcpFlag*, *ProType*),*DR* 包含属性(*Time*, *SIP*, *Sport*, *DIP*, *Dport*, *Dir*, *Content*, *ProType*)。

在 *FR* 中,*Time* 是该数据记录发生的时间,*SIP* 是源主机 IP 地址,*Sport* 是源端口,*DIP* 是目的主机的 IP 地址,*Dport* 是目的端口,能确定目的主机提供的服务,*Dir* 是该网络连接的方向,*IPFlag* 是 IP 包头的特征值是否异常的标志,*SYNFlag* 标识连接建立过程是否正常,*FINFlag* 标识连接终止过程是否正常,*otherTcpFlag* 标识其它 *tcp* 特征值是否异常,*ProType* 标识该数据包的协议类型,如 TCP,UDP 或 ICMP 等。在 *DR* 中,属性 *Content* 是提取出的应用层关键字,如 SMTP 过程中的 HELO,DATA,EXPN 等命令;其它属性与 *FR* 中的对应属性相同。

在不同方向的网络通信中,发生攻击的可能性和危害程度不一样。在检测模型中,网络分为内部网络和外部网络,被保护网络定义为内部网络,其它网络定义为外部网络。网络通信的 Dir 属性有三种值:1)内部→外部,2)内部→内部,3)外部→内部。

一些网络攻击会引起网络通信的异常。正常的网络通信通常是源主机访问目的主机或服务器的常用服务,具有正常的建立连接和终止连接的过程。在建立连接阶段,非正常连接有拒绝连接、企图连接、半连接等异常情况。在终止连接阶段,非正常连接终止有半关闭、连接中断等异常情况。

IP 包头、TCP 包头也是隐藏攻击信息的重要地方,通过篡改版本号和包头长度亦可实施攻击。这类攻击可通过检查 TCP/IP 包头是否正常检测。

一些攻击属于应用层攻击，无法通过检查网络层和传输层的协议特征检测这些攻击，需要通过分析应用层协议数据中的关键命令进行检测。在检测系统中，将客户端向服务器端发出的请求数据中的第一个单词作为应用层协议中的关键字，组成关键字集合。

3.2 基于数据挖掘的异常检测技术

ADESDM 系统使用的基于强规则的挖掘技术与已有的数据挖掘技术相近[2]，主要增加了对应用层数据的挖掘，因此本文重点介绍基于弱规则的挖掘检测技术。

3.2.1 *基于弱规则的挖掘技术* ADESDM 系统主要采用数据挖掘中的关联规则挖掘[19]方法进行挖掘。设 $I=\{i_1, i_2, \cdots, i_m\}$ 是属性的集合，其中每个事务 T 是属性的集合，满足 $T\subseteq I$；设任务相关的事务集 D 是数据库事务的集合。关联规则是形如 $A\rightarrow B, s, c$ 的蕴涵式，其中 $A\subseteq I, B\subseteq I$，并且 $A\cap B=\phi$，规则 $A\rightarrow B$ 在事务集 D 中成立；s 是规则 $A\rightarrow B$ 在事务集 D 中的支持度，是 D 中事务同时包括 A 和 B 的百分比，即概率 $P(A\cap B)$。c 是规则 $A\rightarrow B$ 在事务集 D 中的置信度，即 D 中事务，在包含 A 的情况下，包含 B 的百分比，即条件概率 $P(B|A)$。

例如，关联规则“SIP=202.119.36.5→DIP=32.119.56.3,0.05,0.3”说明在所有的事务记录中，5%的记录是源主机 202.119.36.5 访问目的主机 32.119.56.3，该记录占以 202.119.36.5 为源主机的所有事务记录的30%。

由于在不同的时间段，同一关联规则的支持度和置信度的阈值不一样，我们在关联规则中加入时间属性，关联规则是形如 $A\rightarrow B$, s, c, time 的蕴涵式，其中 time 指规则发生的时间。

