ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.2, February 2006, pp.295−304
DOI: 10.1360/jos170295

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于特征聚类的路由器异常流量过滤算法*

孙知信+, 唐益慰, 张 伟, 宫 婧, 王汝传

(南京邮电大学 计算机科学与技术系,江苏 南京 210001)

A Router Anomaly Traffic Filter Algorithm Based on Character Aggregation

SUN Zhi-Xin+, TANG Yi-Wei, ZHANG Wei, GONG Jing, WANG Ru-Chuan

(Department of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210001, China)

+ Corresponding author: Phn: +86-25-85198095, E-mail: sunzx@njupt.edu.cn, http://www.njupt.edu.cn

Sun ZX, Tang YW, Zhang W, Gong J, Wang RC. A router anomaly traffic filter algorithm based on character aggregation. *Journal of Software*, 2006,17(2):295−304. http://www.jos.org.cn/1000-9825/17/ 295.htm

Abstract: Under the situation of detecting attacks, current IDSs have no good reacting strategy to filter attack traffic. Based on network attacks' traffic characters, an anomaly traffic character aggregation algorithm (AFCAA) is put forward. Because normal DOS (denial of service)/DDOS (distributed denial of service) attack traffic has some characters in their packets' head, AFCAA uses the center of gravity theory to process statistic aggregation and aggregation partition based on the special field of the destination IP attack traffic in a fixed Euclid distance, and then it distills the center of attack traffic dynamically as the characters of attacks. Afterwards, through transmitting these characters to Net Filter, AFCAA can filter abnormal packets efficiently and protect the normal packet transmission. The experimental results show that the software router using AFCAA can efficiently find useful characters of prevalent DOS/DDOS attacks, reduce the harm of attack packets' spreading, and protect the limited network resources.

Key words: denial of service; distributed denial of service; router; character aggregation; anomaly traffic

摘 要: 基于当前入侵检测技术在检测到攻击的情况下没有良好的反应策略过滤攻击流量这一问题,提出了基于攻击流量特征聚类的特征提取算法 AFCAA(anomaly traffic character aggregation algorithm).针对一般DOS(denial of service)/DDOS(distributed denial of service)攻击流数据包头中具有某些相似的特性,AFCAA通过运用重心原理进行统计聚类,在一定的欧氏距离范围内对基于目的 IP 的攻击流样本相应字段进行聚类划分,动态地提取出攻击流的重心作为攻击的特征.然后,及时地把其特征传输给 Net Filter,可以进行高效的过滤,并保护正常流量的传输.实验结果表明,对当前流行的多种拒绝服务攻击,应用 AFCAA 系统的软件路由器都能够较准确地获取异常流量的特征,从而有效地进行过滤,减少攻击包传播的危害,保护有限的网络资源.

* Supported by the National Natural Science Foundation of China under Grant No.60573141 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2005AA775050 (国家高技术研究发展计划(863)); the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (教育部回国人员基金资助), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, Nanjing Government (南京市回国人员基金资助); the Scientific Research Foundation of Huawei Corporation (华为研究基金)

Received 2005-05-11; Accepted 2005-08-15

关键词: 拒绝服务(denial of service,简称 DOS);分布式拒绝服务(distributed denial of service,简称 DDOS);路由器;特征聚类;异常流量
中图法分类号: TP393 文献标识码: A

随着 Internet 在全球的普及,网络的安全问题日益为人们所重视.目前,全球的网络基本上使用的是基于 IPV4 的 TCP/IP 协议的体系结构,其本身有不足之处,为网络黑客的攻击提供了必要条件.现在比较流行的攻击手段为拒绝服务攻击,包括单机(DOS(denial of service))和分布式(DDOS(distributed denial of service))两种.在攻击时,利用大量的垃圾流量堵塞网络,使正常流量得不到有效、及时的传输,极大地降低了网络的可靠性和可用性.

由于拒绝服务攻击是符合 TCP/IP 协议的,所以现有的协议体系对此攻击的发生毫无免疫力.至今为止,该方法还是作为网络攻击最有效的方法之一,特别是分布使拒绝服务攻击(DDOS)具有 IP 欺骗、身份隐藏、远程控制的功能,使黑客的攻击网络难以追溯.为了保证网络的稳定、信息的安全,网络安全成为计算机行业的热门话题,有更多的人员投入到这个反黑客的斗争中,也相继提出了许多不同的算法和技术,取得了一定的成果.

随着软件技术的进一步发展,黑客技术也更新换代,网络攻击的类型和特性也进一步变得多样化、复杂化,使得网络攻击越来越难于被发现和制止,防止黑客攻击的战争任重而道远.

本文第 1 节是相关领域的研究介绍及其与本文的异同.第 2 节通过对 MIT 实验室所提供攻击数据集的分析,得出常见网络攻击异常流量的特征.第 3 节提出攻击流量特征聚类算法(AFCAA),并指出运用此方法来提取异常流量特征的要点.第 4 节利用 Red Hat Linux 9.0 提供的网络组件,结合对 Bloom Filter 算法的改进,把 AFCAA 算法运用于实际的网络环境中,使其成为检测和过滤为一体的系统,并且将此系统在 Red Hat Linux 9.0 的软件路由器环境下进行模拟,对其性能进行分析.第 5 节总结全文,并提出未来我们所需要做的工作.

