协议分析技术在入侵检测中的应用

蔡 敏[1,2],叶 震[1],徐吉斌[1]
(1. 合肥工业大学 计算机与信息学院,安徽 合肥 230009;
2. 巢湖学院 计算机科学与技术系,安徽 巢湖 238000)

摘 要:入侵检测技术是安全防护的重要手段,但是传统的入侵检测系统在高速网络环境下由于误报率和漏报率过高而难以满足实际需要。文中分析了基于模式匹配的入侵检测系统的不足,提出了把协议分析技术和模式匹配技术相结合的检测模型,最后讨论了一种对入侵检测系统的规则库进行精简的方法。这些方法提高了检测准确率和效率,使得入侵检测系统能够适应高速网络环境。

关键词:入侵检测系统;模式匹配;协议分析;规则库
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2007)02-0239-03

Application of Protocol Analysis Technology in IDS

CAI Min[1,2], YE Zhen[1], XU Ji-bin[1]
(1. School of Computer and Information, Hefei University of Technology, Hefei 230009, China;
2. Department of Computer Science and Technology, Chaohu College, Chaohu 238000, China)

Abstract: Intrusion detection technique is an important safety precaution, but the current intrusion detection system can't meet the actual demands because of the defect of high false alarm and false negative rates in high-speed network. Analyzes the limitations of intrusion detection system which is based on pattern matching, puts forward a model which protocol analysis and pattern matching are combined, and discusses a method to reduce the signature library of intrusion detection system. These methods can enhance the accuracy and efficiency of detection, and make intrusion detection system adapt to high-speed network.

Key words: intrusion detection system; pattern matching; protocol analysis; signature library

0 引 言

入侵检测是指通过从计算机网络或计算机系统中收集信息并对其进行分析,从中发现网络或系统中是否有违反安全策略的行为和遭到攻击的迹象,同时作出响应。入侵检测系统是实现入侵检测功能的一系列的软件、硬件的组合[1]。根据入侵行为的属性一般将入侵检测分为异常检测和误用检测。异常检测试图收集所有正常活动,从中提取正常模式,凡是超出正常模式偏离度的活动就被视为非正常活动。这种方法看似一劳永逸,但是要收集所有正常活动事实上是难以做到的,并且如果偏离度的值选择不恰当,误报率和漏报率就会比较高。误用检测事先定义具有某些特征的行为是入侵行为,把所有已知的入侵行为的特征提取出来形成规则库。判断某个行为是否是入侵行为,就是要判断这个活动是否具有已知入侵行为的特征。目前大多数入侵检测系统产品都是基于误用检测的。

1 基于模式匹配的入侵检测系统

模式匹配是基于误用检测的入侵检测系统中最常用的方法。方法是系统将收集到的数据包与规则库中的规则依次进行匹配,若规则匹配成功则调用相应的告警程序,再处理下一个捕获数据包;若匹配不成功,则匹配规则库中下一条规则,直到匹配成功或所有的规则都匹配完为止。模式匹配最大的优点就是方法简单、误报率低,然而模式匹配同时又存在着两个重大的缺点:

1)计算负荷大。每一个数据包从截获到分析再到规则匹配,需要花费大量的时间和系统资源。每秒最大计算次数 = 攻击特征字节数×网络数据包字节数×每秒数据包数量×攻击特征数量。

2)检测准确率低。使用固定的特征模式来检测入侵,不能对未知攻击进行有效检测。即使是已知攻击,

收稿日期:2006-04-19
作者简介:蔡 敏(1975-),男,安徽巢湖人,助教,硕士研究生,研究方向为网络安全;叶 震,副研究员,硕士生导师,研究方向为网络安全。

只要稍做改变,也无法被检测出来。

基于模式匹配的入侵检测系统能否准确地识别出入侵行为,一方面依赖于规则库的完备性,另一方面也取决于能否对网络上的全部数据包进行监听和分析。随着高速网络的发展,网络传输的高速度与模式匹配的低效率的矛盾日趋突出,如果检测速度跟不上网络数据的传输速度将导致数据包丢失,漏报率将随之增高 此外,入侵手段的发展使得规则库日趋庞大,每出现一种新的攻击手段,或对已知攻击稍作修改,就要在规则库中添加相应的规则,这进一步加大了入侵检测系统的负担。入侵检测系统的检测任务正变得越来越重,如果不能做到检测的准确和实时,入侵检测系统就会失去存在的意义。

