第 34 卷 第 1 期 计 算 机 工 程 2008 年 1 月
Vol.34 No.1 Computer Engineering January 2008
·安全技术· 文章编号：1000—3428(2008)01—0175—03 文献标识码：A 中图分类号：TP309

基于地址聚集防抖动异常流量控制系统

吴国庆
(扬州市信息中心，扬州 225009)

摘 要：提出一种新的异常流量检测方法——基于地址聚集的防抖动异常流量检测系统 AWACS。该系统应用 Adapted-Bloom-Filter 算法对目的 IP 地址聚集，运用防聚集抖动的 CUSUM 算法检测是否有流量抖动的脉冲式攻击发生，使检测的结果更加准确，减少了系统的开销。该检测系统已作为一个独立的模块，成功运行于核心路由器中。

关键词：路由器；拒绝服务；防抖动；聚集

Address-aggregated Wobble-defended Abnormal-traffic Control System

WU Gu-qing
(Information Center of Yangzhou Municipality, Yangzhou 225009)

【Abstract】 This paper presents a new abnormal traffic detection method: address-aggregated wobble-defended abnormal-traffic control system. This system employs adapted-bloom-filter algorithm to assemble purpose IP addresses, then uses wobble-defended CUSUM algorithm to detect pulsing denial of service attack, the method can work exactly and it declines the spending of system resources at maximum. A detection system employing this method has been run successfully in routers as an individual module.

【Key words】 router; denial of service; wobble-defended; aggreation

随着Internet 的日益普及，连入其中的计算机数量激增，网络与人们日常生活的关系也越来越紧密。与此同时，网络安全问题也日益突出，入侵事件发生的频率越来越高，入侵的危害性也越来越大。根据CERT(Computer Emergency Response Team)的统计，拒绝服务攻击(DoS)在 1995 年以前每年的增长率接近 50%，2000 年 2 月，包括Yahoo, eBay 在内的数家大型网站在遭受拒绝服务攻击后瘫痪时间长达数小时[1]，CERT 的统计数字充分说明了拒绝服务攻击巨大的危害性。为保证网络稳定和信息安全，全世界的网络安全人员都投入到了这场反黑客的斗争中，也相继提出了许多不同的算法和技术，取得了一定的成果。文献[2]提出了一种基于目的IP聚集的Bloom Filter算法，简单地对目的IP地址进行聚集，设置门限，对超过其门限的IP地址进行告警。文献[3]提出了一个对SYN FLOOD进行检测的CUSUM(Cumulative SUM algorithm)算法，是基于在正常传输情况下的检测。另一种基于统计特性的分析方法是网络流量自相似性分析[4-5]，运用了网络流量的自相似性特性进行分析。文献[6]提出了一种利用协方差分析的异常流量检测算法，但要对大量数据进行操作，计算协方差、矩阵量化等，实现起来也有一定的困难。据此本文提出了一个路由器端异常流量检测的新方法：基于地址聚集的防抖动的异常流量检测方法。该方法主要分为两部分：(1)运用Adapted-Bloom-Filter算法对数据包的目的IP地址进行聚集；(2)由于Adapted-Bloom-Filter算法存在着检测连续性的问题，运用防聚集抖动的CUSUM算法进一步检测是否有流量抖动的脉冲攻击发生。

1 Adapted-Bloom-Filter 算法

1.1 Bloom Filter算法

最基本的Bloom Filter是 1970 年Burton Bloom提出的用来判断某元素x是否在集合$S = \{s_1, s_2,\cdots, s_n\}$中的一种数据结构。Bloom Filter是一个长度为m的 0-1 向量，初始时Bloom Filter的每一位的值为 0。$h_1, h_2,\cdots,h_k$是k个相互独立的散列函数，它们的值域为$\{0,1,\cdots,m-1\}$。对于每一个元素$w \in S$, Bloom Filter中对应的$h_1(w), h_2(w),\cdots, h_k(w)$位置被置为 1(如图 1 所示)。例如，使用$k = 4$ 个相互独立的散列函数，Bloom Filter为$m = 8$ 位的 0-1 向量，对于元素w，散列值为：$h_1(w) = 2, h_2(w) = 3, h_3(w) = 1, h_{24}(w) = 5$，则第 2、第 3、第 1、第 5 位的值被置为 1。

