文章编号：1005-8451（2010）08-0033-03

高速网络环境下一种对入侵检测系统的改进方法

王阿婷[1]，毕红军[1]，刘 云[1]，司夏萌[2]
（1.北京交通大学 通信与信息系统北京市重点实验室，北京 100044；
2.北京交通大学 网络舆论安全研究中心，北京 100044）

摘 要：高速网络环境下，现有的入侵检测系统存在检测效率低、准确率低和丢包现象等问题。改进后的入侵检测系统先对数据包中的应用协议进行预处理，检测出比较明显的入侵特征。如果没有检测到，就利用基于决策树的模式匹配方法进行更进一步的检测。它可以提高检测速度和降低误报率，更加适应高速网络。

关键词：入侵检测；协议分析；决策树算法
中图分类号：TP393.08 **文献标识码**：A

Improved method of High-speed Network Intrusion Detection System

WANG A-ting[1], BI Hong-jun[1], LIU Yun[1], SI Xia-meng[2]
(1. Key Laboratory of Communication & Information Systems, Beijing Municipal Commission of Education,Beijing Jiaotong University, Beijing 100044, China;
2. Center for Security Studies Public Opinion, Beijing Jiaotong University, Beijing 100044, China)

Abstract: In high-speed network, the Current Intrusion Detection System had some deficiencies ,like low efficiency, low accuracy, losing packets. Considering to improve the System, do pretreatment with the protocol in packet first, in order to detect the obvious features of the invasion. And then, it was adopted the decision tree theory to do further detecting. The proposed method had the higher performance, reduced false positives and made intrusion Detection System more suitable to highspeed network.

Key words: intrusion; protocol analysis; decision tree theory

随着计算机和网络技术的迅速发展，人类正逐步迈向网络化和信息化的时代。与此同时，与人们切身利益相关的网络安全问题越来越受到关注。由于黑客活动频繁，各个网站遭到不同程度的攻击，一些金融机构或者个人也遭受了严重的经济损失。对于防火墙技术往往是解决来自于网络外部的攻击，而对于内部攻击却是无能为力。对于网络安全的另一种保障体系：入侵检测系统因为能同时检测来自网络外部和内部的攻击而得到了广泛研究和应用。

入侵检测系统能够及时发现并报告系统中异常现象。随着高速网络环境的大量出现及快速发展，传统的入侵检测技术开始暴露出一些弊端，其中比较突出的是：漏抓的网络数据报多、误报率高、检测速度慢和检测时出现的丢包现象等。本文所提出的网络入侵检测系统，在基于应用协议分析的基础上将传统的入侵检测系统进行了改进，加入了对数据包的预处理过程，在模式匹配环节中使用决策树算法，以提高入侵检测系统的性能，使其更能适应高速网络环境。

收稿日期：2009-12-30
作者简介：王阿婷，在读硕士研究生；毕红军，副教授。

1 改进的入侵检测系统模型

入侵检测作为一种网络防护的安全措施，影响入侵检测系统性能的主要因素有：对协议的分析能力、单位时间内需要处理的数据包的数量、规则库的规模[1]。为了能够提高网络入侵检测系统在高速网络环境中的性能，应该从协议分析和模式匹配这两个过程入手来解决问题。本文提出了一个入侵检测系统模型[2]，如图1。

