一种多层次的追踪器部署策略 *

徐劲松[a,b]
(南京邮电大学 a. 通达学院; b. 计算机学院, 南京 210003)

摘 要: 结合 K-剪枝算法,提出了一种多层次追踪器部署策略,在第一层的追踪器部署时选择较大的 k 值进行剪枝,以较小的改造代价部署相对较少的追踪器,来进行关键追踪,在第二层通过二次剪枝,以部署剪枝域内追踪器,对无法通过第一层次节点进行追踪的攻击进行再次追踪,确定攻击来源。该方案能够以较低的网络改造代价以及网络性能代价完成准确追踪。理论分析以及仿真结果验证了该方案的正确性和有效性。

关键词: 分布式拒绝服务攻击; 追踪器; 数据包标记; 剪枝
中图分类号: TP393.08 **文献标志码**: A **文章编号**: 1001-3695(2010)03-1039-03
doi:10.3969/j.issn.1001-3695.2010.03.064

Multi-layer tracer deployment scheme

XU Jin-song[a,b]
(*a. Tongda College, b. College of Computer Science, Nanjing University of Posts & Telecommunications, Nanjing* 210003, *China*)

Abstract: This paper proposed a multi-layer tracer deployment scheme with the K-diameter algorithm. In the scheme, chosen a larger k-could to value diameter tracer deployment at the first layer. As a result, needed less cost to deploy comparatively few tracers for the sake of conducting the pivotal tracing. The second diameter at the second layer was to deploy tracers in the area of diameter. Could confirmed the source of attack thus by tracing again the attacks which couldn't be traced at the first layer node. The scheme could trace accurately with lower costs of network transformation and network performance. The theoretical analyses and the simulations verified the scheme is correct and valid.

Key words: distributed denial of service attacks (DDoS); tracer; packet marking(PM); diameter

0 引言

分布式拒绝服务攻击(denial of service attacks/distributed denial of service attacks, DoS/DDoS)的攻击者能够轻易伪造源 IP 地址,使得对该攻击所进行的源地址追踪成为 DoS/DDoS 防御中一个难以解决的问题。如何找出真正的攻击者,即 IP 追踪(IP traceback)问题,成为当前 Internet 安全领域一个比较活跃的课题。

数据包标记(packet marking,PM)是目前引起最广泛关注的追踪技术,它是指路径上的路由器节点对转发的数据包加上本站的地址标记,通过从收到的攻击包中收集和分析这些标记,受害主机可以复原出该数据包的传输路径。这种方法的前提是网络中的所有路由器都可以支持数据包标记的算法(本文中将这些路由器称为追踪器(tracer))。实际上,在 2005 年 5 月所收集到的路由器数目高达 192 244 个[1],在实际的网络中不可能花费巨大的资金和时间成本将这些路由器全部更换为追踪器。因此,需要一种有效的方法指导实际网络中追踪器的部署问题。

1 相关研究

目前,针对 DoS/DDoS 攻击源追踪方案的研究中都是利用路由器来进行追踪。文献[2]在边界路由器上实现确定包标记(deterministic packet marking,DPM)方法,可能由于攻击产生于骨干网而不经过追踪器而难以追踪。由 Savage 等人[3]提出的概率包标记(probabilistic packet marking,PPM)方法,专注与如何降低重组攻击路径所需收集的数据报的数量以及降低误报等方面:文献[4]根据路由器统计信息降低数据报被重新标记的概率;文献[5]使用 TTL 信息加速数据报重构路径并减少重构路径发生错误;文献[6]提出使用概率模型的方法降低重构路径所需数据报数量;文献[7]使用着色的方法降低需要的编码数量,并提出了 2-tier 的追踪架构。这些方法虽然允许部署部分追踪器,但是即使在部署的节点超过总节点数一半的情况下,也可能因为攻击包没有经过任何一个追踪器而很难有效追踪到攻击来源。

为了解决追踪器的部署问题,Wang 等人[8]提出了一种追踪部署的策略来防御 DDoS 攻击,使用 K-剪枝算法满足在部分网络节点部署追踪器的情况下,保证一个 IP 数据包传递路径超过 k 跳时,必定经过一个路由器。文献[9]更是基于这种算法与着色包标记算法的基础上作了一些改进,以降低重构路径的数据包数和误报率。

然而在这种算法中,无法检测到攻击的概率仍然存在,并与 k 值的选取相关。追踪器的数量随着 k 值变大而减少,会使得无法检测到攻击行为的概率上升。而且 k 值越大,则删除追踪器节点后的连通子图中的节点数也越大,配合使用其他检测方法需要的代价也就越大。本文正是基于这样的考虑,提出了一种分层次的追踪器部署方案,以较低的代价标记数据包进行

收稿日期: 2009-07-02; **修回日期**: 2009-08-07 **基金项目**: 南京邮电大学青蓝计划基金资助项目(NY206055)
作者简介:徐劲松(1975-),男,江苏南京人,助理研究员,博士研究生,主要研究方向为网络安全、可信计算(xujs@njupt.edu.cn).

