一种针对拒绝服务攻击的IP逆向追踪方案

李毅超 闵 帆 夏梦芹 杨国纬
(电子科技大学计算机科学与工程学院 成都610054)

An IP Traceback Scheme Combating Denial of Service Attack

LI Yi-Chao MIN Fan XIA Meng-Qin YANG Guo-Wei
(College of Computer Science and Engineering, UEST of China, Chengdu 610054, China)

Abstract IP traceback technology is an important means combating Denial of Service (DoS) attacks in the Internet. Based on Deterministic Packet Marking and Probabilistic Packet Marking, this paper proposes a new IP traceback scheme which is both efficient and robust against mark field spoofing.

Keywords Denial of service, IP spoofing, IP traceback, Deterministic packet marking, Probabilistic packet marking

1 引言

拒绝服务攻击(Denial of Service, DoS)是指攻击者有意阻碍合法用户使用某一服务的行为[1]。DoS攻击者可能采取以下三种方式中的一种或多种:a使服务器崩溃;b用大量数据阻塞用户与服务器之间的通讯链路;c用大量请求消耗服务器资源[2]。拒绝服务攻击是Internet上对各商业站点攻击的重要方式[3]。

DoS攻击者一般使用IP伪造来隐藏其实际地址。入口/出口过滤(Ingress/Egress filtering)是防止IP伪造的强有力措施。入口过滤是指:路由器不允许来自网段S的一个包进入Internet,如果该包的源地址不在S的范围内[4]。出口过滤是指:路由器不允许一个包进入S,如果该包的源地址在S的范围内[3]。入口过滤使得IP伪造只能局限于攻击主机所在的网段,出口过滤则是对入口过滤的补充。要使入口/出口过滤起到实际效果,需要Internet上绝大多数机器配置相关软件,而这需要较长时间,也存在一定困难[5]。

IP逆向追踪[2,5~9]是寻找IP伪造攻击者(也称匿名攻击者)实际地址的一类重要手段,如果能追踪到攻击源,就能封锁来自该源的IP包,从而有效地阻止该攻击,甚至追究相应人员的法律责任。

2. 相关工作

包标记[5~8]是近年来提出的一种IP逆向追踪手段,分为确定性包标记与概率包标记。

2.1 确定性包标记

确定性包标记(Deterministic packet marking, DPM)是指:采用IP包的记录路径选项(IP Record Route option)[10],使路由器将其地址写入所转发IP包的可选项(option field)。这样目标主机只需查看包的可选项,就可知道其所经过的路径。

DPM的缺点在于:a)较大比例地增加了包长度。假设IP包平均长度为500字节,平均经过20个路由器,则其长度增加80字节(IPv4)显然无法容忍[6];b)不能保证IP包有足够的空间容纳增加的字节[5];c)较大程度增加了各路由器的操作。

2.2 概率包标记

概率包标记(Probabilistic packet marking, PPM)[6~8]是针对确定性包标记的缺点提出的,适用于对抗第1节所述b),c)型攻击,其常见方案为:

1 节点标记PPM。路由器在转发一个包时,以概率p将自身IP地址写入包头预留的节点域,如果该域已有数据,则将其覆盖[5]。

2 边标记PPM。各路由器在转发一个包时,以概率p将该包所经过的一条边(两个相邻路由器地址)写入包头预留的相应域,该边到目的主机的距离也能以累加方式获得[6]。

边标记PPM需要两个IP地址和一个距离域的空间,一般为72位,文[5]使用编码方法将其压缩到16位并能放入IP包头中的ID域,但这增加了算法的复杂性。下文提到PPM时,均特指节点标记PPM。

