第35卷 第10期 计 算 机 工 程 2009年5月
Vol.35 No.10 Computer Engineering May 2009
·安全技术· 文章编号：1000—3428(2009)10—0115—03 文献标识码：A 中图分类号：TP309

复合包标记IP追踪算法研究

高大鹏[1]，於时才[2]，闫文芝[2]
(1. 兰州理工大学电子与信息工程学院，兰州 730050；
2. 兰州理工大学计算机与通信学院，兰州 730050)

摘 要：在压缩边分段采样算法研究改进基础上，分析攻击路径距离、路由器节点流量统计对标记概率的影响，提出一种复合包标记方法。该方法可以优化算法收敛性，降低运算复杂度和重构路径的差错率，使受害者在最短时间内推测出主要攻击路径，能够很好地应用于多个分布式拒绝服务攻击的攻击源追踪中。

关键词：拒绝服务攻击；IP追踪；压缩边分段采样算法

Research on Composed Packet Marking for IP Traceback Algorithm

GAO Da-peng[1], YU Shi-cai[2], YAN Wen-zhi[2]
(1. Department of Electronic and Information Engineering, Lanzhou University of Technology, Lanzhou 730050;
2. Department of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050)

【Abstract】Based on the current research on improving the Compressed Edge Fragment Sampling(CEFS) algorithm of Savage, the relations among the distance of the attacking path, the statistics on the traffic of routers, marking probability are analyzed. A new approach of composed packet marking method is proposed. In the new proposal the convergence of mathematic is optimized, computational complexity and the false positive alarm for the victim to reconstruct the attack graph is reduced, a victim can construct major attacking path in minimum time. The method can be used in tracking DDoS attacks of multi-source by establishing a simulated test environment and experiment analysis.

【Key words】Denial of Service(DoS) attack; IP traceback; Compressed Edge Fragment Sampling(CEFS)

1 概述

随着网络的迅速发展，Internet已逐步成为人类社会的基础设施，其地位越来越重要。而针对Internet的攻击也层出不穷，特别是拒绝服务类攻击(DoS, DDoS)使Internet面临巨大的安全威胁。现有的网络安全机制基本还没有考虑网络攻击源的检测和追踪问题，即使检测到攻击也缺乏网络范围的自动响应。针对攻击源的实时反向追踪对于及时处理大规模网络的DDoS攻击和网络总体的抗打击能力具有重要意义。

本文将一种改进的压缩边分段采样算法(Compressed Edge Fragment Sampling, CEFS)为基础与可变概率标记相结合，提出一种复合式标记方法DPCEFS。在提高标记算法收敛性、降低计算复杂度和重构路径的差错率等方面有所突破。

2 压缩边分段采样算法

在PPM算法中最重要的是边采样算法[1]，该算法在标记过程中，路由器以一定概率随机将边信息的某个分段(8 bit)、该分段在原始地址中的偏移(3 bit)以及该边距受害者的距离(5 bit)共16 bit重载到IP包头的identification字段中，这样每个边信息(64 bit=32 bit+32 bit的Hash值)分成8个分段，每个分段连同其距离和偏移储存在1个IP包的包头中。在重构过程中，受害者收集所有边信息分段，并根据距离和偏移值对这些标记进行分组，在相同距离下，根据偏移量组合所有可能的边信息，并通过匹配该边信息的数据部分与Hash部分决定是否接受该边信息作为攻击图的一部分。

该算法的一大问题是当在某个距离同时有 m 个攻击者时，受害者将接收到 $m\times k$ 个分段，其中有 k 个偏移，每个偏移有 m 个分段，在组合成可能的边信息时将有 m^k 种情形。为了找到正确的边信息组合，需要进行 m^k 次Hash和比较，这一过程的运算量相对较大，而且随着 m 数值的增加，差错概率也会迅速增加。

此外，由于路由器以固定概率标记数据包，因此受害者要收集到信息需要的足够数据包数会随着攻击路径长度的增大而大量增加，耗用时间也更长。

假定一条攻击路径：$S=(A, r1, r2, \cdots, rd, V)$，其中，$A$ 表示攻击者；V 表示受害者；攻击路径长度 d 表示攻击路径上的路由器数。若路由器以固定概率 P 标记数据包，攻击规模即攻击者发出的数据包数设为 N，那么，可以得出以下结论：因为受害者至少要收到一个来自距离攻击者最近的路由器标记的数据包，即

$$N\times P(1-P)^{(d-1)}=1 \tag{1}$$

所以可得出重构攻击路径所需数据包的期望值为

$$Ind/\left[p\times(1-p)^{(d-1)}\right] \tag{2}$$

从式(2)可以看出，随着攻击路径长度的增加，重构攻击路径所需的攻击包数目会以指数级急剧增长。

基金项目：甘肃省自然科学基金资助项目(ZS031-A25-015-G)
作者简介：高大鹏((1980－)，男，工程硕士研究生，主研方向：网络安全；於时才，教授；闫文芝，工程硕士研究生
收稿日期：2008-10-27 **E-mail**：gdp19800808@126.com