强规则挖掘 强规则的数据挖掘就是找出大于等于一定支持度和置信度的频繁属性集。对于给定的任务相关的数据集，数据挖掘过程可能发现数以万计的规则，其中许多规则是用户不感兴趣的，因此需要对数据挖掘过程进行约束，使挖掘在用户提供的约束指导下进行，即参考事先确定的规则进行挖掘。

弱规则挖 基于弱规则的数据挖掘，就是寻找支持度低于一定阈值、而置信度大于一定阈值的数据记录集。

根据网络攻击的实际特点，有些攻击是异常行为比较频繁的攻击，如 DDOS 攻击等，通过强规则挖掘能检测出此类攻击；而有些攻击异常行为动作发生不频繁，如慢扫描攻击在单位时间内异常扫描少。如果只检查满足强规则的数据，就会漏掉检测此类攻击。因此，我们提出弱规则挖掘方法，在某些数据属性中设定最大支持度和最小置信度，将满足小于最大支持度和大于最小置信度的规则定义为弱规则。

基于弱规则的挖掘寻找发生概率少的记录。这些记录的概率通常比较小，不便于计算比较，因此需要调整支持度和置信度的值，将这些值乘以固定的调整系数 α：$Adapt\_Para(x) = Para(x)\times\alpha$，如设调整系数 $\alpha=100$，针对某个特定主机和端口的慢扫描，$s(dip=10.10.10.10, dport=80)=10\%$，$s(dip=10.10.10.10, dport=1259)=0.1\%$，其调整值为 $Adapt\_s(dip=10.10.10.10, dport=80)=10$，$Adapt\_s(dip=10.10.10.10, dport=1259)=0.1$。

在弱规则挖掘过程中，根据流量属性弱规则集 $WFAR$ 和应用层数据弱规则集 $WAAR$ 在数据记录集 FR 和 DR 中进行约束挖掘。流量属性弱规则集 $WFAR=\{wfa_1, wfa_2, \cdots, wfa_k\}$，其中 $wfa_i(1\leqslant i\leqslant k)$ 分别表示连接的时间，服务，到某主机的某个端口的连接数等关联规则。应用层数据弱规则集 $WAAR=\{waa_1, waa_2, \cdots, waa_q\}$，其中 $waa_j(1\leqslant j\leqslant q)$ 分别表示试图登录数，试图获得超级用户权次数等关联规则。

挖掘过程采用 Apriori 算法[19]进行关联规则挖掘，Apriori 算法可以有效地挖掘出数据记录不同字段之间的关联关系。

3.2.2 *数据挖掘过程* 训练阶段：首先从经过预处理的无攻击的训练数据记录集 FR 和 DR 中，使用 Apriori 挖掘算法，以时间为主要属性，流量和数据内容为次要属性，挖掘出正常行为的所有关联规则，从这些关联规则中选出正常行为的流量属性强规则集 $SFAR$、应用层数据强规则集 $SAAR$、流量属性弱规则集 $WFAR$ 和应用层数据弱规则集 $WAAR$。测试阶段：根据正常行为规则集中的关联规则中的规则属性，实时在线挖掘测试数据中满足对应规则属性的数据记录的强度（支持度和置信度），生成高强度和低强度的规则集；然后将新产生的规则集和正常行为规则集比较：如果新规则在正常行为规则集中，则比较新规则的强度是否超出正常行为规则集的阈值，如果未超出，则放弃该新规则；如果超出，则将其列为可疑规则。如果新规则不在已有的正常行为规则集中，则判别其是否超出事先统一设定的阈值，如果超出，将其列为可疑规则。将可疑规则及其支持数据记录一起送入异常行为判别过程。当判别过程判定该可疑规则为正常行为时，则用其更新正常行为规则集；当判定可疑规则为异常行为时，发出报警信息。