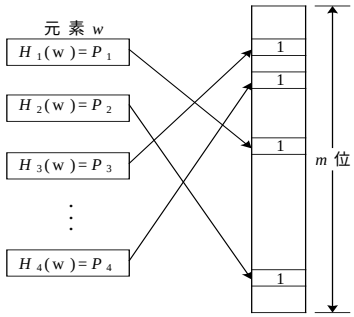

图 1 最基本的 Bloom Filter 算法

对于集合$S = \{S_1, S_2, S_k\}$中的元素S_i将Bloom Filter中的$h_1(s), h_2(s),\cdots, h_k(s)$位的值置为 1。这样，当判断元素$x$是否在$S$中时，可以检查Bloom Filter中的$h_1(x), h_2(x),\cdots, h_k(x)$是否全都等于 1。如果都等于 1，则表示$x$在$S$中，如果$h_1(x), h_2(x),\cdots,$

基金项目：江苏省科技攻关基金资助项目(BE2007058)
作者简介：吴国庆(1974－)，男，硕士，主研方向：计算机软件与通信
收稿日期：2007-01-08 **E-mail：**szx630@sina.com

$h_k(x)$中的某位为 0，则表示x不在S中。

采用Bloom Filter算法来判断一个元素是否在集合S中，可能存在误报的情况。由于存在着hash冲突，一个本来不属于集合S的元素y，其对应的$h_1(y), h_2(y),\cdots, h_k(y)$都因为与别的元素hash冲突，先前已经被置为 1 了，这样按照Bloom Filter的推论，y也是属于集合S的，这时就会出现误报(false positive)。

1.2 Adapted-Bloom-Filter 算法

本文设计了一个基于地址聚集的防抖动异常流量检测系统 AWACS(Address-aggregated Wobble-defended Abnormal-traffic Control System)，在这个系统中，运用 Adapted-Bloom-Filter 算法对数据包的目的 IP 地址进行强度聚集。由于需要分析的信息是 IP 地址，在 Adapted-Bloom-Filter 算法中，采用 4 个独立的 hash 函数 hash1, hash2, hash3, hash4(如图 2 所示)，将 IP 地址点分形式的 4 个域分别映射到这 4 个独立的 hash 函数中，对每个独立的 hash 函数分别建立一张 256 行的表，每一行构建一个计数器。当一个 IP 包进入路由器后，其目的地址被 4 个独立的 hash 函数分别映射到各自不同的域中，如果一个包的到达使某个域中的 $a_{ij}(i<4, j<256)$ 达到上限 TH，则说明到达这个包的目的地址的 IP 包频率已经非常高，对应的 IP 地址为可疑的目的地址。