1 相关性研究

由于在网络流量攻击方法中,拒绝服务攻击(DOS/DDOS)是比较有效的类型,并且发展比较快,所以许多论文都对 DOS/DDOS 攻击方面作了详细的分析,提出了许多不同的检测与预防的方法.下面,我们结合国内外的相关研究成果说明本文算法的不同点,并提出性能的优越性.

由于网络攻击的产生,入侵检测技术成为计算机研究的热门话题,许多研究人员都把重点放在入侵检测上,许多方法都能够在一定程度上判断出网络攻击行为的发生,由于 DOS/DDOS 攻击的包的特征具有可变性能,所以大多数都是基于流量的检测,如文献[1]中提出了累积和改进门限算法来检测网络攻击,文献[2]中通过对网络流量的自相关性进行分析来检测 DOS/DDOS.但是,许多方法只是说明了怎样检测攻击,很少详细地阐述如何对攻击进行反应,若有,也只是简单的说明.

文献[3]详细介绍了在 Linux 环境下,基于规则的分布式网络入侵检测系统 NetNumen 将异常检测和特征检测有机地结合起来.文献[4]提出一种自适应路由器限流算法防御分布拒绝服务攻击的机制,从对 DOS/DDOS 攻击的检测效果来看,比现有方法有明显的改善,具有较好的借鉴意义.

文献[5,6]提出基于服务器的运用认证机制来过滤异常流量.在服务器的前端设置异常流量过滤系统,过滤机制主要为设定信任用户数据库,对包头某些字段进行判别.如果为新用户,则必须运用其设定的特殊协议进行合法性认证,成为信任用户,否则流量被过滤.该方法面临的主要困难在于:攻击者也可以伪造认证报文.而当在路由器之间引入认证机制时,就需要在不同的管理域之间进行协商,实行难度较大.

基于分布式 DDOS 攻击数据包传递过程中的路由信息,文献[7]提出了路由追踪的方法来寻找源攻击地址,运用网络中路由器的合作,在源端过滤攻击包的方法.此方法可以有效地制止攻击的发生,但是一般 DDOS 攻击都是由傀儡机发攻击包,所以对于主控方地址还是无从查寻,且系统需要路由器之间的合同合作,从具体的实施来看比较困难.

在文献[8]中,Schnackenberg 等人设计了熵的方法,主要思想是提取数据包头中的 n 个字段,并且运用装箱方

法,把各个数据包按照字段匹配进入某个箱子,定时地运用熵的方法统计箱子里面的包的比例变化关系来检测攻击.包过滤的依据基于检测算法,对超出正常比例箱子中的包进行丢弃.该算法在监测方面确实比较新颖,但是,其过滤算法只是简单地丢掉超过比例的数据包,有一定的盲目性.

从以上分析可以看到,目前提出的各种方法虽然对 DDOS 检测有一定的效果,但是对于过滤方法都还存在较多的不足,不能有效地解决 DDOS 问题.新的 DDOS 反应机制应能保持与已有的网络机构的兼容,不应对路由器的功能作较大的改动,同时应尽可能地区分正常流量和异常流量.

结合已有的成果,本文提出了在一般 DOS/DDOS 攻击的情况下,区分正常和异常流量的流量特征提取算法 AFCAA(abnormal flow character assemble algorithm).通过在模拟路由环境下实验,该算法取得了很好的效果,并且该算法的结果可以作为路由转发的一项规则加入到路由器中,几乎不影响路由器的效率和功能.

2 常见异常流量特征分析

虽然 DDOS 攻击可以任意地伪造数据包,但是大多数攻击的流量还是有一定的特征可寻的.因为当今的许多 DDOS 攻击都不是由有专业技能的黑客发起的,而是由一些人运用黑客编写的网络攻击程序发起的.由于网络的发展和计算机的普及,使得黑客工具能够很容易在网上下载,使用方式也越来越简单.基于非专业黑客运用黑客工具发起的 DOS/DDOS 攻击,一般不会自己去修改黑客程序的代码,而是以默认的方式去攻击,使得攻击的特征更加有迹可寻.同时,由于 DOS/DDOS 攻击主要是运用大量的数据包来堵塞网络,为了达到攻击的效果,必须使发包的数据足够快,如果攻击方对每个攻击数据包都去做不同设计,就必然会使发包的速度变慢,影响攻击的效果.

我们对文献[9]中的由 MIT 实验室发布的 Darpa 数据集 LLS_DDOS_2.0.2-inside.dump 进行研究,得到了下面的一些特征.

对 TCP 的 DDOS 攻击可能的特征:总是发 ACK 包,为洪水攻击(TCP flag=0x0010);顺序生成源端口;应答序列号一致(Seq=0);随机生成目的端口;所有攻击 TCP 包长度固定(40,即不带数据).对于此类攻击,只要利用以上的其中 1 个或多个特点,就可以对其进行过滤.

同样地,我们对 UDP 和 ICMP 的 DDOS 攻击也作了相似的分析,得到下面一些结论.

