软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(10):2425−2437 [doi: 10.3724/SP.J.1001.2011.03882]

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

随机伪造源地址分布式拒绝服务攻击过滤*

肖 军[1,2+], 云晓春[1], 张永铮[1]

[1](中国科学院 计算技术研究所,北京 100190)

[2](中国科学院 研究生院,北京 100049)

Random Spoofed Source Address Distributed Denial-of-Service Attack Traffic Filter

XIAO Jun[1,2+], YUN Xiao-Chun[1], ZHANG Yong-Zheng[1]

[1](Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

[2](Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: E-mail: xiaojun@software.ict.ac.cn

Xiao J, Yun XC, Zhang YZ. Random spoofed source address distributed denial-of-service attack traffic filter. *Journal of Software,* 2011,22(10):2425−2437. http://www.jos.org.cn/1000-9825/3882.htm

Abstract: Distributed Denial-of-Service (DDoS) attack, using random spoofed source addresses, is popular because it can protect the attacker's anonymity. It is very difficult to defent against this attack because it is very hard to differentiate bad traffic from the normal. In this paper, based on the source addresses distribution statistical feature, an effective defense scheme, which can differentiate vicious traffic from normal traffic, is presented. Based on a novel Extended Counting Bloom Filter (ECBF) data structure, this paper proposes an algorithm to identify normal addresses accurately. Once a normal address is sought out, packets from it will be forwarded with high priority, thus, normal traffic is protected. The simulation results show that this scheme can identify legitimate addresses accurately, protect legitimate traffic effectively, and give better protection to valuable long flows. Because the time complexity of the method is $O(1)$, and it needs several MB memory space, it can be implemented in edge routers or network secure devices like firewalls to defend against random spoofed source address DDoS attacks.

Key words: network security; distributed denial of service; Bloom Filter; random spoofed source address

摘 要: 由于能够有效隐藏攻击者,随机伪造源地址分布式拒绝服务攻击被广泛采用.抵御这种攻击的难点在于无法有效区分合法流量和攻击流量.基于此类攻击发生时攻击包源地址的统计特征,提出了能够有效区分合法流量和攻击流量,并保护合法流量的方法.首先设计了一种用于统计源地址数据包数的高效数据结构 Extended Counting Bloom Filter(ECBF),基于此,提出了随机伪造源地址分布式拒绝服务攻击发生时合法地址识别算法.通过优先转发来自合法地址的数据包,实现对合法流量的有效保护.采用真实互联网流量进行模拟,实验结果表明,所提方法能精确识别合法地址,有效地保护合法流量,尤其能够较好地保护有价值的交易会话.所提方法的时间复杂性为 $O(1)$,并且只需数兆字节的内存开销,可嵌入边界路由器或网络安全设备,如防火墙中,实现随机伪造源地址分布式拒绝服务攻击的在线过滤.

关键词: 网络安全;分布式拒绝服务攻击;Bloom Filter;随机伪造源地址

* 基金项目: 国家自然科学基金(60703021, 61070185); 国家高技术研究发展计划(863)(2007AA010501, 2007AA01Z444)
收稿时间: 2009-10-14; 修改时间: 2010-03-29; 定稿时间: 2010-04-27

中图法分类号: TP309 文献标识码: A

分布式拒绝服务攻击(distributed denial of services,简称 DDoS)是当前互联网最主要的安全威胁之一,严重影响了民众生活和社会经济,甚至于国家安全.近年来,DDoS 攻击事件频繁发生,比较典型的事件是 2009 年 5 月 19 日的“暴风影音”事件[1],造成的经济损失超过 1.6 亿元,其根本原因是暴风影音的域名服务器 DNS pod 遭到了 DDoS 攻击.可见,DDoS 攻击的检测、防御和过滤的研究工作具有重要的理论意义和实际价值.

DDoS 攻击发送大量数据包到被攻击服务器,可以有效阻止合法用户对资源的访问.DDoS 攻击可分为两类:利用协议弱点的攻击和洪泛攻击.SYN Flood 攻击是最典型的协议弱点攻击.TCP 协议收到一个 SYN 包,会开辟一段内存空间.SYN Flood 利用这一弱点,通过发送大量的 SYN 包消耗攻击目标的内存空间.而洪泛攻击则是通过发送大量的数据包来消耗攻击目标的计算能力或带宽等,如 UDP Flood.从攻击数据包源地址真实性的角度来看,DDoS 攻击可以分为采用真实源地址或者伪造源地址.虽然当前僵尸网络规模庞大,攻击者可以利用大量的僵尸主机发动攻击,但是由于伪造源地址能够有效保护攻击者,隐藏僵尸主机,增加攻击防御难度,因此,采用伪造源地址的 DDoS 攻击始终是一种重要的攻击方式[2].

从 DDoS 防御设备部署位置的角度来看,流量过滤可以分为被攻击端过滤(ingress filtering)、攻击端过滤(egress filtering)和骨干网路由器过滤(router-based filtering).Ingress Filtering 是最常见的防御方法,具有易于部署且部署者可以直接受益的优点.但是,几乎当前所有 Ingress Filtering 方法都无法有效抵御随机伪造源地址 DDoS 攻击,最大的困难在于无法有效区分合法流量和攻击流量.基于统计的方法,如 Packetscore 方法[3]和 PCA 方法[4],都只是数据包层面上的过滤,无法为合法流量提供流层面上的保护;而 Hop-Count Filtering[5]和 History-IP Filtering[6]只能保护以前曾出现过的用户,而无法保护新来的用户.

本文关注伪造数据包源地址在整个地址空间($0\sim2^{32}$)内随机分布的 DDoS 攻击.在此类攻击发生时,对一个伪造数据包而言,每一个 IP 地址都有相同的概率被其选中作为其源地址,且概率为 $1/2^{32}$.因此,合法地址可能被伪造数据包选中,作为其源地址.另一方面,合法地址同时也发送了合法数据包到被攻击端.所以,从统计的角度来看,合法地址对应的数据包数大于伪造地址对应的数据包数.

基于上述统计特征,本文提出一种基于源地址分布特征的随机伪造源地址 DDoS 过滤方法(filtering based on the source address distributed feature,简称 FSAD).该方法部署在被攻击端,能够准确识别合法地址,并保护合法流量.该方法具有如下优点:

1) 计算复杂性低,时间复杂性为 $O(1)$,只需消耗数兆字节内存,满足在线处理要求,可以嵌入到边界路由器或者防火墙中;
2) 能够识别出新用户;
3) 只需对历史流量进行学习,无需建立复杂的模型;
4) 不依赖于具体的网络协议,适用范围广.

