

文章编号:1001-9081(2006)08-1831-02

## 一种基于节点采样的包标记追踪方案

揭 摄,孙乐昌

(解放军电子工程学院 网络工程系,安徽 合肥 230037)

(jieshe1996@163.com)

**摘 要:**概率包标记(PPM)是一种有效的 IP 追踪技术,但传统方案基于不现实的假设,存在很多不足,影响了实用性。基于合理的假设条件改进了高级标记方案(AMS),使用可调节的标记概率,根据 TTL 值计算距离和点采样等策略,改善了高级包标记方案存在的弱收敛性,不支持渐进部署,易受伪造标记攻击等缺陷,具有较好的实用性。

**关键词:**IP 追踪;节点采样;最弱链;渐进部署

**中图分类号:**TP393.08 **文献标识码:**A

## Packet marking scheme for IP traceback based on node sampling

JIE She, SUN Le-chang

(Department of Network Engineering of PLA, Electronic Engineering Institute, Hefei Anhui 230037, China)

**Abstract:** Probabilistic Packet Marking(PPM) is an effective technique for IP traceback. However, the traditional PPM schemes are built upon unreal hypothesis, which brings out many limitations and affects their application. Based on reasonable hypothesis, Advanced Marking Scheme (AMS) was improved. With adjustable marking probability, counting distance by TTL value and node sampling, the performance of AMS got improved including weak convergence, inability of incremental deployment and being vulnerable to spoof marking attack.

**Key words:** IP traceback; node sample; weakest link; incremental deployment

### 1 传统概率包标记方案存在的不足

概率包标记(Probabilistic Packet Marking, PPM)<sup>[1]</sup>作为一种有效的 DDoS 攻击追踪技术,具有无需 ISP 干预,避免高昂的管理开销,不产生过高的网络负载,支持事后追踪等优点,引起研究者的关注。但当前的 PPM 方案大多是基于高级标记方案(Advanced Marking Scheme, AMS)<sup>[2]</sup>改进的,往往存在以下不足:1)由于采用固定的标记概率且允许重复标记,因此引发不公平标记概率和最弱链<sup>[1]</sup>问题,进而导致追踪的弱收敛性;2)现有方案多假设路径上的所有路由器都支持标记追踪,即完全部署,这使得在非完全部署时,追踪性能急剧下降;3)由于是概率性地标记转发的数据包,则会出现一部分包在整个路径上都没有标记。而这可能被攻击者所利用,在数据包的标记域中预设虚假的标记信息来误导重构、干扰追踪,即伪造标记攻击。

假设完全部署在当前是不现实的,合理的假设应该是在一个自治域系统内完全部署。在现实中,一个优秀的追踪方案应该是在被逐渐认同后才为 ISP、设备厂商所支持,即具有渐进部署性。因此我们认为一个切实可行的 PPM 方案应该避免弱收敛性,具备渐进部署性,杜绝伪造标记等。

### 2 相关假设及改进思路

紧缩原有 PPM 方案的假设,使之更为合理,并以 AMS 为例进行分析,提出对 AMS 方案的改进思路。

在追踪过程中,假设:1)攻击者可以构造多种包,并会通过伪造源地址及标记域信息来躲避追踪;2)路由器是安全

的;3)受害者可以获取上游网络拓扑。我们认为以上假设是合理的,因为:1)现有很多工具可以构造各式的数据包,如 Iris;2)由于路由器功能单一,漏洞较少,由安全意识很强的专职人员管理,而且即使被攻陷,也是属于“珍贵”资源,可用于更隐蔽更具价值的目的,而不是容易暴露的 DDoS 攻击<sup>[3]</sup>;3)现在已有很多工具可以获取上游网络拓扑,如朗讯贝尔实验室基于 traceroute 的工具和 CAIDA 的 Skitter 工具,它们每天可以获取 10 万个与受害者相连的拓扑,而且短期内,这些拓扑不会显著变动。

