# 网络异常的主动检测与特征分析

赵 辉,张 鹏
(中南大学 信息与通信工程系,湖南 长沙 410083)

**摘 要**:目前入侵检测系统主要使用的技术还是特征检测,它只能检测已知的入侵,而异常检测尽管能检测未知入侵,却无法保证准确性和可靠性。特征检测是建立在对特征的准确定位基础之上的,而异常检测是基于不可靠行为的,只能描述某种行为的趋势。文中对基于异常和特征的入侵检测系统模型做了一定研究,把网络异常特征与异常检测技术结合,提高了入侵检测系统的检测效果。

**关键词**:异常检测;特征提取;主动检测
**中图分类号**:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2009)08-0159-03

# Active Detection and Feature Analysis About Network Anomaly

ZHAO Hui, ZHANG Peng
(Department of Information and Communication Engineering, Central South University, Changsha 410083, China)

**Abstract**: Currently the main technology of intrusion detection system is feature detection, which can only detect the known intrusion, and anomaly detection can be used to detect the unknown intrusion, it is unable to ensure its accuracy and reliability. While anomaly detection is based on uncertain behavior, which can only describe the trend of behavior, feature detection is based on accurate feature locating. In this paper proposed a method which incorporate anomaly detection and feature detection to gain better performance, and also discussed the intrusion detection system model based on feature detection and anomaly detection.

**Key words**: anomaly detection; feature extraction; active intrusion

## 0 引 言

网络安全问题愈来愈受到人们的关注,在此方面首先要做的工作是当网络出现异常时,能够快速发现问题并找出其根源。对网络故障或其它原因引起的网络异常的研究,将有助于发现网络中更细微的变化,提高发现问题的能力,提高警报信息的准确性,进一步提高网络服务质量。而目前所应用的各种检测方法中,都需要对网络中各种攻击行为进行分析,同时进行必要的侦查、取证,而这其中最为关键的就是要从大量数据包文中提取有效的攻击特征,对攻击特征进行准确的定位[1]。根据网络异常时的特征进行检测和诊断,可以保证具有更低虚警率,同时还有更高的检测率。

## 1 网络异常的主动检测

### 1.1 特征检测与异常检测

特征检测又称误用检测,是根据已知的攻击特征建立一个特征库,然后将网络采集的数据与特征库中特征进行一一匹配,若存在匹配的特征,则表明其是一个入侵行为。

异常检测也称为基于行为的入侵检测,一般方法是建立一个对应"正常活动"的系统或用户的正常轮廓。检测入侵活动时,异常检测程序产生当前的活动轮廓并与正常轮廓比较,当活动轮廓与正常轮廓发生显著偏离时即认为是入侵,从而触发相应机制。

特征检测的优势在于能够对已知入侵作出准确判断,异常检测的优势在于对网络异常状况监测的主动性,因此,可以考虑将特征检测中的特征匹配方法与异常检测中对异常数据流行为的监视方法相结合。在以往各种异常检测的方法中,都需要抽取具有正常(或者非正常)行为的样本,每个样本都有其特征定义,即从IP地址、端口号、协议类型等信息得到特征描述,然后归类为各种样本或者搭配成各种组合特征集或者描述为特征向量。异常检测能发现未知入侵,而基于特征的检测能发现已知入侵,因此异常检测与特征检测相结合[2],才能更好地适应复杂多变的网络环境,同时,作为异常检测核心之一的模型建立方式,也离不开对

**收稿日期**:2008-12-03;**修回日期**:2009-02-27
**基金项目**:教育部全国教育科学计划(2006JKS2007)
**作者简介**:赵 辉(1957-),男,副教授,研究方向为网络信息系统、分布式多媒体、现代教育技术等。

特征的分析。

### 1.2 网络异常的常用主动检测方法

在网络监测或管理系统中,网络中的异常通常是通过网络流量异常或者其他设备异常特征来判断。基于流量异常的检测方法有很多,较常用的有基于域值的检测方法、基于统计的检测方法、基于小波的检测方法、基于马尔可夫等随机过程模型的方法和一些基于机器学习、数据挖掘和神经网络等检测方法,这些方法是建立在对特定网络对象的实时监控基础上的[3]。从检测对象来讲,能反映网络异常的方面有很多,比如 NetFlow、日志数据、端口状态以及基本设备状态等,这些都能为我们提供最为直接的信息,只有对这些信息做到较好的统计,才可以利用各种数学的、生物的模型来进行分析与检测。