对 UDP 攻击可能有的特征:攻击数据包的长度固定;随机生成目的端口;随机生成源端口.

对 ICMP 攻击可能有的特征:攻击数据包的长度固定;攻击包的类型固定(ICMP Flag 固定).

虽然对于不同的攻击可能特征不同,但是总可以找到某种特征来过滤攻击,除非是精心设计的攻击方法才能使攻击流和正常流不可区分,所以这些特征对于一般性的攻击都是有效的.

需要特别指出的是,与基于特征的入侵检测判断手段不同,本文所指的特征值都是可变的,是基于统计的,由一定的方法从攻击流量中动态生成,而不是从单个攻击包中寻找到的.

3 异常流量特征聚类算法(AFCAA)

基于上面的分析,在一次网络拒绝服务攻击中,数据包头内容中具有某些相似特性.异常流量聚类算法(AFCAA)就是在这个基础上产生的.

设一个攻击流中有 n 个数据包,我们把各个包看成是一个样本.算法的基本思想是,先将 n 个样品各自看成一个类,然后规定样品之间的距离.起初因为每个样品自成一类,所以类与类之间的距离等于样品之间的距离.选择距离最小的一组并成为一个类,接着计算新类与其他类的距离,再将距离最近的两类合并,这样一直持续下去,直到成为一个类为止.

从物理学中我们可以得到,一个物体的重心与其各个质点大小和位置相关,其中有如下关系:

设 M 为一个物体的重心点,物体中的各个质点的权重为 $\{m_1,m_2,m_3,\ldots,m_n\}$,各个质点到重心的距离为 $l_1,l_2,l_3,\ldots,l_i$,则重心 M 满足:

$$M=\min\left\{\left(\sum_{k=1}^{i}m_k*l_k\right)\Big/\sum_{k=1}^{i}l_k\right\} \tag{1}$$

把数据包中所要聚类的相应值作为权重,由于数据包的重心和各个数据包所在的先后次序无关,所以我们可以得到 $l_1=l_2=l_3=\ldots=l_i$,从而式(1)变为

$$M=\min\left\{\left(\sum_{k=1}^{i}m_k*l_k\right)\Big/\sum_{k=1}^{i}l_k\right\}=\min\left\{\left(l_1\sum_{k=1}^{i}m_k\right)\Big/(k*l_1)\right\}=\sum_{k=1}^{i}m_k/k \tag{2}$$

可以得到:一个数据包类可以用它的重心(该类样品的均值)作代表.这时,类与类之间的距离就可以用重心之间的差所代表的距离来合理地表示.

下面提出基于重心原理的 AFCAA 聚类算法.

定义 **1**. 设两个数据包某个特征字段类 G_p 与 G_q 的重心分别为 $\bar{x}_p$ 与 $\bar{x}_q$,则 G_p 与 G_q 的之间的距离就可以定义为

$$D_{pq}=d_{\bar{x}_p\bar{x}_q} \tag{3}$$

下面给出两个特征类之间欧氏(Euclid)距离的公式.

定义 **2**. 设第 i 个样品数据包字段值 x_i 与第 j 个样品数据包字段值 x_j 之间的距离为 $d(x_i,x_j)$,简记为 d_{ij},如果它满足下列条件:

(1) 非负性: $d_{i,j}\geq 0$, $d_{i,j}=0\Leftrightarrow x_i=x_j$,

(2) 对称性: $d_{i,j}=d_{j,i}$,

(3) 三角不等式:对任意 x_i,x_j,x_t, 有 $d_{i,j}\leq d_{i,t}+d_{t,j}$,

则它们的欧氏距离为

$$d_{i,j}\overset{\Delta}{=}\left[\sum_{i=1}^{k}(x_{i,t}-x_{j,t})^2\right]^{\frac{1}{2}} \tag{4}$$

下面给出 D_{pq} 的距离递推公式.

设某一步中 G_p 与 G_q 的重心分别为 $\bar{x}_p$ 与 $\bar{x}_q$,它们分别有样品数据包 n_p 和 n_q 个,将 G_p 与 G_q 合并为 G_r,则 G_r 内有样品数据包 $n_r=n_p+n_q$ 个,设它的重心为 $\bar{x}_r$,则有

$$\bar{x}_r=\frac{1}{n_r}(n_p\bar{x}_p+n_q\bar{x}_q) \tag{5}$$

某一个 G_t,它的重心是 $\bar{x}_r$, 由欧氏(Euclid)距离公式得到:

$$\begin{aligned}D_{t,r}^2&=d_{\bar{x}_t,\bar{x}_r}^2=(\bar{x}_t-\bar{x}_r)'(\bar{x}_t-\bar{x}_r)\\&=\left[\bar{x}_t-\frac{1}{n_r}(n_p\bar{x}_p+n_q\bar{x}_q)\right]'\left[\bar{x}_t-\frac{1}{n_r}(n_p\bar{x}_p+n_q\bar{x}_q)\right]\\&=\bar{x}_t'\bar{x}_t-2\frac{n_q}{n_r}\bar{x}_t'\bar{x}_p-2\frac{n_q}{n_r}\bar{x}_t'\bar{x}_q+\frac{1}{n_r^2}\left[n_p^2\bar{x}_p'\bar{x}_p+2n_pn_q\bar{x}_p'\bar{x}_q+n_q^2\bar{x}_q'\bar{x}_q\right]\end{aligned} \tag{6}$$

