网络攻击路径重构中的报文标记方案研究

褚为民[1], 陈 波[1,2], 于 泠[1,2]

(1.解放军理工大学 通信工程学院,江苏 南京 210007;
2.南京师范大学 数学与计算机科学学院,江苏 南京 210097)

[摘要] 对目前攻击源追踪中的基于概率的分段报文标记方案进行了分析,指出了这类方法的一些缺陷.给出了利用 IP报文中的选项字段,以概率将流经路由器的地址标注报文,使得受害主机能够根据被标注报文内的地址信息重构出攻击路径的代数方法.重点讨论了路由器在报文中记录流经的地址以及利用报文中记录的信息重构路径的代数方法,并对代数方法的性能作了分析和比较.方案具有很低的网络和路由器开销,也容易扩充到 IPv6和未来的主干网.

[关键词] 攻击源追踪,路由,TCP/IP,网络安全
[中图分类号]TP393 07 [文献标识码]B [文章编号]1672-1292-(2005)01-0061-04

Research on the PacketsMarking Schemes of Reconstructing-Path in Network Attack

CHU Weimin[1], CHEN Bo[1,2], YU Ling[1,2]

(1. School of Communication Engineering, PLA University of Science and Technology, Jiangsu Nanjing 210007, China;
2. School of Mathematics and Computer Science, Nanjing Normal University, Jiangsu Nanjing 210097, China)

Abstract The paper analyzes the probabilistic FMS (FragmentMarking Scheme) of IP traceback, points out some limitations of such schemes, and discusses a scheme based on probabilisticmarking packet by using algebraic coding theory. The victim can use the edges sampled in these packets to reconstruct attack path. The algebraicmethod is analyzed of marking packets and reconstructing attack path algorithm. The performance of thismethod is analyzed and comparedwith FMS. The scheme has very low network and router overhead and support incremental deployment in IPv6.

Key words traceback attack source; router; TCP/IP; network security

0 引言

发起 DDoS(Distributed Denial of Service,分布式拒绝服务)攻击的攻击者通常都采用了地址欺骗,这为追查攻击者的真实位置设置了障碍.而对网络攻击的追踪是对网络攻击做出正确响应的重要前提.因此,需要研究如何重构攻击路径,或对攻击源地址和攻击路径作出尽可能真实的定位.这是当前非常具有研究意义和挑战性的课题.

1 分段报文标记方案分析

1.1 基本思想

文献[1,2]提出了基于概率的分段报文标记策略(FragmentMarking Scheme,FMS).即当报文到达路由器时,把报文所经两个相邻路由器 IP地址形成的"边"的信息和其他相关信息分成 8个数据段,选择其中的一段以概率标记报文.这样虽然每个报文只包含路径的部分信息,但是当 DOS或 DDOS攻击发生时往往会有大量的攻击报文,被攻击主机可以得到足够的信息恢复出完整的攻击路径.

为此,需要在 IP报文中设置 4个域:起始地址域、结束地址域、距离域和标志域,其中前面两个用以表示连接一条边的两端路由器地址,距离域表示一个边样本离受攻击者的距离,标志域表示该报文是否被标注.

收稿日期: 2004-06-29
基金项目: 江苏省政府基金资助项目(BR2003015)和江苏省高校自然科学重点资助项目(03KJA52066).
作者简介: 褚为民(1975-),硕士研究生,主要从事军事运筹学的研究. E-mail: cwmo818@sina.com
通讯联系人: 陈 波(1972-),博士研究生,讲师,主要从事网络信息安全、人工智能的教学与研究. E-mail: bchen@njnu.edu.cn

1.2 标记算法

每个路由器计算其地址的hash值,并与IP地址连接成R',再将R'分成互不覆盖的8个分片.当一个路由器决定标注一个报文,它将随机选择第i个分片写入边分片域,并在位移域写入相应的位移量,同时将距离域赋0,标志域置1.如果距离域已经为0,则表明报文已被前面的路由器标注过.这时,路由器将位于同一位移处的地址分片与前面路由器的分片作异或后写入边分片域,表示这条边连接自身与前面的路由器.如果路由器不标注报文,它将不断增加距离域.

1.3 重构算法

重构过程中,要在被攻击机V处将收到的被标注报文中的边分片标记重组,并进行确认,确认从最靠近V处的边开始,依次上行,直到距离V最远的边为止.最终得到一棵以被攻击机V为树根的树,这就是一个有效的攻击路径集.

1.4 算法分析

这种分段方法具有很高的计算复杂性和很高的误警率.仿真实验表明,当面对25个攻击者的分布式攻击时,FMS约要一天才能构造出完整的攻击树并且最终得到的是数千个可能的攻击节点.

