ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.10, October 2007, pp.2652−2661
DOI: 10.1360/jos182652

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于路由器编码的自适应包标记*

李德全[1+], 苏璞睿[1], 魏东梅[2], 冯登国[1]

[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](西南科技大学 信息工程学院,四川 绵阳 621010)

Router Numbering Based Adaptive Packet Marking

LI De-Quan[1+], SU Pu-Rui[1], WEI Dong-Mei[2], FENG Deng-Guo[1]

[1](State Key Laboratory of Information Security, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

[2](College of Information, Southwest University of Science and Technology, Mianyang 621010, China)

+ Corresponding author: Phn: +86-10-66861794, E-mail: lidequan@is.iscas.ac.cn, http://www.is.iscas.ac.cn/personnel/lidequan/

Li DQ, Su PR, Wei DM, Feng DG. Router numbering based adaptive packet marking. *Journal of Software*, 2007,18(10):2652−2661. http://www.jos.org.cn/1000-9825/18/2652.htm

Abstract: DDoS attack represents a big problem to the Internet community for its high profile, severe damage, and difficult defending. Several countermeasures are proposed for it in the literature, among which, Probabilistic Packet Marking (PPM) is promising. However, all the existing marking schemes are bearing limitations in some aspects. In this paper, a new packet marking scheme is proposed, which is more prompt because of fewer packets needed, more scalable and more efficient in computation compared with other schemes. Furthermore, this scheme limits attackers' ability in spoofing trace message.

Key words: network security; traceback; DoS (denial of service); DDoS (distributed denial of service)

摘 要: 拒绝服务攻击由于其高发性、大危害、难防范而成为因特网上的一大难题.研究人员为此提出了各种各样的对策,其中概率包标记具有较大的潜力.然而,现有的标记方案都存在各种各样的缺点.提出了一个新的标记方案,与其他标记方法相比,该方案具有反映灵敏、误报率低和计算量小的优点.此外,该方法还限制了攻击者伪造追踪信息的能力.

关键词: 网络安全;追踪;拒绝服务;分布式拒绝服务

中图法分类号: TP393 **文献标识码**: A

1 Introduction

Recently, Internet attacks are on the rise—more than 50% increase per year during 1998~2001[1]. Why are so many attacks occurring? Studies reveal that computer attacks have similarities with many other crimes: Perpetrators are motivated with many things, including greed, revenge, and peer pressure. Denial of Service (DoS) attack

* Supported by the National High-Tech Research and Development Plan of China under Grant Nos.2006AA01Z412, 2006AA01Z437, 2006AA01Z433 (国家高技术研究发展计划(863))

Received 2004-11-22; Accepted 2006-03-31

consumes resources associated with various network elements—e.g. web servers, routers, firewalls and end hosts—which impedes the efficient functioning and provisioning of services in accordance with their intended purposes[2]. There are two types of DoS attacks. The first one takes advantage of drawbacks of some implementations or algorithmic deficiencies in some applications by one or more malformed 'killer' packets. The second takes advantage of the fact that the victim is connected to the Internet itself by flooding a deluge of packets to the victim. Since the former type could be solved by patching up vulnerabilities, closing unnecessary and danger services, or filtering out malformed packets, we focus on the latter, the most common reported one[16] called flood type DoS attack. From now on, for simplicity, we refer to flood type DoS attack as DoS attack. DoS attack could be more effective if several attackers at different places conspire since the effect is summed up. This is generally called distributed DoS attack (DDoS attack).

DoS attacks are among the hardest security problems to address because they are easy to launch, difficult to defend and difficult to trace[3]. Firstly, DoS attack tools are available almost everywhere on the Internet; even script kiddies could download these tools and launch attacks at will. Secondly, it is very difficult to differentiate poisonous (attack aiming) requests from legitimate ones, this renders that it's difficult to defend from DoS attack. Thirdly, many DoS attacks do not need two-way communication. So the source addresses of DoS attack packets could be spoofed. In addition, some attacks (e.g. SYN flood attack) will be more effective with source addresses forged and source addresses must be forged for some other attacks (e.g. smurf, fraggle, and the like which adopt reflection) to be effective. This feature gives attackers an opportunity to hide their true identities.

The attack against 13 root servers[4] in October 2002 and the one against '.info' domain system[5] in November 2002 exemplify that the DoS attack trends[6] have shifted from targeting company networks to targeting the infrastructure of the Internet itself. The need to defend DoS attack and to find the attackers (or, at least the attacking source from where packets are spit out) has grown in importance. Until we are able to dedicate attention to these attacks, until we can follow these attacks to their end, we are all vulnerable.

