

文章编号: 1005—8893 (2006) 02—0022—04

# 基于概率包标记算法的安全技术研究<sup>①</sup>

倪彤光, 杨长春, 顾晓清

(江苏工业学院 计算机科学与工程系, 江苏 常州 213164)

**摘要:** 防御分布式拒绝服务攻击 (DDoS) 是当前网络安全中最难解决的问题之一, 在提出的多种对策中, 通过概率包标记算法来进行 IP 跟踪受到广泛重视。这里分析了当攻击者伪造数据包的标记信息时, 概率包标记法及其变种算法的安全性问题, 并提出了一些改进措施, 通过实验证明提高了算法的安全性。

**关键词:** 网络安全; DDoS; 概率包标记; IP 跟踪

**中图分类号:** TP 309

**文献标识码:** A

## Research on the Effectiveness of Probabilistic Packet Marking

NI Tong—guang, YANG Chang—chun, GU Xiao—qing

(Department of Computer Science and Technology, Jiangsu Polytechnic University, Changzhou 213164, China)

**Abstract:** Defending against distributed denial of service attack is one of the hardest security problems on the Internet today. Among several countermeasures, probabilistic packet marking (PPM) used for IP traceback is promising. In this paper, an analysis of the security of some probabilistic packet marking schemes is provided when the marking information is spoofed by the attacker. Some modifications are provided which can improve the security of the scheme.

**Key words:** network security; DDoS; PPM; IP traceback

DDoS 已经成为网络中的主要攻击形式, 这类攻击实施简单, 破坏性强而且攻击者的 IP 地址常常是经过伪装的。针对此类攻击的特点, 研究者提出了很多解决方法<sup>[1, 2]</sup>, 但解决的根本途径是跟踪攻击流, 发现攻击路径, 找出攻击源, 在攻击源阻止进攻, 这就是 IP traceback 方法。IP traceback 方法中有很多策略<sup>[3]</sup>, 其中由 Savage<sup>[4]</sup> 等人提出的概率包标记策略是最有效的方法之一, 之后在它的基础上不断产生出新的改进算法, 如自适应概率包标记<sup>[5, 6]</sup>, 高级包标记和带认证的包标记<sup>[7]</sup>, 以期减少路径重构时所需的数据包, 进一步地压缩标记

信息和认证标记信息。本文首先分析了当存在攻击者伪造的数据包标记信息时, 概率包标记和自适应包标记算法的安全性问题, 并针对其中的安全性漏洞给出了一些改进方案, 通过实验证明该方案有效提高了包标记算法的安全性。

## 1 概率包标记算法

### 1.1 概率包标记原理及自适应概率包标记

概率包标记最早由 Savage<sup>[4]</sup> 等人提出, 它的思想是路由器以固定的概率  $p$  标记数据包, 将路径

① 收稿日期: 2005—12—09

作者简介: 倪彤光 (1978—), 男, 河北邢台人, 硕士, 主要研究方向: 网络安全、虚拟现实。

信息加入到 IP 头中的 16bit 识别域 (ID) 中, 表示为 (start, end, distance)。路由器将自己的 IP 地址填入到 start, 同时 distance 赋值为 0; 否则如果 distance 为已经被标记为 0, 表明上一个路由器对包进行了标记, 则将自己的 IP 地址填入到 end 中; 如果路由器以概率  $1-p$  不对包标记, 则将 distance 值加 1。受害者从这些数据包中提出路径信息, 重构出攻击路径。它最主要的缺点是随着路径的增长, 所需要数据包的数量成指数增长, 这就不可能很快重构路径防御攻击。T. Peng<sup>[5]</sup>、B. Rizvi<sup>[6]</sup> 等人在此基础上提出了自适应算法。该算法中路由器标记数据包的概率是  $p = \frac{1}{c+1-d_v}$ ,  $d_v$  是所在路由器到受害者之间的距离,  $c$  是个常数。当  $d_v$  小于 5 时取  $c=5$ ; 当  $d_v$  大于 5 且小于 10 时取  $c=10$ ; 以此类推直至  $c=30$ , 因为大多数路径的长度不会超过 30。用这样方法受害者只需要更少的数据包就能完成路径重构工作, 从而能在更短时间内找出攻击路径并发现攻击源。

