

IP 追踪技术压缩边分片采样算法的改进

刘 丽, 黄芳芳

(北京科技大学 信息工程学院, 北京 100083)

摘 要: IP 追踪技术是防御拒绝服务攻击 (DoS) 的重要措施, 文中针对用于 IP 追踪的压缩边分片采样算法 (CEFS) 存在的不足, 提出了标注分片自适应概率包标记算法 (LFAPPM), 该算法通过扩大标记空间和采用自适应概率的方法, 减少了重构路径所需包数, 并通过给分片加标注, 减少了重构路径的计算量和误报数, 而且通过初始化没被边界路由器标记的包标记空间, 加强了抗干扰性. 与其他算法进行比较, LFAPPM 算法各种性能指标较优.

关键词: 拒绝服务攻击; IP 追踪; 压缩边分片采样算法; 标注分片自适应概率包标记算法

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)12-0078-04

Improvement of Compressed Edge Fragment Sampling Algorithm for IP Traceback

LIU Li, HUANG Fang-fang

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: IP traceback is an important measure to defend against Denial of Service (DoS) attack. Based on Compressed Edge Fragment Sampling algorithm (CEFS) for IP traceback, a new algorithm is proposed, called Label fragment adaptive probability packet marking (LFAPPM). By enlarging marking space and using adaptive probability for packet marking, the LFAPPM algorithm reduces the number of packets needed for attack path reconstruction. Moreover, the algorithm reduces computation and false positive number in reconstructing multiple attack path by labeling fragment. Then the algorithm strengthens the anti-interference capability by initializing marking space of packets which are not marked by Border Router. Compared with other algorithms, the LFAPPM algorithm has better performance.

Key words: DoS; IP traceback; CEFS; LFAPPM

1 引言

拒绝服务攻击 (Denial of Service attack, DoS) 是当今网络中最为严重的问题之一. 找到攻击源是迅速解决 DoS 攻击的重要方法, 但由于 TCP/IP 协议的源地址可以伪造, 找到真正的攻击源很难. 不过尽管 IP 报头中的源地址可能是虚假的, 每个数据包仍然需要通过攻击者和受害者之间的路由器转发, 因此, 可以借助路由器获得攻击路径, 反向追踪到攻击源, 这就是 IP 追踪技术的主要思想. 目前研究者提出多种 IP 追踪技术, 其中压缩边分片采样算法^[1] (Compressed Edge Fragment Sampling algorithm,

CEFS) 是一种较为有效的 IP 追踪技术, 不论在路由器开销, 还是在实现难易以及与 IP 协议的兼容性等方面都有突出的优势. 但是 CEFS 算法也存在一些问题, 在有多条攻击路径的情况下, 存在重构路径的误报数和计算复杂度高, 重构路径所需数据包数多和抗干扰性差等问题. Song 等改进了 CEFS 算法, 提出了高级包标记算法 (Advanced Marking Scheme, AMS)^[2], AMS 算法降低了重构路径的误报数, 但是在重构路径时需要知道上游的网络拓扑图. Peng 等改进了 CEFS 算法, 提出了自适应概率包标记算法 (Adaptive Probability Packet Marking, APPM)^[3], APPM 算法降低了重构路径所需包数,

收稿日期: 2009-10-15; 修回日期: 2009-12-16

基金项目: 国家自然科学基金项目 (60873193)

但它与 CEFS 算法一样仍然存在路径重构时误报数和计算复杂度高的问题. 文中提出了标注分片自适应概率包标记算法 (Label Fragment Adaptive Probability Packet Marking, LFAPPM). 与同类算法相比, LFAPPM 算法减少了重构路径所需数据包数, 降低了重构路径的误报数和计算复杂度, 加强了抗干扰性, 从而能及早对 DoS 攻击做出响应.