In this paper, an Adaptive packet marking scheme is proposed, which requires fewer packets to reconstruct attack pathes compared with others of the same type. The rest of the paper is organized as follows. We give an overview on some countermeasures to DoS attack in Section 2. Section 3 gives an Adaptive marking scheme. In Section 4, a router numbering based marking scheme is given. Section 5 is about information server needed for path reconstruction. In Section 6, we give some analysis for our scheme and compare it with some others. We conclude the paper in Section 7.

2 Related Work

There are generally two categories of countermeasures to DoS attack: one is mitigating the detrimental impact of the attack on victim (such as Tolerance enhancing at the victim, ingress filtering[7], Route-based distributed packet filtering (DPF)[2]); another is tracing back to the offending parts. The former approach is passive and sometimes expensive (such as hot backups), while the latter is active in that it serves as a deterrent. To learn more about DoS attack countermeasures, please refer to Ref.[16].

Studies suggest that many intruders are deterred by the perceived risks involved. One of the intruders' greatest fears is losing their anonymity. Consequently, if we could traceback to attackers, attacks would be reduced dramatically. Even if we have not traced back to the REAL attackers but just to zombies or bots, it is helpful because we could do some filtering or throttling closer to the packet sources and thus alleviate the attack force more efficiently, we could further track the attackers from zombies too. The aim of tracing in DoS attack is to traceback to packet sources. In recent years, several tracing methods are proposed, such as Packets logging[10],Back

flooding[11], Input debugging[12],ICMP Trace back[13], Centertrack[14], Packet Marking, etc; all have their pros and cons respectively.

2.1 Packet marking

The main idea of packet marking[3,8,9] is to let routers mark packets with partial path information probabilistically. Having received enough packets from attackers or zombies (In this paper, zombies and attackers are treated equally since zombies do what attackers want and our purpose is to find the sources of attack packets), the victim could reconstruct the full path along which attack flow traveled. This field is exploited by Savage, *et al*. and followers. This paper refers the algorithm of Savage, *et al*. as Basic PPM. Following is a brief introduction.

Identification field of IPv4 header is seldom used because very few packets are fragmented on the fly (less than 0.25%[15]). Thus, this field could be used to embed some tracing messages. Let's denote two connected node on the network as an edge. In Ref.[3], the authors define edge-ID as XOR of two IP addresses making up of an edge and let router "mark" this information into packet identification field probabilistically. After receiving enough packets from attacker(s), the victim could use these edges to reconstruct the full path. For this purpose, the victim needs not only edge-IDs but also corresponding distances. Since distance seldom exceeds 30, 5 bits are enough to code it. If we rely only on the Identification field, we have 11 bits left which are not enough for 32bit edge-ID. So edge-ID should be further fragmented into several segments. To ensure reconstructing edge correctly when combining these segments, a simple error detection code is needed. 32bit hash of edge-ID is used as error detector in Ref.[3]. Therefore, 64 bits should be marked for one edge. The authors further interleave the error detection code with edge-ID bit by bit and then separate the 64 bits into 8 blocks. So 5 bits for distance, 3 bits for offset (indicating the place of the block in 64bit full information), and 8 bits for block, are inscribed to identification field, and the total is exactly 16 bits. We reviewed and improved Basic PPM in Ref.[17].

Knowing that Basic PPM is computationally intensive and with high false positive rate, Song, *et al*.[8] propose two algorithms called advanced packet marking and authenticated packet marking scheme respectively. Both make use of Identification field for marking. We refer the advanced marking schemeIIwith m>7 in Ref.[8] as Advanced PPM.

In Advanced PPM, router marks one of several hashes of its IP address into packets each time instead of IP address itself. This method is both computationally efficient and scalable to highly distributed DDoS attack (in the case of 1 500 coordinated attackers, according to the authors). Because the marked information is hashes of IP addresses and hash function is one-way, this necessitates that the victim knows its upstream topology while deriving IP addresses from hashes. Since attack could come from anywhere on the Internet, this means that the victim should know virtually the topology of the whole Internet, which is difficult for all potential victims by themselves since the Internet itself is changing all the time. If all potential victims or their ISPs tries to keep updated topology information by themselves, they have to probe the Internet frequently, thus large volume of extra traffic would be caused.