在数据挖掘过程中，各规则阈值的确定是重点；随时间的不同，阈值的变化很大。本系统根据时间，对同一个规则，制定出该规则在以一个星期为周期，以小时为单位，在工作日和休息日，工作日的工作时间和非工作时间中不同时间段中的阈值。

数据挖掘过程算法：

算法1、无攻击正常数据关联规则集训练算法

输入：经过预处理，待挖掘的流量属性数据记录集：FR_T；

经过预处理，待挖掘的应用层数据记录集：DR_T。

输出：流量属性弱规则集：$WFAR=\{wfa_1, wfa_2, \cdots, wfa_k\}$；

应用层数据弱规则集：$WAAR=\{waa_1, waa_2, \cdots, waa_q\}$；

1）将数据记录集 FR_T 和 DR_T 按照先进先出的方法，分别加入到具有流量属性的数据记录队列 FR_QUEUE 和具有应用层数据内容的数据记录队列 DR_QUEUE；

2）令 $WFAR=\phi$，$WAAR=\phi$；

3）在数据记录队列 FR_QUEUE 和 DR_QUEUE 中，挖掘出所有的关联规则，包括其支持度、置信度和时间，形成临时流量规则集 FAR_TMP 和临时数据规则集 AAR_TMP；

4）对∀ $v1\in FAR\_TMP$，当 $v1$ 低于支持度的阈值，并高于置信度的阈值时，$WFAR=WFAR\cup\{v1\times\alpha\}$；

对∀ $v2\in AAR\_TMP$，当 $v2$ 低于支持度的阈值，并高于置信度的阈值时，$WAAR=WAAR\cup\{v2\times\alpha\}$。

算法2：入侵检测数据挖掘算法

输入：经过预处理，待挖掘的流量属性数据记录集：FR_T；

经过预处理，待挖掘的应用层数据记录集：DR_T；

流量属性弱规则集：$WFAR=\{wfa_1, wfa_2, \cdots, wfa_k\}$；

应用层数据弱规则集：$WAAR=\{waa_1, waa_2, \cdots, waa_q\}$；

输出：SRR，可疑关联规则集；SDR，支持可疑关联规则集的数据记录集。

1）将数据记录集 FR_T 和 DR_T 按照先进先出的方法，

分别加入到具有流量属性的数据记录队列 FR_QUEUE 和具有应用层数据内容的数据记录队列 DR_QUEUE 中。

2)令 $SRR=\phi, SDR=\phi$。

3)For each $wfa_i \in WFAR$,规则 $Rule = wfa_i.Rule$

在数据记录队列 FR_QUEUE 中使用规则 $Rule$ 进行约束挖掘,挖掘出相应数据记录的关联规则及其支持度、置信度和时间,组成临时关联规则集 SC_TMP。

4)For each $v \in SC\_TMP$

if ($v.Time \in wfa_i.Time$)then// 规则 wfa_i 在不同的时间段,有不同的支持度和置信度。

if (v 的支持度小于、置信度大于规则 wfa_i 中同时间段的对应的支持度和置信度)then

$SRR = SRR \cup \{v\}$//将 v 列为可疑规则

$SDR = SDR \cup$ {支持 v 的数据记录}

Endif

Else

if (v 的支持度小于、置信度大于该规则的支持度和置信度的通用阈值)then

$SRR = SRR \cup \{v\}$//将 v 列为可疑规则

$SDR = SDR \cup$ {支持 v 的数据记录}

Endif

Endif

5)让规则集 $WAAR$ 在数据记录队列 DR_QUEUE 中,按照步骤3)和4)的方法,分别挖掘可疑规则集。

6)将可疑关联规则集 SRR,及其支持数据记录集 SDR 输出到异常判定过程。

4 异常行为判别

网络行为的各属性之间相互不独立,具有一定的影响关系[15]。贝叶斯网络描述联合条件概率分布,表示变量之间存在的依赖关系,可以描述网络行为属性之间的关系。贝叶斯网络由两部分组成,一部分是有向无环图(DAG):其每个节点代表一个变量,而每条弧代表一个概率依赖,如果一条弧线由节点 i 到 j,则 i 是 j 的双亲,j 是 i 的后继;当给定其双亲,每个节点变量条件独立于图中的非后继;第二部分是每个节点对应一个条件概率表 CPT。