3 复合包标记算法(DPCEFS)

3.1 改进的压缩边采样算法

针对压缩边采样算法的问题，本文提出一种改进的压缩边采样算法。将边信息存储空间从 16 bit 增加到 31 bit，从而减少 Hash 运算的次数，同时通过 2 个不同的 32 bit Hash 函数共 64 bit Hash 作为误差校验以降低多个攻击者同时存在时重构路径的虚警率。

在 IP 包头内，与数据包的分片和重装有关的字段是标识 identification(16 bit)、标志 flag(3 bit)和片偏移 offset (13 bit)。然而，当标识被重载后，片偏移的保留就没有意义了[2]。因此，可将边信息存储空间从 16 bit 增加到 31 bit(图 1)。

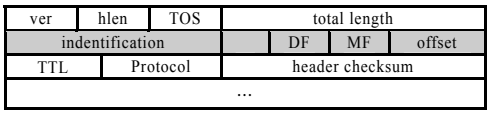

图 1 IP 数据包头结构

重载这 31 bit 对包流量所产生的影响与重载 16 bit 标识位基本一样，只是对占总流量不足 0.25%的分段包有影响[3]。

(1)构建地址。每个路由器首先使用 2 个不同 Hash 函数对其 IP 地址计算 Hashl(32 bit)和 Hash2(32 bit)，并将这 2 个 Hash 值串接得到 64 bit Hash 值，将 64 bit Hash 值直接串接到 32 bit 地址后面构成 96 bit 边信息。该边信息被分成 4 段，每段 24 bit。在标识过程中，路由器以一定概率随机选择这 4 段中的任一段(24 bit)，连同偏移量(因为只有 4 段，所以只需要 2 bit)和距离(5 bit)共 31 bit 重载到 IP 包头中。

(2)IP 包头重载。24 bit 的边信息分段分别重载到 IP 包头的标识 indentification 字段(16 bit)和 IP 包头的分段偏移 offset 字段(占前 8 bit)，5 bit 距离重载到 IP 包头的分段偏移字段(占 9 bit~13 bit)，2 bit 偏移重载到标志中未使用的 1 bit 和标志中的 MF 位，而 DF 位置 1，告知后的路由器不再对本 IP 包进行分段。

(3)减少冲突误报现象。为保证将攻击者的干扰信息降到最低，应定时改变散列函数，如对散列函数Hash()采取预警更新方式，即在通常情况下使网络中的路由器维持同一个Hash函数。一旦发现拒绝服务攻击，可通过广播数据包及人工等方式快速更新为一个随机Hash函数，以免被攻击者利用。

3.2 可变概率包标记算法及改进

可变概率包标记算法的理论根据是：假设从攻击者到受害者的距离为 d(即攻击者到受害者之间经过 d 个路由器)，从攻击者到受害者之间的路由器依次为 $r_1,r_2,\cdots,r_d$。

设 p_i 为路由器 r_i 的标记概率，则：

包只在第 1`个路由器处被标记的概率为 $p_1(1-p_2)\cdots(1-p_{d-1})(1-p_d)$；

包只在第 2 个路由器处被标记的概率为 $p_2(1-p_3)\cdots(1-p_{d-1})(1-p_d)$；

……

包只在第 d-2 个路由器处被标记的概率为 $p_{d-2}(1-p_{d-1})(1-p_d)$；

包只在第 d-1 个路由器处被标记的概率为 $p_{d-1}(1-p_d)$；

包只在第 d 个路由器处被标记的概率为 p_d。

如果每个数据包最终被那个路由器标记的概率都相等(1/d)，则由 Coupon collector[4]理论可知，重构所需的数据包数最少。即

$p_d=1/d, p_{d-1}=1/(d-1), \cdots, p_2=1/2, p_1=1$

当变概率为 $1/i(i\in[1,d])$时，受害端重构攻击路径所需的包的数目最少，所需的数据包平均值为 $d(1+1/2+\cdots+1/d)$，其期望值为 $d\times Ind$，这是理论上的最小值。