追踪,并保证追踪器的部署能够有效追踪到攻击行为。

2 策略描述

在 Wang 等人提出的追踪器部署策略中,当攻击者到受害者(victim)之间的路径没有经过任何一个部署了追踪器的路由器的概率仍然存在,这里将一个被追踪器包围的连通子域称为剪枝域(diameter area)。这个域包含的节点数目与 k 值的选取相关,也与 K-剪枝算法的特点相关。为了对发生在剪枝域中的攻击行为进行追踪,可以针对节点数过多的域再部署一些追踪器。由于 K-剪枝算法求出的部署方案已经被证明是一个 NP 完全问题[8],直观的做法就是在这些剪枝域内再进行一次更小 k 值的 K-剪枝以部署内部的追踪器。

2.1 IP 数据报头编码方案

由于数据包在途中经分段(fragment)处理的情况是很少出现的(不超过 0.25%),IP 头中的标志号(identification)域也很少使用,于是 Savage 等人建议将路径信息写入到 16 bit 的标志号域中。文献[3]指出服务类型(type of services, TOS)域也可以用来作为路由标记使用。在分段包标记算法(fragment marking scheme, FMS)[3]、高级包标记算法(advanced marking scheme, AMS)[10]以及一些改进方案中,都有一个距离(distance)域,用来表示标记数据包的路由器与受害者之间的距离。文献[5]建议使用生存时间(time to live, TTL)域代替 distance 域,以节省更多的空间存储路由器的信息。路径重构时,受害者对具有相同距离信息的标记尝试可能的组合[3]。

使用 1 bit 的 TOS 域来作为标记编号(偏移值)。加上 identification域与 TTL 域,整个利用的 IP 数据包头中的长度为 25 bit。具体编码格式如图 1 所示。

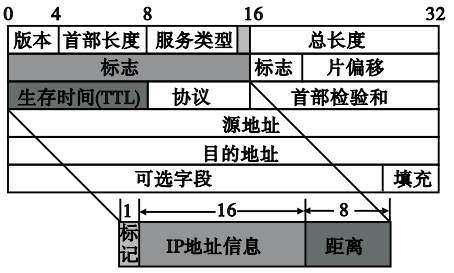


图1 包标记格式

它们的分布和作用如下:

a)标记(MID:marking identification,$0 \leqslant MID \leqslant 1$)。1 bit,用于记录 IP 信息所属的范围。

b)路由器 IP 地址信息(IP fragment)。16 bit,用于记录路由器 IP 标记的信息。

c)路由器与受害者的路由距离(distance)。8 bit,直接使用 TTL 值来标志距离。

当路由器决定对数据包进行标记时,首先查找路由器的转发表,根据数据报从哪一个网络接口进入路由器,将变数 i ($0 \leqslant i \leqslant 1$)写入 MID 位。当数据报来自于邻接的路由器时,标记为 1;当数据报来自于本地时,标记为 0,距离则用 TTL 来代替。假设路由器知道数据报的目的距离为 d,只要将 d 值的两倍写入 TTL 即可。若之后的路由器都没有再对该数据报进行标记,则都会对 TTL 减去 1。当被标记的数据报抵达受害者时,TTL 的值正好显示受害者与标记路由器的距离信息。

2.2 网络模型和定义

为了描述追踪器部署算法以及追踪的方法,定义如下:

定义 1 记网络 $G=(V,E)$ 是一个全连通的无向图。其中:V 为顶点(vertex)或节点(node)的集合;E 为边(edge)的集合。

定义 2 记 $d(u,v)$ 是图形 G 中从 u 到 v 的最短路径长度;否则 $d(u,v)=\infty$。

定义 3 记顶点集合 $V' \subset V$ 为割点集,则 $G-V'$ 将不再连通。

定义 4 记顶点 v 与 u 的最长距离 $e(v)=\max\{d(u,v), u \in V\}$ 为图形 G 中顶点 v 的离心率(eccentricity)。

