# 基于消息认证码与逐跳追溯的确定包标记算法

清华大学计算机系，张晨，2017011307

**摘　要**　只有满足所有边界路由器标记的都是真实地址，所有内部路由器都不修改数据包中的标记信息这两个条件，现有IPV6下的确定包标记策略才能追溯到攻击源，但这两个条件实际上是难以满足的。为解决这一问题，本文提出了一种基于消息认证码与逐跳追溯的确定包标记算法。该算法利用认证概率包标记策略的密钥链思想，使用双密钥链对标记信息进行加密，通过逐跳追溯与路径回传寻找伪造包的来源。分析表明，在攻击包标记未发生伪造时，本算法与现有IPV6 下的确定包标记策略效率相差不大，在攻击包标记被伪造时，本策略能以较小的额外开销追溯到攻击源。

**关键词**　DDoS攻击，攻击溯源，确定包标记，IPV6

## 1　引言

分布式拒绝服务攻击(DDoS)是现今网络攻击的主要组成部分之一。攻击者会通过伪造攻击包的IP地址，达到在网络中隐藏自身的目的，这使得确定DDoS 攻击的来源成为处置DDoS攻击的核心困难之一。目前的攻击溯源技术可分为以下四类：

1. 链路测试技术[1]：从被攻击目标出发，由近及远，依次对被攻击目标的上游路由器进行UDP泛洪。若某条链路上存在攻击流量，泛洪流量将导致攻击流量丢包。根据这一现象，可判断出某条链路上是否存在攻击流量，从而构造出攻击路径。

2. 包标记技术[3]：路由器在进行数据包转发时，将与溯源有关的信息写入数据包头部，标记信息至少包括路由器的标识。受害者综合其收到的数据包内的标记信息，可还原出攻击源或攻击路径。

3. 日志记录技术[4]：路由器记录最近转发的若干数据包, 当攻击发生时, 受害主机向其上游路由器查询是否转发过某个特定的数据包, 通过询问结果重构攻击路径。

4. 基于ICMP的技术[6]：路由器以一定概率随机发送ICMP报文给目的IP主机，告知该路由器的IP地址、前一跳路由器的IP地址、后一跳路由器的IP 地址。主机可通过各个路由器发送过来的ICMP报文重构出攻击数据包的路径。

本文提出了一种基于确定包标记技术的解决方法。这种解决方法利用了消息认证码(MAC)和逐跳追溯，使得受害者能够确定标记的真伪，并在发现标记被篡改时定位进行篡改的路由器。论文的剩余部分结构如下。第二部分对已有的包标记技术进行了综述。第三部分回顾了一种利用消息验证码的概率包标记算法。第四部分对本方案进行了介绍。第五部分是对此方案的分析与评价。第六部分是对本篇论文的总结。

## 2　研究现状

### 2.1　包标记技术

包标记技术主要分为概率包标记(PPM) 与确定包标记(DPM)两类。

概率包标记指路由器按照一定概率将自己的地址信息写入数据包，并在传递数据包的过程中维护当前路由器与进行标记的路由器的距离[3]。其存在的缺陷是，如果网络中的攻击者数目较多，攻击路径的重构需要花费若干天的时间，且还原出的攻击路径假阳性率较高。

Belenky[5]指出对于攻击溯源问题，寻找到攻击包的入口路由器的地址与重构出完整的攻击路径是等价的。他提出的确定包标记算法应用了这

个思想。每个数据包仅在它进入网络的时候被标记，标记的内容为该包进入网络时的边界路由器入口地址。在数据包的传输过程中，标记保持不变。

### 2.2 基于IPV6的确定包标记技术

IPV4数据包包头的空间不足以填写完整的路径信息，因此路由器需要将路径信息进行分段，分别写入多个不同的数据包内。这是攻击路径重构速度慢，假阳性率高的原因之一。IPV6数据包中的逐跳选项头[7, 9]与目的地址选项头[8]可供包标记利用。这两个选项头都有较大空间，可完整储存边界路由器的信息。由于路径信息不再需要分段，IPV6中包标记技术的重构速度与正确率较IPV4网络有明显提升。

Obaid[7]将每个数据包的入口路由器地址写入逐跳选项头。他要求每个潜在受害者保存最近收到的若干个数据包(例如最近的50000 个包)，从而使受害者在攻击结束后仍能进行攻击溯源。

You-ye[8]选择了目的地址选项头，其标记策略与Obaid的策略相似。他认为，在没有攻击的时候无需对数据包进行标记；在路由器已经超负荷时，包标记会将情况恶化。因此他提出，只有当负载位于一个区间$[L_{min}, L_{max}]$ 时，路由器才应对数据包进行标记。