本文第 1 节从总体上介绍 FSAD.第 2 节首先介绍用于统计源 IP 地址对应数据包数的高效数据结构——extended counting Bloom Filter(ECBF),并基于 ECBF 提出合法地址识别算法,详细分析识别算法的误报率和漏报率,并提出识别参数的动态调整算法,最后介绍合法流量的保护方法.第 3 节采用真实网络流量进行模拟实验,验证 FSAD 在地址识别和流量过滤这两方面的准确性和高效性.第 4 节为相关工作.第 5 节对一些相关问题进行讨论.第 6 节总结全文.

1 FSAD 概述

通常,DDoS 攻击工具采用随机函数产生随机数作为伪造源地址[7,8].由于随机数的范围受到处理器和操作系统位数的限制,所以伪造源地址常常局限在部分地址空间.例如,如果操作系统是 32 位的 Linux,则伪造源地址只能分布在 $0\sim2^{31}-1$ 之间.与伪造地址在整个地址空间($0\sim2^{32}$)内分布相比,部分空间分布的伪造方式往往无法最好地发挥攻击效率.例如,对源地址不在伪造源地址覆盖范围内的数据包,安全设备可以直接允许其通过,这样,

被攻击服务器仍然可以为很大一部分用户提供服务.当前,64 位的处理器和操作系统已开始普及,64 位平台生成的伪造源地址将会在整个地址空间内分布.同时,现有其他平台上的 DDoS 工具也很容易修改,实现伪造源地址在整个地址空间内随机分布.由于存在以上两个原因,本文关注伪造源地址在整个地址空间内随机分布的 DDoS 攻击.

FSAD 的过滤包含如下 3 个阶段:

- 检测攻击发生,并根据当前攻击规模、历史流量和源地址识别精度的要求,设定合适的合法地址识别参数;
- 识别出合法源地址,并保存到合法地址表中(legitimate address table,简称 LAT);
- 检查数据包是否来自合法地址,根据检查结果按照不同优先级进行转发.如果数据包转发队列发生溢出,则丢弃数据包.

FSAD 的第 1 步是检测攻击发生.一些过滤方法[9,10]无论攻击是否发生,都进行伪造数据包识别和过滤.与这些方法不同,FSAD 只有在检测到攻击发生时才开启过滤功能.本文采用了基于熵的检测方法[11],能够快速检测到攻击发生.熵能够有效地反映地址分布的聚合情况,同时具有较高的灵敏度.在随机伪造源地址攻击发生时,与正常情况相比,源地址分布更加均匀,熵值相应地有所增加.当熵值超过了一个阈值,可以认为随机伪造源地址 DDoS 攻击发生.由于篇幅限制,本文不对检测方法作详细介绍,重点介绍合法源地址识别算法和合法流量保护这两方面内容.

2 合法地址识别和合法流量保护

本节介绍合法地址识别方法以及在合法地址识别后合法流量的保护策略.首先介绍一个用于统计源地址包数的高效数据结构 Extended Counting Bloom Filter(ECBF),它可以大幅度降低内存开销.基于 ECBF,我们提出了合法地址识别算法,接着分析了识别算法的误报率和漏报率,然后提出了合法地址识别参数动态调整算法,可以保证在不同的攻击规模下满足识别精度的要求,最后介绍了合法流量保护和被保护对象过载控制方法.

2.1 The Extended Counting Bloom Filter (ECBF)

假设待保护地址已经给定,则只需统计源地址的信息.随机伪造源地址攻击数据包的源地址在 0~2^{32} 之间随机分布,如果记录每个源地址对应的数据包数,则内存开销难以承受,需找到合适的方法节省内存开销,并且此方法能够用来识别合法源地址.本文借鉴了龚俭等人[12]构建 Bloom Filter 的方法,采用一种 ECBF 数据结构统计源地址对应包数.与龚俭等人工作的区别在于:1) 使用目的不同,龚俭等人提出的 Bloom Filter 用于异常检测,而本文用于合法地址识别;2) 龚俭等人采用了 3 个哈希函数记录源地址信息,而本文采用了 4 个哈希函数,能够降低地址识别的误报率;3) 龚俭等人提出的方法需要还原 IP 地址,相应的时间复杂度为 $O(M)$,其中,M 为 2^{16},而采用 ECBF 可以直接识别合法地址,相应的时间复杂度为 $O(1)$.ECBF 包含 4 个用于统计源地址数据包数的哈希函数和数组,每个数组分别对应于一个哈希函数,其中,A_1 对应于 IPH,A_2 对应于 IPM,A_3 对应于 IPL,A_4 对应于 IPLH,如图 1 所示.每个哈希函数和统计过程具体介绍如下:

假设收到一个数据包,源地址 Saddr 为($a.b.c.d$),则

- $IPH(Saddr)=256\times a+b$;
- $IPM(Saddr)=256\times b+c$;
- $IPL\ (Saddr)=256\times c+d$;
- $IPLH(Saddr)=256\times d+a$.

然后对 $A_1[256\times a+b]$,$A_2[256\times b+c]$,$A_3[256\times c+d]$ and $A_4[256\times d+a]$分别加 1.

容易看出,ECBF 的每个元素对应于 2^{16} 个源地址.例如,A_1 数组的第 257 个单元对应源地址为(1.1.x.y)的数据包,其中,x 和 y 是 0~255 之间的任意一个数.并且每收到一个数据包,对应于该数据包源地址的 4 个单元值都会被加 1.