1.2 高级包标记和带认证的包标记

为了有足够的空间能在 IP 头中的 ID 域中写入标记的路径信息, 概率包标记算法将路由器的 IP 地址及 32 bit 的校验码共 64 bit 分成 8 块, 以 0 到 7 对应其偏移量, 路由器标记数据包时随机地从 8 块中选取一块连同对应的偏移量存入 ID 域中。在重构路径时必须将这些分块的边信息重新组合, 这是一项十分费时的的工作, 且存在大量误报的现象。为解决这一问题, Song<sup>[7]</sup> 等人提出了高级包标记和带认证的包标记算法。16 bit 的 ID 域表示为三元组 (flag ID, distance, edge), 当路由器标记数据包时随机选择 0~7 之间的整数将它写入 3 bit 的 flag ID 中, 并根据该值选取 hash 函数, 利用该函数将路由器的 IP 地址压缩为 8 bit 写入 edge 字段。当下游路由器不标记数据包时如 distance 值为 0, 根据这种方法将自己的 IP 地址压缩到 8 位后, 与 edge 字段的值进行异或操作并写入 edge 字段。带认证的包标记算法是在高级包标记的基础上使用单向函数和消息认证码 (MAC) 对标记域中的 edge 字段进行认证, 这里就不再详述。通过压缩标记信息和认证标记信息在很大程度上改善了重构路径中误报情况并提高了算法的安全性。

2 概率包标记算法的安全性分析

在图 1 中, 真正的攻击者是 s, s 传送一个伪

造数据包, 其中设置距离域是 0, 边信息是路由器  $u_1$  地址的哈希函数值。当这个数据包到达网络中路由器  $v_1$  的时候, 检查到该包的距离域为 0, 依据高级包标记算法,  $v_1$  将自己地址的 hash 函数值与  $u_1$  地址的 hash 函数做 XOR 运算并写入该数据包的边信息中。如果这个伪造的数据包被下游的路由器标记, 标记信息将被覆盖, 对重构路径不会造成影响, 如果没有被下游路由器标记, 数据包距离域的值就会执行加 1 的操作, 当受害者进行 IP 回溯时, 就认为攻击者是  $u_1$ 。用这种方法攻击者可以伪造多个攻击路径:  $(u_1, v_1 \cdots t) \cdots (u_m, v_1 \cdots t)$ , 所以, 攻击者通过伪造标记信息影响了攻击路径的重构, 当伪造标记信息的攻击者离受害者越近时, 对重构路径的影响就越大。假设攻击者沿着同一条路径进行攻击并设数据包的数量是  $N$ , 路由器以固定概率  $p$  标记数据包, 那么在受害端接收到的一次都没有被标记过的数据包数是  $n_0 = N(1-p)^d$ 。一般取概率  $p=0.04$ , 当攻击路径长度  $d=25$ , 攻击流量  $N=10\,000$  的时候, 有 8 153 个数据包从没有标记过。当路由器以自适应概率  $p = 1/(c+1-d_v)$  标记数据包时,  $n_0 = N(1-p_d) \cdots (1-p_1) = N(1-d/c)$ 。当取  $c=30, d=25$  时, 有 1 666 个的数据包从没有标记过。可见无论使用那种标记概率, 没有标记过的数据包在所有数据包中都占了相当大的比例, 这些数据包极有可能被攻击者利用来伪造虚假的攻击路径。

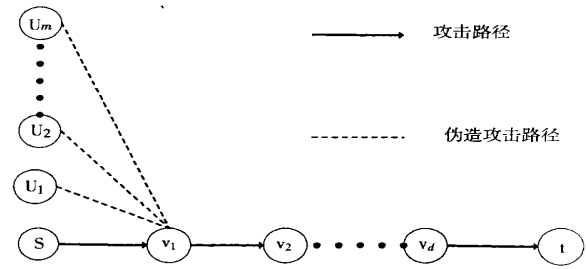


图 1 攻击路径是  $(s, v_1 \cdots v_d, t)$ , 伪造攻击路径是  $(u_i, v_1 \cdots v_d, t), i=1, \cdots, m$

Fig 1 Attack path  $(s, v_1 \cdots v_d, t)$  and a set of  $m$  forgeable paths  $(u_i, v_1 \cdots v_d, t), i=1, \cdots, m$

对于攻击者来说, 希望伪造的攻击路径数  $m$  越多越好, 而受害者则恰好相反, 希望  $m$  值越小越好。在 DDoS 攻击发生和路径的重构实现中, 攻击者可以控制攻击流量  $N$ , 攻击路径长度  $d$  和标记域中的伪造变量  $x_0$ , 而受害者只能控制标记概率  $p$  的选择, 这里本文使用文献 [8] 中的数学模型来讨论  $m$  的取值。假设攻击者  $s$  在  $m$  条伪造路