在被攻击机定位攻击者和DOS攻击的强度之间存在着折衷的关系,该关系可以用标记概率、路径长度和流量特征为参数的函数来表示[4].攻击者和被攻击主机之间的最优决策问题(被攻击主机可以选择标记概率,而攻击者则可以选择伪造标记值,伪造源地址和增大攻击流量)可以表示成受约束的最小最大优化问题.增大标记概率可以提高发现攻击者的概率,但是标记概率受到IP报文中可以利用的空间的限制.在目前的Internet中,采用该方案,单个攻击者可以被定位到2~5个可能的攻击者范围.如果攻击者采用DDOS攻击方式,则攻击者的不确定性就被放大了,也就是降低了本分段报文标记方案的实际效果.

此外,如果报文标记不经过认证则很容易被攻击者利用.如果攻击者控制了网络中的某台路由器,则这台路由器就可以进行虚假的报文标记从而导致被攻击主机不能正确的恢复出攻击路径.

本文在文献[1~4]的基础上,采用代数方法,利用IP报文中的选项字段,以概率将流经路由器的地址标注报文,使得受害主机能够利用被标注报文内的地址信息重构出攻击路径,从而追踪到攻击源点的技术.如何运用代数方法记录报文流经路由器的地址,以及如何利用报文中记录的信息重构路径是讨论的重点.

2 代数编码报文标记方案分析

2.1 基本思想

利用代数方法编码[5]一条完整的路径.在路径开始端设:

$$FullPath_{0,j} = 0 \tag{1}$$

每个路由器首先产生一个[0.1]之间的随机数r,若r小于某个预定的概率值p时,则认为该路由器有权标注报文.若一个有权标注报文的路由器收到一个还未处于"标注状态"的报文时,就认为该路由器是第一个路由器,并为该路径随机选取一个x_j,计算

$$FullPath_{i,j} = (FullPath_{i-1,j} * x_j + A_j) \bmod p \tag{2}$$

其中,$A_1, A_2, \cdots A_n$为路径P上各路由器的32位IP地址,$A_{j,1}, A_{j,2}, \cdots, A_{j,k}$是$A_j$的分块;$x_j$为第$j$个报文的一个随机数;$p$为比$2^{32}$小的最大的素数.

同时将报文标记为"标注状态",然后将$FullPath_{i,j}$值和x_j一起记录在报文中传送到路径上的下一个路由器.

若一个有权标注报文的路由器收到处于"标注状态"的报文时,用(2)式计算,将自己的IP地址加入$FullPath_{i,j}$中.路由器不必知道路径的总长和在路径上的位置便可计算路径编码.

为了减少标注位,将一个路由器IP地址分成k个信息块,并附加$\lceil \log_2 k \rceil$个比特位来表明在一个给定的报文中记录哪个信息块.每个路由器将其所有的信息块加到同一个报文中.方法是为每个报文中的多项式添加k个系数,即每个路由器更新FullPath共k次,依次替换它们IP地址的每个信息块.以将路由器j的IP地址A_j分成$k=4$块(每块长8位)为例:

(1)取小于2^8的最大的4个素数:$m_1 = 233$, $m_2 = 239$, $m_3 = 241$, $m_4 = 251$;

(2)计算$A_{j,1} = A_j \bmod m_1$, $A_{j,2} = A_j \bmod m_2$, $A_{j,3} = A_j \bmod m_3$, $A_{j,4} = A_j \bmod m_4$

(3)若路由器j有权标注报文,则计算$((((((FullPath * x_j) + A_{j,4}) * x_j + A_{j,3}) * x_j + A_{j,2}) * x_j + A_{j,1}) \bmod m_4$

当$k=4$,此方法只需要$\log_2 2^8 + \log_2 16 + 1 = 13$位和$4d$个报文.

重构路径时,对于

$$f_j(x) = A_{1,1} + A_{1,2}x + \cdots + A_{1,k}x^{k-1} + A_{2,1}x^k +$$

$$\cdots + A_{n,k}x^{nk-1} \quad (3)$$

可以通过解下面的 GF(P) 域上的线性方程组得到 $A_{j,1}$、$A_{j,2}$、$A_{j,3}$、$A_{j,4}$.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{nk-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{nk-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{nk-1} \end{bmatrix} \begin{bmatrix} A_{1,1} \\ A_{1,2} \\ \vdots \\ A_{n,k} \end{bmatrix} = \begin{bmatrix} \mathrm{FullPath}_{n,1} \\ \mathrm{FullPath}_{n,2} \\ \vdots \\ \mathrm{FullPath}_{n,3} \end{bmatrix}$$

该方程组的系数矩阵的行列式经过行列互换后, 是一个范德蒙行列式, 只要所有的 x_j 互不相同, 该行列式值必不为 0, 该方程组必有解.

