基于有序标记的 IP包追踪方案

曲海鹏[1,2], 冯登国[1], 苏璞睿[1,2]
(1. 中国科学院软件研究所信息安全国家重点实验室, 北京 100080; 2 中国海洋大学计算机系, 山东青岛 266071)

摘 要: 包标记方案是一种针对 DoS攻击提出的数据包追踪方案, 由于其具有响应时间快、占用资源少的特点, 近年来受到了研究者的广泛关注. 但由于包标记方案标记过程的随机性, 使得受害者进行路径重构时所需收到的数据包数目大大超过了进行重构所必需收到的最小数据包数目, 从而导致重构误报率的提高和响应时间的增长. 本文提出了一种基于有序标记的 IP包追踪方案, 该方案通过存储每个目标 IP地址的标记状态, 对包标记的分片进行有序发送, 使得在 DoS发生时, 受害者重构路径所需收到的标记包的数目大大降低, 从而提高了对 DoS攻击的响应时间和追踪准确度. 该算法的提出进一步提高了包标记方案在实际应用中的可行性.

关键词: 网络追踪; 拒绝服务; DoS; 分布式拒绝服务; DDoS; 包标记
中图分类号: TP309 文献标识码: A 文章编号: 0372-2112 (2006) 01-0173-04

An IP Traceback Scheme Based on Marking-in-Order

QU Hai-peng[1,2], FENG Deng-guo[1], SU Pu-rui[1,2]
(1. State Key Laboratory of Information Security, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China;
2. Department of Computer, Ocean University of China, Qingdao, Shandong 266071, China)

Abstract Packet Marking Scheme have been proposed for achieving traceback of DoS attacks, which has several advantages such as short responding time and small resource consuming. Since the randomicity of the packet marking procedure, the number of the packets needed by the victim in reconstructing is much more than the minimum number of packets required, therefore resulting in the increase of the rate of false positives and the responding time. In this paper, a marking-in-order scheme is proposed, which need less packets in reconstructing procedure because of marking the packets in order by storing every target IP's marking state in a hash table. The scheme is more practical for its high efficiency.

Key words IP Traceback; denial of service; DoS; distributed denial of service; DDoS; Packet marking traceback

1 引言

DoS(拒绝服务)攻击具有易于实施, 难以防范和追踪的特点, 一直是 Internet安全的一个严重的威胁. DoS防御中, 对攻击源的追踪既可以为防御机制提供有价值的信息, 又可以作为事后追究攻击者责任的法律证据. 但是, 由于 DoS攻击大多结合了 IP地址欺骗技术进行攻击, 使得对 DoS攻击的追踪成为 DoS防御中一个难以解决的问题. 为了解决这一问题, 研究人员提出了多种追踪方案, 如包标记[1~4]、日志记录[5]、连接测试[6]、ICMP追踪、覆盖网络[7]等, 这些方案都存在各自的优缺点. 其中, Savage等人提出的概率包标记方案[2], 通过路由器对 IP包的标记, 使受害者可以分析一定数目的攻击包中包含的标记信息, 来完成对攻击路径的部分重构, 并达到对风暴型拒绝服务攻击的路由进行追踪的目的. 这一方案以其实现技术简单、不增加网络负载、效率高等优点, 近年来得到了广泛研究和探讨. 研究者们在此基础上, 提出了多种改进包标记方案. Song等人提出了一种高级包标记方案[8], 在数据包中存储的不再是完整的节点 IP地址, 而是节点 IP地址的 Hash值, 以牺牲存储空间为代价, 获取较高的重构准确度.

在概率包标记方案和高级包标记方案中, 由于受到 IP包存储标记的空间限制, 路由器地址信息都被分成 8个分片, 每个 IP包只能存储其中一个分片. 当路由器决定对通过该路由器的数据包进行标记时, 就会随机选择 8个分片中的一个, 将其保存到该数据包的标记字段中. 这种对标记的随机选择方式同样被用于其他包标记方案中. 由于使用了随机标记的方式, 在标记过程中, 对于若干目标 IP地址相同的数据包, 每次选取的分片与以前选取的分片发生重复的概率不断增大, 而如果发生重复, 则重复的标记分片不能给受害者提供任何有益的信息, 而只会增加受害者重构过程的计算量和受害者重构所需收到的标记包个数, 从而增长了受害者对攻击进行响应的时间, 同时降低了响应的有效性. 图 1是对随机标记过程中受害者收到不重复数据包的仿真结果, 可以看出, 随着收到数据包数目的增

收稿日期: 2005-02-04; 修回日期: 2005-09-05
基金项目: 国家自然科学基金 (No 60273027)、国家杰出青年基金 (No 60025205); 国家 973重点基础研究发展规划项目 (No G1999035802)

长,其中包含的不重复包的个数的增加速度逐渐变慢.