利用

$$\bar{x}_t'\bar{x}_t=\frac{1}{n_r}(n_p\bar{x}_t'\bar{x}_t+n_q\bar{x}_t'\bar{x}_t) \tag{7}$$

则有

$$\begin{aligned}D_{tr}^2&=\frac{n_p}{n_r}(\bar{x}_t'\bar{x}_t-2\bar{x}_t'\bar{x}_p+\bar{x}_p'\bar{x}_p)+\frac{n_q}{n_r}(\bar{x}_t'\bar{x}_t-2\bar{x}_t'\bar{x}_t+\bar{x}_q'\bar{x}_q)-\frac{n_pn_q}{n_r^2}(\bar{x}_p'\bar{x}_p-2\bar{x}_p'\bar{x}_q+\bar{x}_q'\bar{x}_q)\\&=\frac{n_p}{n_r}D_{tp}^2+\frac{n_q}{n_r}D_{tq}^2-\frac{n_pn_q}{n_rn_r}D_{pq}^2\end{aligned} \tag{8}$$

以上就是基于重心原理的 AFCAA 聚类算法的距离递推公式.

并类的原则是,域与域之间的距离最近的两域合并.其聚类步骤如下:

(1) 规定样品数据包之间的距离,计算出 n 个样品数据包的距离 d_{ij}(某字段的值差),i, j=1,2,3,…,n,得到对称矩阵 $D(0)$:

$$D^2(0)=\begin{bmatrix} 0 & & \text{对} & \\ d_{2,1}^2 & 0 & \text{称} & \\ d_{3,1}^2 & d_{3,2}^2 & 0 & \\ \cdots & \cdots & \cdots & \cdots \\ d_{n,1}^2 & d_{n,2}^2 & \cdots & d_{n,n-1}^2 \quad 0 \end{bmatrix}_{n\times n}.$$

开始,每个样品数据包都自成一类,所以 $D_{p,q}^2=d_{p,q}^2$.

(2) 选择 $D^2(0)$ 中除对角线外的最小的元素,设为 $d_{p,q}^2(=D_{p,q}^2)$,则将 G_p 与 G_q 合并成一个新类,记为

$$G_r=\{G_p,G_q\} \tag{9}$$

(3) 利用递推公式(8)计算新类 G_r 与其他类 $G_t(t\neq p,q)$ 的平方距离.

将 $D^2(0)$中第 p,q 行及第 p,q 列用公式(5)并成一个新行新列,新行新列对应于 G_r,所得到的矩阵记为 $D^2(1)$.

(4) 对 $D^2(1)$重复上述对 $D^2(0)$的第 2 步和第 3 步的作法,得到 $D^2(2)$.如此下去,直到按照某个字段所分的各个数据包类之间的距离达到一定的阈值为止.

运用此算法的主要目的是为了提取出异常流量的特征,如第 2 节所述.由于提取的是流量的特征,不是单个数据包的特征,所以算法必须缓存一定数量的异常流量数据包,并取出它们的相关字段进行比较分析,才能比较准确地确定是否有相应的特征.

由于缓存包个数多少直接影响到系统计算量的大小(计算量以缓存包数的平方增长),所以,确定缓存包个数的多少也是决定算法有效性的关键差数之一:

(1) 缓存包过少会使特征提取不准确,严重影响过滤效率,并且使系统频繁地进行特征提取,产生波动;

(2) 缓存包过多会使系统计算量过大,运行速度变慢,并且使特征确定变得复杂且不准确.

所以,要准确地确定异常流量的特征,就要确定适合的缓存包数目.本文对正常网络和受到攻击中的包特征变化规律进行了分析,如图 1 所示.

从图 1 可以明显地看到下面的特点:

(1) 在正常网络中,包总是一段一段进行传输的,一般为 10~15 个左右为一组,不会超过 30 个(分组包数大于 20 个的几率小于 5%),在同一个组中包的许多特征值是相同的.

(2) 在受到攻击时,对于一般的拒绝服务攻击,其包组数目明显地变大,一般都超过 50 个.不同强度的攻击包组数目分布不同,越强烈的攻击,其包组数目就越大.对于上文所分析的 Darpa 数据集,包组数目甚至达到几百.

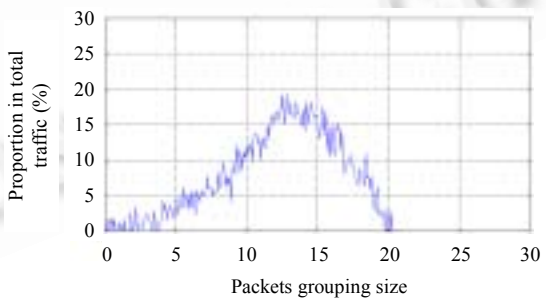

Fig.1 Traffic packets grouping state in normal condition

图 1 正常情况下流量包组大小分布情况

(3) 对于某些精心构造的攻击,会使得数据包没有特征,从而也无从得到包组大小分布,但这种攻击由于攻击方构造包过于复杂,其强度必然受到制约.