虽然可变概率包标记算法有很多优势，但也存在明显不足：

(1)在实现过程中，主要的困难是 i 值的确定。这里可以应用 Liu Jenshiuh 等人提出的思想：利用 IP 报文头部的 TTL 域信息确定当前的 i 值。另外，由于 TTL 初始值是由操作系统和协议决定的，因此不会被攻击者篡改，提高了标记信息的准确性。他们统计了目前存在的不同操作系统和协议的 TTL 的初始值，得到了一个集合{32, 64, 126, 255}。可以根据当前的 TTL 值和路径长度不会大于 32 的结论[5]判断其初始值，从而计算出 i 值。例如，当前 TTL 为 49，那么其初始值只能为 64，可以计算出 i 值为 64-49=15。这个数据包已经过 15 个节点转发，该节点是数据包经过路径上的第 16 个节点，于是可以依概率 1/16 决定是否需要标记该数据包。

(2)路由器越靠近攻击者，其标记数据包的概率越大，尤其是边界路由器的标记概率为 100%，即边界路由器要对所有转发的数据包进行标记，这将导致路由器的负担过重。如果该处网络流量太大，会导致网络拥塞甚至路由器服务瘫痪。为了减小边界路由器的负担，本文提出如下改进方案：把边界路由器上的标记概率定为 50%，而在边界路由器的下一跳路由器上执行标记算法时，先判断数据包是否已被标记，然后继续相应的操作，以避免覆盖边界路由器的标记信息。这实际上是对经过边界路由器的数据包进行了分流(对标记负载的分担)。这种做法在标记结果和时间效率上并没有折损，却使边界路由器的标记负担减少了 50%。

(3)当 i 较大时，p_i 会很小。那么受害者就需要得到更多来自于攻击者的数据包才能重构出攻击路径，造成整个过程耗用时间更长，而且这一点很可能会被攻击者利用，攻击者通过伪造距离域值并将它置为一个较大的数值，使后续路由器以小概率标记数据包，那么被攻击者从一定量的数据包得到的标记信息就会少很多。克服这一问题最直接的方法是采用认证方式确保该距离值不被伪造。但是考虑到其实现的难易程度，可以采取一种折中的办法，即在概率 p_i 变得很小之前就将其固定为某个值，如 p=0.04。文献[1]已说明采用固定概率时取概率为 0.04 是一个较好的选择。

4 复合包标记算法的性能分析

4.1 收敛性

重构攻击路径所需要的包个数决定了算法的收敛性，复合概率包标记算法路径重构所需的数据包数期望理论值为 $d\times Ind$。因为所需包数最少，所以重构时间也最短。

设攻击者发出 N 个攻击数据包，在到达攻击者所在网络的边界路由器时，路由器启动标记算法，标记概率为 1/2，被标记的数据包数 $N_1=N/2$；到达下一跳路由器时，路由器执行标记算法，被该路由器标记的数据包数 $N_2=N/2$。所以，到达第 3 跳时的 N 个攻击包都被标记且被第 1 跳路由器和第 2 跳路由器标记的数目均为 $N/2$。这一结果与标记过程改进前基本一样，只是在末减少被标记数据包数目的前提下减少了边界路由器的负载。所以，依然保持了良好的收敛性，减少了网络开销。

4.2 算法时间复杂度

当有多个攻击者同时存在时，会导致在相同距离下的多

个具有相同偏移量的分段，因此，需要计算多次 Hash 验证真实的边信息。假设在距离 d 处有 m 个路由器在 DDoS 的攻击路径中，则文献[1]中的压缩边分段采样算法在重构路径时所需要的计算量(所需要计算的 Hash 次数)为 $O_{CEFS}=O(m^8)$，而复合概率包标记方法在重构路径时所需要的计算量为 $O_{DPCEFS}=O(m^4)$。显然，m 越大，这种差距会越大，复合算法的计算效率优势更明显。因此，复合包标记算法更适用于大范围网络攻击的追踪。

4.3 重构路径的差错率

对于长度为 h 的随机 Hash 函数，接收任意构造的候选边信息的概率为 $1/2^h$，若在任意距离 d 处有 m 个攻击者，则在距离 d 处错误地接收一个边信息的最大概率(即差错率)[6]为

$$p_i=1-\left(1-1/2^h\right)m^k \tag{3}$$

当 m^k 《2^h 时，式(3)约等于 $m^k/2^h$。对于 m 取不同值，压缩边采样算法与动态概率包标记方法的差错率进行比较。容易得出，对于在同一距离同时攻击者越多，复合概率包标记方法的差错率比压缩边分段采样算法的差错率越低。因此，复合概率包标记方法适用于多个攻击者同时存在的 DDoS 攻击。