在拒绝服务攻击的防御中,对攻击源追踪的快速性和有效性直接决定了防御措施能否有效实施.而减少重构时所收到数据包数目,不仅可以有效增加追踪的能力,而且对于及时发现攻击源,缩短对攻击采取防御措施的时间,也都非常重要.如果能够使发往相同目标 IP的数据包中的标记值有序排列,就可以大大提高攻击追踪的响应速度,但是由于路由器存储空间的局限性,保存已经转发包的地址信息又是很困难的.所以,研究如何对相同目标 IP地址的包进行有序标记,是非常重要的.

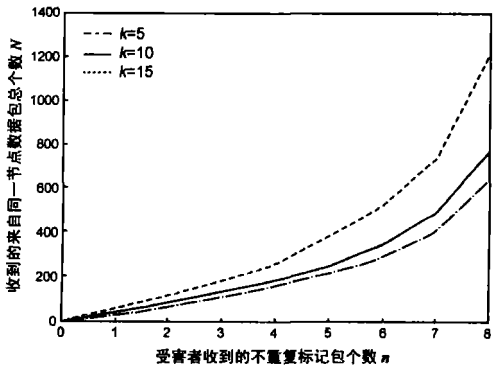


图 1 随着受害者收到来自同一节点数据包的增长,受害者收到的不重复标记包数目并没有呈线性增加,而是增加速度逐渐变慢(k为路由器距离受害者跳数)

在本文中,我们提出了一种有序标记方案,通过在每个标记节点对每个目标 IP的标记状态进行存储,可以有效减少受害者重构攻击路径时所需收到的标记包平均数目.这种方案可实施于目前提出的大多数数据包标记方案,比如 Savage提出的概率包标记方案、Song提出的高级包标记方案等,在实施该标记方案后受害者重构攻击路径时所需收到的标记包数目都有显著降低.本文其他部分安排如下:第 2节对有序标记追踪方案进行介绍;第 3节是对该方案的仿真实验结果;第 4节讨论了几个主要因素对该方案实现结果的影响;第 5节总结全文.

2 有序标记追踪方案

我们提出的有序标记追踪方案由状态转移算法、标记算法和重构算法三部分组成.

2.1 状态转移算法

在 Internet上,共有 2^{32}种可能的 IP地址.对应每个目标 IP地址,共存在 8个需要存储的标记状态,因而需要分配 $\log_2 8 = 3$比特的存储空间.但是,在每个路由器节点存储一个所有 IP地址的标记状态列表在存储空间上是不可行的.在我们提出的方案中,路由器使用一个长度为 2^n的 Hash表 T来存储所有目标 IP地址的标记状态,以减少存储状态标记需要的存储空间.其中,n可以根据路由器的可用存储空间大小和路由器的数据包流量等参数进行设定.数据包的目标 IP地址通过 3个相互独立、输入长度为 32、输出长度为 n的 Hash函数对应到 Hash表 T上的相互独立的 3比特存储空间.当路由器决定对一个数据包进行标记时,采取算法 a设置该目标 IP的标记状态:

算法 a 状态转移算法

令 $T(x)$表示在 Hash表 T 中第 x 位的值,A 为所要标记的数据包的目标 IP地址.

1. 计算$\begin{Bmatrix} H_1(A) \\ H_2(A) \\ H_3(A) \end{Bmatrix}$,令 $\alpha = (\alpha_1, \alpha_2, \alpha_3) = (T(H_1(p)), T(H_2(p)), T(H_3(p)))$, $\beta = \sum_{i=1}^{3} 2^{i-1}\alpha_i$;

2. 若 $\beta = 7$,令 $\alpha_i = 0$, $i = 1, 2, 3$;否则,改变 α_i 的值,令 $\beta + 1 = \sum_{i=1}^{3} 2^{i-1}\alpha_i$, $\alpha_i \in \{0, 1\}$, $i = 1, 2, 3$成立;

3. 在标记算法 b中,选择分片偏移为 β的标记.标记状态的转移过程如图 2所示.