以 AMS 方案为基础,提出改进思路:1)AMS 方案假设路径上的路由器均参与标记追踪,并以固定标记概率可重复地进行携带距离标识的边采样。为了解决 AMS 方案存在的弱收敛性问题,借鉴文献[4]提出的可调节标记概率的思想来改善最弱链问题。2)为了使改进方案支持渐进部署,就要解决改进方案在非完全部署下的可用性和健壮性问题。因为 AMS 方案假设路径上的每个路由器都具备标记功能,而且采用了边采样标记。在完全部署时,构成一条边的两个节点是相邻的,但在非完全部署,特别是少量部署时,一条边的两个节点可能是相隔多个保留路由器(不支持追踪的路由器)的标记路由器。因为仅有标记路由器可以修改距离信息,那么采用边采样的 AMS 方案中的距离信息仅对标记路由器计数,而并非真实距离,这将产生大量的误报,从而影响路径的重构。因此我们考虑采取节点采样来代替 AMS 方案中的边采样。3)为了杜绝伪造标记攻击,我们考虑让自治域边界路由器对接收包的标记域进行初始化,即使数据包原本携带了攻击者预设的伪造标记(标记和距离信息),也会在自治域边界

收稿日期:2006-02-16;修订日期:2006-04-11

作者简介:揭摄(1977-),男,安徽六安人,博士研究生,主要研究方向:信息与网络安全; 孙乐昌(1951-),男,教授,博士生导师,主要研究方向:信息与网络安全、分布式系统。

被初始化内容所覆盖。4) 为了获取真实的距离,我们放弃 AMS 等传统方案中 0/1 递增加的计量方法,而是采取生存时间(Time to Live, TTL)计数,因为 TTL 计数不受路由器是否支持标记的影响。

### 3 具体改进及算法描述

#### 3.1 改善 AMS 方案的重构弱收敛性

AMS 等传统方案的弱收敛性根源于采取固定的标记概率并且允许重复标记。文献[4]根据标记路由器在路径中位置来调节标记概率以解决最弱链问题。对于固定标记概率  $p$ , 标记路由器到攻击源距离为  $d$ , 则重构攻击路径所需标记包的数学期望为  $N(d) = \frac{\ln(d) + O(1)}{p(1-p)^{d-1}}$ 。根据 coupon collecting 问题, 可知当标记概率与  $d$  互为倒数时, 即  $p_d = 1/d$  时,  $N(d)$  值最小。但这是完全部署时的结果, 在非完全部署时, 特别是少量部署时, 如果第一个标记路由器的  $d$  较大, 超出一定阈值  $L$ , 为了增大该标记被接收的概率, 可考虑  $p_d$  乘上一个放大系数  $\beta$ , 其中  $\beta < d$  以确保  $p_d < 1$ , 这是为了防止第一个标记路由器的标记概率过小而引发类似的最弱链问题。 $L$  和  $\beta$  的具体取值取决于自治域的部署情况和网络管理员对可能进行的攻击所要求的追踪精度。

#### 3.2 标记域的初始化

由于要支持非完全部署, 为了鼓励 ISP 部署, 赋予改进方案“谁参与, 谁受益”的部署特性, 即在自治域边界处对流入包的标记域进行初始化, 这杜绝了伪造标记攻击。对于大自治域而言, 至少可以追踪到自治域边界, 如果其中还有小自治域, 基于安全和快速响应的角度考虑, 继续追踪仍然必要。与 AMS 方案类似, 将 16bit 标识域和 1bit 的 RF 位重载为标记域, 划分为 6bit 的距离字段和 11bit 的标记字段。 $d$  是变量, 而标记路由器是如何获知  $d$  的值呢? 在自治域边界处将包的  $TTL_{LSB6}$  ( $TTL$  值的 6 个最低有效位) 复制到距离字段, 并使用全局 hash 函数  $h(\cdot)$  计算边界路由器 IP 的 11bit 摘要写入标记域, 由于  $TTL$  值每经一个路由器就递减 1, 因此根据距离域值与当前  $TTL_{LSB6}$  的差值可知距离  $d$ 。使用 6bit 的距离字段是为了确保至少能追踪 32 跳, 因为攻击者能预设  $TTL$  初始值, 可能出现  $00100000_2 - 1 = 00011111$  的情况, 如果用 5bit 表示距离字段, 得到  $d = 11111_2 = 31$ , 而这是错误的。