$CPT_{i,j}=P(node=j|parent\_node=i)$,$CPT$ 表的每行是每个子节点对其父节点离散值的概率,其中 $CPT_{i,j} \geqslant 0(\forall i, j)$,$\sum_j CPT_{i,j}=1(\forall j)$,父节点的状态值可以是多元的。子节点的概率由父节点的概率和对应的 CPT 值决定:$P(node=j)=P(parenet\_node=i)\times CPT_{i,j}$。

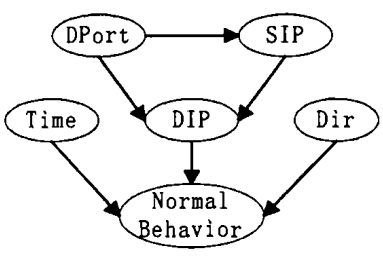

图3 判别异常行为的贝叶斯网络

利用贝叶斯网络判别异常行为,需要建立网络连接的各个属性间的关系。由于不可能对所有的网络结构进行计算,特别是当变量增多时,可能的网络结构成倍增加,因此必须在已有的先验知识下进行网络选择。通过对各种攻击方法和训练数据的分析,我们建立了网络连接的主要属性依赖关系图(图3),通过各属性之间的依赖关系,判别数据挖掘过程得到的可疑规则为正常行为的概率。

根据贝叶斯网络 DAG 图蕴含的偏序一致性原则,节点 $DPort$ 对节点 $Normal\ Behavior$ 的影响可以通过节点 DIP 传递。在该图中,节点 $Time$ 的值以周为周期,根据其中的工作时间和非工作时间,高峰期、非高峰期等特性进行分段的时间段;节点 Dir 是网络连接的方向,根据网络连接建立时,源主机和目的主机的方向关系确定,有三个值:内网→外网,外网→内网,内网→内网;$DPort$,SIP,DIP 的值分别为其实际的值,在训练阶段确定它们之间的影响率。

该网络的结构已知,并且节点变量是可见的,可以通过计算 CPT 值来直接训练网络;CPT 值通过在训练数据集中计算各节点不同状态值的条件概率得到。

当计算正常行为的概率 $p(Normal\ behavior)$时,首先将数据挖掘过程输出的可疑行为规则及其支持数据作为输入,将其代入该网络进行计算,最后得出 $p(Normal\ behavior)$的值。当该概率低于一定阈值时,认为其是异常行为(即攻击行为),生成报警信息;否则,认为其是正常行为,并使用这些数据调整数据挖掘阶段关联规则集的阈值。

算法3:判别可疑行为是 $Normal\ behavior$ 的概率

输入:SRR,可疑关联规则集;SDR,支持可疑关联规则集的数据记录集

输出:AlarmInfo,报警信息

1)For each $Rule \in SRR$

1.1)使用 SDR 中支持 Rule 的数据以及 CPT 值,计算概率:

$P(SIP)=P(SIP|\ DPort)\times P(DPort)$;

$P\ (DIP)=P(DIP\ |\ DPort,\ SIP)\times P(DPort)\times P(SIP)$;

$P\ (Normal\ behavior)=P(Normal\ behavior\ |\ Time,\ DIP,\ Dir)\times P(Time)\times P(DIP)\times P(Dir)$;

1.2)根据概率判定是否是异常行为

If $P(Normal\ behavior) >= P\_Threshold$ then

{ //$P\_Threshold$ 是正常行为的概率阈值

该可疑行为是正常行为(误报)

调整数据挖掘过程中的阈值

{ //将调整的阈值以加权(权值为 β)的方式,和旧阈值一起计算的新的阈值)