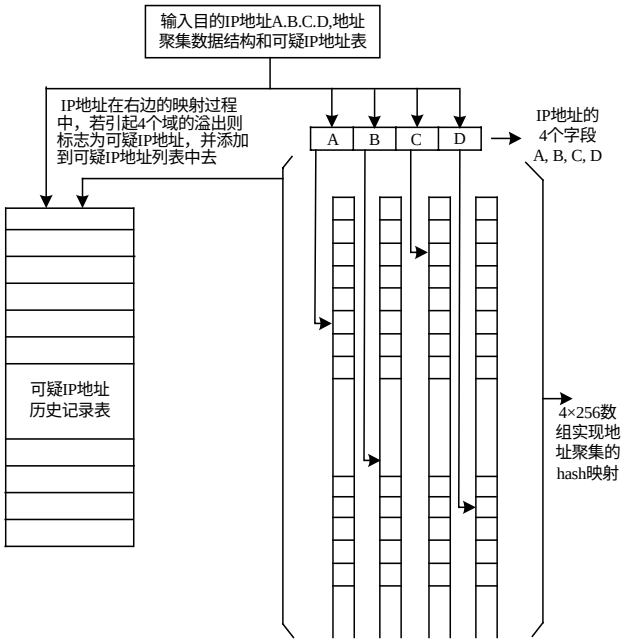

图 2　地址聚集算法

Adapted-Bloom-Filter 算法还增加了一个历史记录表结构，记录一些必要的历史信息，一方面过滤经过 hash 映射表没有引起溢出的流量，保持 hash 映射历史的延续性，另一方面过滤经过 hash 映射表引起溢出的流量，放行假性聚集流量。地址聚集算法的主要思想，如图 2 所示。

算法主要的流程如下：

首先进入包的目的地址被提取，然后目的地址分别送给 hash 映射表做分析处理，同时 hash 映射表的输出结果要送给可疑地址记录表进行处理。最后算法根据目的地址，hash 映射表结果和当前的历史记录更新记录表的数据。

对记录的操作完全取决于进入包的目的地址。如果当前进入包的目的地址没有引起溢出，也没有历史记录，那么根本不对记录进行处理。在各个流量正常的情况下，本文的算法大部分都处于休眠状态，以保证正常流量的转发。地址聚集模块处理流程如图 3 所示，增加可疑 IP 地址流程见图 4。

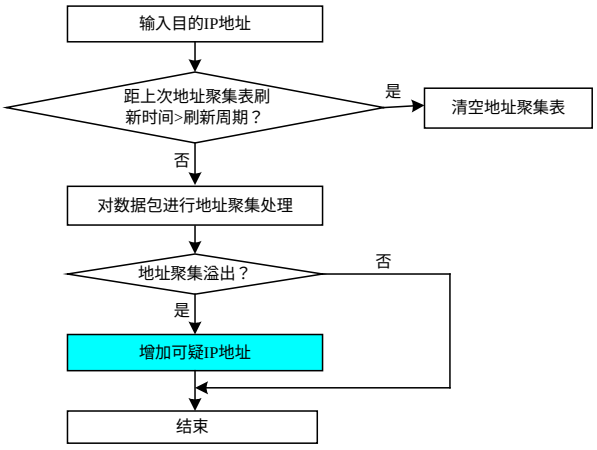

图 3　地址聚集流程

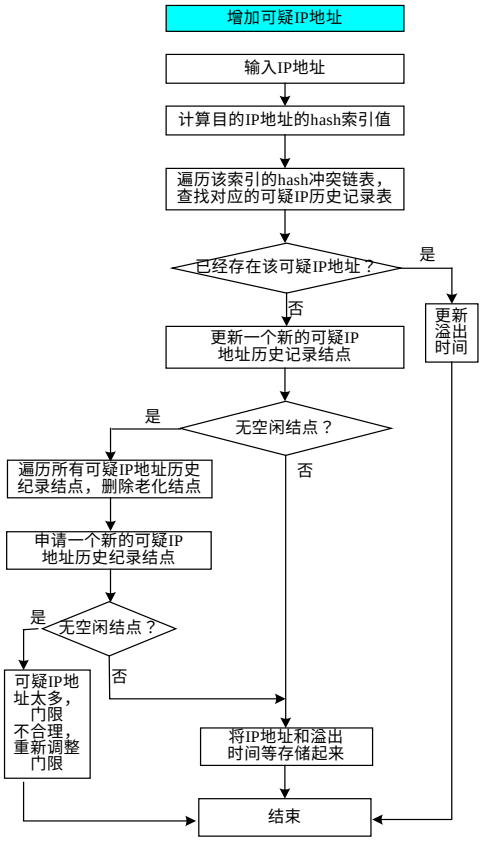

图 4　增加可疑 IP 地址流程

地址聚集时，首先判断是否需要对地址聚集 hash 表进行更新，如果需要更新，则将 hash 表全部清空。地址聚集处理是指对 IP 数据包 hash 处理和强度统计，这里的 IP 数据包 hash 处理是指将数据包头中的目的 IP 地址点分形式的 4 个值分别映射到 hash 表的 4 个域中，对每个域的地址分别计数，如果累加的值超过预定的阈值，则 hash 表溢出，说明该 IP 地址是高强度的 IP 地址，需要将该 IP 地址信息加入到可疑地址信息表中，以便进行流量模式聚集等处理；如果没有溢出，则表明流向该目的地的包无异常。