只有满足所有边界路由器标记的都是真实地址，所有内部路由器都不修改数据包中的标记信息这两个条件，上述策略才能正常工作。但是，由于攻击者可以入侵并操纵路由器，这两个条件实际上是难以满足的。因此，在设计包标记策略时，还应考虑包标记真伪性的验证方法。

## 3 认证概率包标记

在部署了概率包标记的网络中，受攻击的路由器同样会通过伪造包标记达到阻止受害者还原攻击路径的目的。为解决这一问题，需要对数据包中的标记进行认证。文献[2]提出了一种较为有效的认证策略，被称为认证概率包标记策略。下文将对此种策略进行简要介绍。

首先，网络中的每个路由器$R_i$分别生成一个密钥链$K_{j,i}$。产生方法为，随机选择一个种子$K_{N,i}$，按照公式$K_{j,i} = g(K_{j+1,i})$依次生成密钥链中的其他元素。其中，函数$g$为一个单向哈希函数。此函数可以由$K_{j+1,i}$ 计算出$K_{0,i}, ..., K_{j,i}$，却无法根据$K_{0,i}, ..., K_{j,i}$还原出$K_{j+1,i}$。

将时间分成若干个区间。在第$t$个时间区间，路由器$R_i$使用密钥$K_{t,i}$对在这个时间区间内收到的需要标记的包进行标记。标记方法为，使用带密钥的哈希函数$f_K$(其中K为密钥，该函数被称为消息认证码，简称MAC)，将数据包的源地址、目的地址及要写入数据包的标记信息进行加密，把标记信息与加密后的值一同写入数据包。在时间区间$t$结束后，每个路由器$R_i$都将以$\delta_r$的延迟在网站上公布此时间区间的密钥$K_{t,i}$.延迟时间$\delta_r$ 应大于网络中的可能延迟时间与路由器与受害者的时间同步误差之和。

受害者收到数据包后，保存数据包的到达时间。若要对某个数据包进行验证，则首先需要使用包的到达时间确定包发送时所处的时间区间。假设包的到达时间为$T_a$，受害者与路由器的同步时间误差为$\pm\delta_s$，网络中的最大延迟为$\delta_d$，则包的可能发送时间$T_s$ 满足$T_a - \delta_s - \delta_d < T_s < T_a + \delta_s$。因此，如果时间区间的长度远大于$2\delta_s + \delta_d$，受害者有很大的概率能够确定该数据包发送的具体时间区间。在确定了发送的时间区间后，受害者便可利用相应的密钥验证标记真伪。

## 4 基于消息认证码与逐跳追溯的确定包标记算法

### 4.1 数据包的标记

本策略中，数据包传输路径上的每一个路由器都有可能对数据包进行处理，故选用逐跳选项头作为标记域。

每个路由器$R_i$分别生成两个密钥链$A_{j,i}$与$B_{j,i}$。产生方法为随机选择两个种子$A_{N,i}$与$B_{N,i}$，按$A_{j,i} = g(A_{j+1,i})$，$B_{j,i} = g(B_{j+1,i})$生成密钥链中的其他元素。其中，函数$g$为单向哈希函数。

将时间分成若干个区间，每个时间区间的长度$L$满足$L = 2\delta_s + \delta_d$。与认证概率包标记密钥与时间区间一一对应不同，本策略中，每个密钥链对应两个连续的时间区间——$A_{t,i}$对应第$2*t-1$与第$2*t$个时间区间，$B_{t,i}$对应第$2*t$与第$2*t+1$个时间区间。图1 展示了这种对应关系。

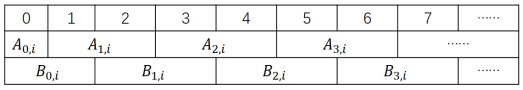


Figure 1: 时间区间与密钥链的对应关系

在第t个时间区间，边界路由器$R_i$分别使用两个密钥链中与这个时间区间对应的密钥$A_{(t+1)/2,i}$、$B_{t/2,i}$对将要进入网络的数据包进行标记。假设标记过程使用的带密钥哈希函数为$f$。密钥$A_{(t+1)/2,i}$、$B_{t/2,i}$对应的$f$函数分别为$f_A$、$f_B$，该数据包进入网络时的边界路由器入口地址为$IP_{in}$，该数据包的内容为$M_p$。边界路由器分别使用$f_A$、$f_B$对二元组$(IP_{in}, M_p)$进行加密(结果分别记为$E_A$、$E_B$)，再将三元组$(IP_{in}, E_A, E_B)$写入数据包的逐跳选项头。

