基于集成神经网络的智能决策入侵检测系统①

陈 晓[1] 夏 威[2] 包 文[1] (1.后勤工程学院 后勤信息工程系 重庆 400016;
2.后勤科学研究所 北京 100071)

摘 要： 针对传统入侵检测系统存在误报率、漏检率较高的问题，提出了一种将误用入侵检测和异常入侵检测相结合的智能决策入侵检测系统，该系统基于集成神经网络技术，通过 D-S 证据理论可以将两种技术很好地结合起来，提高入侵检测系统的效率。阐述了该入侵检测系统的总体结构部署以及各组成模块的相应结构设计。

关键词： 入侵检测系统；神经网络；误用；异常；D-S 证据理论

Intrusion Detection System Combining Misuse and Anomaly Based on Neural Network Ensemble

CHEN Xiao[1], XIA Wei[2], BAO Wen[1] (1.Department of Logistical Information Engineering, Logistic Engineering University, Chongqing 400016, China; 2. Logistics Science Institute, Beijing 100071, China)

Abstract: Traditional intrusion detection systems always have such problems as distortion and leakage. To solve these problems, this paper puts forward a new intrusion detection system which could combine misuse detection and anomaly detection. The system is based on neural network ensemble and it uses D-S evidence theory to combine the two intrusion detection technologies. The paper also expatiates on the main structure of the intrusion detection system and the composing module designation.

Keywords: intrusion detection system; neural network; misuse; anomaly; D-S evidence theory

1 引言

随着计算机和网络技术应用的日益普及，计算机网络安全越来越受到人们的重视。入侵检测作为网络安全研究的重要内容，更是引起了国内外学者的广泛关注。目前，对入侵检测技术的研究主要集中在入侵检测系统的体系结构研究和入侵检测算法的研究两个方面，分布式入侵检测系统是体系结构研究的热点，而入侵检测算法研究的趋势则是把人工智能方法应用于入侵检测中来[1]。

2 集成神经网络

神经网络尤其是近年来发展起来的一门技术，它具有自学习自适应的能力，只要提供相关数据，神经网络就会通过自学习从中提取正常的用户或系统活动的特征模式，不需要进行大量的统计分析[2]。集成神经网络作为机器学习中的一个研究热点，能较好地解决入侵检测的智能化问题，使入侵检测系统具有自学习功能。研究证明集成神经网络可以提高系统的泛化能力，而且在集成神经网络的构建中，各网络间的差异越大，集成的效果越好[3]。

目前，入侵检测技术按照检测原理可分为误用检测(Misuse Detection)和异常检测(Anomaly Detection)[4]。误用检测是对利用已知的系统缺陷和已知的入侵方法进行入侵活动的检测，检测准确度很高并且因为检测结果有明确的参照，但它不能检测未知的入侵，也不能检测已知入侵的变种，因此可能发生漏报；异常检测需要建立目标系统及其用户的正常活动模型，然后基于这个模型对系统和用户的实际活动进行

① 收稿时间:2009-09-09;收到修改稿时间:2009-10-27

审计，以判定用户的行为是否对系统构成威胁，其优点是它不需要有系统缺陷的知识，且具有较强的适应性和通用性，缺点是难于提取完整的用户正常行为特征且用户行为可能发生巨大变化，误报率较高。

由上述介绍可以看出，两种入侵检测技术在原理上有着明显的差异、在应用中具有一定的互补性，通过集成神经网络将二者结合起来，可以结合二者之长，提高入侵检测系统的检测准确性。

3　系统总体结构部署

由于原理的不同，异常入侵检测是基于主机的，一般部署于主机上，误用入侵检测是基于网络数据包的，一般部署于网络中，因此，提出系统的总体部署框图如下：

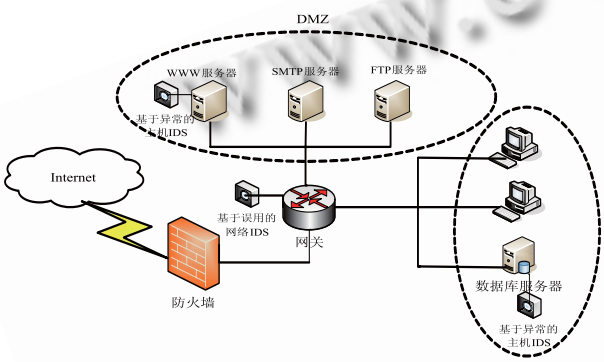

图 1　系统总体部署框图

如上图所示，在一个典型的网络拓扑结构中，将误用入侵检测部署于网关中，用于捕获整个网络进出口的数据包，而将异常入侵检测部署于网络中的关键主机，如 DMZ 区的服务器、数据库服务器等等，用于检测关键主机是否被入侵，这样实际上构成了一种分布式入侵检测系统。