#### 3.3 标记过程

标记路由器  $R$  根据  $d$  计算标记概率  $p_d$ , 并以此来决定是否标记包, 但是对于任何包, 如果它遇到的第一个标记路由器的  $d$  值超过某个阈值  $L$ , 为了防止标记概率  $p_d$  过小, 乘上一个放大系数  $\beta$ 。如果决定标记, 则计算  $h(R)$  并写入标记字段, 同时将当前的  $TTL_{LSB6}$  写入距离字段, 否则不对标记域进行操作。标记算法描述如下:

在 AS 边界路由器处:  $m = h(IP_{AS})$

for each incoming packet  $p$

$p.distance \leftarrow TTL_{LSB6}$

$p.marking \leftarrow m$

在每个标记路由器  $R$  处:  $m' = h(IP_R)$

for each a incoming packet  $p$

$d = p.distance - TTL_{LSB6}$

$p_d = 1/d$

if  $(p.marking = m) \& (d > L)$  then  $p_d = p_d \cdot \beta$

let  $x$  be a random number from  $[0..1)$

if  $x \leq p_d$  then

$p.distance \leftarrow TTL_{LSB6}$

$p.marking \leftarrow m'$

#### 3.4 路径重构

在路径重构过程, 受害者通过工具获得其上游拓扑图  $G'$ ,  $G'$  是一个以受害者  $t$  为根的树形结构。我们的重构目标是得到由攻击路径构成的攻击图  $G$ , 且  $G \subset G'$ 。在受害者处, 根据包的距离字段值与  $TTL_{LSB6}$  的差值表示标记路由器到受害者的距离  $d$ , 得到一个有  $l$  个不同  $d$  值的集合  $\alpha$ , 将得到的标记根据相同  $d_i (1 \leq i \leq l)$  划分为不同的集合  $D_i$ 。设  $G'$  中第  $d_i$  层节点的集合为  $B_i'$ 。计算  $B_i'$  中每个地址的 hash 值得到集合  $D_i'$ , 将  $D_i$  中的每个元素与对应的  $D_i'$  中的元素进行比较, 查找匹配的值, 从而得到对应的 IP 集合  $B_i$ 。因为  $D_i'$  中的元素可能远多于  $D_i$ , 所以将  $D_i$  中的元素与  $D_i'$  的进行匹配。重构算法描述如下:

for each received packet  $p$

$d = p.distance - TTL_{LSB6}$

将  $d$  加入集合  $a$

$\forall d_i \in a$

$B_0 = t$

$\forall m \in B_i'$

将  $h(m)$  加入集合  $D_i'$

$\forall y \in D_i$

if  $y \in D_i'$  then

将  $y$  对应的 IP 加入到集合  $B_i$

输出  $B_i$

对应  $G'$  进行剪枝提纯得到攻击图  $G$

### 4 改进方案的性能分析

#### 1) 防止伪造标记攻击

由于改进方案是基于 AMS 方案的, 我们在自治域边界处采取了确定性地标记(初始化), 从而杜绝了伪造标记和距离值的攻击, 而且赋予了改进方案“谁参与, 谁受益”的部署特性, 确保了至少可以追踪到自治域边界。

#### 2) 渐进部署性

由于采取基于  $TTL$  值每经过转发路由器(不管是标记路由器还是保留路由器)就递减 1 的属性, 可以准确地进行距离标识, 解决了 AMS 等 PPM 方案在非完全部署下的错误距离计数问题, 从而支持了渐进部署, 使得改进方案更向实际应用迈进了一步。