2 压缩边分片采样算法研究

2.1 压缩边分片采样算法

IP 报文被分片的可能性小于 0.25%^[4], 因此可利用标识域来标记信息, CEFS 算法利用标识域作为标记空间, 采用三种技术减少了对 IP 报头的存储要求. 首先, 边用相邻路由器的异或值表示, 节省一半的存储空间. 其次, 把每个边分成 8 个不重叠的分片, 额外用 $\log_2 8$ 位存储分片在原始地址中的偏移. 最后, 增加了误差校验, 每个边 ID 由其地址和该地址的 32 位 hash 值逐位交叉构成. 当路由器标记 IP 包时, 其随机地从 8 个分片中选一个, 连同对应的偏移, 以及距离 (在标记时置为 0) 填入到该包的标识域中. 当路由器选择不标记包时, 它先检查距离域的值是否为 0, 如距离为 0, 则将其距离加 1, 并将其地址与标识域中偏移量对应的分片与包中已有的分片异或再重新填入; 如果距离不是 0, 只需将距离加 1. 路径重构时, 通过重组距离值相同而偏移值不同的分片建立候选边, 并使用由边计算出来的 hash 值与分片所携带的 hash 值比较, 相同则为合法边. 距离为 0 时, 计算出的边就是路由器的 IP 地址, 所以通过与上游邻近边相异或, 逐步得出上游路由器的 IP 地址, 从而重构出完整的路径. CEFS 算法中标记信息的编码如图 1 所示.

0	23	78	15
偏移	距离	边分片	

图 1 CEFS 算法中标记信息的编码

2.2 压缩边分片采样算法存在的问题

- (1) 受害者重构路径所需数据包数目大;
- (2) 有多条攻击路径时, 存在重构路径时检查包分片组合数目大和误报数高的问题;
- (3) 重复标记引起路由器开销大;
- (4) 攻击者对重构攻击路径易进行干扰.

3 标注分片自适应概率包标记算法

针对 CEFS 算法存在的问题, 提出了标注分片自适应概率包标记算法 (LFAPPM), 具体方案和算

法如下介绍.

3.1 LFAPPM 算法设计方案

(1) 扩大标记空间, 采用点边分片分开存储的方法, 并给属于同一个路由器的分片加上相同的标注. 扩大标记空间, 减少了分片数目, 从而降低了重构路径所需包数. 通过给属于同一路由器的分片加上相同的标注 (分片所属路由器地址的 15 位 hash 值), 重构路径时只需把相同距离、相同标注的分片按照偏移量的顺序组合起来即可, 大大减少了重构路径检查包分片组合数目. 同时点分片 (路由器地址分片)、边分片 (路由器地址分片与同偏移域的下游相邻路由器地址分片的异或值) 分开存放, 重构路径时不需要事先了解网络拓扑图就可以验证所得攻击路径中节点的正确性.

标记空间除了和 CEFS 算法类似用标识域外, 增加了保留位、片偏移字段和服务类型域作为标记域. 既然标识域被用来标记信息, 片偏移字段也就失去了意义, 也可被用来标记信息. RFC791 定义保留位为 0, 但当该位为 1 时, 包的转发不会受到影响, 该位也可用来标记信息. 目前服务类型域没有统一的使用规范, 也可用来作标记域, durreesi 等^[5]证实了使用该域不会对包的转发产生影响. 这样就有 38 位标记空间. 扩大标记空间后, LFAPPM 算法中标记信息的编码如图 2 所示.

0	12	67	14 15	22 23	37
偏移	距离	点分片	边分片	标注	

图 2 LFAPPM 算法中标记信息的编码

(2) 采用自适应概率来标记包, 标记包时采用非重复标记, 即如果包已被前面的路由器标记过, 则后面路由器对其不再标记. 假设攻击者到受害者的距离是 $d+1$, 从攻击者到受害者需经过 d 个路由器到达受害端, A 为攻击者, V 为受害者, 攻击路径为 $(A, R_1, R_2, \dots, R_d, V)$. 根据自适应概率标记策略^[3]的思想, 保证受害者收到每个路由器上的标记包的概率都是 $1/d$ 时, 重构路径所需包数最少. 设 R_i 标记数据包的概率为 p_i , 数据包被 R_1 标记后不再被后面的路由器标记, 所以它到达受害者的概率为 $p_1 = 1/d$, 数据包被 R_2 标记到达受害者的概率是 $(1 - p_1)p_2 = 1/d$, 对于其他路由器作类似推导, 得以下方程组:

$$\begin{aligned}
 p_1 &= 1/d \\
 (1 - p_1)p_2 &= 1/d \\
 (1 - p_1)(1 - p_2)p_3 &= 1/d \\
 &\dots\dots
 \end{aligned}$$

$$(1 - p_1)(1 - p_2) \cdots (1 - p_{d-1})p_d = 1/d$$

解得: $p_1 = 1/d, p_2 = 1/(d-1), p_3 = 1/(d-2), \dots, p_d = 1$. 即 $p_i = 1/(d+1-i) = 1/d_v$, d_v 为路由器到受害者的跳数. Internet 路由协议是基于目的地址的, 当路由器传送数据包时, 查询路由表就可知道自己到目的地的距离, 因此该方法易于实施.

(3) 如果数据包在经过边界路由器时没有被标记, 则将其标记空间置为 0, 这样攻击者伪造的信息就会被覆盖掉, 对重构攻击路径不会产生干扰.

3.2 LFAPPM 算法

NodeFragment 代表点分片; EdgeFragment 代表边分片; 片偏移、距离、标注、标记空间分别用 offset、distance、label、MarkSpace 表示. 具体算法如下:

(1) 路由器 R_i 的标记过程

```

let L=Hash(R)
let k be the number of non-overlapping fragment in R
for each packet w
let x be a random number from[0..1)
let o be random integer from[0..k-1]
let f be the fragment of R at offset o
在边界路由器处:
根据 w 的目的 IP 地址查路由表取得  $d_v$ 
if(  $x < 1/d_v$  ) then
write f into w. NodeFragment
write 0 into w. EdgeFragment
write o into w. offset
write L into w. label
write 0 into w. distance
else
write 0 into w. MarkSpace
在中间路由器处:
if(w. MarkSpace= 0) //包没被标记过
根据 w 的目的 IP 地址查路由表取得  $d_v$ 
if(  $x < 1/d_v$  ) then
write f into w. NodeFragment
write 0 into w. EdgeFragment
write o into w. offset
write L into w. label
write 0 into w. distance
else
increment w. distance
else if(w. MarkSpace! =0) //包已被标记过
if( $d = 0$ )
let f be the fragment of R at offset w. offset
write  $f \oplus w. NodeFrag$  into w. EdgeFrag
increment w. distance

```

else

increment w. distance

(2) 受害者 v 重构过程

```

let FragTbl be a table of tuples(offset, distance, Node-
Fragment, EdgeFragment, label)
let G be a tree with root v
let edges in G be tuples (start, end, distance)
let maxd:=0 //置 maxd 为 0
let last:=v //置 last=v
for each packet w from attacker
FragTbl.Insert(w. offset, w. distance, w. NodeFrag-
ment, w. EdgeFragment, w. label)
if (w. distance>maxd) then
maxd:=w. distance
for d:=0 to maxd
for all ordered combinations of fragments with the same
label at distance d
//对于距离为 d 具有相同标注的分片按照片偏移量顺
序组合
construct node  $u_d$  and edge  $z_d$ 
//构造节点  $u_d$  和边  $z_d$ 
remove these combinations with hash( $u_d$ ) $\neq$ label
if ( $d=0$ ) then
insert edge ( $u_d$ , last, d) into G
last:= $u_d$ 
if ( $d\neq 0$ ) then
 $u_d := z_d \oplus last$ 
if ( $u_d = u_d$ ) then
insert edge ( $u_d$ , last, d) into G
last:= $u_d$ 
extract path( $R_1 \cdots R_j$ ) by enumerating acyclic paths in G

```

4 算法比较

从重构路径所需包数、误报数、路径重构时^[6-7]计算量、路由器开销、抗干扰性等几个方面, 对文中提出的 LFAPPM 算法与其他包标记算法 CEFS、AMS、APPM 进行比较.

通过模拟实验来比较几种算法重构路径所需包的个数. 采用 1000 次实验的平均值作为实验结果.

首先验证了 CEFS 算法在概率 p 固定的情况下重构路径所需包的个数, 实验中攻击者到受害者的距离从 1 增加到 30, 标记概率 p 的选择分别为 0.01、0.03、0.04、0.08、0.12. 从图 3 中可以看出, CEFS 算法在 $p=0.04$ 时随着路径的增长, 重构路径所需包数增加的速度最为缓慢, 重构路径所需包的个数最少.