定义 5 记图形 G 的半径为 $\gamma(G)=\min(e(v))$;反之记 $\mathrm{diam}(G)=\max(e(v))$ 为其直径。

定义 6 若图形 G 中存在顶点 v 有 $e(v)=\gamma(G)$,则记该顶点 v 为图形 G 的中心点 c;记中心点的集合为 C。

定义 7 若图形 G 中删除边 e 并将 e 连接的两个顶点合并为一个新顶点,则称为图形收缩(contracted)。若图形 G 对所有可以收缩的边进行收缩,所得到的图形成为收缩子图。

定义 8 若图形 G 中顶点子集 $S \subset V$,且 $\forall v \in S$ 两两互不相邻,则称 S 为独立集。

定义 9 记顶点 v 的度数(degree)为 $d(v)$,指与顶点 v 相关联的边的条数。

定义 10 记顶点 v 相邻顶点的集合为 $P(v)=\{u \in V \mid d(v,u)=1\}$。

定义 11 记用 K-剪枝算法分割出的剪枝域(diameter area)为 G_d,记算法得出的割点集为 T,记剪枝域连通子图的节点数为 num_G。

2.3 多层次的追踪器部署算法

多层次的追踪器部署算法的目的是为了在图 G 中设法找出最少的节点数来部署追踪器,使得网络满足无论攻击者来自于网络何方,攻击数据报都会经过一个追踪器,或者保证攻击源经过的第一个追踪器一定在 k 跳之内。首先使用 K-剪枝算法对网络拓扑图进行分割,第一次分割后产生数个剪枝域,如果这些剪枝域中节点的数目超过设定的值 σ,则对该剪枝域再用 K-剪枝算法进行分割,其中两次选取的 k 值第一次可以较大,第二次选取的 k 值要比第一次选取的 k 值小。算法伪代码描述如下:

```
/* 多层次的追踪器部署,输入拓扑图,输出一个多层次的追踪器部署节点集 T */
/* 以 i 作为层次的标记 */
i←0 G_i←G
validate1 : /* 选自一个合适的 k 值并进行剪枝 */
/* 输入全连通的无向拓扑图,输出本层追踪器部署节点集 T_i */
T_i←∅ if k = 1 then
/* 当 k = 1 时,使用贪心算法求出独立集 S */
/* 输出该层的部署节点集 T_i */
S←∅
validate2 : if G_i ≠ ∅ then
for whole V search v when d(v) = min d(u), ∀ u ∈ V G_i←G_i - P(v) ∪ {v} S←S + {v}
goto validate2
else break
T_i←V - S
return T_i
else /* 当 k≠1 时,进行剪枝 */
```

validate3: $C \leftarrow \varnothing$ for whole V count $e(v)$, $\forall v \in V$
if $e(v) = \gamma(G)$ then $C \leftarrow C + \{v\}$
select $\forall c \in C$ /* 任选一个中心点 */
if $k \bmod 2 = 1$ then $W \leftarrow \{v \in V \mid d(v,c) \leqslant (k-1)/2\}$
else $W \leftarrow \{v \in V \mid d(v,c) \leqslant \lfloor (k-1)/2 \rfloor\}$ $Y \leftarrow \{v \in V \mid d(v,c) \leqslant \lfloor (k-1)/2 \rfloor + 1\}$
validate4: if $Y \neq \varnothing$ then
select $\forall v_y \in Y$/ * 任选一个顶点 */
$Y \leftarrow Y - \{v_y\}$ if $\mathrm{diam}(G_{W \cup \{v_y\}}) < k$ then $W \leftarrow W + \{v_y\}$
goto validate4
else break
contract G_W to a point v_w get G^* / * G^* 是收缩后的连通图 */
$T_i \leftarrow T_i + \{u \in V \mid d(u, v_w) = 1\}$ $G' \leftarrow G^* - T_i - \{v_w\}$ / * G'是删除中心点以及追踪器的子图集合 */
for each subplot $\forall G_s \subset G'$/ * 对每一个子图进行剪枝 */
if $\mathrm{diam}(G_s) \geqslant k$ then
goto validate3
else break return T_i
$G'' \leftarrow G_i - T_i$/ * 删除追踪器形成剪枝域集合 */
$i \leftarrow i+1$ for each diameter subplot G_i'/ * 针对每一个剪枝域的节点数进行运算 */
if $\mathrm{num}_{G_i'} > \sigma$, $\forall G_i' \subset G''$ then $G_i \leftarrow G_i'$
goto validate1
else break