4.4 模拟结果

为验证复合包标记策略的优越性，本文以 NS2.31 为仿真实验平台，采用来自 CAI-DA(Cooperative Association for Internet Data Analysis)收集的 Internet 环境下跟踪路由数据库的实验数据进行仿真攻击实验。选取攻击距离的长度为 1~30，进行 100 次实验取平均值。实验表明，本文的复合包标记方法明显优于原有的压缩边采样算法，见图 2、图 3。

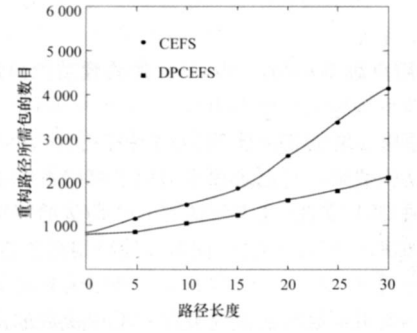

图 2　2 种策略重构路径所需包数目的比较

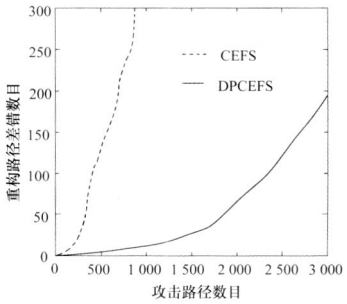

图 3　2 种策略差错率的比较

5　结束语

复合概率包标记方法是压缩边采样算法和可变概率包标记算法的有机结合。改进的压缩边采样算法增加了边信息存储所需要的空间，明显降低了重构过程的计算复杂度，通过采用 64 bit Hash(由 2 个不同的 32 bit Hash 函数得到)作为校验明显降低了多个攻击者同时存在时重构路径的差错率；以路由器距离跳数为因子的可变概率包标记算法降低了重构路径所需要的标记包数量要求，大大提高了追踪算法的实时性。仿真实验结果表明，复合包标记方法能够更有效地应用于大规模 DDoS 攻击源追踪。

参考文献

[1] Savage S, Wetherall D, Karlin A, et al. Practical Network Support for IP Trac-eback[C]//Proceedings of the 2000 ACM SIGCOMM Conference. Stockholm, Sweden: [s. n.], 2000.

[2] Liang Fen. Real Time IP Traceback with Adaptive Probabilistic Packet Marking[J]. Journal of Software, 2003, 14(5): l005-l010.

[3] Tanenbaum A S. Computer Networks[M]. 4th ed. Indianapois: Prentice Hall PTR, 2003.

[4] Boneh A, Hofri M. The Coupon Collector Problem Revisited Commun[J]. Statist-Stochastic Models, 1997, 13(l): 39-66.

[5] 李德全. 拒绝服务攻击对策及网络追踪的研究[D]. 北京: 中国科学院研究生院, 2004-06.

[6] Ferguson P, Senie D. Network Ingress Filtering: Defeating Denial-of-Service Attacks which Employ IP Source Address Spoofing[S]. RFC 2827, 2000.

编辑　张正兴

(上接第 108 页)

参考文献

[1] 钟怀东, 徐　慨, 项顺祥. Ka 波段卫星通信降雨衰减特性分析[J]. 广播与电视技术, 2008, (4): 91-94.

[2] 唐映得. Ka 频段卫星通信中的雨衰与抗雨衰技术[D]. 西安: 西安电子科技大学, 2001.

[3] 陈　刚, 朱诗兵.卫星通信发趋势初探[J]. 山西电子技术, 2004, (1): 33-35.

[4] Lin K T, Zaks C, Dissanayake A W. Results of an Experiment to Demostrate the Effectiveness of Open-loop Up-link Power Control for Ku-band Satellite Links[C]//Proc. of Intl Conf. on Antennas Propagation. [S. 1.]: IEEE Press, 1993: 234-238.

[5] International Telecommunications Union ITU-R PN.618-5-1997 Propagation Data and Prediction Methods Required for the Design of Earth-space Telecommunications Systems[S]. 1997.

[6] 汪荣鑫. 随机过程[M]. 2 版. 西安: 西安交通大学出版社, 2006.

[7] 沈福民. 自适应信号处理[M]. 西安: 西安电子科技大学出版社, 2001.

编辑　索书志