2 改进系统中部分模块的功能

2.1 协议分析预处理模块

对于一些简单或容易辨认的可疑协议，在这一步就能够被识别出来。可以对包中的数据各部

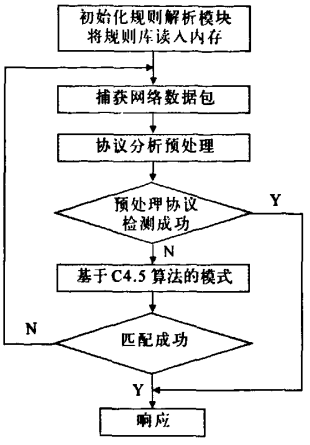

图1 改进的入侵检测系统模型

分进行全面的协议规则检查，如果检测到与协议规则不符的数据包时就直接进入响应模块。比如FTP协议，假设数据包中的协议为FTP协议，在接到这个数据包之后，先对它进行最直接的判断，检查它是不是有很特殊的特征，这时可以立刻判断出包中的协议类型。而对于FTP协议来说，当加入了由用户定义的一些操作或参数时，可以通过计算这些操作或参数的长度作为判断是否出现缓冲区溢出的根据，此时可以直接判断是否发生了攻击。直接将这个数据包送入相关FTP协议模式匹配的模块中，而不需要再进行协议分析。这样的改进对于高速网络环境来说，可以有效地改善入侵检测系统的实时性，提高检测速度，减少检测时出现的丢包现象，同时也可以有效地减少模式匹配中的资源。

协议分析[3]的数据包预处理模块的另一个功能就是进行协议分析。采用基于应用协议的网络入侵检测方法，它将捕获的数据包进行处理，形成属性-值的数据形式，直接用于下一步的模式匹配。这样做的好处：

（1）能够根据特定的应用协议再进行细致的分析。例如，可以通过检测状态码中是否含有代表服务被拒绝的“403”来发现是否有未授权访问网页[3]。然而简单模式匹配方法则可能在非状态码的其他区域搜索到“403”，就会将正常的数据误判成入侵数据。在高速网络环境下，这将会导致系统的误报次数增加，准确率下降。而基于应用的网络入侵检测方法来说，能够获取与目的主机相同的内容，可以更准确的判断是否发生了入侵行为。

（2）对于不同的应用协议，比如HTTP、FTP、SMTP、TELNET、POP等，可以根据协议的不同，形成不同类型的属性-值，便于下一步模式匹配的使用。这样能够有效地提高检测的准确性、减小模式匹配的计算量、减少误报率，更能适应高速网络。

2.2 入侵检测模式匹配模块

在入侵检测模式匹配过程中引入了决策树，对检测的规则进行优化。它与协议分析相互结合，其核心是选择决策树算法利用样本数据生成决策树模型，然后用生成的决策树对未知数据进行分类预测。这种方法可提高检测速度并降低误判率[4]。

文中选用了ID3算法的改进算法C4.5[5]，这种算法可以选择具有最高信息增益率的属性作为测试属性，能够有效地处理连续属性。

2.2.1 入侵检测决策树的生成过程

首先建立数据分类和属性文件。数据分类为intrusion和normal，表示入侵和正常两种状态。属性文件用于存放属性名、属性类型以及离散性属性的所有可能取值。其中属性类型数据根据每个属性对应的协议字段的特点，标记为离散型或者连续型。再准备训练数据。一般训练数据都是以tcpdump格式存储的。经过协议分析预处理后，生成属性-值的二维表形式数据。最后利用C4.5算法，由信息增益率选择测试属性创建决策树。

2.2.2 模式匹配过程

利用协议分析预处理模块处理后的结果作为输入，将属性-值型记录中与根节点测试属性对应的属性值提取出来，并将此值与分支值比较。如果找不到匹配的分支，检测结束，此记录的分类为该节点的默认最佳分类。如果默认分类为入侵，就立即响应。如果此属性值与某个分支值匹配，该分支将此记录指向下一个子节点，再对下一个节点属性进行比较，直至到达叶节点为止。叶节点所标记的分类为该记录的分类。模式匹配过程如图2。在高速网络环境下，可以在检测的过程中只检测相关的规则集，而不是遍历整个规则集，可以大大的减少匹配计算量，提高匹配效率；同时针对某一种协议进行匹配可以提高匹配的正确率。