3 On the Marking Probability

In the existing packet marking schemes[3,8], the marking probability p for any router is fixed and uniform. After a router marks a packet, the packet might be re-marked by downstream routers. If a router R is k hops away from the victim, the probability that the information it marked survives through the whole journey to the victim is $(1-p)^{k-1}$. So, while a packet arrives at the victim, the probability that it is marked with router R but nowhere thereafter is $p(1-p)^{k-1}$. The longer the distance k is the smaller the probability is. The link nearest to the attacker is the 'weakest' link[9]. This result is that many packets are needed for path reconstructing. Furthermore, about $(1-p)^d$ of all packets

will be intact for a journey of d hops. This leaves much room for spoofing.

In this section, we will give an Adaptive marking policy to reduce the number of packets needed in reconstructing the attack path. This makes sense for several reasons: (a) The fewer packets needed from the attacker, the quicker the reconstruction is and the more prompt the reaction could be; (b) Less room is left for attacker to spoof; (c) If the attacker attempts to evade from tracing, he must send fewer packets from each point. This weakens the attack force for the same number of bots or requires more bots for intended attack force.

Suppose that the attacker is d+1 hops away from the victim, such as in Figure 1. If all packets from the attacker are marked by some intermediate routers, there would be little room for the attacker to spoof. If packets are last marked by each router equally likely (that is with probability 1/d), the fewest packets are required in reconstructing the attack path. To achieve this, every router should mark each packet with a probability that is the reciprocal of the path length that the packet has traveled[18,19]. Let's denote the marking probability for router r_i with p_i, then p_i=1/i, (i=1,2,…,d).

$$Attacker \rightarrow r_1 \rightarrow r_2 \rightarrow \ldots \rightarrow r_d \rightarrow victim$$

Fig.1 An attack path

Nevertheless, routers have no way to get the parameter i, so it could not decide how much p_i is. We have developed an adaptive marking method[18,19] in which a router marks a packet with a probability q_j according to the distance j in the marking field of the packet. Here,

$$q_{-1}=1,\ q_0=1/2,\ q_1=1/6,\ q_2=1/10,\ q_3=q_4=...=0.04.$$

While a packet arrives at its edge router, since no router has marked it, the router should mark it with probability 1 to counter spoofing at the end system where the attacker locates, and we denote the probability in this case as q_{-1}.

If some router R_a under the attacker's control forges the distance field, the best it could do is to set the distance larger than 2. Figure 1 is about the average numbers of packets required in reconstructing paths of varying lengths for Advanced PPM with fixed probability p=0.04 and with adaptive probabilities. We could see that without router spoofing, using adaptive probability is much better than the fixed one in that fewer packets are needed for path reconstruction (about 64% in average). With some routers spoofing, more packets are needed while adaptive probability is employed (about 67% more in average). On a whole, the adaptive probability is more preferable than the fixed ones for the following reasons:

(1) Generally, routers are better secured than most of the end systems, so incidents of router compromise are much fewer than that of the end system compromise.

(2) Even if some routers are compromised by attacker, the attacker may use them in some more "important" and stealthy scenarios since they get them with hard work.

(3) Even if attacker summons routers in a DoS attack, compared with end systems available for the attacker, these routers are much fewer, thus, on the whole, the victim needs fewer packets to reconstruct the attack tree.

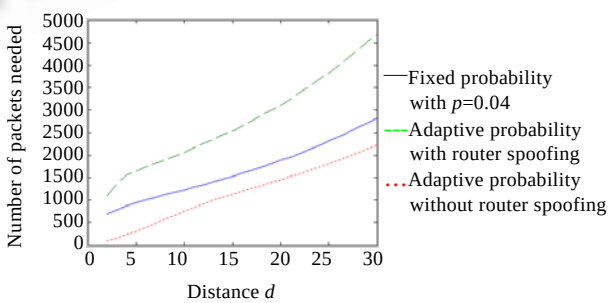

Fig.2 Number of packets needed for fixed and adaptive probability

4 Router Numbering Based Packet Marking

The idea behind our router numbering comes from the following two observations.

(1) The size of IP address is one of the key reasons that Basic PPM has high false positive rate and needs a large number of packets for the victim to reconstruct the attack path. There are 32 bits for an IPv4 address. While keeping an IP address and a 32bit error checking code into blocks of 8 bits, we need 8 blocks. This results in a large amount of combinations to check, hence many false positives. If we could reduce the number space needed to identify a router, we could reduce the number of packets needed and the number of combinations for path reconstruction. Reducing of combinations further reduces false positives.