4　系统逻辑结构设计

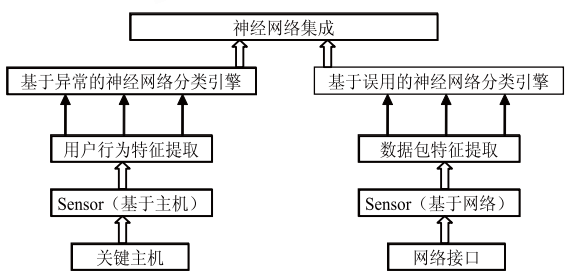

图 2　系统工作流程图

采用的集成神经网络对网络入侵的学习分为两个步骤，首先，采用单个神经网络对样本分别进行训练，然后通过相应的算法对神经网络进行集成。系统工作原理如图 2 所示，传感器分别收集来自网络和主机的数据信息，交由特征提取模块进行预处理提取特征向量，将提取出的特征向量分别作为神经网络的输入，最后由神经网络集成模块进行集成，判断是否入侵。

(1) 传感器(Sensor)，用于原始数据的搜集，针对网络安全的不同要素设置。对于本系统包含了两种类型的传感器：基于主机的传感器和基于网络的传感器。

其中基于主机的传感器主要对系统审计数据进行收集，审计数据是操作系统对用户所做操作的一种统一表示和分类，提供了比较精确的用户操作信息，用户可以用不同的命令做相同的操作，但表示成审计事件流的格式就有相同的事件序列。基于主机的传感器将审计数据收集后用 ASCII 码表示，该数据可以是操作系统调用过程或用户操作行为等等。

基于网络的传感器负责抓取网络中的数据包，部署时需将网络接口设备设置成混杂模式，侦听流过该设备的所有数据，通过对 Ethernet，SLIP 等底层协议的分析，得到网络上传输信息的内容。

(2) 特征提取模块，也可看作数据预处理模块，其作用是将传感器收集到的数据解析和转化成神经网络可识别的输入，将整理好的数据输入神经网络分类引擎。若数据是训练样本集，则作为算法输入，从该样本集中提取入侵攻击模式或特征。

基于主机的传感器对应用户行为特征提取，根据获取需要的用户活动信息来归纳出系统用户的行为特征：包括用户的正常活动时间、用户通常登录使用的计算机集合、用户通过系统命令进行正常系统存取的外部主机的集合、用户正常使用的命令集、用户的 CPU 使用模式等等，以用户的正常行为特征信息的量度值作为神经网络的输入。

基于网络的传感器对应数据包特征提取，通过协议分析技术对捕获的数据包进行解析，提取出代表网络数据流的特征向量，采用核主成分分析(KFCA)对特征向量进行降维处理，把降维后的向量送入集成神经网络分类引擎，作为神经网络分类引擎的输入向量。在综合权衡各种攻击手段后，选取的入侵特征主要包括协议码、源地址、目的地址、源端口、目的端口、数据报的类型码、数据报的代码、报文头部长度、数据报长度。这 9 个特征组成了入侵检测中的一条特征，用它们可以描述网络中出现的攻击行为。

(3) 神经网络分类引擎。入侵检测实际是对入侵

行为和用户行为进行分类和识别，为了识别可能的入侵者，系统需要具有入侵行为模式特征的知识，而用户行为和入侵行为模式的动态性要求入侵检测系统具有自学习、自适应的功能，充分利用神经网络所具有的识别、分类和归纳能力，可以使入侵检测系统适应入侵特征的可变性。