$Threadnew = Threadadapt \times \beta + Threadold \times (1-\beta)$

}

}

else

该可疑行为是异常行为,生成报警信息

Endif

5 实验结果

我们采用美国国防部高级计划研究署(Defense Advanced Research Projects Agency, DARPA)1999年[9]提供的评估入侵检测系统的测试数据集来测试系统的性能。1999年的测试数据集在1998年的测试基础上,增加了新的测试内容,主要测试 IDS 系统发现新的未知攻击的能力。该测试数据集收集了模拟网络5个星期的运行数据,其中第1个和第3个星期

是没有攻击的训练数据，用来训练异常检测系统；第2个星期的数据是含有43个攻击例子的测试数据（整个测试数据集有64种攻击方法），用来训练误用检测系统；第4和第5星期是测试数据。

ADESDM 系统是异常检测系统，所以我们选择第3个星期的数据（无攻击）作为训练数据，第5个星期的数据作为测试数据来测试我们的系统，实验结果见表1。在我们使用的测试数据中，共有57种攻击方法，180次攻击，系统检测出136次攻击，检测率为78.3%，误报46次。该测试结果优于DARPA1999测试的IDS性能。为了测试异常判别器对降低误报率的作用，我们将数据挖掘模块的输出作为报警，系统总共产生了291次报警，其中误报次数是155；经过异常判别器的判别后，系统误报次数从155次降低到46次，另外系统利用弱规则挖掘技术能较好地检测 IP sweep 和 Port sweep 攻击。DARPA1999测试结果[13]显示几个参加测试的 IDS 系统对新攻击的检测率只有19%，对较难检测的隐秘探测攻击（如 IP-Sweep，PortSweep）和 U2R 攻击的检测效果比较差。

表1 DSBODM 系统测试结果

攻击类型	新攻击数目	检测数目	检测率
Probe	37	33	89.2%
DOS	59	52	88.1%
R2L	49	36	73.5%
U2R	31	19	61.3%
Data	4	1	25%
总数	180	141	78.3%

在测试“弱规则数据挖掘”方法的检测性能时，我们以测试数据中网络数据包的正常速度，10倍正常速度和1/10正常速度，共3种不同的速度在网络中回放测试数据。该方法在10倍正常速度时，漏检部分 IP sweep 和 Port sweep 等攻击；在1/10正常速度时，检测出慢扫描等攻击，但误报率明显增高，在通过异常判别器的判别后，才降低了误报率。

6 相关工作比较

NIDS 通常采用基于规则的误用检测方法，如 SNORT[13]、BRO[11]和 Lee[18]等，Lee[18]中提出的 MADAM ID 系统在网络入侵检测中，通过采用数据挖掘中的 RIPPER 分类方法，找出常见攻击方法的特征，并根据该特征实时检测网络入侵行为。另外一些系统如 NIDES[1]，ADAM[2]和 SPADE[14]等采用异常检测方法。NIDES[1]主要监测网络数据的端口号和地址，通过分析一段较长时间（如几天或几小时）无攻击或很少攻击的数据，建立正常行为模式，然后判断短时间（几秒或几个数据包）里的数据是否明显不同，如果明显不同，则将其作为异常报警。ADAM[2]通过检查端口号、主机地址、子网号和 TCP 的状态来检测异常行为，通过概率分布，判断异常行为属于（正常，已知攻击，未知攻击）中的哪一类。本系统在数据挖掘阶段同时采用强关联规则和弱关联规则挖掘，不但挖掘端口号、主机地址、协议状态等信息，而且同时挖掘应用层的数据，提高检测的范围和能力。在分类判别阶段，本系统充分利用网络各属性之间的关系来判别可疑行为，比其它种系统具有更高的准确率，更低的误报率。