径上伪造了相同数量的数据包, 这样①在路径  $(u_i, v_1)$  上, 伪造的数据包插入到真实的数据包中根本无法分辨, 受害者无法从数据流量上判断这  $m$  条路径的真伪。②在  $v_1$  处接受到的伪造数据包是相同的, 那么它们被标记的概率也是相同的。定义  $\alpha_i(p)$  是数据包最后在路由器  $v_i$  被标记而没有被下游路由器标记的概率, 可得  $\alpha_1(p) = \alpha_1^s(p) = \alpha_2^s(p) = \dots = \alpha_m^s(p)$ 。这样得到了一个目标函数和两个约束条件:

$$\min_p \max_{x_0} m(p, x_0) \quad (1)$$

$$\alpha_1(p) = \alpha_1^s(p) = \alpha_2^s(p) = \dots = \alpha_m^s(p) \quad (2)$$

$$N\alpha_1(p) = Np(1-p)^{d-1} \quad (3)$$

目标函数 (1) 中的  $\max$  是考虑到攻击者的因素,  $\min$  是考虑到受害者的因素, 约束条件 (2) 是指这  $m$  条伪造路径信息仅在  $v_1$  处被标记的概率是相同的, 约束条件 (3) 是指在  $v_1$  处标记的数据包数必须大于等于 1, 将攻击者可控量  $N$  引入模型。从约束条件 (2) 可得

$$m\alpha_1(p) = \alpha_0(p) \Rightarrow m = 1/p - 1 \quad (4)$$

其中  $\alpha_0$  是数据包从没有被标记过的概率。当选用固定概率的时, 通过目标函数和约束条件的运算得到

$$\frac{1}{N} \leq p \leq 1 - \left(\frac{1}{N}\right)^{1/(d-1)}, \quad m \approx \frac{N^{-1/(d-1)}}{1 - N^{-1/(d-1)}} \quad (5)$$

当选用自适应概率

$$p = \frac{1}{(c+1-d_v)}, \quad m = c - d_v \quad (6)$$

从图 2 中可以看到, 在包标记算法中,  $m$  的取值是与标记概率  $p$  成反比的, 取固定概率  $p$  时,  $m$  值的范围很广  $1 < m < 10^5$ ; 而取自适应概率  $p = 1/(c+1-d_v)$  时,  $m$  的最大值是 5, 最小值是 1。从上述分析中可以看到攻击者利用从未标记过的数据包伪造路径信息, 使受害者不能重构正确的攻击路径, 而且还应考虑到另一种情况是: 如果攻击者控制了攻击路径上的一个路由器, 利用路由器修改上游路由器标记过的数据包, 那么受害端使用修改过的数据包信息也不能准确重构攻击路径。所以下面本文从提出了两个改进方案来解决这些问题。

### 3 包标记算法的改进方案

本文使用文献 [9] 中提出了一种计算各个攻

击路径的权重的方法, 把权重信息加入到候选攻击路径中, 通过和正常情况下的候选攻击路径权重进行比较, 可更好地分析真正的攻击源, 起到降低误报率的效果。

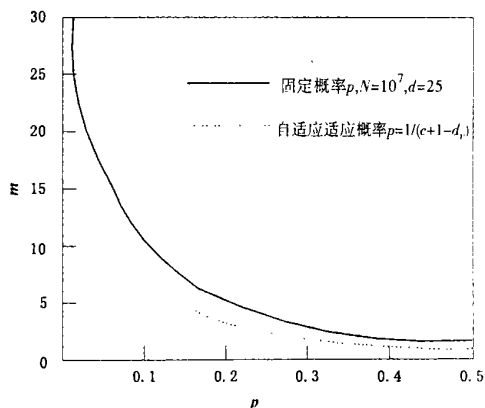


图 2 在固定概率与自适应概率标记时伪造路径数  $m$  的取值

Fig 2 The forgeable paths  $m$  as fixed packet marking and adjusted probabilistic packet marking

设一候选攻击路径  $l = \{R_1 \rightarrow R_2 \rightarrow \dots \rightarrow R_d \rightarrow V\}$ , 路由器  $R_i$  以概率  $p$  标记的数据包有个转发到  $R_{i+1}$ , 则称边  $R_i \rightarrow R_{i+1}$  的权重为  $x_{i \rightarrow i+1}$  ( $R_{d+1} = V$ ), 记为  $W_{R_i \rightarrow R_{i+1}} = x_{i \rightarrow i+1}$ 。候选攻击路径  $l$  的权重记为:

$$W_l = \frac{\sum_{i=1}^d x_{i \rightarrow i+1}}{d} \quad (7)$$

利用每一个路径的权重值  $W_l$  与正常情况下的权重  $W'_l$  相比较, 如果  $\frac{W_l - W'_l}{W'_l} < \sigma$  (设定的一个限值), 则认为路径不是攻击路径。