根据校园网的实际情况,在流量统计方面主要应用以下几个手段来做到主动、实时地对校园网络进行检测,分别为:

(1) NetFlow 的网络异常流量检测。它的主要优势在于首先 NetFlow 直接在路由器上采集流的信息。只要是通过这个路由器的 IP 流,它都可以进行检测;其次,NetFlow 的流记录信息不涉及到高层信息;第三,异常检测不需要庞大且不断增长的特征库,对于数据记录只需对照正常的数据行为基准,就可以进行检测各行为是否偏离这个基准即可。

(2)Syslog 日志检测。通过设置 Syslog 日志服务器,对日志数据接收之后,然后对 Syslog 包进行解析及预处理如果检测到可疑的异常行为,发出告警信息。

(3)端口定时扫描。通过 MIB(Management Information Base,管理信息库)对网络设备端口状态进行读取,并且对不同主机的同一端口以及特定主机的不同端口进行统计,对转发包数量,以及丢包率等特征进行抽取。这种方法主要是基于端口的统计特征[4],当发现有较大浮动时,发出警告,并抓取相应数据包进行分析。主要统计的数据有如下几方面(见表 1)。

另外,在对网络设备的检测中,由于很多的网络攻击行为必然导致对于网络设备的异常,而对于校园网中,最明显的影响表现在 CPU 的利用率上,根据实际的经验,在网络受到攻击时,通常不会仅仅存在少量的数据包,而是在短时间内数据包的数量剧增,特别是 ARP 攻击和其他不符合标准 IP 格式的数据包,会直接使 CPU 的利用率在几秒钟内由平时的 10% 左右上升至 35% 以上。由此,通过实时的检测,分别获取每 5 秒、1 分钟、10 分钟内,CPU 利用率情况,如果发现异常,则根据级别发出不同警报,同时抓取该时段的数据包,如图 1 所示。

表 1 端口信息统计表

| 统计项 | 说明 |
|---|---|
| 端口索引 | 标记每个端口 |
| 端口状态 | 显示端口是否开启或关闭 |
| 每秒流入/出包数 | 对每个端口的进出数据包进行统计,通过其数据量可以了解某个端口的负载情况 |
| 每秒流入/出包变化率 | 包的变化浮动能够说明网络的实时状态,特别是在遭受攻击时,变化率会更加明显 |
| 每秒丢包率 | 如果丢包率出现较大偏差,则从一个方面说明网络设备的负载存在问题或者被人侵 |

查询CPU利用率 冲突mac分析 网关冲突分析 拓扑环路分析 [redacted] Syslo

检索条件: ALL 排序方式: 1分钟数据 确 定

共326521条信息 1008/页 页码: 1 每页显示的条数: 1008 修 改

共324页 当前第1页 首页 上一页 下一页 尾页

| 序号 | 设备名 | 设备IP | 模块号 | vlan号 | 最近5秒钟内利用率 | 最近1分钟内利用率 | 最近5分钟内利用率 | 取值时间 |
|---|---|---|---|---|---|---|---|---|
| 1 | protal_sl | 202.197.48.70 | 100 | 55 - 65 | 95% | [redacted] | 53% | 2007-10-11 15:46:01 |
| 2 | S6506:STV | 202.197.48.30 | 1 | 2 - 8 | 78% | [redacted] | 77% | 2008-05-11 03:25:01 |
| 3 | S6506:STV | 202.197.48.30 | 1 | 2 - 8 | 78% | [redacted] | 36% | 2008-10-18 20:05:01 |
| 4 | S6506:STV | 202.197.48.30 | 2 | 9, 11 - 14, 16 | 78% | [redacted] | 77% | 2008-09-29 01:45:02 |
| 5 | S6506:STV | 202.197.48.30 | 2 | 9, 11 - 14, 16 | 78% | [redacted] | 78% | 2008-09-29 01:55:01 |
| 6 | S6506:STV | 202.197.48.30 | 2 | 9, 11 - 14, 16 | 77% | [redacted] | 78% | 2008-09-29 02:25:01 |