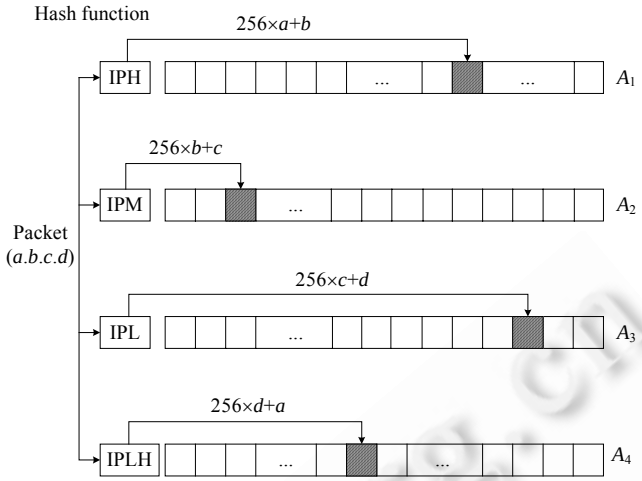


Fig.1 Extended counting Bloom Filter (ECBF)

图 1 扩展的计数 Bloom Filter 结构

2.2 合法地址识别算法

基于前文提出的ECBF结构,本节给出一种随机伪造源地址DDoS攻击时合法地址的识别算法,如图2所示.

```
Set identifying time interval and threshold T;
while(1)
    Receive a packet;
    Get source ip address sip;
    Record sip in ecbf;

    If (every element's value of sip in 4 arrays>T)
            Sip is a legitimate address;
    fi;

    if (time interval is over)
        Empty 4 arrays;
        Start a new time interval;
    fi;
End while;
```

Fig.2 Legitimate address identifying algorithm under random spoofed source address DDoS attacks

图 2 随机伪造源地址 DDoS 攻击合法地址识别算法

算法的第 1 步是设置合适的识别时隙 interval 和识别阈值 T.由于时隙和阈值的设定与识别精度要求以及当前的攻击规模相关,如何设置会在对识别算法的误差分析之后加以阐述.

每收到一个数据包,需进行 4 次哈希操作,由于每个哈希操作的时间复杂性都是 $O(1)$,所以算法的时间复杂性是 $O(1)$.本识别算法基于 ECBF,采用了 4 个数组,每个数组共 2^{16} 个元素,若每个元素大小以 4B(字节)为例,则本识别算法的内存开销为 1MB.可见,本文提出的合法地址识别算法具有很好的时间复杂性和空间复杂性.

2.3 误报率和漏报率

为了深入分析识别算法的准确性,首先将 ECBF 的元素分为两类:仅记录了伪造包数的元素和同时记录了合法包和伪造包数的元素;接着给出了元素值的分布特性;然后从误报率(false positive rate)和漏报率(false negative rate)两方面分析识别算法的误差.

2.3.1 元素值分布

参数 X 表示一个时隙后,一个元素的伪造攻击包计数.参数 X_i 表示,当收到此时隙中第 i 个攻击包时,该元素的累加值.显然,X_i=0 或者 X_i=1.假设在一个时隙内共有 m 个伪造数据包到达被攻击端,则

$$X=\sum_{i=1}^{m}X_i.$$

X 服从二项分布,不易计算.但由于伪造攻击数据包较多,二项分布可用泊松分布来近似.任意一个 IP 地址被选中的期望值是 $m/2^{32}$,每个单元对应 2^{16} 个 IP 地址,所以每个单元的攻击包计数服从泊松分布,期望值是 $m/2^{16}$.所以,

$$p(X>T)=\sum_{k=T+1}^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda},$$

其中,$\lambda=m/2^{16}$.

为了验证上述推断,用 ECBF 记录随机生成的 20×64K 个伪造源地址.任意选择 ECBF 中 5 000 个元素,元素值分布如图 3 所示,圆点表示实际元素值分布,而曲线为推断的泊松分布曲线.可见,元素值的分布情况与泊松分布基本接近.

假设一个时隙内一个合法地址的数据包数服从泊松分布.由于两个泊松分布之和仍然是泊松分布,并且期望值是两者期望值之和,所以元素的合法数据包计数也服从泊松分布,假设其期望值为 n.此元素同时记录了攻击数据包数和正常数据包数,两者均服从泊松分布,期望值分别是 $m/2^{16}$ 和 n.所以,此元素值也服从泊松分布,期望值是 $m/2^{16}+n$.图 4 显示了两类元素值的分布情况,其中,左边的曲线对应仅记录了伪造包的元素值,右边的曲线对应记录了合法数据包的元素值,两者期望值分别是λ_1 和λ_2.这里,$\lambda_1=m/2^{16}$,$\lambda_2=m/2^{16}+n$.

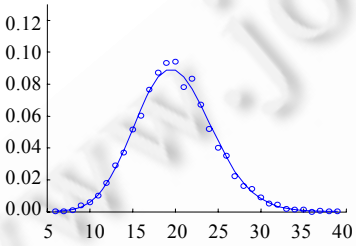


Fig.3 Elements' values distribution

图 3 元素值分布

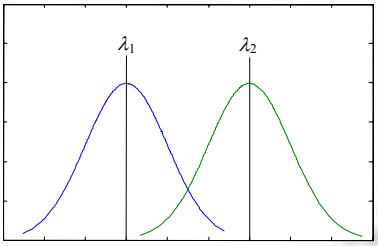


Fig.4 Two kinds of Poisson distribution

图 4 两类元素值的泊松分布

2.3.2 误报率和漏报率

本文通过一个阈值 T 来区分上文所提及的两类元素,在此基础上分析识别算法的误报率和漏报率.

误报率为一个伪造地址被识别为合法地址的概率.不失一般性,假设伪造源地址($x.y.z.w$)被识别为合法源地址,根据识别算法,它的 4 个对应元素值需均大于阈值 T,所以误报率为

$$p(A_1[256\times x+y]>T)\times p(A_2[256\times y+z]>T)\times p(A_3[256\times z+w]>T)\times p(A_4[256\times w+x]>T).$$

$p(A_1[256\times x+y]>T)$表示 $A_1[256\times x+y]>T$ 的概率,剩下 3 项的含义与之类似.

$p(A_1[256\times x+y]>T)$,$p(A_2[256\times y+z]>T)$,$p(A_3[256\times z+w]>T)$和 $p(A_4[256\times w+x]>T)$的分析相似,为便于分析,用 $p(X>T)$分别代替它们,误报率可表示为 $p^4(X>T)$.

$p_a(X>T)$表示伪造包计数 X 大于 T 的概率,$p_m(X>T)$表示伪造包计数和合法包计数之和 X 大于 T 的概率.如图 5 所示,区域 A 表示 $p_a(X>T)$,区域 B 表示 $p_m(X>T)$.