结论 随着网络的发展，不断出现新的攻击方法，要求 IDS 不但能及时检测新攻击，同时还要有较低的误报率。异常检测是该领域中研究的难点和重点。近年来，数据挖掘方法正成为入侵检测的有效手段，本系统综合利用数据挖掘中的关联规则挖掘和分类方法，有效地提高了对网络行为的异常检测能力。在实际的网络活动中，网络各属性之间的影响较大，本文提出和利用了部分属性影响的简单模型，需要进一步研究网络各属性之间的关系。

参考文献

1 Anderson, Debra, Lunt T F, Javitz H, Tamaru A, Valdes A. Detecting unusual program behavior using the statistical component of the Next-generation Intrusion Detection Expert System (NIDES), Computer Science Laboratory SRI-CSL 95-06 May 1995. http://www.sdl.sri.com/papers/5/s/5sri/5sri.pdf

2 Barbará D, Wu N, Jajodia S. Detecting Novel Network Intrusions using Bayes Estimators. First SIAM International Conference on Data Mining, 2001. http://www.siam.org/meetings/sdm01/pdf/sdm01_29.pdf

3 Cooper G, Herskovits E. A Bayesian method for the introduction of probabilistic networks from data. Machine Learning, 1992, 9(4): 309～347

4 Debar H, Becker M, Siboni D. A neural network compnent for an intrusion detectionsSystem [A]. IEEE Symposium on Security and Privacy [C], Okland: IEEE Computer Society, 1992. 256～266

5 Ghosh A K, Schwartzbard A, Schatz M. Learning Program Behavior Profiles for Intrusion Detection. In: Proc. of the 1st USENIX Workshop on Intrusion Detection and Network Monitoring, April, 1999, Santa Clara, CA. http://www.cigital.com/～anup/usenix_id99.pdf

6 Javitz H, Valdes A. The SRI IDES statistical anomaly detector. In: Proc. of IEEE Symposium on Security and Privacy, 1991

7 Heckman D, Mandani A, Wellman M. Real-World applications of Bayesian networks. Communications of the ACM, 1995, 38(3): 38～45

8 Kendall, Kristopher. A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems. Masters Thesis, MIT, 1999

9 Lippmann R, et al. The 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks, 2000, 34(4): 579～595

10 Mahoney M, Chan P K. Learning Nonstationary Models of Normal Network Traffic for Detectiong Novel Attacks. SIGKDD'02, Edmonton, Alberta, Canada, July 2002

11 Paxson, Vern. Bro: A System for Detecting Network Intruders in Real-Time. Lawrence Berkeley National Laboratory Proceedings. In: 7'th USENIX Security Symposium, Jan. 1998, San Antonio TX. http://www.usenix.org/publications/library/proceedings/sec98/paxson.html

12 Lippmann R P, Fried D J, Graf I, et al. Evaluating Intrusion Detection Systems: the 1998 DARPA Off-Line Intrusion Detection Evaluation. In: Proc. of the 20000 DARPA Information Survivability Conf. and Expositon (DISCEX), Vol. 2, IEEE press, Jan. 2000

13 Roesch, Martin. Snort - Lightweight Intrusion Detection for Networks. In: Proc. USENIX Lisa '99, Seattle: Nov. 1999

14 SPADE, Silicon Defense. http://www.silicondefense.com/software/spice/

15 Staniford S, Hoagland J A. Practical Automated Detection of Stealthy Portscan

16 Forrest F, Hofmeyr S A, Somayaji A, et al. A sense of self for UNIX processes. In: Proc. of IEEE Symposium on Security and Privacy 1996

17 Fan W, Lee W, Miller M, et al. Using Artificial Anomalies to Detect Unknown and Known Network Intrusion. http://www.cs.columbia.edu/～wfan

18 Lee W. A Data Mining Framework for Constructing Features and Models for Intrusion Detection Systems: [Ph. D thesis]. Columbia University, June 1999

19 Han J, Kamber M. DATA MINING: Concepts and Techniques. Higher Education Press, 2001