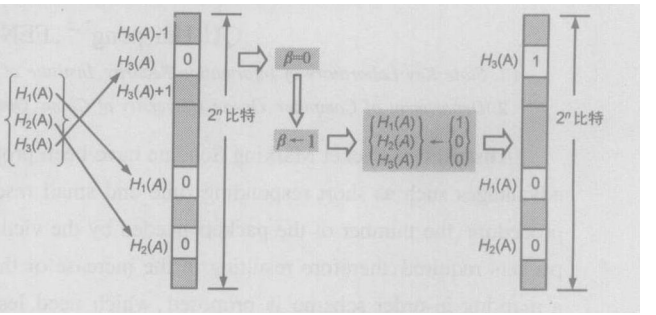


图 2 状态转移过程

2.2 标记算法

每个路由器对经过该路由器的标记算法如下:

算法 b 路由器 R 的标记过程

令 E为所要标记到数据包中的边信息

令 k表示 E 中分片的数量对每一个数据包 w

在 $[0, 1)$中取随机数 x

if $x < p$ then

调用算法 a,获得 β的值

设 f为 E 中偏移为 β的分片

将 f写入 $w.frag$

将 0写入 $w.distance$

将 β写入 $w.offset$

else

if $w.distance = 0$ then

求 α、β的值

if $\beta = k$ then

$\beta \leftarrow 1$

else

$\beta \leftarrow \beta + 1$

改变 α各元素值,使等式 $\beta = \sum_{i=1}^{3} 2^{i-1}\alpha_i$ 成立

设 f为 E 中偏移为 β的分片

将 $f \oplus w.frag$写到 $w.frag$

将 $w.distance$ 增 1

注: 参数 p 表示的是路由器 R 处的标记概率, 一般为 0.04

2.3 重构算法

当受害者收到所有的标记包后, 使用与非有序标记时相同的攻击路径重构算法进行路径重构

3 仿真实验与结果

由标记算法可以看出, 当路由器对转发目标为某 IP地址的包进行标记时, 由于在 Hash表中存储了标记状态, 每次对同一目标 IP地址的包进行标记时, 就会顺序取 8个标记中的一个标记分片进行标记. 在这种情况下, 当对某 IP地址的攻击包经过路由器 R时, 只要 R对这些数据包的标记次数达到 8次, 那么在受害者处就可以获得足够的标记包用来重构路径中节点 R的 IP地址.

我们在实验环境下对有序包标记和普通包标记在重构时所要收到的标记包数目进行统计分析, 结果如图 3所示. 在图 3中, X轴表示实验结果的序号, Y轴是进行重构时收到的包的总数目. 在实验中, 同时对非攻击包 /攻击包比例对于有序包标记在重构时需要的包的数量的影响进行考察, 在实验进行的整个时间区间内, 非攻击包占流经路由器的所有转发包的比例在区间 [0.0001, 0.1] 上均匀递增.

由图 3可以看出, 普通包标记对某个节点进行重构时, 所需收到的包的总数目在 1000个左右波动, 且震荡较大; 而有序包标记所需要收到包的总数目基本在 500个左右, 且震荡较小. 非攻击包占所有数据包比例在区间 [0.0001, 0.1] 上变化时, 对重构所需包的数目有一定影响, 但影响不很明显.

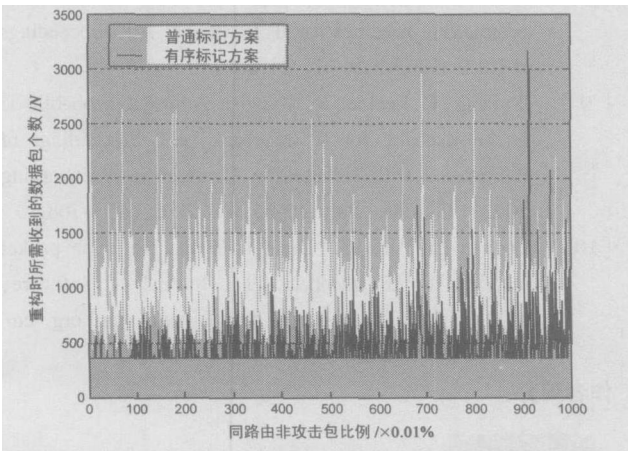


图 3　普通包标记方案与有序包标记重构时所需包的数量对比

4 讨论

有序包标记方案对重构时所需包数目的提高依赖于对标记状态的存储, 有三方面的因素会对提高的程度产生影响, 包括: Hash函数的碰撞, 同目标 IP非攻击数据包 /攻击数据包的比例, 以及后续路由器的重复标记. 下面我们对这三方面影响逐一讨论.