显然,$p(X>T)=(1-p)\times p_a(X>T)+p\times p_m(X>T)$,其中,$p$ 是 ECBF 任意一个数组中记录了合法包数的元素比率.假设共有 C'个元素记录了合法包,则 $p=C'/N$,其中,N=65536.所以,误报率为

$$((1-p)\times p_a(X>T)+p\times p_m(X>T))^4=\left(\left(1-\frac{C'}{N}\right)\times p_a(X>T)+\frac{C'}{N}\times p_m(X>T)\right)^4 \quad (1)$$

从图 4 可以看出,$p_m(X>T)>p_a(X>T)$,所以误报率随 C'的增加而增加.通常,同时与一个服务器交互的 IP 地址不会超过 10K,为便于计算,本文用 10K 代替 C'.相应地,计算出的误报率大于实际误报率.

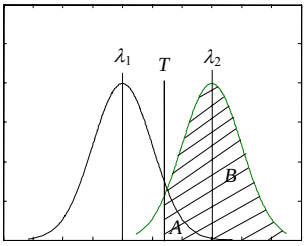


Fig.5 Two kinds of probability of X greater than T

图 5 两类 X 大于 T 的概率分布

漏报率是一个合法 IP 地址未被识别的概率.根据合法地址识别算法,如果对应的 4 个元素值中有一个或多个小于阈值 T,则这个地址不会被识别.所以,漏报率为

$$1-p^4(X>T).$$

对记录了合法数据包数的元素,$p(X>T)$即为 $p_m(X>T)$,则漏报率为

$$1-p_m^4(X>T) \tag{2}$$

2.4 识别参数调整算法

从图 5 可以看出,$p_m(X>T)$和 $p_a(X>T)$都随着阈值 T 的增加而降低.由公式(1)和公式(2)可知,随着阈值 T 的增加,误报率降低而漏报率增加.基于这一结论,本节提出一种识别参数动态调整算法,以满足在不同攻击规模下识别精度的要求,如图 6 所示.

M: The predicted packet number per second without attacks;
N: The current packets number per second under an attack;
R_n, R_p: The false negative and false positive requirement values;
$f_n(T,\lambda)$: The function used to calculate the false negative, T is
the Poisson random variable and λ is the average rate of success;
$f_p(T,\lambda)$: The function used to calculate the false positive, T is
the Poisson random variable and λ is the average rate of success;
$\Delta\lambda_3$:=min(the expected value of normal packets arrival rate in a second);
$\Delta\lambda_1$:=$(N-M)/65536$;
λ_1:=$\Delta\lambda_1$;
λ_3:=$\Delta\lambda_3$;
interval:=1;

```
while(1)
    T:=min(T'|f_n(T',λ1)<R_n)
    λ2=λ3+λ1
    if (f_p(T,λ2)<R_p)
        break;
    else
        interval:=interval+1;
        λ1:=λ1+Δλ1
        λ3:=λ3+Δλ3;
    fi;
End while;
```

Fig.6 Identifying variable modifying algorithm

图 6 识别算法参数调整算法

算法的核心思想是,选择最小的可满足误报率 R_p 要求的阈值 T,分析其是否满足漏报率 R_n 的要求.如果满足,则相应的时隙即为识别时所用时隙;否则,延长检测时隙 $interval$,直到找到合适的阈值 T,同时满足误报率和漏报率的要求.

参数 N 为攻击发生时每秒数据包(包括攻击数据包和合法数据包)到达数目.$\Delta\lambda_1$ 和 $\Delta\lambda_3$ 分别为元素记录的每秒攻击数据包期望值和合法数据包期望值;λ_1,λ_2 和 λ_3 分别为一个时隙记录的攻击数据包期望值、攻击数据包和合法数据包期望值之和以及合法数据包期望值.参数 M 为攻击发生时到达攻击端的每秒合法数据包数预测值,本文使用 ARMA[13]模型来根据历史流量预测当前流量.流量预测一直是研究的热点,相关的研究成果也较多,由于篇幅所限,本文不作详细讨论.

网络中的大部分流只包含一个数据包,对一个服务器而言,这类流与服务器不进行交互,往往没有价值.本文只关注合法用户与服务器交互的数据流.用户与服务器建立连接后,往往发送几百或者数千的数据包到服务器[14].这样,识别参数调整算法能够在一个或者几个循环内设定出识别参数;同时,合法地址也很容易被识别出来.$\Delta\lambda_3$ 可以依据用户的访问情况提前设定.

2.5 包过滤和过载控制

FSAD 通过一个合法地址表(legitimate address table,简称 LAT)来保存识别出来的 IP 地址.LAT 每项记录一个被识别的 IP 地址,包含两个部分:IP 地址和记录了该 IP 地址最近被识别时间的时间戳.由于 Open Hash 比 Closed Hash 在 DDoS 攻击时更加鲁棒[15],LAT 以 Open Hash 方式组织,如图 7 所示.LAT 的每一项消耗 12 字节,以记录 10K 个合法地址为例,LAT 的内存开销为 120K 字节.可见,LAT 的内存开销极小.由于采用了哈希表,所以查找一个地址是否为合法地址的时间复杂性为 $O(1)$.

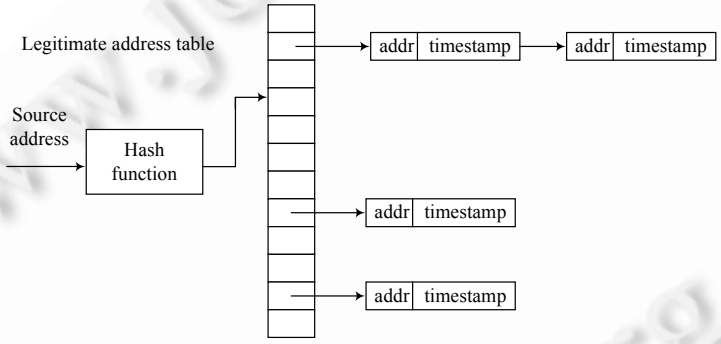


Fig.7 Legitimate address table

图 7 合法地址表

如果一个合法用户停止访问被保护服务器,则其 IP 地址应从 LAT 中移除,否则,以此地址作为源地址的伪造数据包仍然可以进入服务器.FSAD 依据时间戳移除过期的地址,一旦一个合法地址的时间戳超过了设定的阈值,将会从 LAT 中移除.通常,用户浏览一个页面的时间(thinking time)是 7s~70s[16],为了避免淘汰某些浏览速度较慢的用户,选取 75s 作为淘汰阈值.

