文章编号:1001 - 9081(2006)04 - 0806 - 03

基于前馈多层感知器的网络入侵检测的多数据包分析

周炎涛[1,2],郭如冰[1],李肯立[1],吴正国[2]
(1. 湖南大学 计算机与通信学院,湖南 长沙 410082; 2. 海军工程大学 信息与电气学院,湖北 武汉 430033)
(yantao_z@hnu.cn)

摘 要:提出了一种新型网络入侵检测模型,在该模型中,首先将截获的数据包结合历史数据包数据库进行协议分析,找出可能存在的入侵行为的相关数据包,然后采用前馈多层感知器神经网络对这些相关的数据包进行回归分析,最终获得检测结果。该模型与传统采用单数据包检测方式的网络入侵检测系统(NIDS)模型相比,具有更低的漏检率。

关键词:网络入侵检测系统;数据挖掘;前馈多层感知器;协议分析
中图分类号: TP393.08 **文献标识码**:A

Multiple data packet analysis of network intrusion detection system based on multilayer forward neural network

ZHOU Yan-tao[1,2], GUO Ru-bing[1], LI Ken-li[1], WU Zheng-guo[2]
(1. *School of Computer and Communication, Hunan University, Changsha Hunan* 410082, *China*;
2. *College of Information and Electrical Engineering, Naval Engineering University, Wuhan Hubei* 430033, *China*)

Abstract: A new network intrusion detection model was proposed. Based on the model, the currently data packet was integraed with historical data packet to process a protocol analysis, then the data packet that correlated with possiblely intrusion affair could be found out. The multilayer forward neural network was used to process a regression analysis to such data packets and got the results of intrusion detection. With the new model based on multiple data packet, the missed rate of traditional network intrusion detection system was decreased.

Key words: Network Intrusion Detection System(NIDS); data mining; Multiple-Level Perception(MLP); protocol analysis

0 引言

随着 Internet/Intranet 技术日趋成熟,通过 Internet 进行的各种电子商务和电子政务活动日益增多,伴随而来的网络安全问题逐渐成为 Internet 及各项网络服务和应用进一步发展所需解决的关键问题。入侵检测系统是网络安全方面的重要课题。基于网络的入侵检测通过监听网络中的数据包来获得必要的数据来源,并通过协议分析,特征匹配,统计分析等手段发现当前发生的攻击行为。由于网络入侵检测的分析对象是网络协议,通常而言是标准化的,独立于主机的操作系统类型,因此,一般没有移植问题。同时基于网络的入侵检测系统通常采取独立主机和被动监听的工作模式,所以,它的运行丝毫不影响主机或服务器的自身运行[1]。但是,传统的网络入侵检测通常仅对单个的网络数据包进行分析,这导致它仅能对孤立的网络事件进行分类[2]。然而,当前的网络入侵事件通常采用多个数据包协助或者分布式的方式完成,对于这部分入侵,很可能从单个数据包来看均是正常的,而由于传统的网络入侵检测系统(Network Intrusion Detection System,NIDS)仅检测单个数据包,这就造成了传统的网络入侵检测系统对于该种入侵方式的漏检,从而导致整个系统的漏检率升高。

本文提出了一种结合有协议分析的基于前馈多层感知器(Multiple-Level Perception,MLP)的多数据包分析网络入侵检测模型。该模型在采用传统的对孤立的单个数据包分析的基础上,对当前截获的数据包与最近的历史数据包进行协议分析,发现潜在的入侵威胁,然后利用 MLP 神经网络对当前数据包和最近的历史数据包进行回归分析,对多数据协作入侵和分布式入侵进行识别,从而降低整个系统的漏检率。

1 基于 MLP 多数据包分析的 NIDS 模型

本文建立的检测模型如图 1 所示。在该模型中,数据采集模块从网络数据流中截取数据包,经数据预处理模块之后分别送入两个检测系统,分别进行单数据包入侵检测和多数据包入侵检测。对于传统的单数据包入侵检测,可以采用协议分析,特征匹配,统计分析等手段,本文不作重点讨论。对于多数据包的入侵检测将是本文要讨论的重点。

2 数据的采集和预处理

数据采集模块负责抓取网络中的数据包,并将其送入下一级的数据预处理模块。网络数据截获可以通过两种方法实现,一种是利用以太网络的广播特性,另一种方式是通过设置路由器监听端口实现。两种方式适用于不同的工作情况。