对于基于异常的神经网络分类引擎，将用户特征行为提取模块所提取的用户行为特征作为输入[5]。神经网络的隐节点采用线性阈值单元对输入层得到的用户行为特征的量度信息进行判决，线性阈值单元的阈值可先由有关专家给出初始值，其后，在系统的学习训练阶段再自动进行调整。处理时，线性阈值单元根据对量度信息的判决结果是 0 还是 1 来确定下一步到达哪个输出层节点。r_{ij} 的值表示使用输入样本对系统进行训练时，从网络隐节点 i 到输出节点 j 的频度。这些频度值可以作为一个用户行为是否是入侵活动的评测标准，即这些频度值可用来描述用户行为的轮廓特征。在输出节点进行综合评判，得到每个输出层节点的值 y1 与 y2，分别表示由当前输入的用户行为的量度信息推测该用户行为属于用户正常行为和行为表现为异常的概率。

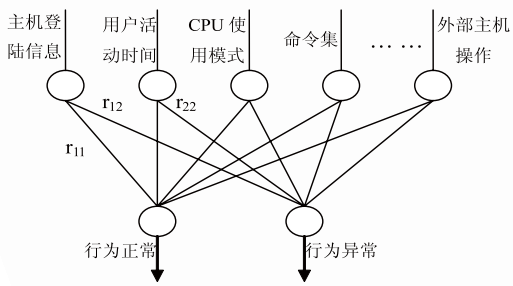

图 3　基于异常的神经网络分类引擎

对于基于误用的神经网络，接受入侵特征提取模块的输出，选用前面所述的 9 个基本的 TCP 特征作为 BP 网络的输入层，因此输入层的神经元个数为 9，输入层与隐含层的传递函数选择正切 Sigmoid 函数。检测的目的是为了区分是正常数据还是入侵数据，因此输出层只需一个即可，输出为 0 表示正常数据，输出为 1 表示入侵数据，所以隐含层与输出层的传递函数选择对数 Sigmoid 函数。

(4) 神经网络集成模块。D-S 证据理论是目前用于信息融合的主要方法之一，它是一种推理方法，用于人工智能中的不精确推理，能够处理由不知道引起的不确定性。用证据理论进行目标识别时，要求综合有关领域专家的知识和经验，由于神经网络事先经过了大量样本的学习，使神经网络的识别结果具有一定的可信度，若把多个神经网络的输出作为一条证据，利用证据理论方法把由此得到的证据不断结合起来，可以实现神经网络的集成，取得更好的检测效果。

结合神经网络和证据理论的入侵检测系统的结构图如图 4 所示，图中根据 D-S 证据结合规则，可以把神经网络传来的证据一次次地结合起来。当信任函数值达到一定的门限时输出结果；否则，不足以判断，继续获取证据以进一步判断。

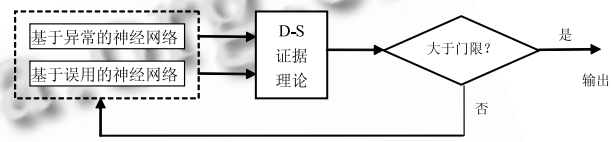

图 4　利用 D-S 证据理论实现神经网络集成

5　结论

本文提出了一种将误用入侵检测和异常入侵检测相结合的智能决策入侵检测系统模型，该模型基于集成神经网络技术，通过 D-S 证据理论实现两种入侵检测技术的结合。总的来说，该模型具有以下特点：

(1) 前端采用人工神经网络技术，具有学习功能，并且把繁琐的人工统计由神经网络去完成，大大减轻了人工劳动。

(2) 在神经网络集成中采用了数据融合的 D-S 证据理论，很好地结合了误用和异常两种入侵检测技术，具有较高的检测正确率。

(3) 根据两种入侵检测技术原理各自特点，将底层传感器在网络中实行分布式部署，降低了整个入侵检测系统漏报率。

参考文献

1 吉根林,凌霄汉,程学云.神经网络集成的分布式入侵检测方法.南京航空航天大学学报, 2007,39(2):231－234.
2 郑成兴.网络入侵防范的理论与实践.北京:机械工业出版社, 2006. 178.
3 常卫东,王正华,鄢喜爱.基于集成神经网络入侵检测系统的研究与实现.计算机仿真, 2007,3:134－136.
4 高小伟,蒋晓芸. BP 神经网络在入侵检测系统中的应用及优化.山东大学学报(工学版), 2006,12:107－108.
5 李鸿培,王新梅.基于神经网络的入侵检测系统模型.西安电子科技大学学报, 1999,10:668－669.