FSAD 根据数据包的源地址合法性进行数据包转发,图 8 描述了转发过程.收到一个数据包后,FSAD 查找该数据包的源地址是否包含在 LAT 中,如果包含,则该数据包被放入一个具有高转发优先级的队列缓冲区(high priority queue,简称 HPQ)中;如果没有包含,则放入一个低转发优先级的队列缓冲区(low priority queue,简称 LPQ)中.HPQ 和 LPQ 都使用先进先出(first in first out)策略.如果缓冲区已满,后来的数据包将会被丢弃,直到缓冲区再次有空间容纳新来的数据包.FSAD 优先转发 HPQ 中的数据包,直到 HPQ 为空,才会转发 LPQ 中的数据包.FSAD 的转发能力根据被保护对象的处理能力预先设定,以实现对被保护对象的过载控制.

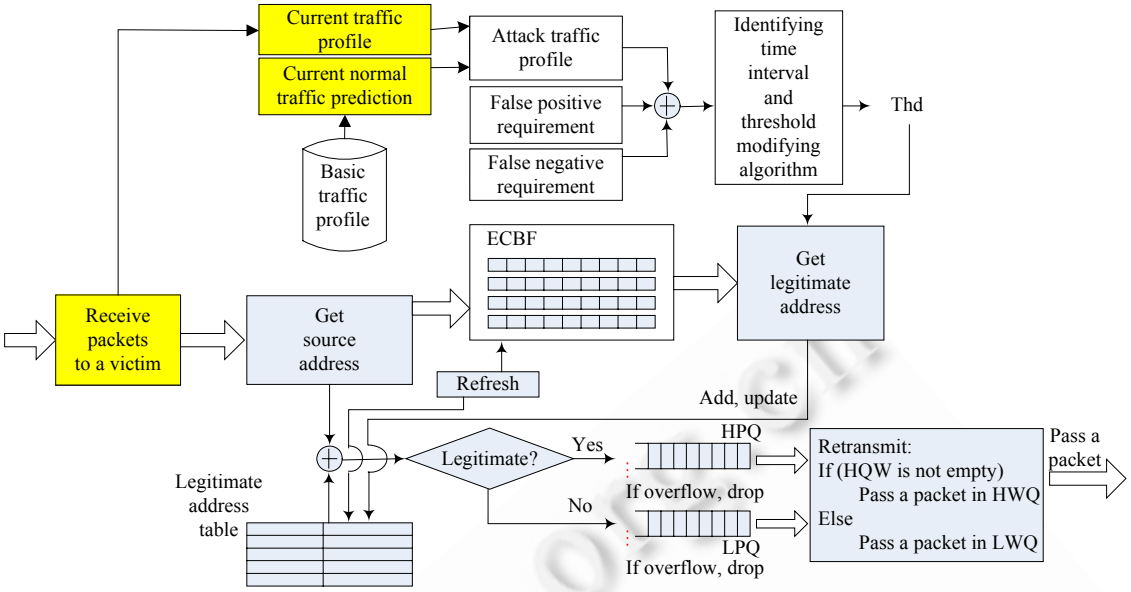


Fig.8 Packet retransmitting process

图 8 数据包转发过程

3 实 验

本文采用真实网络流量,通过模拟来验证 FSAD 的性能.相关的模拟参数见表 1.

Table 1 Setting of simulation

表 1 模拟参数

Nominal profile	15 minutes traffic to a HTTP server, between 14:15~14:30 from Mar 30, 2009 to April 1, 2009, from the WIDE project[17]
Legitimate traffic	Use the trace of the same target and of the same link, between 14:15~14:30, April 2, 2009 and the average arrival rate is about 600 pps
Attack type	TCP-SYN flood attack, destination port=80, packet size=64, other attribute values of the forged packets are uniformly randomized in the range of the corresponding allowable space
Attack intensity	The attacking packets arrival rate is from 2.5K pps to 30K pps

由于 FSAD 不依赖于具体的网络协议,实验中采用了最常见的随机伪造源地址 TCP-SYN Flood 攻击来检测 FSAD 的性能.

3.1 性能指标

合法数据包的通过率是衡量一个 DDoS 过滤系统的重要指标,实验比较了 FSAD 和另一种统计方法——Packetscore[3]在合法流量保护的性能.

由于 FSAD 基于地址识别进行过滤,实验中也考察了 FSAD 对合法地址识别的准确性,并考察了数据包数属于 6 个不同范围的合法地址,其数据包的通过率.

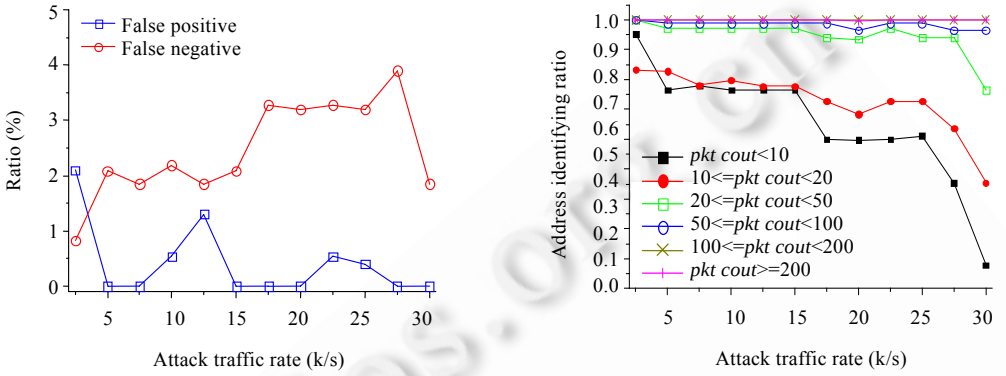
总结起来,实验考察了 FSAD 如下性能:

- 不同攻击规模下,合法地址识别的误报率和漏报率;以及对应于 6 个不同包数范围的合法地址,其被识别出来的比率;

- 不同攻击规模下,识别合法数据包的误报率和漏报率;
- 不同攻击规模下,包数属于 6 个不同范围的合法地址,其数据包的通过率.

3.2 实验结果

合法地址识别性能如图 9 所示.图 9(a)表明,在不同的攻击规模下,FSAD 地址识别的误报率和漏报率均低于 4%.图 9(b)为 6 个数据包数属于不同范围的合法地址识别比率.在相同的攻击规模下,随着对应数据包数的增加,合法地址识别比率也相应地增加.随着攻击规模的增加,FSAD 的合法源地址识别精度也随合法地址数据包的减少而降低,但对拥有较多数据包的合法源地址仍然维持了高识别精度.