数据预处理模块负责接受数据采集模块传送过来的原始

收稿日期:2005 - 10 - 21;**修订日期**:2005 - 12 - 20 **基金项目**:湖南省自然科学基金委资助项目(03JJY3104)

作者简介:周炎涛(1963 -),男,湖南汉寿人,副教授,博士研究生,主要研究方向:计算机网络通信、网络安全; 郭如冰(1980 -),男,湖南长沙人,硕士研究生,主要研究方向:网络安全、数据挖掘; 李肯立(1971 -),男,湖南涟源人,副教授,博士,主要研究方向:组合优化、并行处理; 吴正国(1943 -),男,湖北汉川人,教授,博士生导师,主要研究方向:数字信号处理、网络故障诊断.

网络数据包,并对其进行必要数据预处理,产生下一步检测所需的数据,送入检测引擎。在我们的模型中,该部分仍然保留这些功能。但值得注意的是,由于下一步需要多数据包进行协议分析,同时在协议分析时我们需要最近的历史数据包,因此,在该模型中数据预处理模块还需要完成下面的任务:

1)将处理过的数据包加上一个时间戳压入历史数据包数据库;

2)将带时间戳的数据包送入下一层的数据包协议分析模块。

由于在下一步的数据包协议分析时,需要比对最近历史数据包,所以在将数据包送入数据包协议分析模块和压入历史数据包数据库时,需要在其中加上时间戳以确认最近的历史数据包。

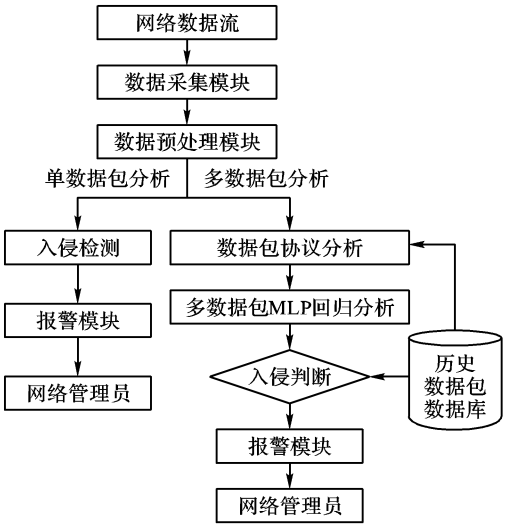

图 1　多数据包分析网络入侵检测模型

3　数据包协议分析

协议分析是用来检测攻击特征存在的技术,其特点是充分利用网络协议的高度有序性。网络协议并非随机变化的字节流,协议数据包是一个高度有序的系统,其中的结构是完全可知的,且与一系列协议规则紧密联系。协议分析技术正是利用这一知识来快速检测某个攻击特征的存在。

在我们的模型中,数据包协议分析模块主要任务是发现可能出现的采用多数据包方式的入侵。在数据包协议分析模块中包含有一系列的检测规则,这些检测规则都是针对一些多数据包方式入侵或分布式入侵的特点所制定的。例如,目前分布式的 DoS 攻击对于服务器的危害相当严重,这一类攻击通常都是在短时间内从网络上的不同主机上对同一主机的同一端口发送大量的垃圾数据包,从而造成该主机的瘫痪。针对这一特性,我们制定一条检测规则:当数据包协议分析模块从数据包预处理模块接受到数据后,从中提取出源主机地址,目标主机地址和目标主机端口号等信息,观察历史数据包数据库中是否存在与当前记录具有相同的目标主机地址,目标主机端口以及相异的源主机地址同时时间戳与当前记录相个很小的若干条记录,如果存在这样的若干条记录的话,将这些相关的历史数据记录和该条记录一起送入下一级的回归分析模块处理。

需要注意的问题是:1)对于需要考察的历史数据包与当前数据包的最大时间间隔应该设多大;2)需要考察的历史数据包的条数。对于考察的数据包时间间隔问题,应该针对不同的规则设定不同的时间间隙,例如对于上文提到的针对 DoS 攻击的检测就应该设置较小的测试时间间隙;而对于象测试从同一源主机对同一目标主机采用多数据包协作攻击的问题,由于各个关联数据包到达的间隔时间可能会较大,所以应当设置一个较大的测试时间间隙。而对于考察的历史数据包的条数问题,需要根据所监测网络的规模,所需要达到的报警级以及实际使用过程中的经验来选取。