(2) According to our analysis, the reason that the Advanced PPM has fewer false positives than the Basic PPM is essentially that the former makes use of the topology information of the Internet. We may also employ topology information to the same end.

Knowing that only very few routers are involved (compared with the whole Internet) in attacking and tracing and that the tracing operation is local (from one router to its neighbors). Fewer bits may work other than 32 bits in tracing. In this section, we propose some simplified numbering methods to identify routers. We also exploit the Identification field in IP header similar to Basic PPM and Advanced PPM. We also use 5 bits to store distance. So, 11 bits are left. We here propose 3 numbering methods.

4.1 Numbering scheme I

One way to reduce false positives is to reduce combinations at the victim. If a packet could convey full node information or full edge information, there would be no combinations needed in revealing a node. Since the tracing process is local and recursive, we could not get the node without edge information or other surrounding information. So to embed full node information is not a viable approach to our problem. In numbering scheme I, routers embed full edge information into a single packet, that is, embed full information of two adjacent nodes into a packet. In this way, only 5 bits could be used to number routers (only 11 bits are left for two nodes because 5 bits are reserved for distance), thus only 32 possible numbers are available. In the case of DDoS attack, it is very likely to get several nodes at the same distance with the same number. So the false positive rate would be high. Numbering schemeIis not a good solution.

4.2 Numbering scheme II

There is one bit left unused in numbering scheme I. We may exploit it to reduce false positives. Since one bit represents 2 statuses, we could number router with 10 bits and separate them into 2 halves, each with 5 bits. The marking field is shown in Fig.3. Each time a router marks a packet, it chooses one half randomly and embeds it into the '*Frag*0' field of the packet. It also sets the offset bit to indicate which half is chosen. If a router chooses not to mark the packet and the distance field is zero, it embeds one half of its number into the '*Frag*1' field according to the offset bit and then increments the distance. This method is also not good enough in terms of false positive as we will show later in Section 6.2, so we will not go further about this method.

Distance (5 bits)	*Offset* (1 bit)	*Frag*0 (5 bits)	*Frag*1 (5 bits)

Fig.3 Numbering scheme II based marking form

4.3 Numbering scheme III

In the former numbering schemes and those in Ref.[3,8], information about two adjacent nodes is embedded into a packet. In fact, we may also embed information about three adjacent nodes too. With 11bit space, we could embed 3 bits for each of the 3 nodes. 2 bits represent 4 statuses, so we could code routers with 3×4=12 bits. We

number each router on the Internet with a 12bit number randomly and equi-probably chosen from $\{0,1,\ldots,2^{12}-1\}$. Of cause, this causes collisions. Our purpose is to reduce false positives other than eradicate them, so, some collisions does not matter. At the victim, the marking field is shown in Fig.4.

Distance (5 bits)	*Offset* (2 bits)	*Frag*0 (3 bits)	*Frag*1 (3 bits)	*Frag*2 (3 bits)

Fig.4 Numbering scheme III based marking form

While a router marks a packet, it could embed one of the 4 blocks (3 bits each) of its number into the packet, and could also embed 2 or 3 blocks. In path reconstruction, the victim collects all the marked information with distance 0 which indicates that the information is marked by the very upstream routers of the victim. To get the number of every upstream router directly connected with the victim in the attack tree, the victim has to check many combinations. If he could extract only one block from a packet, it's likely that there would be some false positives while there are more than one router at the same distance in the REAL attack tree because there is no more information other than the topology to check the combination. The more the victim could extract from a packet, the fewer false positives. For example, suppose that R_1, R_2 and R_3 are upstream routers connected directly with the victim and that R_1 and R_2 are in the attack tree (Fig.5). R_1, R_2 and R_3 are numbered with 751, 1 309 and 733 respectively. These numbers could be divided into (1,3,5,7), (2,4,3,5), (1,3,3,5) respectively. If router embeds only one block a time while marking a packet, the victim will get four block sets {1,2}, {3,4}, {5,3} and {7,5} with distance 0. Then, router R_3 will be a false positive. If router embeds three blocks a time, R_3 won't be a false positive because there will be no packet containing triple blocks (1,3,3) with distance 0 and offset 0. So, we choose the method that a router marks 3 blocks to a packet if it chooses to mark it.