(a) Normal address identifying accuracy of FSAD (b) Normal address identifying ratio of 6 different packets number ranges

(a) FSAD 合法地址识别精度 (b) 对应包数在 6 个不同范围的合法源地址识别比率

Fig.9 Normal address identifying performance of FSAD

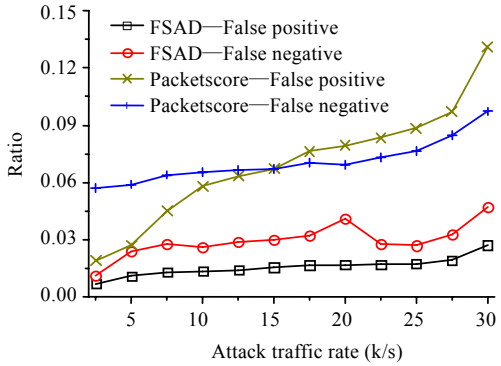
图 9 FSAD 的合法地址识别性能

图 10(a)显示了 FSAD 和 Packetscore[3]两种方法的合法包通过情况.

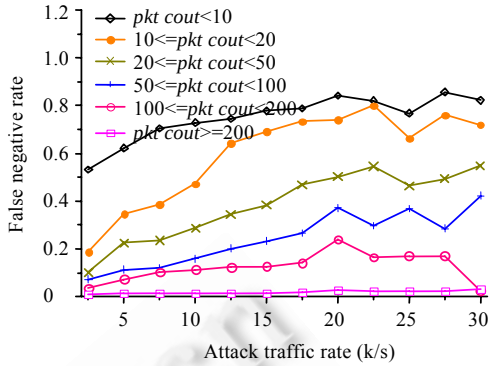
随着攻击规模的增加,两种方法的误报率和漏报率都相应地有所增加.与 Packetscore 相比,FSAD 能够更好地保护合法流量.FSAD 基于源地址分布统计特征进行过滤,一个地址一旦被识别出来,除非此 IP 地址被提前从 LAT 移除,否则,这个地址发送的后续数据包将会直接允许通过.Packetscore 基于计算出的 CLP 值过滤数据包,因而对数据包属性值的分布比较敏感.FSAD 能够有效地保护合法流量.虽然随着攻击规模的增加,FSAD 对对应于较少数据包的合法地址识别精度下降,但 FSAD 对拥有较多数据包的合法地址仍然具有高识别精度,因此能够有效保护这些包含较多数据包的大流;又因为少数的大流通常占据了总流量的大部分,所以,即使攻击规模增加,合法数据包的通过率仍然较高.如图 10(a)所示,即使攻击流量达到了合法流量的 60 倍,FSAD 仍然可以保护 96%的合法数据包.

图 10(b)和图 10(c)显示了 FSAD 和 Packetscore 对数据包数属于 6 个不同范围的合法地址,其合法数据包的通过情况.

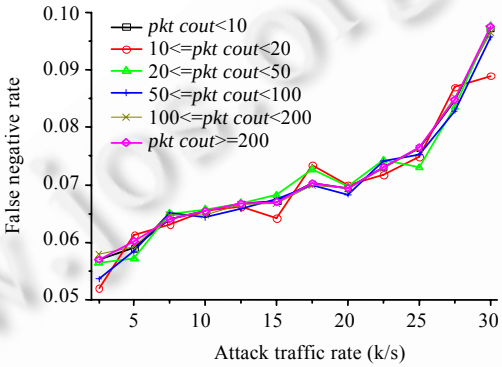
在 FSAD 过滤方法中,一个合法地址被识别后,将会被保存下来,其后续的数据包被允许通过.因此,一个流包含的数据包越多,就越能获得更好的保护.Packetscore 根据 CLP 值决定一个数据包是通过还是丢弃,因而包数属于不同范围的合法地址,其数据包通过率基本接近.显然,与 Packetscore 方法相比,FSAD 更倾向于保护包数较多的流.对一些交易网站或拍卖网站,如 eBay 等,交易数据流往往包含了更多的数据包,因此,FSAD 能够更好地保护有价值的交易流.



(a) FSAD and Packetscore's false positive rate and false negative rate of normal packets passing
(a) FSAD 和 Packetscore 合法包识别的错报率和漏报率



(b) FSAD packet passing ratio of address whose corresponding packets number are in different ranges
(b) FSAD 方法中,处于不同包数范围合法地址的包通过率



(c) Packetscore packet passing ratio of addresses whose corresponding packets number are in different ranges
(c) Packetscore 中,处于不同包数范围合法地址的包通过率

Fig.10 Normal packets passing ration performance of FSAD and Packetscore
图 10 FSAD 和 Packetscore 合法数据包通过情况

4 相关工作

依据过滤位置的不同,DDoS 攻击过滤可以分为被攻击端过滤、攻击端过滤和路由器过滤.

Ingress Filtering 是一类通过过滤可疑数据包或者非法数据包抵御 DDoS 攻击的常用方法.通常只有在一个数据包的源地址在过滤系统许可的范围内,才允许其进入子网或服务器.Wang 等人[5]基于数据包在网络传递过程中网络跳数不易被攻击者伪造这一原理,提出了一种基于跳数过滤(hop-count filtering,简称 HCF)的伪造数据包 DDoS 攻击过滤方法,基于数据包首部的 *TTL*(time-to-live)值判断数据包是否伪造,从而过滤 DDoS 攻击.Tao 等人[6]基于以往出现用户还会再次访问这一特点,提出了一种基于历史 IP(history-IP)的 DDoS 防御方法,该方法对未出现在历史 IP 表中的源地址进行过滤,以实现对 DDoS 攻击的防御.Ingress Filtering 往往需要复杂的离线或者在线学习,并且无法为新到用户提供保护.Ingress Filtering 的主要局限性在于攻击者能够利用的网络资源或者计算资源远远大于防御设备拥有的资源,常常超过防御或者过滤设备处理的能力.

Egress Filtering 是另一种常用的过滤手段,不同点在于过滤设备部署在源端,如果数据包的源地址属于本地子网,数据包将被允许通过.这种过滤方法需要路由器开启过滤功能,但由于 ISP 往往各自为战,因而该方法无法普遍实施.D-WARD[18]是一种新颖的 Egress Filtering 方法,通过比较正常流量模型和当前流量来识别是否有

攻击发生,比如进出当前子网的数据包比例或者目的地址的连接数等.如果当前流量状况偏离正常流量模型,则认为攻击发生,进行过滤.