#### 3) 计算开销

改进方案比 AMS 方案多使用了 1bit 的 RF 位, RF 位是未用的保留标志位, 使用它不会影响网络应用。写入标记和距离信息可以在路由器更新  $TTL$  值和 IP 首部校验和时完成。改进的标记算法在不标记包时无需递增距离值和进行地址信息异或和写入操作, 节省了计算开销。因为更新  $TTL$  值和首部校验和是协议必须的操作, 所以标记算法不会显著的增加路由器上的开销, 易为厂商和 ISP 所接受。由于我们采用的是节点标记, 重构的得到的也是一个标记路由器的 IP 地址的集合, 然后对照上游拓扑图对它进行剪枝提纯, 重构的计算复杂度相当于重复地对树形结构进行宽度优先搜索, 并不显著地高于 AMS 方案。

#### 4) 收敛时间和误报率

由于改进方案正是通过解决最弱链问题来改善 AMS 等 PPM 方案的弱收敛性。而且在非完全部署时, 考虑第一个标记路由器都距离攻击源较远的情况, 可以在根据距离计算所

(下转第 1841 页)

的残差图像(图 1 和图 2 中的(g))。表 1 给出图 1 和图 2 中的有损采样重构图像的压缩比率(Bit Rate)、均方误差(MSE)、峰值信噪比(PSNR)及图像逼真度(NMSE)数据的比较结果。从实验结果看, $r = 0$  时的各种指标都较  $r = 1$  时要

好,但是从重构图像的整体信息的保留情况看, $r = 1$  时的重构图像要较  $r = 0$  时更加完善,且各指标的差距并不是太大。因此,可以说广义形态变换作为一种图像的有损采样方法是具有一定研究价值的。

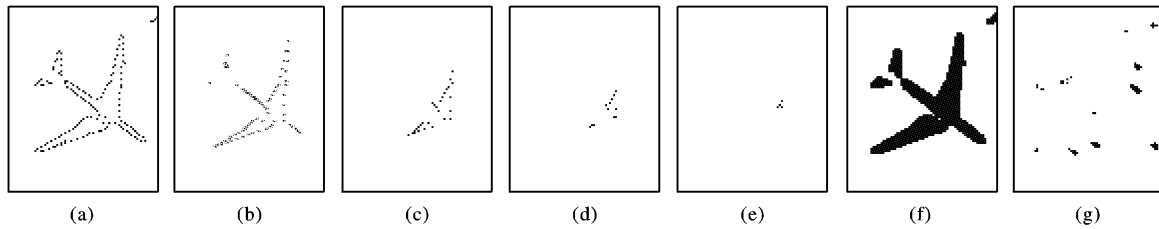


图 1  $r = 0$  时的形态采样序列((a)~(e))及压缩重建图像(f)和残差图像(g)

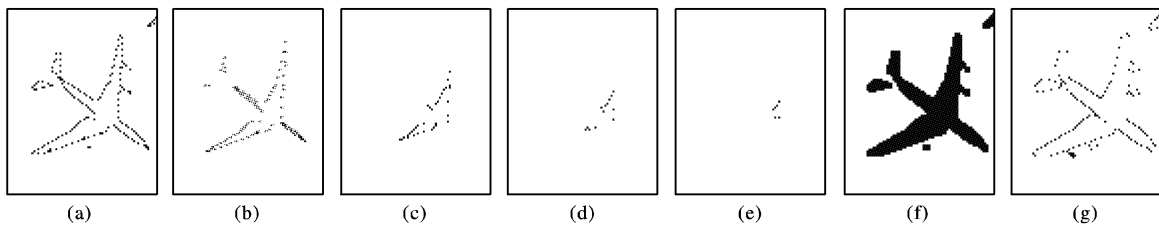


图 2  $r = 1$  时的形态采样序列((a)~(e))及重建图像(f)和残差图像(g)

表 1 基于形态采样方法的相关数据对照

	有损采样	有损采样
Bit Rate	0.2283	0.2805
MSE	0.0107	0.0200
PSNR	25.7334	23.0033
NMSE	0.9326	0.8736