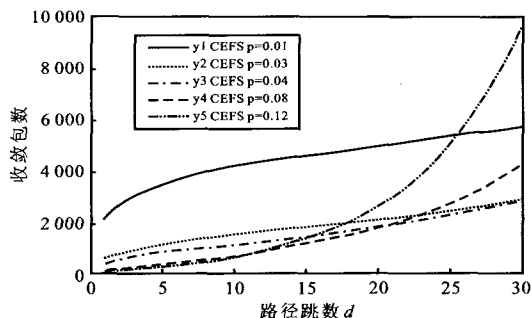


图3 CEFS算法概率不同时重构路径所需包数的比较

比较这几种算法重构路径所需包数,CEFS算法和AMS算法的概率选取 $p=0.04$,APPM算法的概率选取 $1/d_A$,LFAPPM算法的概率选取 $p=1/d_v$ 。从图4中可以看出,LFAPPM算法与同类算法相比,重构路径需要较少的包。

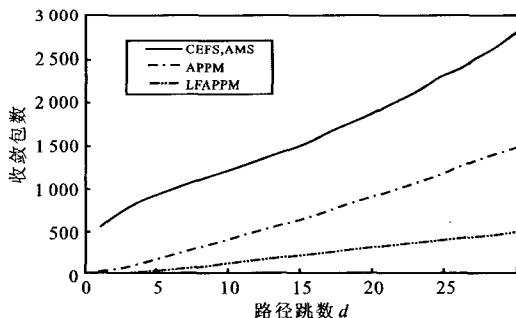


图4 各种算法重构路径所需包数的比较

由上面分析可知,LFAPPM算法在不需要上游网络拓扑图的情况下,重构路径所需包数比同类算法都少。重构路径所需包数和工作量越少,就能更快地重构出攻击路径,从而更快地找到攻击源。此外,LFAPPM算法降低了误报数和攻击者的干扰因素,这样就能及时更好的采取防御措施,减少攻击所带来的危害。

5 结束语

IP追踪^[8]是追踪攻击源普遍采用的方法。文中提出了LFAPPM算法,与其他包标记算法相比,在

重构路径所需包数、误报数、路径重构时计算量、路由器开销、抗干扰性等方面,LFAPPM算法表现出较好的性能。

参考文献:

- [1] Savage S, Wetherall D, Karlin A, et al. Network support for IP traceback[J]. IEEE /ACM Transactions on Networking,2001, 9(3): 226—237.
- [2] Song D, Perrig A. Advanced and authenticated marking schemes for IP traceback [C]// Proceedings of the Twentieth Annual Joint Conference on IEEE Computer and Communications Societies. Anchorage, IEEE INFOCOM,2001(2):878—886.
- [3] Peng T, Leck C, Ramamohanrao K. Adjusted probabilistic packet marking [C]// Proceedings of Networking 2002. Pisa ;IEEE Computer Society, 2002:697—708.
- [4] Stoica I, Zhang H. Providing guaranteed services without per flow management[C]// Proceedings of the 1999 ACM SIGCOMM Conference. USA:ACM,1999,29(4): 81—94.
- [5] Durrresi A, Paruchuri V, Barolli L, et al. Efficient and secure autonomous system based traceback[J]. Journal of interconnection Networks,2004(5): 151—164.
- [6] Boneh A, Hofri M. The coupon—collector problem revisited[J]. Commun Statist Stochastic Models, 1997,13 (1):39—66.
- [7] Theilmann W, Rothermel K. Dynamic distance maps of the internet [C]// Proceedings of IEEE INFOCOM 2000. Piscataway, NJ, USA:IEEE, 2000:275—284.
- [8] 揭振,孙乐昌. 基于 Tabu 标记改进的 IP 追踪[J]. 微电子学与计算机,2007,24(2):85—88.

作者简介:

刘丽女,(1968—),博士,副教授,研究生导师。研究方向为网络计算、通信网络安全等。

黄芳芳女,(1980—),硕士。研究方向为通信网络安全。