4　多数据包 MLP 神经网络回归分析

4.1　MLP 算法描述

人工神经网络(Artificial Neural Nets,ANN)属于高度参数化的统计模型,经常用于回归或分类的预测建模。它采用多层的结构,每一层的输出(基本元素的线性组合的转换结果)作为下一层的输入,在下一层又以同样的方式来组合输入(对元素进行加权线性组合),最后做非线性转换,其基本的模型公式如式(1):

$$y = \sum_k \omega_k^{(2)} f_k\left(\sum_j \omega_j^{(1)} x_j\right) \tag{1}$$

其中 ω 是线性组合的权,f_k 是非线性变换。

MLP 是人工神经网络中应用最广泛的模型,它提供了从实数的输入向量 x,到实数的输出向量 y 的非线性映射。其基本的思想是,将一个 k 维的向量输入值乘以一个 $k \times d_1$ 的权矩阵,然后对得到的 d_1 个值进行一个非线性变换,得到第一层的 d_1 个输出作为第二层的输入,在第二层对 d_1 个乘以一个 $d_1 \times d_s$ 的权矩阵,对得到的 d_s 个值做一个非线性变换,得到第二层的输出,按此逐层变换,直至最后得到一个输出结果 y。

这个算法的最显著的特点是模型结构的多层非线性性,这里不但输出结果 y 是关于输入数据的非线性函数,而且参数在评分函数里也是非线性的。这就造成了网络生成的不同类之间的决策边界也可能是高度非线性的。

4.2　模型中 MLP 算法的应用

MLP 算法作为入侵检测算法,其神经网络的输入为网络数据包中采集到的数据。对于单个数据包分析,仅需从当前数据包中提取相关数据作为 MLP 神经网络的输入,MLP 网络的输出作为入侵检测的评分值,将其与设定阈值比较,即可得到检测的结果[4]。

本文讨论的模型要求对多数据包进行分析,因此 MLP 网络的输入为模型的上一层,即数据包协议分析层所传来的当前数据包和相关历史数据包。算法首先从所有这些数据包中提取相关数据,然后将这些数据按数据包的时间序列,以及提取的数据排序,组成个多维向量,作为 MLP 网络的输入。网络的输出仍然是入侵检测的评分值。

下面考虑一个简单的例子:假定在数据包协议分析模块中,选定的相关数据包数为 3,在每个数据包中提取的数据为 2。采用的 MLP 网络为包含一个隐藏层,层中包含 3 个神经元的简单网络模型,网络中采用 S 函数 $1/(1+\exp(-x))$ 作为非线性变换函数。则其网络的结构如图 2 所示。

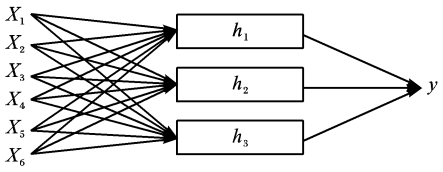

图2 网络结构实例

这里输入 $X_1, X_2, X_3, X_4, X_5, X_6$ 是当前数据包和相关最近历史数据包中提取的有用数据的有序组。其经过第一层加权计算 3 个内积：

$$S_1 = \sum_{i=1}^{6} \alpha_i x_i$$

$$S_2 = \sum_{i=1}^{6} \beta_i x_i$$

$$S_3 = \sum_{i=1}^{6} \gamma_i x_i$$

对 S_1, S_2, S_3 分别进非线性变换，得到：

$$h_1 = \frac{1}{1 + e^{-S_1}}$$

$$h_2 = \frac{1}{1 + e^{-S_2}}$$

$$h_3 = \frac{1}{1 + e^{-S_3}}$$

其中 h_1, h_2, h_3 作为第一层的输出，由于该例子是只有一层的简单网络，因此对以上输出做线性组合得到：

$$y = \sum_{i=1}^{3} \omega_i h_i$$

这里 y 就是由 MLP 网络得到的 NIDS 评分值，将其与由经验得到的阈值比较即可获得入侵检测的结果。

4.3 算法参数的估计

MLP 算法的参数估计，就是 MLP 网络模型的建立过程。具体到入侵检测中，对于滥用检测也就是对入侵模式的学习过程；对于异常模式检测也就是网络中用户行为特征的提取过程。