根据求出的 $A_{j,1}$、$A_{j,2}$、$A_{j,3}$、$A_{j,4}$, 利用中国剩余定理, 就可以还原出路由器 IP 地址 A_j, 步骤如下:

令 $M = \prod_{i=1}^{k} m_i = m_1 m_2 \cdots m_k$, $M_j = M/m_j = \prod_{\substack{i=1 \\ i \neq j}}^{k} m_i$, 令 y_j 满足 $M_j y_j \equiv 1 (\mathrm{mod} m_j)$, $j = 1, 2, \cdots, k$.

由于 $(M_j, m_j) = 1$, 故 y_j 存在, 则 $A_j \equiv A_{j,1}M_1 y_1 + A_{j,2}M_2 y_2 + \cdots + A_{j,k}M_k y_k (\mathrm{mod} M)$ 就是问题的解, 而且是唯一的.

地址分片后, 对于长度为 d 的路径, 最后所要恢复的多项式 (3) 中将含有 kn 个未知量, 这将增加重构的复杂性, 因此我们把对全路径的编码简化为对相邻两个路由器形成的边的编码.

例如相邻两个路由器的 IP地址分别是 Z 和 Y, 各分成 4片: z_1, z_2, z_3, z_4 和 y_1, y_2, y_3, y_4, 边的代数编码则为: $\mathrm{Edge}(x) = (z_1 + z_2x + z_3x^2 + z_4x^3 + y_1x^4 + y_2x^5 + y_3x^6 + y_4x^7) \bmod 251$. 要求该边两端的路由器地址, 只需恢复多项式 $f_{\mathrm{Edge}}(x) = z_1 + z_2x + z_3x^2 + z_4x^3 + y_1x^4 + y_2x^5 + y_3x^6 + y_4x^7$ 中的各项系数, 再利用中国剩余定理就可以还原出边上的两个路由器的 IP地址 Z、Y. 为了便于边的重构, 我们将“标注状态”标记位修改为记录边样本离受攻击机的距离.

IP选项字段最适合添加信息, 本方案用此字段来存放路径上各边的代数编码. 其中设置 3个域: 累加器域 (accumulator)、随机数域 (random) 和距离域 (distance). 其中累加器域占 8位, 用来存放边代数编码; 随机数域占 4位, 用来存放随机数 x; 4位距离域表示一个边样本离受攻击机的距离, 如图 1所示.

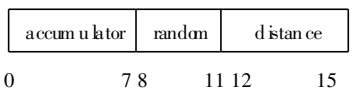

图 1 记录域的设置

2.2 路由器标注报文过程

标注过程伪代码如下, 仍以 k= 4为例.

```
at each router:
Z= the router's IP address
//每个路由器将其 IP地址分成 4片
Z1= Z mod 233
Z2= Z mod 239
Z3= Z mod 241
Z4= Z mod 251
for each packet w
r= random (0..1)
if r < p then //当一个路由器 j有权标注一个报文时
  x= random integer (0..7)
//产生一个随机数 xj存入随机数域
  w. accumulator= (z1 + z2 x+ z3 x^2 + z4 x^3) mod509
//将其 IP地址的 4个分片加到累加器域中
  w. random= x
  w. distance= 0 //将距离域赋 0
else
  if w. distance= 0 then //如果距离域已经为 0, 则表明报文已被前面的路由器标注过
  w. accumulator= (w. accumulator* w. random) + z4
//路由器将其 IP地址的 4个分片加到累加器域即可
  w. accumulator= (w. accumulator* w. random) + z3
  w. accumulator= (w. accumulator* w. random) + z2
  w. accumulator= (w. accumulator* w. random) + z1
increment w. distance
//其他情况下路由器不标注报文, 只不断增加距离域的值
```

考虑到路由器有可能被攻击者所控制, 在标记算法时可以加入认证功能, 如 MAC、HMAC-MD5等, 提高标记的可靠性. 然而基于公钥的数字签名有两个问题, 首先是签名非常费时 (如 PII 500 MHz 的机器对 1 024 bit的 RSA 每秒只能签 100个左右), 其次是空间开销大 (1 024 bit的 RSA签名需 128字节). 因此我们使用 MAC对标记部分进行密钥认证, 如 HMAC-MD5效率是 1 024 bit RSA 认证的 103到 104倍, 尤其是适合于仅有 32 bit存储空间上进行的认证方案.