### 4.2 包标记的验证

收到数据包后，受害者需要保存数据包的来源及到达时间。设某个数据包的到达时间为$T_a$，则它的可能发送时间$T_s$满足$T_a - \delta_s - \delta_d < T_s < T_a + \delta_s$，该区间的长度恰好为L，其在密钥链A、B中对应的密钥至少有一个保持不变。例如，在图2中所示的4个可能发送时间段中，上面两个时间段在密钥链A中所对应的密钥保持不变，下面两个时间段在密钥链B中所对应的密钥保持不变。受害者利用保持不变的密钥检验包标记是否被篡改。若包标记没有被篡改，则包标记中的入口地址即为攻击包的来源，若包标记被篡改，则需使用下面两小节所描述的方法进行溯源。

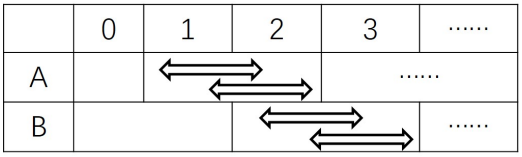


Figure 2: 每个时间区间对应一个确定密钥

### 4.3 伪造包的溯源

受害者发现虚假的攻击包后，向该包的上游路由器发送一个数据包，来要求上游路由器寻找攻击包的来源。收到溯源请求后，上游路由器会检测每个经过它的包的标记是否被篡改，检测方式与受害者的检测方式相同。如果发现了虚假包，则继续向包的来源发送溯源请求。如此递归，便可追溯到伪造包标记的路由器。收到溯源请求的路由器可能永远无法监测到虚假包，因此它只会在收到请求后的一段时间进行监测，以减少开销。若这段时间内无法监测到虚假包，则可以认为这条链路上已不存在虚假包。

路由器无法监测到虚假包导致溯源失败，有三种可能的原因：

1. 此次DDoS攻击已经结束
2. 此次攻击的攻击路径发生了改变
3. 下游路由器被攻击，在没有收到虚假包的情况下向上游路由器发送了溯源请求。

如果是由于原因2，受害者会在之后发起新一轮的攻击溯源，这轮攻击溯源将沿新的攻击路径进行回溯，从而追溯到攻击源。对于原因3，为防止受攻击的路由器不断向上游路由器发送溯源请求，造成上游路由器负担过重，可以引入“冷却时间”的概念——若一个路由器收到了下游路由器的溯源请求，却没有检测到被篡改的包，则会在之后的一段时间内，忽略此下游路由器的溯源请求。冷却时间应设为随机值，以防止攻击者通过精准计算冷却时间避开溯源。

### 4.4 溯源信息的回传

追溯到的攻击路径应传回受害者进行存储分析。在本方案中，每确定一个新的攻击路径上的路由器，该路由器便向受害者发送一个包含自己IP地址的反馈包。这个反馈包沿重构出的攻击路径进行传递，每经过一个路由器，那个路由器便将自己的IP地址信息加到这个包的最后。例如，在图3 中，A收到了一个来自B 的被篡改的包，于是向B发送了溯源请求。B收到溯源请求后，对所有到达它的数据包进行监测，发现了一个来自C的篡改包。它将执行两个操作，一是向C发送溯源请求，要求C继续追溯攻击路径，二是向A报告最新发现的路由器C，实现方式为向A发送一个内容为”CB”的反馈包。路由器C收到溯源请求后，监测到来自D的篡改包，故向D发送了溯源请求，并向B发送一个内容为”DC”的反馈包，来报告路径中的新路由器D。B收到内容为”DC” 的包后，将自己的信息附加到这个包之后，形成内容为“DCB“的反馈包，把它发送给A。若D能够监测到来自E的篡改包，则向E发送溯源请求，并向C 发送反馈包”ED”，C向B发送反馈包”EDC”，B向A发送反馈包”EDCB”……

为提高攻击路径回传的安全性，可以定义一个时间间隔$\Delta T$。如果一个路由器在收到上一个反馈包后的$\Delta T$的时间内，没有收到新的反馈包，则认为此次溯源已经结束，它会向其下游路由器回

传一个终止包，报告溯源结束，并忽略之后收到的反馈包。如果路由器收到了终止包，则也会认为溯源结束，忽略反馈包并回传终止包。受害者收到终止包时，此次溯源宣告结束。

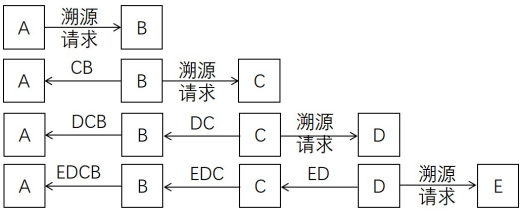


Figure 3: 溯源信息回传过程

# 5 分析与评价

## 5.1 额外开销

与现有的IPV6下确定包标记策略相比，包标记的加密与验证仅会引入少量的额外开销。加密信息会导致数据包长度增长，但由于数据包包头的长度远长于本策略中加密信息的长度，这部分额外开销可以忽略不计。包标记的加密运算与解密运算会增加相关路由器的负担，然而，因为每个数据包仅需加密一次、验证一次，路由器的额外计算开销是可以接受的。