为了不让攻击者利用那些没被标记过的数据包, 最简单的方法是在边界路由器上标记所有的数据包, 并设定所有数据包的 distance 值为 0。这样攻击者伪造的信息全被覆盖, 但边界路由器的处理负担比较大。第 2 种方法是在边界路由器上把 edge (8 bit) 字段的值全设为 1, 在标记包信息时对即将存入 edge 的值做 253 取模的运算, 因为 8 个 bit 全为 1 时值为 255, 用 253 取模就可以避免在 edge 字段出现全为 1 的情况, 而且 253 是个很大的质数它的取模结果分布性也较好。受害者检查接收到的数据包时, 如果 edge 字段全为 1 的话就说明该包从没有标记过, 不需做任何处理。

### 4 模拟实验

本文进行了一系列的实验来验证当存在路由器

被攻破的情况下改进方案的可行性。为检验该方案在真实环境中的有效性, 实验数据利用由 CAIDA<sup>[19]</sup> (Cooperative association for internet data analysis) 提供的跟踪路由数据库进行仿真攻击实验, 由跟踪路由数据库构建跟踪拓扑树。

图3中选取的攻击路径长度从1到30, 路由器标记数据包的概率是选用的自适应概率  $p = \frac{1}{c+1-d_v}$ , 进行1000次实验取平均值。实验表明, 当存在路由器被攻破的情况下, 通过使用文中提出的改进方案, 重构路径所需的数据包的量明显小于不使用改进方案时所需的数据包量, 所需数据包数平均减少了50%。

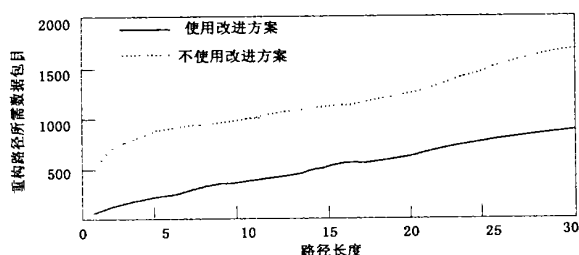


图3 重构路径所需数据包的比较

Fig 3 Comparison of the number of packets needed for reconstruction paths

## 5 结 论

利用概率包标记技术可以有效地跟踪 DDOS 攻击者, 即时对攻击者进行 IP 回溯找出攻击路径并发现攻击源。本文主要研究了概率包标记的安全性问题, 受害者在进行 IP 回溯时, 无法辨别数据包是否被攻击者伪造, 而且路由器也有可能被攻破, 攻击者利用此漏洞可以大量伪造标记数据包来隐藏自己的真正位置。针对这一问题, 本文提出了概率包标记的改进方案, 通过实验证明该改进方案可以有效提高算法的安全性。

## 参考文献:

- [1] Chen L, Longstaff T, Carley K. Characterization of Defense Mechanisms Against Distributed Denial of Services Attacks [J]. Computers & Security, 2004, 23: 665—678.
- [2] Douligenis C, Mitrokotsa A. DDoS Attacks and Defense Mechanism: Classification and State-of-the Art [J]. Computer Networks, 2004, 44: 643—666.
- [3] Hsu F, Chiueh T. A Path Information Caching and Aggregation Approach to Traffic Source Identification [A]. In Proc of the IEEE International Conference on Distributed System (ICDCS 2003) [C]. Rhode Island: IEEE Computer Society, 2003. 332—339.
- [4] Savage S, Wetherall D, Karlin A, et al. Network Support for IP Traceback [A]. In Proc of the IEEE/ACM Transactions on Networking [C]. Stockholm: IEEE Communications Society, 2001, 9 (3): 226—237.
- [5] Peng T, Leckie C, Ramamohanrao K. Adjusted Probabilistic Packet Marking [A]. In Proc of Networking 2002 [C]. Pisa: IEEE Computer Society, 2002. 697—708.
- [6] Rizvi B, Gaucherand E. Analysis of Adjusted Probabilistic Packet Marking [A]. In Proc of IP Operation and Management (IPOM 2003) [C]. Kansascity: IEEE Communication Society, 2003. 9—13.
- [7] Dawn X Song, Perrig A. Advanced and Authenticated Marking Schemes for IP Traceback [A]. In Proc of the IEEE INFOCOM 2001 [C]. Anchorage: Institute of Electrical & Electronics Engineering, 2001. 878—886.
- [8] Park K, Lee H. On the Effectiveness of Probabilistic Packet Marking for IP Traceback under Denial of Service Attack [A]. In Proc of the IEEE INFOCOM 2001 [C]. Anchorage: Institute of Electrical & Electronics Engineering, 2001. 338—347.
- [9] 徐永红, 杨云, 刘凤玉, 等. 基于权重包标记策略的 IP 跟踪技术研究 [J]. 计算机学报, 2003, 11: 1597—1603.
- [10] David Moore. CAIDA Cooperative Association for Internet Data Analysis [EB/OL]. <http://www.caida.org>. 2004—08—22.