从上面的分析可以得到:缓存包数目可以设定的范围为一个大于 20 的值.由于缓存数据的交换是基于滑动窗口机制的,如果设置得大,只会过多地消耗资源并产生负面影响(不能很好地得到强度小的攻击的特征),所以,一般情况下设置为 20 左右,本系统在实现时设置为 20.但是,对于具体的网络情况可能会有所不同,需要根据具体情况进行调节.

结合第 2 节所提出的攻击字段特征,AFCAA 算法的主要功能是:对于确定固定字段的特征,即用来判断一定数量的字段是否具有一个汇聚类,也即有足够比例的包在这个汇聚类内.对于某个字段的随机特征,AFFCAA

算法就是确定一定数量的某字段在一定的距离内可以分成多少类,如果数量多于某个门限值,则认为此字段为随机的.由于随机主要是针对端口的,正常流量情况下都是一组数据包,其端口都是相等的,下一组到达时再变化为另一个端口,而许多异常流量数据包的端口则是随机或者线性变化的,所以,可以设定距离为 0 时 AFFCAA 的分类情况,以确定端口是否随机.

4 AFCAA 算法实现和性能分析

4.1 AFCAA算法的实现

4.1.1 基于目的地址聚集的 Bloom Filter 算法

为了运用 AFCAA 算法,首先必须获得异常流量的目的 IP 地址.文献[10]提出了基于目的地址聚集的 Bloom Filter 算法来检测 DDOS 攻击的方法,符合本文对确定攻击目的地址的要求,所以,本文采用此算法来获得受攻击 IP 地址.

为了适应本文的 ACCF 算法和改善 Bloom Filter 算法的性能.我们对此算法作了改进:

(1) 基于 TCP/IP 协议,我们在 Bloom Filter 算法的每个域 $a_{ij}(i<k,j<m)$设置了 4 个值 a_{ij}(TCP,UDP,ICMP,TOTAL),其中 TCP,UDP,ICMP 分别表示 TCP 包、UDP 包、IMCP 包,TOTAL 为 3 类包的总和.这样,我们在得到某个域溢出的同时,可以通过其 IP 包类型分布进行分析,同时得到是哪几类包对溢出起主要作用.

(2) 阈值问题是 Bloom Filter 算法的主要问题.对于阈值的设定,由于在网络中的数据流量都是符合一定的随机分布的,在一般情况下,对于核心路由器中网络流量的变化,在一定时间内可以作为一个正态分布来处理,从概率论原理可以得到其 99.73%的分布在以下范围:

$$x_n \in (\mu - 3\delta/\sqrt{n}, \mu + 3\delta/\sqrt{n}) \tag{10}$$

其中 $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 为其均值和方差值.

当发生攻击的时候,特别是拒绝服务(DOS/DDOS)攻击发生的时候,其统计特性就被破坏了,即产生了所谓的重心漂移.由于受到攻击时门限的值明显比正常的网络流量要大,为向右漂移,所以取式(10)中的上限值.考虑到门限过大会使低强度的攻击漏检,本文取门限为 $th=\mu+3\delta$,由式(10)得到 $x_n \in (\mu - 3\delta/\sqrt{n}, th)$ 的概率大于 99.73%,所以在理论上对于中高强度攻击的漏检率小于 1%,门限可以在初始无攻击的状态下,系统经过学习动态地获得,并且需要根据不同时间段进行动态的更新.

(3) 由于网络中 IP 地址是由 4 个 8bit 的 32bit 数值组成,所以我们自然地把 Hash 函数设为 4 个.本文设定 Hash 函数为其 8bit 的本身数值,每个数组总共为 $2^8 = 256$.由此可知,两个不同的 IP 地址映射到相同域的概率为 $p = (1/256)^4 = 2.328^{-10}$, 其 Hash 冲突的概率非常小,应该能够满足要求.

本文选择 Bloom Filter 只作为测试 AFCAA 算法之用.之所以选择 Bloom Filter 算法作为系统的检测部分,是因为其实现方便、内存使用率低,并且其对目标 IP 分离.在许多文献中,提出了基于统计的检测算法,有的算法可能比 Bloom Filter 算法要好,AFCAA 系统完全可以移植到其他检测系统中,作为检测完后的包过滤模块运行.

4.1.2 系统实现结构

本文基于 Red Hat Linux 9.0 系统的软件路由器做了原型系统,并通过 Net Filter 组件来实现对软件路由器的取包和过滤操作.具体的系统结构如图 2 所示.

由图 2 可以看出,程序实现分为用户空间部分和内核空间部分,中间通过 NetLink 进行消息通信,并且在每个模块设置了日志输出模块,以记录系统整个模块的运行情况.

由于本系统应用的环境为网络中的核心路由器,其端口在单位时间内通过的数据包数量一定很大.考虑到本系统对路由器包处理能力的影响,所以在检测的时候运用了一定的抽样算法,在不失去网络流量原有特征的同时,减少系统开销,而过滤的时候则对每个包进行识别,以保证不漏掉攻击包.且系统在用户和内核之间,基于消息控制机制、效果反馈模块记录过滤的效果,并由特征反馈门限控制 AFCAA 算法进程的启动与休眠,进一步

减少资源的消耗,提高运行效率.