4.1 碰撞问题和碰撞上界

由于对某一目标 IP地址的标记状态的存储是使用 Hash函数映射到长度为 2n的 Hash表 T中, 那么首先需要考虑的就是 Hash函数的碰撞问题, 和 Hash函数碰撞带来的影响.

下面对碰撞上界进行估算.

为估算状态标记的碰撞, 我们首先证明引理 1.

引理 1　设 $0<q<1$, 则有 $(1-q)^n > 1-nq$.

证明从略.

假设 H 是输出长度为 m 的理想 Hash函数, 那么, $H(x)$ 与任意 n 个与 x 不同的数的 Hash输出碰撞的概率为 $(1-(1-p)^n)$, 其中 $p = 1/2^m$. 由引理 1, 我们可以求出 $H(x)$ 与任意 n 个与 x 不同的数的 Hash输出碰撞的概率 $P_{collision}$ 的一个上界 B:

$$B = np > 1-(1-p)^n = P_{collision} \quad (2)$$

由引理 1可以得到定理 1:

定理 1　设一定时间内转发 IP包的目标地址数目为 n, 那么当 Hash表 T 的长度大于 A 时, 在该段时间内, 每个状态标记与其他状态标记发生碰撞的概率 p 小于 $3n/A$.

证明: 转发的 IP包目标地址数目为 n, 每个地址对应 3个标记, 所以标记的总数目最多为 $3n$ 个. 因为 Hash表 T 的长度大于 A, 所以两个不同 Hash值 $H(x), H(y)$ 的碰撞率小于 $1/A$. 由式 (2) 可得, 在该段时间内每个状态标记与其他状态标记发生碰撞的概率

$$p = 1-(1-(1-(1-1/A)^n))^3 < 1-(1-n/A)^3 < 3n/A \quad (3)$$

由定理 1可以看出, 状态标记之间的碰撞概率随着 Hash表长度的增加而下降. 例如, 当一定时间 t 内转发包的目标地址最多为 10000时, 如果 Hash表 T 的长度为 2^{24}, 则每个状态标记的碰撞的概率低于 0.2%.

另外, 概率包标记追踪方案主要是用于风暴式拒绝服务攻击的追踪. 当拒绝服务攻击发生时, 发往受害者处的攻击包应该在短时间内具有较大流量. 由于在攻击发生的时间内, 有足够数量的发往受害者处的攻击包通过标记路由器, 所以, 攻击包的状态标记会大大低于定理 1给出的碰撞下界, 从而进一步减少 Hash函数碰撞带来的影响.

4.2 同目标攻击包 /非攻击包比例对有序标记方案的影响

标记状态的存储有助于对 IP包的顺序标记, 从而减少受害者重构时需要的标记包的数目. 但是, 当对发往某一目标 IP地址的攻击包进行从 1到 8的顺序标记时, 其中出现了发往该攻击地址的非攻击包, 若非攻击包出现的次序为 k, 那么只有当标记到第 $n+k$ 个发往该 IP地址的数据包时, 受害者才可以收到包含在攻击包中的该路由器的全部地址标识.

但是, 由于概率包标记追踪方案主要是用于风暴式拒绝服务攻击的追踪, 而在攻击包与非攻击包出现的频率相当的情况下, 风暴式拒绝服务攻击并不能对目标系统形成明显损害. 所以, 当一个目标受到来自某攻击路径的攻击时, 在同一路径上, 攻击包的数目会远远高于非攻击包的数目. 在这种情况下, 有序标记方案与普通方案相比较, 重构时所需包的总数目是有明显减少的.

图 4是路由器转发到某个 IP地址的攻击包 /非攻击包比例的变化对受害者重构该主机所需数据包总数目的影响, 实验采用数据中, 非攻击包在整个转发的包中是呈均匀的随机分布. X轴是非攻击包在所有数据包中的比例, 范围是从 0% 到 20%. Y轴是受害者重构该节点所需要收到的攻击包数目.