与DPM相比,PPM的优点在于:a)只需1个IP地址的空间;b)传输过程中不增加包长度;c)各路由器仅以概率p对包进行标记,减少了路由器操作。

PPM的缺点在于:a)推导出一条完整的路径是一个比较缓慢的过程;b)对于分布式DoS,该方法很可能失效[5]。

3. 复合的标记与追踪方案

本节结合确定性包标记与概率包标记方案,提出一种复合的标记与追踪方案。

3.1 假设条件

记攻击主机为A(Attacker,A可以不唯一);记被攻击主机为V(Victim,V唯一);记使用IP包中记录路径选项的包为RR包。

本文方案依赖于下列假设条件,在实际情况中,这些假设条件通常成立。

假设1 攻击包的源地址和标记域被伪造。

源地址被伪造是使用包标记方案的根本原因。除分布性很好的DoS攻击外,攻击者一般会伪造IP包源地址。伪造标记域可能较大程度干扰逆向追踪[6]。

假设2 网络寻径在较短时间内不发生改变。

理论上,Internet节点的拓扑结构随时可能发生改变,但路由器等网络重要组件的改变并不经常发生,在较短的时间(如几分钟)内,网络寻径改变的概率相当小。

李毅超 硕士,讲师,研究方向:计算机网络、网络信息安全。闵 帆 硕士,博士研究生,主要研究方向:分布式系统、主动网络。夏梦芹 硕士,博士研究生。杨国纬 教授,主要研究方向:分布式系统、主动网络。

假设3 令正常主机向 V 的最大发包速度为 NS，单个 A 向 V 的发包速度均为 P，且有 $P>NS\cdot K$。

对于 A 唯一(单点攻击)及 A 数量较少的分布式 DoS 攻击，K 为不小于100的正整数。

假设4 V 无法根据某个包内数据判断其是否攻击包。

根据假设3和假设4，能且仅能根据流量判断哪些路径是攻击路径。

3.2 方案描述

方案由本节密切相关的三个算法组成。

当路由器接收到包时，运行如下算法：

算法1 路由器运行的标记算法

1 如果该包为一般的 IP 包，且目标地址不为本机，则以概率 p 将本路由器 IP 地址填入包头的节点域并转发；(相当于节点标记 PPM)

2 如果该包为 I 类 ICMP 包，且目标地址不为本机，直接转发；

3 如果该包为 I 类 ICMP 包，且目标地址为本机，则将源地址与目标地址对换，生成一个 RR 包，并将 ICMP 包中的 *Key* 拷贝到该 RR 包相应域中，发送；

4 如果该包为 RR 包，且目标地址不为本机，将本机 IP 地址写入该包可选项中并转发。(相当于 DPM)

当一个主机发现通讯量超过一定阈值时，认为自己处于被攻击状态，运行如下算法：

算法2 被攻击主机运行的追踪算法

1 暂停其服务；

2 生成密钥 *Key*；

3 在 Δt 时间内，以到达 IP 包的标记域为目标地址，发送包含 *Key* 的 I 类 ICMP 包，并等待其回应；

4 运行算法3，分析到达的包含 *Key* 的 RR 包；

5 根据分析结果，向离 A 最近的路由器发送 II 类 ICMP 包，以使后者丢弃以 V 为目标地址的 IP 包。

接收到各路由器反馈的 RR 包后，V 就得到了一系列路径，这时需要分析哪些路径为攻击路径，以及哪些路由器(为与相关术语匹配，本算法中称之为"节点")离 A 最近。V 可以根据 RR 包得到以 V 为根的树(如图1所示，不包括 A)。

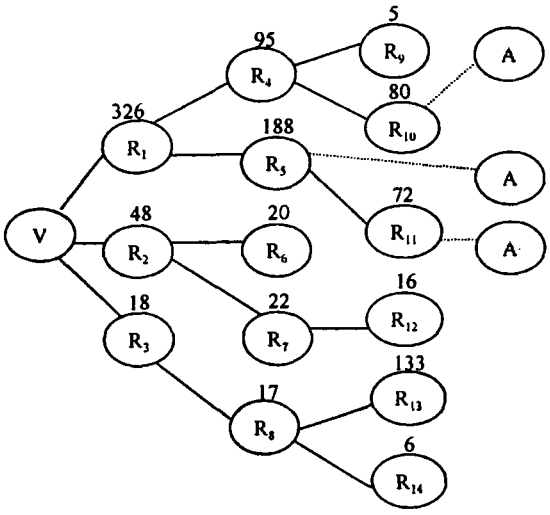

图1 拓扑结构示例

图1中，各节点上面的数据(称为标记数)表示以该节点为源向 V 发送 RR 包的数量。下面以图1为例描述 RR 包分析算法。

算法3 被攻击主机运行 RR 包分析算法

1 按节点层次从小到大的次序：记当前节点的标记数为 MP，其子节点数为 C，各子节点标记数为 MC_i 且 $MC_1\geqslant MC_2\geqslant\cdots\geqslant MC_C$。如果 $MP<\sum_{i=1}^{C}MC_i$，依次减小 $MC_1, MC_2, \cdots, MC_C$ 直到 $MP=\sum_{i=1}^{C}MC_i$。如图中应将 R_{13} 的标记数改为11。另例：节点 N 标记数为40，其四个子节点标记数依次为110，98，8，6，则应将各子节点标记数依次改为13，13，8，6；