该算法与文献[10]提出的算法最大的不同是可以在上一层次的追踪器部署时选择较大的 k 值进行剪枝,有效降低了追踪器部署的数量。网络拓扑被第一层的追踪器 T_0 划分为数个剪枝域。平时第一层的追踪器只要以极低的概率对经过的数据报进行标记,当受害者发现攻击行为时,才要求下层的追踪器进行追踪。$k=3$ 时第一层剪枝的范例如图 2 所示。

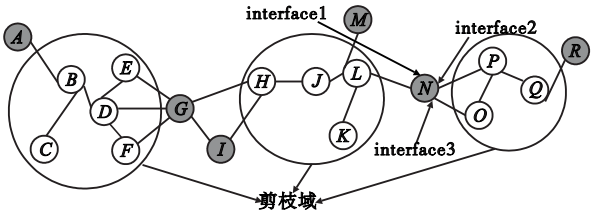


图2　$k=3$时第一层剪枝的范例

3　数据报来源追踪

3.1　来源追踪算法

当受害者发现有攻击行为时,会对来源的数据报进行分析。数据报所标记的信息包含追踪器的地址以及 MID 值。当 MID 值为 0 时,表示数据报来自于追踪器的本地网络。这时只要配合使用其他工具就可以对攻击源进行定位。当 MID 值为 1 时,表示数据报可能来自于其他路由器。一般情况下,可以保证该数据报应该是在路由器邻接的剪枝域。虽然标记方法是以概率的方式标记数据报,事实上被标记过的数据报可以通过一定的方法避免被再次标记。假设数据报是按照最短路径的顺序进行传递,基于路由器知道数据报的目的距离。当被标记过的数据报经过另一个追踪器时,该追踪器可以通过 $2d$ – TTL 的值判定该值是否在其他追踪器距离的范围内。如果落在该距离范围内,则表示该数据报已经过其他追踪器。

再回到图 2,追踪器 N 进行数据报标记前,会先判断数据报的来源。如果是本地网络,则标记 MID 值为 0,并进行地址与距离的标记动作。假如数据包是从 interface1、interface2、interface3 上过来,首先根据目的地址的距离 d 与收到数据报的 TTL 值,求出 $f(d) = 2d - \mathrm{TTL}$。由于追踪器 N 知道邻接追踪器 I、G、M、R 的距离 $T_h = \{2,3,4\}$,只要判定 $f(d)$ 是否落在 T_h 之内,就可以判定出数据报是否被标记过,也就是说已经过其他追踪器,否则数据报来自于邻接的剪枝域内。当受害者判定攻击者来自于剪枝域内,就会发起下一层的追踪器进行追踪。如果下一层没有部署追踪器时,就需要对剪枝域进行线性查找,直到找到攻击源为止。追踪算法的伪代码如下:

/ * 受害者检测到攻击并发起追踪 */
$i \leftarrow 0$
validate: if MID = 0 then find attack source
else $i \leftarrow i+1$ / * call the tracer awake PM of i level */
if $T_i = \varnothing$ then
find attack source by liner search
else PM with highly probability by i level tracer
goto validate

当受害者收到攻击数据报后,会使用概率包标记的一般方法先重构路径,以找到距离最远的追踪器,并认为该攻击数据报来自于该追踪器或该追踪器相邻的剪枝域内。追踪器根据数据报发送的网络接口判断是来自于追踪器本地网络还是剪枝域。如果来自本地网络,直接找到攻击源;如果来自于剪枝域(该域由网络接口直接判断),首先查找下层有没有部署追踪器。如果没有部署,则有必要作线性查找;否则通知下一层的追踪器以较高的概率标记数据报,以查找攻击源。这里有一种特殊的情况存在,就是攻击源与受害者都落在同一个剪枝域内,这时候受害者需要通知本剪枝域邻接的追踪器发起本域内的追踪。