2 协议分析检测方法

2.1 协议分析的基本思想

单纯的模式匹配检测方法存在缺陷的主要原因是它把网络数据包看作是无序随意的字节流,对该网络数据包的内部结构完全不了解。为了减少匹配算法的计算量和得到更准确的检测结果,使入侵检测系统适用高速网络的要求,一些学者将协议分析应用到入侵检测领域。所谓协议分析简单地说就是检测数据包是否符合协议的规范[2],任何违背 RFC 规定的数据包均视为协议异常。

网络通信的核心协议是 TCP/IP 协议,TCP/IP 协议作为国际互联网 Internet 的基础,已成为事实上的工业标准,它与网络的结构、类型无关,这使得协议分析具有很广泛的适用性。TCP/IP 协议模型共有五层[3],它们从下向上依次为 硬件层、链路层、网络层、传输层和应用层。每个协议的数据包都各自具有标准的格式和明确的含义,易于提取正常的行为模式。协议分析技术有效利用了网络协议的层次性和相关协议的知识来快速地判断攻击特征是否存在。协议分析技术的关键技术包括协议解码、数据重组、命令解析等技术。协议解码就是观察并检测所有网络数据包,如果数据包不符合预期的标准,就发出报警。数据重组在协议解码的基础上对数据包进行 TCP 流重组及 IP 分片重组,从整体上来检测一次会话,它充分利用了协议运行信息来检测协议相关的异常和多步骤攻击、分布式攻击等复杂攻击。命令解析可针对各种高层协议,分析出攻击串以及各种可能的变体。例如 Web 服务器会对一个 URL 有多种不同的等价表达式,命令解析技术把各种 URL 表达式转化成一种规范化的形式,从而可以检测出企图利用 Web 服务器的灵活性来躲避检测的攻击。

2.2 协议分析过程

对一个网络数据包进行协议分析的过程就是一条从协议树[4]根结点到某个叶子结点的路径(如图 1 所示)。

由协议标识和端口号来决定进入哪个协议分析器,例如以太帧的第 13,14 字节处包含了两个字节的网络层协议标识,0800 为 IP 协议,0806 为 ARP 协议,0835 为 RARP 协议等。在 IP 数据包的格式定义中,第 10 个字节为传输层协议标识,如 TCP 为 06,UDP 为 11,ICMP 为 01 等。而 TCP 数据包的第 13 第 14 个字节为应用层协议标识(端口号),如 80 为 HTTP 协议,21 为 FTP 协议,23 为 TELNET 协议等。例如现在捕获到一个以太网数据包,以 16 进制表示为:00 AO C9 8F FF Cl 00 00 E8 6F AD 59 08 00 45 00 01 90 85 00 40 00 80 06 E3 4E C8 00 00 C8 00 00 00 50 04 0B 00 50 00 15 33 FB 6F CC C1 E0 50 18 22 38 28 D1 00 00 47 45 54 20 2F 2E 2E 25 63 31 65 6D 33 32 2F 63 6D 64 2E 65 78 65 20 48 54 54,协议分析流程如下:

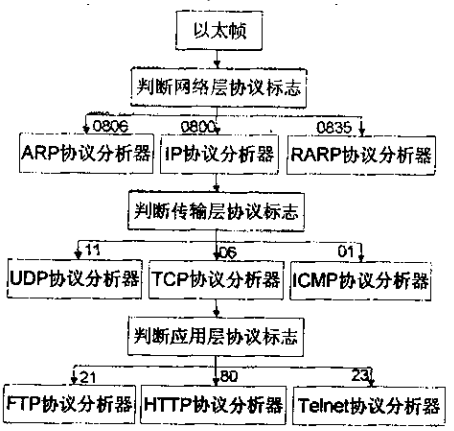

图 1 协议分析流程

① 根据以太网协议,数据包第 13、14 两字节为 0800 判断此包封装 IP 包;

② 再检查 IP 包第 10 个字节,发现是 06,所以这个 IP 数据包封装了 TCP 包;

③ TCP 包第 13、14 字节为 0050,即十进制 80,该数据包封装的是一个 HTTP 协议的数据包;

④ HTTP 数据包从第 17 个字节是 URL 开始处,进入模式匹配模块。

从该例可以看出,协议分析技术利用协议规则寻找攻击,只需检查特定字段,而不是整个数据包,因此极大地减少了运算量。与传统的模式匹配技术相比,每个数据包的比对次数由上亿次减少到几百次或几十次,这样 IDS 就可以处理更多的数据包,从而解决了传统的网络入侵检测系统在高速网络环境下的丢包问题[4]。

2.3　引入协议分析的检测模型

虽然基于协议分析技术的入侵检测有着巨大的优势,但目前基于协议分析的入侵检测系统还不够成熟,许多协议分析器的研发还处于实验阶段[5],另一方面,协议分析技术在最后检测时很多时候还需要用到模式匹配,因此把协议分析技术和模式匹配技术相结合显然是个更好的方法,它们相互补充,发挥各自的长处。引入协议分析技术的入侵检测模型如图 2 所示。