采用 MLP 中使用最广泛的评分函数，误差平方和（SSE）作为参数估计函数：

$$S_{SSE} = \sum_{i}^{n} (y(i) - \hat{y}(i))^2 \qquad (2)$$

其中 $y(i)$ 是第 i 个数据点真实的目标值，$\hat{y}(i)$ 是网络的输入值，它是关于输入向量 $x(i)$ 和 MLP 参数 ω 的函数。将式（1）带到式（2）中有：

$$S_{SSE} = \sum_{i}^{n} (y(i) - \sum_{k} \omega_j^{(2)} f_k(\sum_{j} w_j^{(1)} x_j^i))^2 \qquad (3)$$

其中 x_j^i 表示第 i 组输入向量的第 j 个分量，ω 是线性组合的权，f_k 是非线性变换。

考虑式(3)，S_{SSE} 是关于未知参数 ω 的函数，使其最小化。由于式(3) 是一个高度非线性函数，所以不存在使 S_{SSE} 最小化的参数 ω 的闭合形式解，因此这里的参数估计问题是一种非平凡的多元优化问题，要找到满意的局部最小值需要迭代的局部搜索技术。常用的方法有误差反向传播算法（BP 算法）等。

4.4 训练数据的选择

想要获得较好的参数估计，训练数据的选择十分重要，在这一模型中，训练数据的选择需要根据多数据包回归分析中所采用的检测方式而定。

对于采用滥用方式的回归分析，我们需要在模型运行的网络中，模拟出各种网络入侵活动，然后采用数据采集的相通原理捕获入侵时的网络数据包，并加盖时间戳，然后按协议分析的规则分析数据包，并提取数据包中的相应数据，作为训练数据。

对于采用异常方式的回归分析，需要捕捉到一定量的带有时间戳的用户正常使用条件下的数据包，并将其作为训练数据。

5 结语

本文构建了一种针对相关的几组数据包进行 MLP 回归分析的入侵检测模型，在模型中同时结合了数据挖掘方法和协议分析方法来分析数据包。可以预见，由于协议分析的加入，可以较为迅速地找出网络中潜在的多数据包协作和分布式入侵。由于检测是对多数据包的协同检测，因此对于多数据包协作以及分布式入侵要比传统的入侵检测系统更为有效，从理论上来说该模型与一些传统的采用单数据包检测方式的 NIDS 模型[8] 相比，具有更低的漏检率。

对于该模型的下一步研究可以集中在历史数据包数据库上。由于该模型需要额外建立一个储存历史数据包的数据库，由此会给检测主机的存储空间有一定的要求。解决这一问题的方法可以采用定期检测数据库的方法，对于具有较远时间戳的历史数据包，可以丢弃。另外，在改进数据库时，可以考虑定期采用历史数据库中数据包作为训练数据来进一步改善 MLP 网络的性能。

参考文献：

[1] MUKHERJEE B, HEBERLEIN LT, LEVITT KN. Network intrusion detection[J]. IEEE Network, 1994, 8(3): 26 – 41.

[2] WANG L, YU G, WANG G, *et al*. Method of evolutionary neural network-based intrusion detection[A]. International Conferences on Info-tech and Info-net, ICII 2001[C]. Beijing, 2001. 13 – 18.

[3] LEE W, STOLFO SJ, MOK KW. A Data Mining Framework for Building Intrusion Detection Models[A]. IEEE Symposium on Security and Privacy[C]. 1999. 120 – 132.

[4] CANNADY J. Neural networks for misuse detection: Initial results [A]. Proceedings of intrusion detection 98 Conference[C]. Louvain-la-Neuve: IEEE Press, 1998. 31 – 47.

[5] HOFMANN A, SCHMITZ C, SICK B. Rule extraction from neural networks for intrusion detection in computer networks Systems[A]. IEEE International Conference on Man and Cybernetics[C]. 2003. 1259 – 1265.

[6] MUKKAMALA S, JANOSKI G, SUNG A. Intrusion detection using neural networks and support vector machines Neural Networks[A]. Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN '02[C]. 2002. 1702 – 1707.

[7] KANG B-D, LEE J-W, KIM J-H, *et al*. An intrusion detction system using principal component analysis and time delay neural network [Z]. Enterprise networking and Computing in Healthcare Industry, HEALTHCOM, 2005.

[8] WANG J, WANG Z, DAI K. A network intrusion detection system based on the artificial neural networks[A]. Proceedings of the 3rd international conference on Information security[C]. 2004.