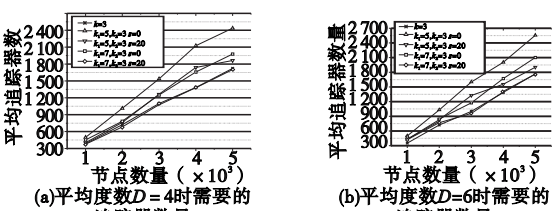
3.2　追踪成本分析

系统进行追踪的成本主要体现在攻击发生后进行搜索产生的数据成本。

当数据报直接来自于追踪器,在本方案中无须进行其他工作,可以直接找到攻击源,因此认为无须定义追踪成本。当攻击者落在剪枝域中,若是单层剪枝或剪枝域内没有部署此层的追踪器,则搜索成本定义为:对剪枝域内网络节点作线性搜索的数量;当此层有追踪器存在,额外的追踪成本需要定义为:攻击者所在的剪枝域邻接路由器与受害者的距离(通知成本)及该剪枝域的直径长度(广播成本)。

4　系统仿真

为了测试该方案的性能,实验数据采用了从 CAIDA[1] 获得的真实因特网数据。实验中取平均度数 $D=4$ 和 $D=6$ 两种情况分别对需要部署的追踪器数量与单层结构的追踪器部署进行比较。实验中剪枝参数分别选为 $K=3, K_1=5, K_2=3, \sigma=0, K_1=5, K_2=3, \sigma=20, K_1=7, K_2=3, \sigma=0, K_1=7, K_2=3, \sigma=20$ 来比较它们所需要部署的追踪器数量以及查找的成本。图 3 为需要部署的追踪器数量的比较。



(a)平均度数$D=4$时需要的追踪器数量　　(b)平均度数$D=6$时需要的追踪器数量

图3　追踪器部署数量的比较

图 3 显示,使用单层追踪部署方案,设定剪枝值 $k=3$ 和多层部署时与多层追踪部署方案中设定剪枝值 $k_1=7, k_2=3, \sigma=20$ 所需要部署的追踪器数目并没有太大 (下转第 1044 页)

$$A=[\underbrace{255\ \cdots\ 255}_{255}\ 0\ 1\ 2\ \cdots\ \underbrace{255\ \cdots\ 255}_{(NM-2\times255)}]^{\mathrm{T}} \quad (23)$$

得到其对应的加密图像 C_A。然后结合式(18)和 S 得到用于置乱的随机数 $t_{256\sim510}$。

重复上面的步骤,直至取

$$A=\begin{bmatrix}\underbrace{255\ \cdots\ 255}_{n\times255}\ 0\ 1\ 2\ \cdots\\ (N+M-1-255n)255\ \cdots\ \underbrace{255}_{NM-(N+M)}\end{bmatrix}^{\mathrm{T}} \quad (24)$$

得到其对应的加密图像 C_A。然后结合式(18)和 S 得到用于置乱的随机数 $t_{(N\times255+1)\sim(N+M)}$。

综合上述步骤,确定用于置乱的 $N+M$ 互不相等的随机整数 $t_{1\sim N+M}$,至 d)。

d)计算

$$t_1=\begin{cases}\mathrm{mod}\ (t_i,M) & 1\leqslant i\leqslant M\\ \mathrm{mod}\ (t_i,N) & (M+1)\leqslant i\leqslant (N+M)\end{cases} \quad (25)$$

完毕。

此方法不需恢复密钥的任何部分,通过选择明文攻击确定 S 和用于置乱的随机整数 $t_{1\sim N+M}$,则可对任意大小为 $I\times J$ (I 能整除 N, J 能整除 M)的加密图像成功解密(说明:用 I 和 J 分别代替式(25)中的 N 和 M,得到用于置乱的随机整数 $t_{1\sim(I+J)}$)。若加密图像的大小不满足"I 能整除 N, J 能整除 M"的条件,则令 $N=I$,$M=J$,再执行上述攻击步骤。

2.4 增强安全性的建议

原加密算法的优点在于,采用多混沌系统,密钥空间大;采用超混沌系统,其混沌轨道复杂。但是。算法在结构设计上存在致命弱点,即加密过程与明文无关,使得算法无法抵御已知明文攻击和选择明文攻击。

建议解决方案是采用密文反馈和多轮次加密,这样可以使明文图像中每个像素的变化扩散影响到尽可能多的密文像素,使算法具有很好的扩散和混淆特性,增强对各种攻击方法的抵御能力。

3 结束语

对文献[2]提出的一种基于超混沌系统的图像加密算法进行了安全性分析。本文提出的选择明文攻击和已知明文攻击方法可实现对任意大小加密图像的准确破译。为了增强文献[2]中算法的抗攻击能力,建议在原算法的基础上增加密文反馈或采用多轮加密的形式,以提高其安全性。

参考文献:

[1] WONG K W, KWOK B S K, LAW W S. A fast image encryption scheme based on chaotic standard map[J]. Physics Letters A, 2008, 372(15): 2645-2652.