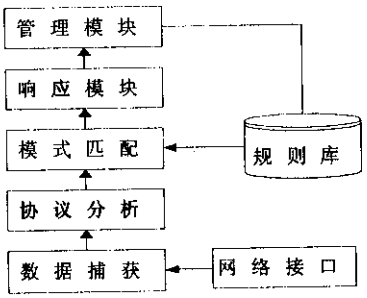

图 2　引入协议分析技术的入侵检测模型

其中数据捕获模块负责获取网络数据包;协议分析模块包括各种协议分析器,由于协议类型较多,一般采用模块化的插件技术[6]。使用这种技术可以方便地增加、删除和修改检测处理方法。对一个新的协议分析只需编写这个协议分析器并调用注册函数在系统注册就可以使用了。协议分析模块对捕获的数据包进行协议分析,如果协议分析器发现有协议异常,就调用告警程序,如果没有发现异常,再交由模式匹配模块进行检测;模式匹配模块将当前数据包与规则库中规则进行匹配,判断是否存在入侵及入侵类型。协议分析模块减少了模式匹配模块需要匹配的数据包的数量和长度,从而提高了判断效率。规则库可以手工添加规则,也可以运用数据挖掘等方法产生规则,还可以将运行过程中发现的新规则加入到规则库中;响应模块对检测到的入侵行为采取一系列措施,如发出警报,切断连接等;管理模块是管理人员与机器交互的界面,方便管理员使用和配置系统。

3　规则库的精简

协议分析技术弥补了基于模式匹配的入侵检测系统的不足,但是协议分析器本身当然也会消耗性能,每加入一个协议分析器,就会不同程度地降低系统性能。在具体实际环境中,应根据需要选择相应服务的协议分析器。另一方面,庞大的规则库也需要精简,在某个特定环境下,不一定使用到规则库中所有的规则。比如系统并未开放 FTP 服务,那么 FTP 协议分析器和所有与 FTP 攻击规则都是不必要的,这时它们的存在反而无形中会白白浪费检测时间,降低检测效率。多余的规则有可能会引起错误告警,所以应该根据实际需要选择所需要的协议分析器和规则集。

协议类型是创建规则库的重要条件属性。可以依据规则所依赖的协议类型将规则分类,可以分成 ICMP 规则集、TCP 规则集、IP 规则集、HTTP 规则集、FTP 规则集等等。在系统中维护一张按协议类型建立的规则索引表,索引表结构示意图如图 3 所示。

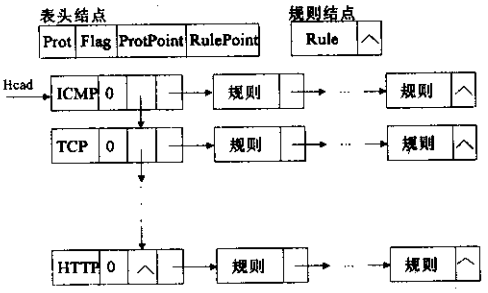

图 3　按协议类型对规则进行分类

索引表中有两类结点:表头结点和规则结点。表头结点包括协议类型(Prot)、是否已打开标志(Flag)、指向下一个表头结点的指针(ProtPoint)、指向第一个依赖该协议的规则结点的指针(RulePoint)。规则结点包括:规则(Rule)、指向下一个依赖该协议类型的规则结点的指针(RulePoint)。索引表初始状态为空。入侵检测系统启动后,依次将每条规则调入内存时,获取该规则所依赖的协议类型,若该协议的表头结点已经存在,则产生规则结点接入该协议的链表表尾;若该协议类型不在索引表中,则创建该协议的表头结点并接入表头链表表尾,再产生规则结点加入该表头结点的表尾。索引表建成后,系统管理员将依据网络状况决定启动哪些协议分析器和规则集,启动规则集就是将指定协议的 Flag 标志位置 1。当有数据包到达匹配模块时,先判断该数据包的协议类型,匹配时只需匹配相应协议中的规则,不必考虑其他协议的规则,从而有效减少了需要匹配的规则数。如果某数据包的协议类型 Flag 标志为 0,则认为该数据包异常。

4　小　结

单纯的基于模式匹配的入侵检测系统已越来越不能满足高速网络的要求,必须对现有的入侵检测系统加以改造。协议分析技术与模式匹配技术相结合,是异常检测与误用检测的一种结合方式,减少了需要匹配的数据包数量和长度,既提高了检测效率和准确率,又保留了模式匹配的优点。由于协议分析器采用模块化插件技术,规则库也按协议类型进行了分类,所以系统管理员可以对系统进行优化配置,依据实际需要选