3 实验测试

实验环境为：CPU：Intel Core Duo 1.80 GHZ，

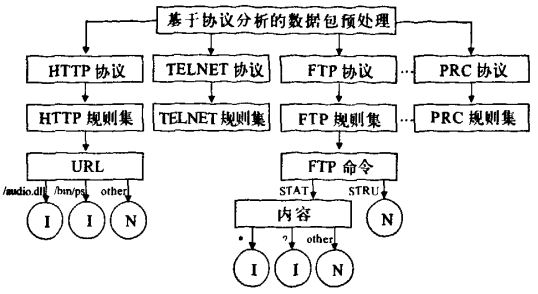

图2　模式匹配过程

操作系统：Windows XP Service Pack 3，内存DDR 1 GB，训练和测试数据为采用WinPcap4.1.1和Wireshark1.0.10捕获的数据包。

先利用Wireshark捕获一定数量的网络数据包，手工地向某些包中添加入侵特征值，作为测试使用的数据；然后再训练这些数据生成C4.5决策树，根据不同的属性值生成各自的检测规则。

实验1：检验协议分析预处理功能是否有效。

前面提到过有一些入侵检测特征相对比较容易被检测出来，不用再进行模式匹配。在捕获的数据包中随机选取了50条，并手工向20条中分别添加了比较明显的入侵特征值，作为测试使用的数据。检测时间如表1。

表1　检验协议预处理结果　　单位：ms

匹配规则	进行协议预处理	不进行协议预处理
25	2.104 433	3.081 248
50	2.104 500	3.440 018
100	2.104 896	3.579 416

实验结果：当规则数量较少时，两种情况下使用的时间相当；随着规则数量开始增加，进行协议预处理所用的检测时间相对要少。

实验2：比较简单模式匹配方法和改进后的方法。

使用简单模式匹配方法与使用本文提出的方法的检测时间比较：实验过程中，规则数从10条开始，一直增加到200条。用两种待检测的方法进行检测，记录结果，如表2。

实验结果：当测试数据量和匹配规则数较少时，两种情况下使用的时间相当；随着测试数据量和匹配规则数开始增加时，改进后的方法所用的检测时间相对要少。

实验3：检测的正确率比较。

表2　两种模式下检测时间比较　　单位：ms

测试数据量	匹配规则数	使用简单模式匹配方法	使用本文提出的方法
50	20	5.372 290	3.069 558
50	40	9.815 234	3.391 358
50	120	19.873 429	3.553 982
50	240	38.987 094	3.99 980
100	240	75.231 043	6.683 296
200	240	149.239 901	11.899 002

首先选取25条测试数据，针对性设置10条检测规则；手工将规则的入侵检测值添加到测试数据包的数据部分；用两种待检测的方法进行检测。

实验结果：使用简单模式匹配方法只对3条进行正确的判断，而本文提出的基于协议分析的决策树方法对20条进行了正确的判断。

4　结束语

本文采用C4.5算法实现了基于决策树的协议分析网络入侵检测，将决策树与协议分析技术结合的同时还对数据包进行了预处理。面对高速大流量的网络环境，利用C4.5算法能够减少系统资源，可以提高检测效率，减少匹配计算量，使模式匹配技术在一定程度上实现智能化，而不仅仅是单纯的字节匹配；利用协议分析预处理可以提高匹配的精确度，减少误报率。使入侵检测系统更加适应高速网络。

参考文献：

[1] 凌　宇，徐　雄，石林安. NIDS中协议分析和模式匹配的研究[J]. 信息技术，2006（5）：82-84.

[2] Ulf lindvist,Phillip Brentano, Doug Mansur. IDS Motivation, architecture, and An Early Prototype[J]. Comupter Security Laboratory, US Davis, 2007 (2): 160-171.

[3] 蔡　敏，叶　震，徐吉斌. 协议分析技术在入侵检测中的应用[J]. 计算机技术与发展，2007，17（2）：240-241.

[4] LEE W. A data mining framework for constructing features and models for intrusion detection systems[D].New York:Columbia University, 1999.

[5] Quinlan J R. C4.5:Programs for Machine Learning[J]. New York:Morgan Kaufman, 1993 (2): 25-26.

[6] HAN Jian-wei,KAMBER M. Data mining concepts and techniques[M].Beijing:China Machine Press,2000:188-194.