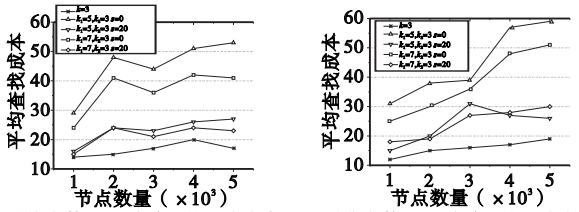
[2] GAO Tie-gang, CHEN Zeng-qiang. A new image encryption algorithm based on hyper-chaos [J]. Physics Letters A, 2008, 372(4): 394-400.

[3] RHOUMA R, BELGHITH S. Cryptanalysis of a new image encryption algorithm based on hyper-chaos [J]. Physics Letters A, 2008, 372(38): 5973-5978.

[4] GUAN Zhi-hong, HUANG Fang-jun, GUAN Wen-jie. Chaos-based image encryption algorithm [J]. Physics Letters A, 2005, 346(1-3): 153-157.

[5] COKAL C, SOLAK E. Cryptanalysis of a chaos-based image encryption algorithm [J]. Physics Letters A, 2009, 373(15): 1357-1360.

(上接第 1041 页)差距。但是值得注意的是,在多层部署方式下,由于平时只有第一层节点在进行标记动作,对于网络的效率影响较小,系统成本比较低。图 4 是各种部署方案下追踪器响应追踪时追踪成本的比较。



(a)平均度数$D=4$时需要的追踪成本 (b)平均度数$D=6$时需要的追踪成本

图4 追踪响应的查找成本比较

从图4看出,多层架构的追踪器部署方案在响应查找时会有通知以及广播成本存在。但是由于部署的方案在平时不作追踪时需要动作的追踪器较少,也对系统性能造成的影响较小。

5 结束语

本文提出了一种使用多层次的方式对网络中的攻击源追踪进行部署的方案。理论分析与实验结果表明,改进后的方案可以使网络以较小的改造代价部署相对较少的追踪器来进行攻击源的追踪。在以后的工作中,如何设计一个较好的追踪算法以配合该部署方案进行追踪仍然是一个重点,并且在标记方案上,除了需要设计一个完善的标记方法外,如何有效压缩标记信息的大小也需要仔细斟酌。

参考文献:

[1] CAIDA. Skitter[EB/OL]. http://www.caida.org/tools/measurement/skitter/router_topology/.

[2] BELENKY A, ANSARI N. IP traceback with deterministic packet marking[J]. Communications Letters, 2003, 7(4):162-164.

[3] SAVAGE S, WETHERALL D, KARLIN A, *et al*. Network support for IP traceback[J]. ACM/IEEE Trans on Networking, 2001, 9(3): 226-237.

[4] SOZAKI H, ATA S, OKA I, *et al*. Performance improvement on probabilistic packet marking by using history caching[C]//Proc of the 6th Asia-Pacific Symposium on Information and Telecommunication Technologies. 2005:381-386.

[5] YAAR A, PERRIG A, SONG D. FIT: fast Internet traceback[C]// Proc of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies. 2005:1395-1406.

[6] LIU Wu, DUAN Hai-xin, LI Xing. Improved marking model ERPPM tracing back to DDoS attacker[C]//Proc of the 3rd International Conference on Information Technology and Applications. 2005:759-762.

[7] MUTHUPRASANNA M, MANIMARAN G, ALICHERRY M, *et al*. Coloring the Internet: IP traceback[C]//Proc of the 12th International Conference on Parallel and Distributed Systems. 2006:12-15.

[8] WANG C H, YU C W, YU K M *et al*. Tracers placement for IP traceback against DDoS attacks[C]//Proc of International Conference on Communications and Mobile Computing. 2006.

[9] 刘渊, 陈彦, 李秀珍. 基于追踪部署的着色包标记算法的研究[J]. 计算机应用研究, 2008, 25(10):3102-3104.

[10] SONG D X, PERRIG A. Advanced and authenticated marking schemes for IP traceback[C]//Proc of the 20th Annual Joint Conference of Computer and Communition Societies. 2001:878-886.