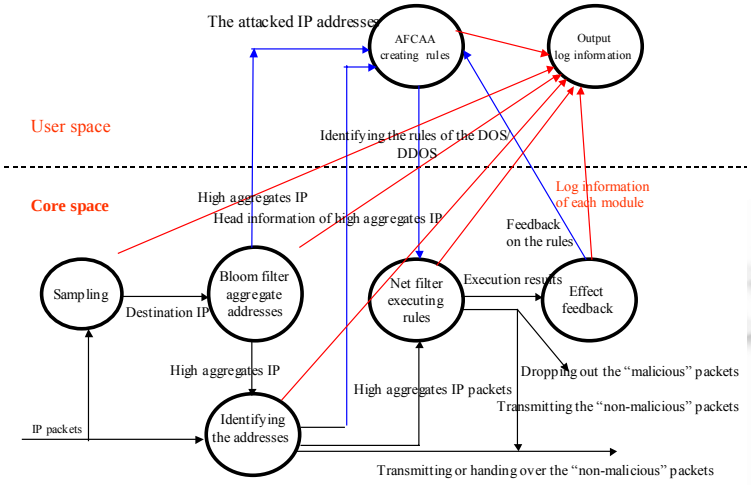

Fig.2 AFCAA system structure chart

图 2 AFCAA 系统结构图

4.2 系统测试

4.2.1 攻击性能测试

本文在模拟的路由环境中对本算法的原型系统进行了测试,攻击机经过加载 AFCAA 系统的软件路由器,对目标 IP 地址发动 DOS/DDOS 攻击.

对 AFCAA 算法的参数设置情况为:缓存 20 个数据包的信息进行 AFCAA 算法分析,对于某个被攻击的目的 IP 地址和攻击类型,分别提取符合条件的数据包的包头相关字段进入 AFCAA 算法队列;判断某个字段固定的聚类算法距离门限为 1;判断端口随机的聚集算法距离门限为 0,随机门限为 50%(如果不同类的数量超过统计量总数的 50%,则确定为随机);特征反馈门限为 80%(如果过滤水平不到 80%,则系统需要重新激活 AFCAA 算法,再次判断流量特征).下面是对具体测试情况的分析.

(1) 纯拒绝服务攻击

首先,我们用纯攻击来测试本系统对攻击的识别和过滤情况.从图 3 可以看出,在 T1 时候攻击方对目标 IP 发起了典型的 TCP SYN Flood 攻击;在 T2 时刻,AFCAA 系统检测到了攻击,并给出了这个攻击的特征,把特征传递给内核的 Net Filter 后,Net Filter 开始根据特征来过滤异常流量.由于是对某个目标 IP 的攻击,所以不影响其他 IP 数据包在路由器的正常转发;在 T3 时刻,攻击停止,相应的过滤也就停止了.

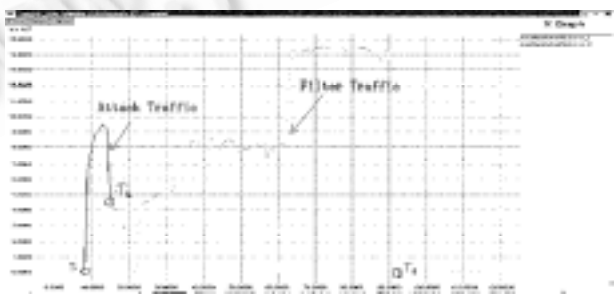

Fig.3 TCP SYN Flood attack traffic filter test chart

图 3 TCP SYN Flood 攻击测试流量过滤图

为了验证算法对其他拒绝服务攻击的有效性,我们不但做了不同类型的 TCP 攻击实验,而且还做了

UDP,ICMP 和 MIX(TCP+UDP+ICMP)等常见拒绝服务攻击实验,都获得了与 TCP SYN Flood 几乎类似的效果,证明了本算法确实能够得到拒绝服务攻击的明确特征.

(2) 加载正常流量的 ICMP Flood 攻击

下面,我们对加载一定正常流量的网络发起拒绝服务攻击,以说明本算法对区分正常流量和异常流量的有效性.从图 4 可以看到,在 T1 时刻,攻击方对目标 IP 发起了 ICMP Flood 攻击,并且与目的 IP 进行正常的文件传输;在 T2 时刻,AFCAA 发现了该攻击,并且递交流量特征给 Net Filter.通过 Net Filter 的过滤,大部分异常流量都被过滤掉了,余下的只是正常的流量,效果非常明显.