2 按节点层次从大到小的次序：如果某节点的标记数小于直接兄弟节点最大标记数的1/5，删除以该节点为根的子树，如图中应删除 R_9；

3 按节点层次从大到小的次序：如果某节点已是叶节点，且其标记数小于同级节点平均标记数的1/2，删除该节点，如图中应依次删除 $R_{12}, R_{13}, R_{14}, R_6, R_7, R_8, R_2, R_3$；

4 所剩叶节点一定是与 A 相邻的路由器，如图中 R_{10}，R_{11}；

5 按节点层次从大到小的次序：如果某节点的标记数大于其子节点标记和的2倍，则该节点也一定是与 A 相邻的路由器，如图中 R_5。

3.3 方案说明

该方案实际上是 PPM 与 DPM 的综合，在正常情况下使用 PPM，而发现 DoS 攻击后，路由器可以根据 V 的要求使用 DPM。

本小节主要对算法3进行说明。将上下文所指节点简记为 N；由于节点与其地址存在一一对应关系，也将节点 N 的地址简记为 N；N 到 V 的跳数简记为 d。

1 由算法1可知：对于来自同一源地址的 IP 包，V 收到包的标记地址为 N 的概率为 $p\cdot(1-p)^{d-1}$，为 N 某子节点的概率则为 $p\cdot(1-p)^{d}$。因此有

$$MP\cdot(1-p)=\sum_{i=1}^{C}MC_i \tag{1}$$

显然 $p>0$，但考虑到 MP 与 MC_i 均为整数，所以有

$$MP\geqslant\sum_{i=1}^{C}MC_i \tag{2}$$

这样 $\sum_{i=1}^{C}MC_i>MP$ 有两种可能：a)由概率的不确定性，使得对少量较大标记值进行较小修改，不会对方案的结果产生实质影响；b)攻击者以 N 的某些子节点为标记地址伪造了大量包，其中最值得怀疑的是那些较大标记数。因此，算法3第1步能很大程度消除攻击者伪造标记地址的影响；

2 由假设3，能根据 N 向 V 的发包速度判断其是否节于 A 与 V，这是算法3中2，3两步的依据；

3 一个路由器可能直接或间接(通过另外的路由器)地接收到攻击包，这是算法3第5步的原因。

3.4 性能分析

从发现攻击到锁定离 A 最近的路由器所需时间 T_L 由三部分组成：I)搜集足够攻击包的时间 Δt；II)发送 I 类 ICMP 包到接收反馈 RR 包的间隔时间；III)算法3的运行时间。下面依次分析这三个部分。

I．不失一般性，令 A_i 表示某一确定的攻击源。为确认某一路由器 N 为离 A_i 最近的路由器，V 需要获得一定数量标记域为 N 的包。令需要相应包的数量为 R_N，接收到来自 A_i 的包数量为 R_A，则有