图 1 CPU 状态显示

## 2 特征的描述和提取

特征是对已知的异常行为的描述。特征必须能准确描述某一种网络异常的特点,对于每种网络异常现象,都应该对其相应的恶意数据包进行分析,提取攻击特征。一个特征应该是一个数据包或数据包序列的独有特性。理想情况下,通过特征应该总是能够发现网络异常的源头,同时应该不会对正常的网络流量造成影响,而异常检测的预处理则是根据一定的要求通过数据包的包头获取所需的网络连接信息。在假设绝对安全、完全正常的环境中收集网络连接信息,并按照一定的采样间隔,在一定时间内形成正常网络连接库,假定基于特征的检测能够立即以某种方式对攻击进行特征描述,那么通过匹配已知的入侵来识别模式,包括协议分析和特征分析,最终可以完成对网络异常的正确判断。

可以把异常检测所建立的用户异常行为特征建立数据库,把相同或者相似的用户行为归结为一类,把按时段划分的各范围内网络业务量统计并建立特征,在对网络数据流进行一系列的数据预处理,经过原始报文的捕获,将会话还原形成数据流,也就是把网络业务

量数据流转化成为特征向量的形式。此外,需要进一步对数据流进行特征提取,即用数据流特征代替数据流进行异常分析。并根据当前行为对异常行为特征库的匹配程度判断用户行为异常的根源。

主要的特征分析及提取方法可以有以下几种:

(1)基本网络协议方面的特征描述。

由于 TCP/IP 协议是目前使用最广泛的网络互联协议,也是最经常被攻击破坏的一个大目标,因此,在研究网络特征描述和提取的问题时,根据 TCP/IP 协议体系,主要考虑的数据包是 TCP 数据包、UDP 数据包和 ARP 数据包等。如何在大量的数据中提取出具有代表性的特征模式,用于对程序或用户行为做出描述,是实现检测的关键。在对由 Winpcap 截获的 TCP-Dump 格式数据进行分析和待征提取,从包含大量冗余信息的数据中提取出尽可能多的安全信息,抽象出有利于进行判断和比较的特征集合,这种特征可以用于基于特征检测的特征向量模型[5],也可以用于基于异常检测的行为描述模型。

针对网络攻击,可以对 IP、TCP、UDP、ARP 几种常用的网络协议进行分析,以其数据包的包头中得到能够表征攻击特征的不同字段的特殊取值作为攻击特征的基本网络协议特征[6]。例如从 IP 包头中提取有用的几个特征:源 IP 地址、目的 IP 地址、协议类型、时间戳等。通过对攻击进行网络协议分析,用协议包头的特殊字段值来标示网络攻击特征,例如,在中南大学网络中心出现的一种异常现象,每次出现该异常的主要表现就是 CPU 会告警,由此抓取该时段的数据包,通过分析找到其中多条记录的协议类型有问题,如图 2 所示。

```
00 e0 4c 91 ed 17 00 e0  fc 61 3b 83 08 00 45 00
00 8a 00 00 40 00 2f 11  a6 4e 7c 59 1e 1d ca c5
3f d9 23 28 07 41 00 76  54 6e 02 04 00 6e 09 07
06 77 03 0f 00 03 04 e7  0c ae 00 18 d1 77 52 07
02 00 00 00 61 05 0a d5  b3 dd 19 3e 03 e0 3f a4
c0 ce dc 72 57 2b 1c dd  47 48 fe f6 70 68 f5 28
c7 1a 3d 5c b9 29 1d 26  75 8b 00 00 00 00 00 00
00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00
```

图 2　有问题的数据包

将该数据包以十六进制显示后,从图中可以看出,该记录外层为 IP 包头(0800),然后是 UDP(11),但是再往下层分析,其协议类型为“e7”,无法判断该类型为何种协议,通过对多次数据的分析,该记录源端口总是为“2328”,源 IP 地址为 “00 18 d1 77”,而这正是路由器的一个端口,可以把这类特征作为一种判断的标准,在以后发现与此匹配的异常特征,则可以直接对该端口作处理,而不必从大量数据包中查找根源。