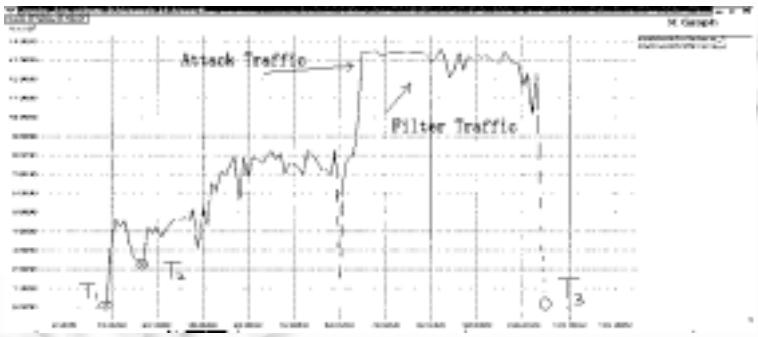

Fig.4 ICMP Flood attack with normal flow traffic filter test chart

图 4 带正常流量的 ICMP Flood 攻击测试流量过滤图

算法的另一个特点是对协议的隔离性.对于一种协议的攻击,算法不会过滤其他协议的数据包,所以就不会影响其他协议数据包的正常传输.本算法除了那些对数据包精心设置的攻击以外,对于当前网络中普通的 DOS/DDOS 攻击总是有效的,而且算法简单、易操作,其综合效应令人满意.

4.2.2 算法对比测试

现有的许多关于 DDOS/DOS 的算法对于怎样检测叙述得比较多,也提出了许多高效的检测方法(如文献[1,2]中的检测算法),而对于检测之后怎样过滤攻击流量、保护正常流量,则提出得比较少,有的甚至一笔带过.为了验证算法在包过滤方面的优越性,本文在 Bloom Filter 算法的基础上,参考了文献[8]中 Feinstein 等人提出的包比例失衡过滤方法,与本文的 AFCAA 算法进行比较.

本文所提到的流量特征主要分为两类:某个字段(包长、应答序列、TCP flag 标记等)的固定和端口字段的随机.在过滤时,分别为过滤某个字段固定的包和过滤端口不连续的包.根据不同的流量特征、文本,分别在不同攻击包比例的情况下对其过滤的效果与 Feinstein 等人所提出的包比例失衡过滤方法进行了比较.

(1) 字段固定特征包过滤(如图 5 所示)

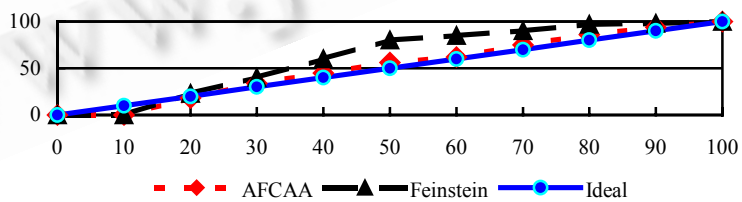

Fig.5 The contrast chart of filter performance in the character of field static attack

图 5 字段固定特征的攻击下包过滤性能对比图

(2) 端口随机特征包过滤(如图 6 所示)

从图 5 和图 6 中可以看出,由于基于比例失衡的 Feinstein 算法只是根据数据包不同的协议类型来丢弃某一种协议的包,因而具有一定的盲目性.在攻击包比例占总包数 50%以上的时候,其丢包率就达到了 80%以上,而

其中有大约 30%的包并不是攻击包,虽然一样地阻止了 DDOS/DOS 攻击,但却牺牲了其他网络服务.

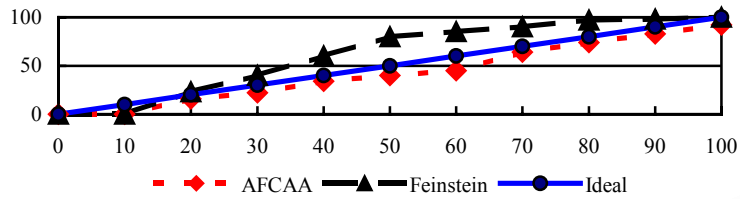

Fig.6 The contrast chart of filter performance in the character of random port attack

图 6 端口随机特征的攻击下包过滤性能对比图

而 AFCAA 算法主要是基于包特征对比下的过滤,减少了过滤策略的盲目性.图 5 是以流量某字段固定为特征的攻击(典型的攻击为 TCP SYN flood)的过滤方法.由于一次攻击产生的数据包之间某个字段是固定的,系统用 AFCAA 算法得到这个字段的值后,对每个包的相关字段进行对比,所以理论上的漏检率为 0%.由于网络中的正常包也有一定的几率正好和攻击包的特征相同,从而被系统过滤掉(在大型的路由器中因数据量较大,这是经常发生的),所以在图 5 中,实际过滤的值要比理想情况下稍微大一点.根据一般情况下 IP 包长的变化范围,其理论上的发生概率为 $1/(65535-40)\approx 0.0015\%$,相比之下,还是非常小的.对于有明显固定值特征的拒绝服务攻击,其过滤精确度达到 95%以上.

对于图 6 中端口随机特征攻击(典型的为 UDP flood 攻击)的过滤,采用的原则是:宁可放弃过滤少量攻击包,保护正常包的通过(其误检率理论上达到 0%).采取的方法是:在经过路由器的包序列中,放过所有相关端口与其相邻包的端口相同的包,其余的过滤.由于运用随机进行过滤的方法具有不确定性,所以,其总体效果没有固定字段的方法效率高.但是,对于 Feinstein 算法,其盲目性还是比较小的,在实际检测中,一般都能过滤 80%以上的攻击包.