对于地址聚集时，hash 表溢出门限 TH 的设定，由于在单位时间内的网络数据包流量 x_n 都是符合一定的随机分布的，因此其 99.73%分布在一定范围内，见式(1)。

$$x_n \in (\mu - 3\delta/\sqrt{n}, \mu + 3\delta/\sqrt{n}) \quad , \quad n = 1,2,3 \tag{1}$$

其中，$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ ；$\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 。

发生攻击的时候，特别是拒绝服务攻击发生的时候，其统计特性就被破坏了，表现为流量突然猛烈增大，有 $x_n >> \mu + 3\delta/\sqrt{n}$ ，根据式(1)，取 $TH = \mu + 3\delta$ ，TH 的值是要在系统测试时在实际的网络环境中经过学习得到的，对于不同的网络环境，门限 TH 的数值是不同的。

对于 hash 函数的选取问题，由于本文所面对的环境是网络中的路由器，经过的 IP 地址必定很多，因此就设定 hash 函数为其 8 bit 的本身数值，每个数组总共为 $2^8 = 256$ 个。2 个不同的 IP 地址映射到相同域的可能性为：$P = (1/256)^4 = 2.328^{-10}$，所以其 hash 冲突的概率非常小。

2 防聚集抖动的 CUSUM 算法

由于黑客技术的发展和黑客软件的泛滥，网络攻击也变得多样化、复杂化。在拒绝服务攻击中，攻击端可以控制攻击的强弱变化，产生脉冲方式的攻击，从而使原来的 IDS 系统检测的结果产生抖动，严重的情况下就失去了其检测的能力。基于以上所述的脉冲攻击的特性，提出了一个防抖动的累积算法来更好地防止脉冲式攻击，即防聚集抖动的累积和算法(Cumulative Sum, CUSUM)。

CUSUM 算法的设计思想是对信息进行累加，将过程中的小偏移累积起来，达到放大的效果，以便提高检测过程中小偏移的灵敏度。CUSUM 在检测均值的小偏移时，比较有效，而且可以根据点的倾斜程度的改变，方便、直观地检测到变化。根据 CUSUM 的这种特征，在攻击的早期，就可以检测到攻击，而且它能以连续方式监控输入随机变量，从而达到实时检测的目的。本文对该算法进行了自调整，一旦出现 CUSUM 累加值下降的情况，就开始在下一个检测周期开始进行检验。把一个检测周期分为 n 个相同时间间隔的小段，则每个小段的时间间隔为

$$T = \Delta t / n$$

用 t 检验法来辅助判定是否发生了脉冲攻击，需要检验：

$$H_0 : \mu \le \mu_0 = S/n$$

$$H_1 : \mu > \mu_0 = S/n$$

则拒绝域为

$$t(n) = \frac{M_i - \mu_0}{S/\sqrt{n}} \ge t(n-1)$$

如果H_1可以接受，则认为此IP地址是可疑IP地址，如果H_1没有接受，则认为此IP地址不是可疑IP地址，需要调整 sum_n 的值，以减少累积造成的影响。

$$sum_n = sum_{\max} + m_i - S$$

$sum_{\max}$ 的值为连续告警之前最近且大于 F 的最大 CUSUM 累积值。以这个位置为起点是因为未来的统计值是由攻击者决定的。如果把 $sum_{\max}$ 的值设置得太小，例如设置到连续报警之前的最低谷的位置，那么在该时刻只是攻击者进行短暂的“休息”，就会容易引起漏报的现象，而现在所采用的 $sum_{\max}$ 的值则避免了上面所述的问题，从而可以检测出脉冲式的攻击。