(2)从网络流量特征描述异常特征。

很多网络异常并不是通过发送一两个数据包就可以表现出来的,而是通过频繁发送大量数据包才会引起的,因此需分析一段时间内的网络流量来描述这些异常的特征。实时从网络流量中提取一些网络流量的基本特征数据,比如流量的大小、包长的信息、协议的信息、端口流量的信息、TCP 标志位的信息等,根据实际需要选取不同方面组合,从而得到各类异常特征的模型,而这种模型是根据需要实时设置和改变的[7]。

比如现在比较常见的 ARP 欺骗,其常用的方式可以简单归结为 3 种:(1)拒绝服务攻击;(2)冒充网关;(3)ARP 溢出攻击。这几种攻击行为都需要发送大量的数据包才能够实现,因此,可以通过设置时间窗口来构造流量特征,然后实时抓取数据包,提取出一组数据包的特征,从而描述网络异常特征并找出是哪些设备或主机是攻击源头。

以下是读取 ARP 数据包的包头,提取各条记录的信息:

```
my( $seeks, %arrp);
$seeks=0;
%arrp=(); # reset
$arrp{'arp_time'} = &returntime( $_[0],0,8); # time
$arrp{'arp_rel_length'} = &repack10( $_[0],0,4); # real length
$arrp{'arp_net_length'} = &repack10( $_[0],0,4); # net length
$arrp{'arp_des_mac'} = &repack16( $_[0],0,6); # des mac
$arrp{'arp_res_mac'} = &repack16( $_[0],0,6); # res mac
$arrp{'arp_type'} = &repack16( $_[0],0,2); # arp type 0806 for arp pack
$arrp{'arp_hard_type'} = &repack16( $_[0],0,2); # hard type 0001 for 以太网
$arrp{'arp_proc_type'} = &repack16( $_[0],0,2); # proc type 0800 for ip
$arrp{'arp_hard_length'} = &repack16( $_[0],0,1); # hard length 06 fo 以太网
$arrp{'arp_proc_length'} = &repack16( $_[0],0,1); # proc length 04for ip
$arrp{'arp_reply_flag'} = &repack16( $_[0],0,2); # quest/reply $arrp{'arp_des_ip'} = &anip(&repack16( $_[0],6,4)); # des ip
$arrp{'arp_res_ip'} = &anip(&repack16( $_[0],6,4)); # res ip
$arrp{'arp_pack_id'} = $arrp{'arp_time'};
$seeks= $arrp{'arp_rel_length'} -42;
seek( $_[0], $seeks,1); #
```

(3)攻击的其他特征的描述。

面对复杂的网络环境,仅从几个方面就能有效地描述网络异常的各类特征,显然是不能完全实现的,还

(下转第 165 页)

## 3 结束语

描述了引入时间约束的角色访问控制,更好地满足了访问控制最小权限的原则 ,可以有效地解决时间敏感活动的访问控制问题,增强了访问控制的力度。引入时间后的系统有着更全面、更具体的安全属性描述能力。但是仍然存在许多值得研究的问题,关于时间约束在角色访问控制中的实际应用应进一步进行研究,文中在时间约束的角色访问控制中的整体方面还有待进一步完善。

**参考文献:**


[1] Sandhu R S, Coyne E J, Feinstein H L, et al. Role—Based Access Control Models[J]. IEEE Computer, 1996, 29(2): 38 - 47.

[2] 张新华,陈军冰. 时间约束的 RBAC 模型及应用[J]. 计算机技术与发展,2007,17(6):246 - 249.

[3] Bertino E, Bonatti P, Ferrar E. TRBAC: A temporal Role - based Access Control Model[J]. ACM Transactions on Informarion and Systems Security, 2001, 4(3): 191 - 233.

[4] Ferraiolo D F, Sandhu R, Gavrila S. Proposed NIST standard for role - based access control[J]. ACM Transactions on Information and System Security, 2001, 4(3): 234 - 274.

[5] 黄　建,卿斯汉,温红子. 带时间特性的角色访问控制[J]. 软件学报,2003,14(11):1944 - 1949.