#### 4 结语

广义形态变换理论的建立对于发展和扩展形态分析的理论和应用方法具有重要的意义。文中通过对形态算子理论的推广,研究了更具一般性的形态分析方法及其变换性质,并以此为基础提出了基于广义形态变换的形态采样方法,其理论及实验结果表明该采样方法能保留图像的重要信息,实现由采样图像有损重构形态滤波图像,采样图像本身则是实现对原始图像有损重构的最小图像集合。

#### 参考文献:

- [1] HARAICKRM, ZHUANG XH, LIN C, *et al.* The Digital Morphological Sampling Theorem [J]. IEEE Transactions on Acoustics Speech

and Signal Processing, 1989, 37 (12): 2067 – 2090.

- [2] HEIJMANS HJAM. Morphological Image Operators [M]. Boston, Massachusetts: Academic Press, 1994. 71 – 117.
- [3] WANG DM, LABIT C. A Lossless Morphological Sampling Scheme For Segmented Image Compression [A]. IEEE Transactions, Proceedings in International Conference on Image Processing [C]. 1995, vol 1. 23 – 26.
- [4] AGAM G, DINSTEIN I. Adaptive Directional Morphology with Application to Document Analysis [A]. Mathematical Morphology and its Applications to Image and Signal Processing [C]. Kluwer Academic Publishers, 1996. 401 – 408.
- [5] KRESCH R, MALAH D. New Morphological Skeleton Properties Leading to Its Efficient Coding [A]. IEEE Workshop on Non – Linear Signal and Image Processing [C]. Neos Marmaras, Halkidiki, Greece, 1995. 995 – 999.
- [6] MARAGOS PA, SCHAFER RW. Morphological Skeleton Representation and Coding of Binary Image [A]. IEEE Transactions on Acoustics Speech and Signal Processing [C]. 1986, vol 34. 1128 – 1244.
- [7] SALEMBIER P. Morphological Multiscale Segmentation for Image Coding [J]. Signal Processing, 1994, 38(3): 359 – 386.

(上接第 1832 页)

得的标记概率上乘上一个放大系数以增大该路由器及其后续路由器地址信息被受害者接受的概率。由于我们采取了在自治域边界初始化的策略,杜绝了伪造标记和距离信息的干扰,可以显著地降低误报率。当然,上游拓扑图的精确度也是导致误报的重要因素,但是 AMS 也面临相同的问题,因此在自治域中要保持路由的相对稳定或者及时更新上游拓扑图可以降低误报率。

#### 5 结语

现有的大多数 PPM 攻击追踪方案往往基于过于理想的假设条件,削弱了实用性。针对经典的 AMS 方案进行分析,紧缩追踪的假设条件,着重针对其存在的弱收敛性、不支持非完全隶属、易受伪造标记和距离信息干扰等问题进行改进。改进方案采用非固定的标记概率改善了 AMS 重构的弱收敛

性,而且具有很好的抗伪造标记攻击和支持渐进部署的能力,具有较好的实用性。

#### 参考文献:

- [1] SAVAGE S, WETHERALL D, KARLIN A, *et al.* Practical Network Support for IP Traceback [A]. Proceeding of ACM SIGCOMM [C]. Stockholm, Sweden, 2000. 295 – 306.
- [2] SONG DX, PERRIG A. Advanced and Authenticated Marking Schemes for IP Traceback [A]. Proceeding of IEEE Computer and Communications Societies [C]. Stockholm, Sweden, 2001. 878 – 886.
- [3] 李德全, 徐一丁, 苏璞睿, 等. IP 追踪中的自适应包标记 [J]. 电子学报, 2004, 32(8): 1334 – 1337.
- [4] PENG T, LECKIE C, KOTAGIRI R. Adjusted Probabilistic Packet Marking for IP Traceback [A]. Proceeding of IFIP Networking 2002 [C]. Pisa, Italy, 2002. 697 – 708.