Router-Based Filtering 在骨干网路由器上实现对攻击流量的过滤,可以有效阻止伪造攻击包到达目的地址.Spoofing Prevention Mothod[19]和 Passport[20]方法均采取了让数据包携带 secrete key 的策略,通过 secrete key,路由器可以验证数据包源地址的真实性.Distributed Packet Filtering[21]和 Source Address Validity Enforcement Protocol(SAVE)[9]均通过数据包进入路由器的接口验证数据包是否伪造,从而实现过滤.此外,SAVE 方法提供了一种发现数据包所进入接口的方法,而 BGP Anti-Spoofing Extension[10]则采取让数据包携带表示数据包所进入方向的信息.Pushback[22,23]是另一类重要的 Router-Based Filtering 方法,监测到攻击发生之后,在靠近攻击端的路由器进行流量过滤,可以在攻击流量刚进入骨干网时就实现过滤,可有效降低 DDoS 攻击危害.Router-Based Filtering 方法需要消耗路由器的计算资源,并且需要路由器之间的密切协作.由于 Router-Based Filtering 需要 ISP 的大规模投入,具有较大的实施难度,并且 ISP 各自为战,因而使得 Router-Based Filtering 难以广泛普及.

5 讨 论

前文对合法地址识别和合法流量保护的原理进行了阐述,下面有必要对如下相关问题作进一步的讨论.

5.1 哈希函数选取

第 2.1 节中提到,龚俭等人的工作采用 3 个哈希函数,本文所提出的 ECBF 采用 4 个哈希函数,能够降低合法地址识别误报率,下面对此作进一步分析.

源地址为($a.b.c.d$)的攻击数据包,依据合法地址识别算法,被识别为合法数据包的概率为

$$P=p(A_1[256\times a+b]>T)\times p(A_2[256\times b+c]>T)\times p(A_3[256\times c+d]>T)\times p(A_4[256\times d+a]>T).$$

如果采用 3 个哈希函数,地址识别算法不变,则上述攻击数据包被识别为合法数据包的概率为

$$P'=p(A_1[256\times a+b]>T)\times p(A_2[256\times b+c]>T)\times p(A_3[256\times c+d]>T).$$

在同样的合法访问和攻击下,$P<P'$.可见,采用 4 个哈希函数与采用 3 个哈希函数相比,能够降低误报率.下面给出一个误报实例.

采用 3 个哈希函数记录每个源地址对应的数据包数.现有两个合法地址($a.b.c.d$)和($e.b.c.f$),两者对应的单元值均大于阈值 T.此时,源地址为($a.b.c.f$)和($e.b.c.d$)的攻击包均会被识别为合法数据包.如图 11 所示,两个曲线框分别对应($a.b.c.f$)和($e.b.c.d$)这两个误报地址.采用 4 个哈希函数能够有效避免此类误报.

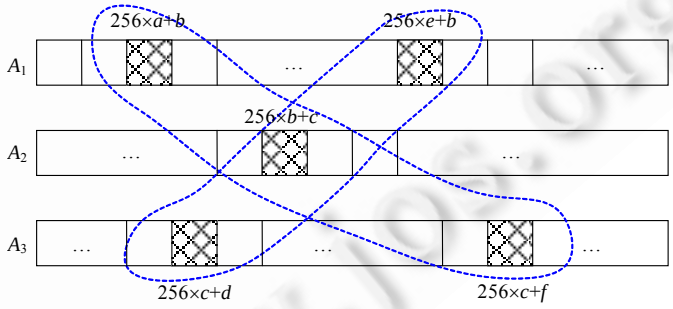


Fig.11 A normal address identifying error

图 11 合法地址识别错误实例

5.2 合法地址识别算法在IPV6地址空间的适用性

本文所提合法地址识别算法基于如下规律:随机伪造源地址攻击包在整个地址空间内均匀分布,从统计角度来看,每个单元记录的伪造攻击包数相同.合法地址对应的单元同时记录了伪造包数和合法包数,因而具有比伪造地址更多的数据包数.在 IPV6 地址空间下,这一规律保持不变,因而合法地址识别算法在 IPV6 空间下仍然

适用.由于地址识别准确性与 ECBF 的哈希个数相关,为了减少误报,需对本文所提出的 ECBF 结构进行扩充,增加更多的哈希函数.

5.3 局限性

由攻击包源地址的真实性角度来看,DDoS 攻击可分为采用真实地址和采用伪造地址两类.伪造地址可以进一步分为伪造固定地址、伪造子网地址、随机伪造源地址 3 种.本文所提合法地址识别算法依赖于攻击包源地址在地址空间内的随机分布特性,因而不适用于真实地址、伪造固定地址和伪造子网地址这 3 类攻击.

随机伪造源地址的分布范围与操作系统、CPU 和攻击代码相关,可分为在整个地址空间内分布和在部分地址空间内分布.虽然本文只关注攻击包源地址在整个地址空间内分布的情况,但由于攻击源地址在部分地址空间范围内分布时,伪造源地址在此范围内仍然服从随机分布,因而本文的合法地址识别算法可识别此范围内的合法地址,但需识别出攻击包源地址的分布范围.

合法地址识别算法利用了攻击包源地址分布的统计特性,而与具体协议无关,即识别算法适用于多种协议的随机伪造源地址 DDoS 攻击,如 SYN Flood 攻击、UDP Flood、ICMP Flood 或者多种协议的混合攻击.

6 结 论

本文针对随机伪造源地址 DDoS 攻击,基于其源地址分布的统计特征提出了一种过滤方法,能够有效识别合法地址,为合法流量提供保护.首先,提出了一个高效的数据包数统计结构 ECBF,基于此,提出了合法地址识别算法;然后分析了识别算法的准确性,并提出了一种识别参数调整算法,可以在不同的攻击规模下满足识别精度的要求.FSAD 依据数据包源地址的合法性,给数据包以不同的优先级转发.采用了真实网络数据进行模拟,结果表明,本文所提方法能够高精度地识别合法地址,有效保护合法流量,并可为有价值的长流提供更好的保护.