2.3 攻击路径重构过程

攻击路径重构时采用的拓扑结构见图 2 当 DoS或 DDoS攻击发生时, 流量监测机收集报文, 结合该机定期收集到的上游拓扑结构图, 再配合可视化的方法, 可以使受攻击机快速重构出攻击路径.

重构过程中, 在受攻击机处, 将收到的被标注报文中的边代数编码重构, 最终得到一个有效的攻

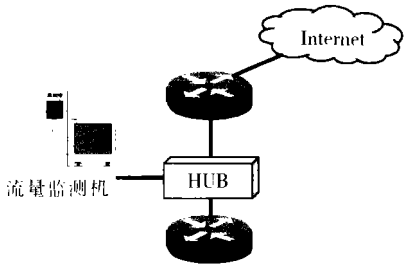

图 2　重构路径时采用的拓扑结构

击路径.重构算法的伪代码如下:

```
//首先将这些边代数编码按距离分类,
//将距离同为 d(d= 1, 2, …)的组成一组线性方程
for each packet w
    group by w. distance into matrix equation
//求解 d= 1的线性方程组,得到距离受攻击机 d跳处的边两端的路由器 IP地址分片,
//再利用中国剩余定理求出两个路由器的 IP地址,并存入攻击路径集合 Path
d= 1
while (d< = maxd)
    solving matrix equation which w. diatance= d
    if (d= 1) then
        let results Z and Y into PathSet
    else
        //若求出的两个 IP地址中有一个属于 Path集合,则将另一个也加入 Path集合,
        if (Z or Y in PathSet) then
            let Z or Y into PathSet
    increament d //逐次增加 d值,求出距离受攻击机 d跳处的路由器 IP地址,最终得到一条有效攻击路径.
```

2.4　算法分析

可行性分析,假设每个报文以等概率进行标注,那么收到一个来自 d跳处路由器的被标注报文的概率为$P(1-P)^{d-1}$. 接收者在收到来自 d跳处路由器的单一样本前,至少收到$\frac{1}{P(1-P)^{d-1}}$个报文. 由于每条边标识分成 k个分片,因此要重构距离被攻击者 d跳处的边,至少需要收$\frac{k}{P(1-P)^{d-1}}$个报文. 所以重构整条路径(最大距离为 maxd),至少需收到$\sum_{d=1}^{maxd}\frac{k}{P(1-P)^{d-1}}$个报文. 这在拒绝服务攻击的情况下是不难做到的.

仿真实验表明,采用代数编码的概率报文标记和重构算法时间主要在于稠密线性方程组的求解上,用有回代的高斯的消去法或选主元的高斯消去法,时间复杂度限定在 $O(n^3)$. 此外该算法易于并行化,如果假定使用 n^2+n个处理器,并排列成 $n*(n-1)$的阵列,在 PRAM上采用高斯-约旦消去法[6],算法的时间复杂度将为 $O(n)$.

3　小结

本文研究的报文概率标记技术可以重构出攻击路径的候选攻击路径集,追踪到实施了追踪功能的网络边界,可以使 ISP部署相关策略对这些范围的主机加以防范. 网络入侵追踪仅仅是保证网络安全的第一步. 通过有效的入侵追踪,应该采取进一步的措施,如阻断或抑制入侵连接,这样通过动态调整网络保障边界,以更好地保障网络安全. 这也是未来网络安全模型的发展方向.

[参考文献]

[1] Stefan Savage, David Wetherall, Anna Karlin, *et al*. Network support for IP traceback[J]. IEEE/ACM Transactions on Networking, 2001, 9(3): 226-237.

[2] Song X D, Perrig A. Advanced and authenticated marking schemes for IP traceback[A]. In Proceeding of IEEE INFOCOM[C]. 2001. 878-886.

[3] Drew Dean, Matt Franklin and Adam Stubblefield. An algebraic approach to IP traceback[A]. In Proceedings of NDSS'01, 2001. 318-326.

[4] Kihong Park, Heejo Lee. On the effectiveness of probabilistic packet marking for IP tracback under deial of service attack[A]. In Proceeding of IEEE INFOCOM[C]. 2001. 338-347.

[5] 冯贵良, 吴新文. 代数几何码[M]. 北京: 科学出版社, 2000. 72-103.

[6] 陈国良. 并行算法的设计与分析[M]. 北京: 高等教育出版社, 2002. 133-149.

[责任编辑: 刘健]