(下转第 244 页)

2) 数据包浏览器,能够从日志的数据包解码并显示第三层和第四层的信息。

3) 根据指定的参数生成饼图和条状图,并统计数据。

在这一模块中涉及大量的数据库操作。概括起来有以下 3 个主要部分:

* 连接数据库。

```
$ db = mysql_ connect (" localhost "," root "," lusnort ") ;
mysql_ select_ db(" mydb ", $ db) ;
```

* 查询数据库。

```
$ temp_ sql =" SELECT DISTINCT layer4_ sport, COUNT(event. cid) ".
" FROM event ". $ criteria[0] .
" WHERE ". $ criteria [1] ." AND layer4_ sport is NOT NULL ".
" GROUP BY layer4_ sport ORDER BY layer4_ sport ";
$ result = mysql_ query( $ temp_ sql, $ db) ;
```

* 按照需要取回数据的两种方式。

```
a. $ myrow = mysql_ fetch_ row( $ result)) ;/* 结果按数组下标取回:$ myrow[i] ;
b. $ myrow = mysql_ fetch_ array( $ result) ;/* 结果按字段取回:$ myrow[" 字段名 "] ;
```

另外,在该模块中提供绘图功能,可以在网页中通过设定特定的参数,绘制统计图 style[pie | bar]。绘图步骤:

(1) 创建基本 PS 对象(假设为 $ image),填充背景,以后的全部 PS 操作都是基于这个背景图像的;

(2) 在 $ image 上作图;

(3) 输出这个图像。

以饼图为例,其设计思想是:首先以用 imagecreate()来生成一个空白图形,然后在空白图形中用 imagearc()圆弧函数先画圆弧,再画两条线连接圆心和圆弧端点,再用 imagefilltoborder 函数来填充扇形。

2 结束语

从网络安全多层次的防御的角度出发,入侵检测已经受到越来越多的关注,入侵检测技术也有了长足的发展。然而,入侵检测依然面临着许多问题:随着能力的提高,入侵者会研制更多的攻击工具,使用更为复杂精致的攻击手段;攻击者采用加密手段传输攻击信息;入侵检测系统自身的安全性也面临考验;过高的错报率和误报率;日益增长的网络流量导致检测分析难度加大,高速的网络环境导致很难对所有数据进行高效实时分析;响应单元的具体职能的定义和实现方法还不完善等[5]。

文中提出了一个以 Snort 为内核的轻量型入侵检测系统。Snort 采用 Libpcap 捕包,但这并不是最有效的方法。因为它一次只能处理一个包,成了制约入侵检测对高带宽(1 Gbit/s)网络进行监控的瓶颈。将来打算加载捕包模块到 Linux,使其内核本身进行捕包。同时,该系统没有提供网络设备联动,而仅仅是对入侵事件统计分析。这一方面也有待完善。

参考文献:

[1] 蒋国春,冯登国. 网络入侵检测原理与技术[M]. 北京:国防工业出版社,2001.

[2] 刘 武,段海新,杨 路,等. 基于 Web 的网络入侵检测取证系统的设计与实现[J]. 计算机应用,2003,23(5):50 - 52.

[3] DuBois P. MySQL 网络数据库指南[M]. 北京:机械工业出版社,2000.

[4] Koziol J. Snort 入侵检测实用解决方案[M]. 北京:机械工业出版社,2005.

[5] 俞晓雯,高 强,丁 杰. 一种入侵检测取证系统模型的设计[J]. 微机发展,2004,14(8):117 - 119.

(上接第 241 页)

择相应协议分析器和规则集,减少了需要匹配规则数。通过这些方法,拓宽了模式匹配模块的瓶颈,达到了提高检测效率和准确性的目的。

参考文献:

[1] 斯海飞,赵国庆. 入侵检测技术分析概述[J]. 电子对抗技术,2002,17(2):31 - 34.

[2] 宋劲松. 网络入侵检测——分析发现和报告攻击[M]. 北京:国防工业出版社,2004:77 - 94.

[3] Stevens W R. TCP/IP 详解(卷一:协议)[M]. 范建华等译. 北京:机械工业出版社,2002.

[4] 刘学波,孟丽荣. 高速网络环境下的网络入侵检测系统的研究[J]. 计算机工程与设计,2005,26(5):1236 - 1238.

[5] 侯方明,李大兴. 一种新的基于协议树的入侵检测系统的设计[J]. 计算机应用研究,2005,22(7):150 - 152.

[6] 罗桂琼. 基于协议分析的入侵检测系统[J]. 电脑与信息技术,2005,13(4):56 - 59.