3 基于地址聚集的 AWACS 系统的实现与测试

3.1 AWACS 系统组成及仿真测试环境

AWACS 系统设计如图 5 所示。

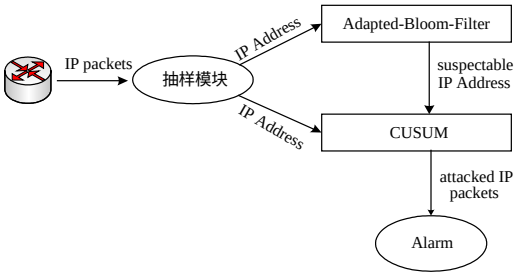

图 5 AWACS 系统设计

3.2 实验数据分析

笔者在网络环境中对此系统进行了实验，设定对 IP:172.16.115.20 在时刻 50 到时刻 110 进行分布式拒绝服务攻击(DDoS)，并设置系统的各参数如下：累积门限为 S=5.0 × 10^3，定时检测频率为 1 s，刷新频率为 5 s，Adapted-Bloom-Filter 的门限为 TH=5.5× 10^3，累积最大值 max=10× 10^3。

系统用 Adapted-Bloom-Filter 对通过路由器的 IP:172.16. 115.20 数据包的频率进行了跟踪，具体数据见图 6。从图 6 中可以看到，当 T(1)=50 时刻，目的 IP:172.16. 115.20 在单位时间内聚集的数值 M_{t1} >TH，此时 Adapted- Bloom-Filter 产生了可疑 IP 地址，然后系统就启动了 CUSUM 进行累积操作，在 T(2)=51 时刻，由于 M_{t2} >S，则对其大于 S 的数据进行累积，并产生攻击报警，由于异常的数据量随着时间产生抖动，在 T(3)=60 时刻，M_{t3} <TH，这时 Adapted- Bloom-Filter 算法无法对其产生作用，但此时 $D_n(sum_{t3}) = 1$，所以累积算法认为此时该目的 IP 的流量还处于异常状态，如图 7 所示。

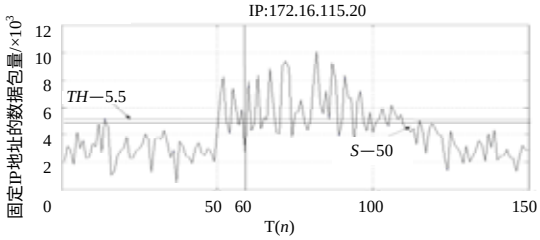

图 6 对固定 IP 地址数据包数的跟踪

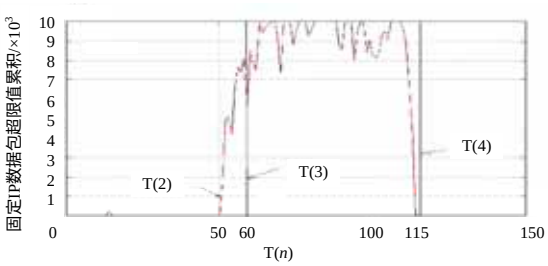

图 7 M-MULTOPS 中固定 IP 地址的超限累积

从图 7 可以看到，在时刻 100 以后，攻击就停止了，系统直到 T(4)=115 时刻，对于该目的 IP 地址的流量在连续的几个时间段内都是正常的 $M_i < TH$ ，使得 sum_i 的值迅速减小，T(4)时 $D_n(sum_{t4})$ =0，此时才认为该 IP 地址为正常地址，取消报警操作，并从相应节点中消除其累积项，达到期望的结果。

4 结束语

本文介绍了一种新的异常流量检测方法——防抖动的地址聚集算法，该方法运用 Adapted-Bloom-Filter 算法对目的 IP 地址进行聚集，由于这种算法不能检测出脉冲式的 DoS/DDoS 攻击，又结合防聚集抖动的 CUSUM 算法设计出

（下转第 180 页）