本文所提过滤方法无需建立复杂的训练模型,能够为新的合法用户提供流量保护,不依赖于具体协议,实用范围广.由于对地址的识别和合法地址的查询均采用哈希结构,仅需数兆字节内存开销,且时间复杂性为 $O(1)$,可实时在线过滤,或嵌入到边界路由器或者现有的网络安全设备中,如防火墙等,因而具有较好的应用前景.

References:

[1] http://net.chinabyte.com/519DNS/

[2] Ehrkenkranz T, Li J. On the state of IP spoofing defense. ACM Trans. on Internet Technology, 2009,9(2):6(1)−6(29).

[3] Kim Y, Lau WC, Chuah MC, Chao HJ. Packetscore: Statistic-Based overload control against distributed denial-of-service attacks. In: Proc. of the INFOCOM. HongKong, 2004. 141−155. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1354679 [doi: 10.1109/INFCOM.2004.1354679]

[4] Sun H, Zhaung Y, Chao HJ. A principal components analysis-based robust DDoS defense system. In: Proc. of the IEEE Int'l Conf. on Communications. Beijing, 2008. 1663−1669. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4533357 [doi: 10.1109/ICC.2008.321]

[5] Jin C, Wang HN, Shin KG. Hop-Count filtering: An effective defense against spoofed DdoS traffic. In: Proc. of the 10th ACM Conf. on Computer and Communications Security. Washington, 2003. 30−41. http://dl.acm.org/citation.cfm?id=948116 [doi: 10.1145/948109.948116]

[6] Peng T, Leckie C, Ramamohanarao K. Protection from distributed denial of service attacks using history-based IP filtering. In: Proc. of the IEEE Int'l Conf. on Communications. Anchorage, 2003. 482−486. http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F8564%2F27113%2F01204223.pdf%3Farnumber%3D1204223&authDecision=-203 [doi: 10.1109/ICC.2003.1204223]

[7] Dittrich D. http://staff.washington.edu/dittrich/misc/tfn.analysis.txt. 1999.

[8] Barlow J. http://packetstormsecurity.org/distributed/TFN2k_Analysis-1.3.txt. 2000.

[9] Li J, Mirkovic J, Wang MQ, Reiher P, Zhang LX. SAVE: Source address validity enforcement protocol. In: Proc. of the INFOCOM. New York, 2002. 1557−1566. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1019407 [doi: 10.1109/INFCOM.2002.1019407]

[10] Lee H, Kwon M, Hasker G, Perrig A. BASE: An incrementally deployable mechanism for viable IP spoofing prevention. In: Proc. of the 2nd ACM Symp. on Information, Computer and Communication Security. Singapore, 2007. 20−31. http://dl.acm.org/citation.cfm?id=1229293 [doi: 10.1145/1229285.1229293]

[11] Cormen TH, Leierson CE, Rivest RL, Stein C. Introduction to Algorithms. 2nd ed., Cambridge: MIT Press and McGraw-Hill, 2001. 101−122.

[12] Gong J, Peng YB, Yang W, Liu WJ. Reconstructing the parameter for massive abnormal TCP connections with bloom filter. Journal of Software, 2006,17(3):434−444 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/434.htm [doi: 10.1360/jos170434]

[13] Zou BX, Liu Q. ARMA-Based traffic prediction and overload detection of network. Journal of Computer Research and Development, 2002,39(12):1645−1652 (in Chinese with English abstract).

[14] Xu J, Lee W. Sustaining availability of Web services under distributed denial of service attacks. IEEE Trans. on Computers, 2003, 52(2):195−208. [doi: 10.1109/TC.2003.1176986]

[15] Ben-Porat U, Bremler-Barr A, Levy H. Evaluating the vulnerability of network mechanisms to sophisticated DDoS attacks. In: Proc. of the INFOCOM. Phoenix, 2008. 2297−2305. http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F4509594%2F4509595%2F04509893.pdf%3Farnumber%3D4509893&authDecision=-203 [doi: 10.1109/INFOCOM.2008.298]

[16] Arlitt M, Jin T. Workload characterization of the 1998 World Cup Web site. HP Labs Technical Report, 1999.

[17] http://mawi.wide.ad.jp/

[18] Mirković J, Prier G, Reiher P. Attacking DDoS at the source. In: Proc. of the IEEE Int'l Conf. on Network Protocols. Paris, 2002. 312−321. http://dl.acm.org/citation.cfm?id=656169

[19] Bremler-Barr A, Levy H. Spoofing prevention method. In: Proc. of the INFOCOM. 2005. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1497921 [doi: 10.1109/INFCOM.2005.1497921]

[20] Liu X, Li A, Yang XW, Wetherall D. Passport: Secure and adoptable source authentication. In: Proc. of the USENIX Symp. on Networked Systems Design and Implementation. 2008. 365−378. http://dl.acm.org/citation.cfm?id=1387615

[21] Park K, Lee H. On the effectiveness of router-based packet filtering for distributed DoS attack prevention in power-law Internets. In: Proc. of the ACM SIGCOMM. San Diego, 2001. 15−26. http://dl.acm.org/citation.cfm?id=383061 [doi: 10.1145/383059.383061]

[22] Ioannidis J, Bellovin SM. Implementing pushback: Router-Based defense against DDoS attacks. In: Proc. of the Network and Distributed System Security (NDSS). 2002.

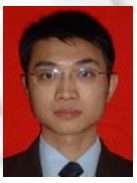
[23] Yau DKY, Lui JCS, Liang F, Yam Y. Defending against distributed denial-of-service attacks with max-min fair server-centric router throttles. IEEE Trans. on Networking, 2005,13(1):29−42. [doi: 10.1109/TNET.2004.842221]

附中文参考文献:

[12] 龚俭,彭艳兵,杨望,刘卫江.基于 Bloom Filter 的大规模异常 TCP 连接参数再现方法.软件学报,2006,17(3):434−444. http://www.jos.org.cn/1000-9825/17/434.htm [doi: 10.1360/jos170434]

[13] 邹柏贤,刘强.基于 ARMA 模型的网络流量预测.计算机研究与发展,2002,39(12):1645−1652.

肖军(1979－),男,江苏大丰人,博士生,CCF 会员,主要研究领域为 DDoS 攻击检测,DDoS 攻击过滤.



张永铮(1978－),男,博士,副研究员,CCF 会员,主要研究领域为网络安全.

云晓春(1971－),男,博士,教授,博士生导师,主要研究领域为网络安全,互联网建模.