当且仅当数据包的标记被伪造时，才需要进行“伪造包的溯源”与“溯源信息的回传”两个步骤。“伪造包的溯源”仅增添了正在进行追溯的路由器的运算与存储负担，而不会增加其它路由器的负担。另外，仅需传输少量的数据包，便可满足溯源过程中路由器间的通信需求，因此溯源过程几乎不会增添网络的通信负担。

## 5.2 追溯效率

若攻击包的包标记是真实的，则仅需一个攻击包便可确定攻击来源，这一点与已有的IPV6下确定包标记策略是相同的。若包标记是伪造的，则需进行逐跳追溯。逐跳追溯过程中每一跳所花费的时间，与路由器的包标记验证效率成正比。包标记验证的效率瓶颈为，需等待进行包标记的路由器上传相应的密钥,等待时间的上界为$2 * L$，其中L为时间区间的长度。若网络的延迟较低，且路由器间的时间误差较小，则可以通过设定一个较小的L值提高追溯的效率。

## 5.3 安全性

由于每个数据包都会被标记，攻击者在数据包进入网络前制造的虚假包标记会被边界路由器的真实包标记覆盖。若攻击者试图在数据包进入网络后对包标记进行篡改，则该包标记将无法通过受害者的验证。因此，攻击者无法通过虚假包标记对受害者进行误导。另外，受害者发现攻击者伪造的包标记后，会启动“伪造包的溯源”过程，锁定攻击来源，所以攻击者无法在网络中隐藏自身。

# 6 总结与反思

目前，IPV6下的确定包标记技术依赖于中间路由器不会对数据包进行篡改，边界路由器不会伪造包标记两条假设，但在实际情况下，这两条假设并不成立。因此，本策略将概率包标记中基于消息认证码的认证方式迁移到DPM中，以判断包标记的真实性。同时，还设计了一个在发现虚假数据包后进行溯源的算法，使得在发生伪造的时候，能追溯到最近的伪造包标记的路由器。

此策略在解决包标记的验证问题的同时，也引入了新的问题，例如，仅在攻击路径上的所有路由器都部署此项策略时，虚假包的回溯才能成功进行。另外，虚假包的回溯仅支持对正在发生的攻击的溯源，且回溯的效率有待进一步提高。因此，在未来的工作中，还需对此策略进行进一步的优化。

# References


[1] H. Burch, B. Cheswick, "Tracing anonymous packets to their approximate source", *Proc. 14th Systems Administration Conf. (LISA '00)*, pp. 319-327, 2000.

[2] D. Song, A. Perrig, "Advanced and authenticated marking schemes for ip traceback", *Proceedings of IEEE INFOCOM 2001: Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 878-886, April 2001.

[3] S. Savage, D. Wetherall, A. Karlin, T. Anderson, "Network support for IP traceback", *IEEE/ACM Trans. Networking*, vol. 9, pp. 226-237, June 2001.

[4] A. C. Soneren et al., "Single-packet IP Traceback," *IEEE/ACM Transactions on Networking*, vol. 10, December 2002, pp.721-34.

[5] A. Belenky and N. Ansari, IP traceback with deterministic packet marking, *IEEE Communications Letters*, vol. 7, no. 4, pp. 162-164, April 2003.

[6] Bellovin, S., Leech, M., and Taylor, T., (2003), 'ICMP Traceback Messages,' *draft-ietf-itrace-04*,2003, work in progress.

[7] Syed Obaid Amin, Choong Seon Hong, "On ipv6 traceback", *The 8th International Conference on Advanced Communication Technology (ICACT)*, pp. 2139-2143, Feb. 2006.

[8] You ye Sun, Cui Zhang, Shao qing Meng, Kai ning Lu, "Modified deterministic packet marking for ddos attack traceback in ipv6 network", *11th IEEE International Conference on Computer and Information Technology*, pp. 245-248, Aug. 2011.

[9] Ashwani Parashar, Ramaswamy Radhakrishnan, "Improved deterministic packet marking algorithm for ipv6 traceback", *International Conference on Electronics and Communication Systems (ICECS)*, pp. 1-4, Feb. 2014.