由于 AFCAA 算法的检测效率与攻击流的特征有密切关系,所以并不是固定的:好的特征过滤效率高,差的特征过滤效率也差.每一个攻击都有其不同的效率,对于某些特殊环境下精心构建的攻击,甚至失去流量特征,所以此算法并没有一个确定的效率理论值.但对于当前出现的一般攻击方法,此算法还是相当有效的.如何进一步挖掘拒绝服务攻击的有效特征,是我们下一步所要做的工作.

5 总 结

本文介绍了一种新的异常流量特征提取方法——基于流量聚类的特征提取算法 AFCAA.该算法能在攻击检测的基础上动态地生成攻击流量的特征,以此可以用统计的方法来区分攻击和正常数据包,以便于过滤攻击包,同时保护正常流量传输.通过测试,本方法能够有效地区分正常流量和攻击流量,减少攻击包传播的危害,从根本上来抵御 DOS/DDOS 攻击.

与基于特征的入侵检测手段不同,AFCAA 所提取的特征是基于统计的,是从攻击流量中动态生成的,而不是从单个攻击包中提取的.由于是在不开包的情况下来过滤数据,所以效率较高.并且,因其在低于 3 层的 TCP/IP 协议范围内进行检测和过滤,故对于未来的 IPV6 协议,本算法也同样是适用的.

今后要做的工作是发现更多、更明显的攻击流特征,使系统能够更准确地区分攻击包和正常包,更好地保护正常流量的传输,维护网络安全.

References:

[1] Siris VA, Papagalou F. Application of anomaly detection algorithms for detecting SYN flooding attacks. In: Regency H, ed. Global Telecommunications Conf. (GLOBECOM 2004). Dallas: IEEE, 2004. 2050−2054.

[2] Jin SY, Yeung DS. A covariance analysis model for DDoS attack detection. In: Baal-Schem J, Bregni S, eds. Communications, 2004 IEEE Int'l Conf. Paris: IEEE Communications Society, 2004. 1882−1886.

[3] Li W, Wu LF, Hu GY. Design and implementation of distributed intrusion detection system NetNumen. Journal of Software, 2002,13(8):1723−1728 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/1723.pdf

[4] Liang F, Yau D. Using adaptive router throttles against distributed denial-of-service attacks. Journal of Software, 2002,13(7): 1120−1127 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/1120.pdf

[5] Thomas R, Mark B, Johnson T, Croall J. NetBouncer: Client-Legitimacy-Based high-performance DDoS filtering. In: Wermer B, ed. DARPA Information Survivability Conf. and Exposition 2003. Washington: Institute of Electrical and Electronics Engineers, Inc., 2003. 14−25.

[6] Kim YH, Jo JY, Chao HJ, Merat F. High-Speed router filter for blocking TCP flooding under DDoS attack Performance. In: George AD, Johnson E, Richard GG, Xue GL, eds. Computing and Communications Conf. 2003. Phoenix: IEEE Computer Society, 2003. 183−190.

[7] Sung M, Xu J. IP traceback-based intelligent packet filtering: A novel technique for defending against Internet DDoS attacks. IEEE Trans. on Parallel and Distributed Systems, 2003,14(9):861−872.

[8] Feinstein L, Schnackenberg D, Balupari R, Kindred D. DDoS tolerant networks. In: Wermer B, ed. DARPA Information Survivability Conf. and Exposition, 2003. Washington: Institute of Electrical and Electronics Engineers, Inc., 2003. 73−75.

[9] 2000 DARPA intrusion detection scenario specific data sets. 2000. http://www.ll.mit.edu/ll.mit.edu/SST/ideval/data/2000/2000_data_index.html

[10] Chan EYK, Chan HW, Chan KM, Chan VPS, Chanson ST, Cheung MMH, Chong CF, Chow KP, Hui AKT, Hui LCK, Lam LCK, Lau WC, Pun KKH, Tsang AYF, Tsang WW, Tso SCW, Yeung DY, Yu KY. IDR: An intrusion detection router for defending against distributed denial-of-service (DDOS) attacks. In: Hsu FD, Ibarra OH, Saldana RP, eds. Proc. of the 7th Int'l Symp. on Parallel Architectures, Algorithms and Networks. Los Alamitos: Ateneo de Manila University, 2004. 581−586.

附中文参考文献:

[3] 李旺,吴礼发,胡谷雨.分布式网络入侵检测系统NetNumen的设计与实现.软件学报,2002,13(8):1723−1728. http://www.jos.org.cn/1000-9825/13/1723.pdf

[4] 梁丰,YAU D.利用路由器自适应限流防御分布拒绝服务攻击.软件学报,2002,13(7):1120−1127. http://www.jos.org.cn/1000-9825/13/1120.pdf

孙知信(1964－),男,江苏南京人,博士,教授,主要研究领域为计算机网络与安全,计算机仿真,软件工程.

宫婧(1978－),女,讲师,主要研究领域为计算机网络与安全.

唐益慰(1982－),男,硕士生,主要研究领域为计算机网络与安全.

王汝传(1944－),教授,博士生导师,CCF 高级会员,主要研究领域为计算机网络与安全.

张伟(1973－),男,博士生,讲师,主要研究领域为计算机网络与安全.