[6] 胡程瑜, 李大兴. 带时间约束和角色控制的工作流系统授权模型[J]. 山东大学学报,2006,36(3):39 - 43.

[7] 杨　珍,刘连忠. 时间约束的角色访问控制系统的设计与实现[J]. 计算机应用研究,2008,25(1):195 - 199.

[8] 张少敏,王宝义,周利华. 一种具有时间约束的基于角色的授权管理模型[J]. 武汉大学学报,2006,52(5):578 - 581.


(上接第 158 页)

**参考文献:**


[1] Sahami M, Dumais S, Heckerman D, et al. A Bayesian Approach to Filtering Junk E - Mail. In Learning for Text Categorization[C] // 1998 Workshop. Madison, Wisconsin: [s. n.], 1998.

[2] Duda R O, Hart P E, Stork D G. Pattern Classification 2nd [M]. [s. l.]: Wiley, 2002.

[3] 杨　斌, 路　游. 基于统计学习理论的支持向量机的分类方法[J]. 计算机技术与发展, 2006,16(11):56 - 58.

[4] 张　丽, 黄　东. 基于 Winnow 算法的反垃圾邮件引擎的设计与实现[J]. 计算机技术与发展, 2006,16(4):170 - 175.

[5] 成宝国, 冯宏伟. 一个基于 Naive Bayesian 垃圾邮件过滤器的改进[J]. 计算机技术与发展, 2006,16(2):98 - 99.

[6] 戴劲松,白英彩. 基于贝叶斯理论的垃圾邮件过滤技术[J]. 计算机应用与软件, 2006(1):110 - 111.

[7] 汤　伟, 程家兴, 纪　霞. 一种基于概率推理的邮件过滤系统的研究与设计[J]. 计算机技术与发展, 2008,18(8): 76 - 79.

[8] 龚　伟, 李柳柏. 基于 IDSS 的中文垃圾邮件过滤模型设计[J]. 计算机技术与发展, 2007,17(3):163 - 165.

[9] Linguistic Data Consortium(LDC)[EB/OL]. 2007. http://www.ldc.upenn.edu.

[10] Oxford Text Archive(OTA)[EB/OL]. 2007. http://ota.ahds.ac.uk/.


(上接第 161 页)

需要综合其他多渠道的信息并抽象其特征。另外可以通过对 Snort 规则库的分析,按照协议包头、协议类型、端口等划分、归纳出更多的网络异常特征,然后可以选择适当的串匹配算法,或者特征匹配算法来快速达到定位。

## 3 结束语

特征检测与异常检测是不能分割开来的,如何能更加有效地通过异常检测的手段来发现问题,结合特征检测的方式来提取异常特征,从而更加快速、准确地查找出异常网络的根源,是我们最终要实现的目标。文中主要是将两种检测手段的优势提取出来,并且把各自的特点相结合,在面对更加复杂的网络异常问题时,可以灵活地作出选择与判断,从而为应用各种模型来实现入侵检测作了比较充足的工作。

**参考文献:**


[1] 肖海军,王小非,洪　帆,等. 基于特征选择和支持向量机的异常检测[J]. 华中科技大学学报,2008,36(3):99 - 102.

[2] David F,王建新,王　斌. 基于异常和特征的入侵检测系统模型[J]. 计算技术与自动化,2004,23(3):19 - 22.

[3] Sielken R S. Application Intrusion Detection[R]. Virginia: University of Virginia, 1999.

[4] 王平辉,郑庆华,牛国林,等. 基于流量统计特征的端口扫描检测算法[J]. 通信学报,2007,28(12):14 - 18.

[5] 曹苏来. 典型攻击行为描述及特征向量提取[J]. 科技信息,2008(2):37 - 39.

[6] 李韦韦,陈　海,徐振朋. 基于多层特征匹配的网络入侵检测系统[J]. 计算机应用与软件,2008,25(2):278 - 280.

[7] Guimaraes M, Murray M. Overview of intrusion detection and intrusion prevention[C] // Proceedings of the 5th annual conference on Information security curriculum development. [s. l.]: [s. n.], 2008.