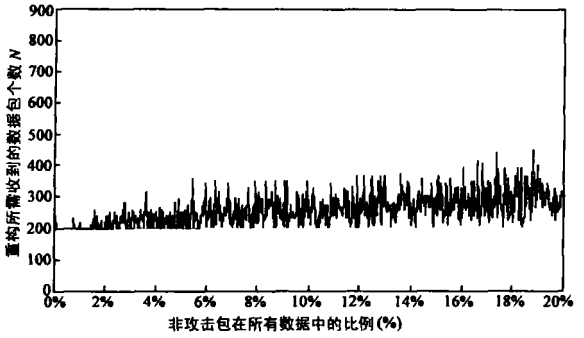


图 4 非攻击包 /攻击包比例对重构所需数据包数目的影响

4.3 路由器重复标记对有序标记方案的影响

使用有序包标记时, 由于中间路由器可能对已经做过标记的包再次标记, 会导致受害者收到的包的标记不是连续分布的. 但由于标记包被再次标记的可能性比较小, 所以, 即使距离达到 20跳, 使用有序标记的包进行重构时, 所需的包数目仍然明显低于一般的标记方案. 表 1所示, 为当距离分别为 10, 15, 20时, 随非攻击包比例变化, 有序包标记所需收到的数据包数目, 和普通包标记方案所需收到的数据包数目的对比. 由表 1可以看出, 即使是距离受害者较远的路由器, 实施有序标记方案后对于受害者重构攻击路径所需数据包数目的减少仍然是显著的.

表 1 受害者重构不同距离路由器所需收到的数据包总数目

非攻击包占总转发包的比例	有序标记方案			普通标记方案		
	10	15	20	10	15	20
1%	308	387	485	727	980	1221
2%	334	412	509			
5%	352	452	571			
10%	532	686	888			

5 结论

本文分析了已经提出的各种包标记方案进行重构时所需要收到的标记包的数目; 然后提出了一种有序标记的包标记方案, 并将该方案在重构时所需要包的数目上与以往方案进行了比较. 比较结果说明, 该方案的提高是显著的. 该方案对于已经部署了 IP包标记方案的路由器来说, 只需要增加一定大小的存储空间, 就可以大大降低重构所需要的包的数目, 使受害者可以更早地采取应对措施, 从而可以有效减轻 DoS攻击和 DDoS攻击的危害性. 以较小的存储代价换取在实时防御上的快速的响应时间, 在包标记方案的实际应用上是有一定意义的.

参考文献:

[1] Alex C Snoeren, et al. Hash-based IP traceback[A]. Proceedings of the ACM SIGCOMM 2001[C]. San Diego, California, USA, August 27-31, 2001. 3-14.

[2] Hal Burch and Bill Cheswick. Tracing anonymous packets to their approximate source[A]. Usenix LISA[C]. New Orleans, December 3-8, 2000. 313-321.

[3] Steven Bellovin, Marcus Leech, Tom Taylor. ICMP Traceback Messages[R]. work in progress, Internet Draft, draft-ietf-itrace-02. txt, 2001.

[4] Robert Stone. Centertrack: an IP overlay network for tracking DoS floods[A]. Proceedings of 9th USENIX Security Symposium[C]. 2000.

[5] Kihong Park, Heejo Lee. A Proactive Approach to Distributed DoS Attack Prevention using Route-Based Packet Filtering[R]. Technical Report, CSD00-017, Department of Computer Sciences, Purdue University, 2000.

[6] Stefan Savage, David Wetherall, Anna Karlin and Tom Anderson. Network support for IP traceback[J]. ACM/IEEE Transactions on Networking, June 2001, 9(3): 226-237.

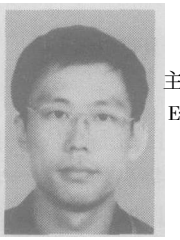
[7] K Park, H Lee. On the effectiveness of probabilistic packetmarking for IP traceback under denial of service attack[A]. Proceedings of IEEE INFOCOM 01[C]. 2001. 338-347.

[8] Dawn X Song, Adrian Perrig. Advanced and authenticated marking schemes for IP traceback[A]. Proceedings of IEEE INFOCOM'01[C]. 2001.

[9] T Peng, C Leckie, R Kotagiri. Adjusted probabilistic packetmarking for IP traceback[A]. Proceedings of the Second IFIP Networking Conference (Networking 2002)[C]. Pisa, Italy, 19-24 May 2002. 697-708.

[10] Dequan Li, Purui Su, Dengguo Feng. Notes on packet marking for IP traceback[J]. Journal of Software, 2004, 15(2): 250~258. http://www.jos.org.cn/1000-9825/15/250.htm.

作者简介:



曲海鹏 男, 1972年出生于山东莱西, 博士, 主要研究领域为信息安全、网络安全. E-mail: haipeng@acm.org

冯登国 男, 1965年生, 研究员, 博士生导师, 主要研究领域为信息安全、网络安全. E-mail: feng@is.iscas.ac.cn.

苏璞睿 男, 1976年生, 博士, 主要研究领域为信息安全、网络安全. E-mail: supurui@is.iscas.ac.cn