$$R_N = R_A \cdot p(1-p)^{d-1} \quad (3)$$

令 A_i 生成包的速度为 P，则

$$\Delta t = \frac{R_A}{P} = \frac{R_N}{P \cdot p(1-p)^{d-1}} \quad (4)$$

其中，R_N 与定位 A_i 的精确度有关，R_N 由 V 决定，路由器可以设置 p 以使 Δt 最小。

令 $$f(p) = p(1-p)^{d-1} \quad (5)$$

$$f'(p) = (1-p-d \cdot p+p)(1-p)^{d-2} = (1-d \cdot p)(1-p)^{d-2} \quad (6)$$

所以当 $$p = \frac{1}{d} \quad (7)$$

时，$f(p)$取最大值，Δt 取最小值。

在 Internet 环境下，一般设置 $d=20$[8]，这时应设置 $p = 0.05$；若进一步地设置 $R_N = 50$，则

$$\Delta t = \frac{2650}{P} \quad (8)$$

特别地，$P=500$(个/秒)时，$\Delta t = 5.3$ (秒)；$P=30$(个/秒)时，$\Delta t = 88.3$(秒)。

若 V 能缓存最近一段时间所收到的包，则 Δt 应相应地减小。

Ⅰ. 第二部分时间与信道的传输速率有关，一般情况下小于1秒。

Ⅱ. 算法4只涉及单重循环，其时间复杂度为

$$O(ND) \quad (9)$$

其中 ND 为向 V 发 RR 包的路由器的数量。算法4运行时间不会超过数秒。

综上所述，T_L 主要由 Δt 决定，而后者主要由 P 决定，当 A 数量较少时，为保持攻击的有效性，P 一般较大，这时 T_L 可控制在10秒内；而当 A 数量较多(分布性较好的 DDoS)时，P 可取较小值，这时 T_L 可能为几十秒或更大。P 较小所带来的另一重要问题是：此时 K 也为较小值(如小于10)，用本方案无法有效地区分攻击包与正常包。

3.5 伪造标记域的影响

A 伪造的标记域到达 V 不被修改的概率为

$$(1-p)^{d-1} \quad (10)$$

特别地，当 $d=20$，$p=0.05$时，该概率为0.3585。

PPM 无法获知包标记域是否被伪造，因此，大量的标记域伪造包将对逆向追踪起到干扰甚至误导作用[8]。

但要对本方案起到真正的干扰作用，攻击者必须对标记域进行如下伪造：a)使用路由器的 IP 地址，否则 V 根据该包发送的 I 类 ICMP 没有回应的 RR 包，该包被“过滤”掉；b)相应路由器地址的比例应大致符合(2)式。这两个条件要求攻击者对被攻击主机周围网络的拓扑结构有很好的了解，而这在实际中通常无法做到。

3.6 方案优点

本方案的主要优点在于：

1. 能有效对抗 K 较大的分布式 DoS 攻击；

2. 能快速反应，一般不超过几十秒；

3. 攻击停止后仍可进行，只需要 Δt 时间就能搜集足够的攻击包；

4. 比 DPM 针对性更强，只是在遇到攻击的时候对 RR 包使用 DPM；

5. 与 PPM 相比，能有效消除 A 伪造标记域的影响；

6. 无需人工干预，所有操作都可由路由器和主机 V 自动完成；

7. 算法1中 RR 包本身数据量非常少，加入沿途路由器地址后其长度不会超过 IP 包长度上限；

8. 由于采用了密钥 Key(虽然对密钥的使用相当简单)，攻击者难于伪造 RR 包。另外，对于攻击者来说，伪造 RR 包本身就是很容易被追踪的操作。

小结 本文结合 DPM 与 PPM 提出了一种新的逆向追踪方案，并对其性能进行了分析。该方案具有快速反应、抗干扰力强等优点，能有效对抗攻击节点数量不是很大的 DoS 攻击。如果取消假设4的限制，本方案可获得更好性能。

参 考 文 献

1 Denial of Service Attacks. CERT Coordination Center, Oct. 1997. Available at: http://www.cert.org/tech_tips/denial_of_service.html

2 Houle K J, Weaver G M, Long N, Thomas R. Trends in Denial of Service Attack Technology. CERT Coordination Center, Oct. 2001

3 Garber L. Denial-of-service attacks rip the Internet. Computer, Apr. 2000. 12～17

4 Ferguson P, Senie D. Network ingress filtering: Defeating denial-of-service attacks which employ IP source address spoofing. RFC 2827, 2000

5 Savage S, Wetherall D, Karlin A, Anderson T. Practical Network Support for IP Trace-back. In: Proc. 2000 ACM SIGCOMM, vol. 30, no. 4, ACM Press, New York, Aug. 2000. 295～306

6 Burch H, Cheswick B. Tracing anonymous packets to their approximate source. In: Proc. 2000 USENIX LISA Conf. Dec. 2000. 319～327

7 Baba T, Matsuda S. Tracing Network Attacks to Their Sources. IEEE Internet Computing March. April 2002. 20～26

8 Park K, Lee H. On the Effectiveness of Probabilistic Packet Marking for IP Traceback under Denial of Service Attack. In: Proc. IEEE INFOCOM '01, 2001

9 Bellovin S, Leech M, Taylor T. ICMP Traceback Messages. Internet draft, work in progress, Jan. 2003; available online at http://www.ietf.org/internet-drafts/draft-ietf-itrace-03.txt (expires July 2003)

10 Postel J. Internet protocol. RFC 791, 1981