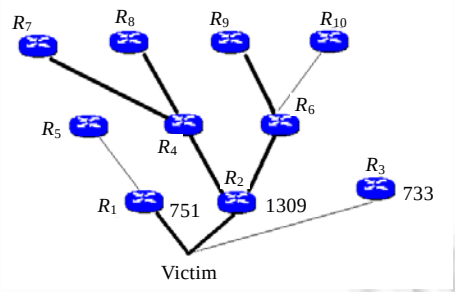

Fig.5 Attack tree, thick lines are parts of the attack paths while thin ones are not

We depict the marking algorithm and the path reconstruction algorithm based on Numbering scheme III in Figs.6 and 7.

```
Get (distance,offset,frag0,frag1,frag2) from the packet
If not the edge-router of the packet
{
    If distance=31
      Drop the packet
    Choose a random number α from [0,1]
    If α>q_distance
      {
        If distance=0
              frag1←block[offset]
        If distance=1
            frag2←block[offset]
           distance←distance+1
           }
      else
              call marking
}
else
        call marking

marking proc:
{
        Choose a random number β from [0,1]
        let j=int(4β)
        distance←0
        offset←j
        frag0←block[j]
        frag1←block[(j+1)mod4]
        frag2←block[(j+2)mod4]
      }
```

Fig.6 Marking scheme at router

```
Let T be the upstream topology graph of victim v
let G be a tree with root v
let set[0],set[1],set[2],…,be sets of tuples(distance,offset,
frag0,frag1,frag2) with distance 0,1,2,…respectively.
for each packet w from attackers
    Insert (distance,offset,frag0,frag1,frag2) from w to
    set[distance]
In set[0]
For all combinations {(0,0,frag00,frag01,frag02), (0,1,frag10,
frag11,frag12), (0,2,frag20,frag21,frag22), (0,3,frag30,
frag31,frag32)}
If  (
      frag00=frag22=frag31
      frag01=frag10=frag32
      frag02=frag11=frag20
      frag12=frag21=frag30
      )
And (frag00.frag01.frag02.frag21) is the number of some
node N in G connected directly with v
  Insert node N and the corresponding edge N–v into G.
In set[1]
For all combinations {(1,0,frag00,frag01,frag02), (1,1,frag10,
frag11,frag12), (1,2,frag20,frag21,frag22), (1,3,frag30,
frag31,frag32)}
If  (frag02=frag20
      frag12=frag30
      frag22=frag00
      frag32=frag10)
And (frag01.frag11.frag21.frag31) is the number of some
node N1 in G at distance1,
(frag00.frag10.frag20.frag30) is number of some node N2
which is a upstream router of and connected directly with
N1
  Insert node N2 and the corresponding edge N2–N1 into
  G.
For i from 2 to max-distance
{
  In set[i]
    For all combinations {(i,0,frag00,frag01,frag02),
    (i,1,frag10,frag11,frag12), (i,2,frag20,frag21,
    frag22), (i,3,frag30,frag31,frag32)}

      If (frag02.frag12.frag22.frag32) and(frag01.
      frag11.frag21.frag31) make up of an edge at
      distance i+1 in G and (frag00.frag10.frag20.
      frag30) is the number of some node N connected
      with the node (frag01.frag11.frag21.frag31)

          Insert node N with number (frag00.frag10.
          frag20.frag30) and the corresponding edge
          into G.
}
```

Fig.7 Attack path (tree) reconstruction procedure at victim

5 Topology Information Server

The Internet authority may number all routers connected to the Internet or these routers may be numbered with the hashes of their IP addresses. A topology information server may exist and host the topology information about the whole Internet. Topology information contains all routers, with their IP address, their tracing number, and their linking status among others needed. Several topology information servers may be scattered on the Internet for convenience. If the topology is changed at any local place, the change must be sent to all these servers for updating, or, these servers may probe the Internet to keep their information updated. Generally, victims may keep some local topology information. If the tracing procedure reaches some edge of which the victim has not the topology information out side, the victim fetches further information needed from some convenient server and then continues the tracing procedure.

6 Analysis

To evaluate the tracing methods, three dimensions must be considered: (1) Number of packets needed for path reconstruction; (2) False positive rate; (3) Workload in path reconstruction. The fewer these parameters are, the better the method is. In this section, we will compare the router numbering-based marking scheme with others according to these parameters. Before any evaluation, let's give some assumptions first. Suppose that:

(1) While reconstructing the attack tree, the victim gets k nodes at distance d (some may be false positives).

(2) The average number of upstream routers directly connected to a router on the Internet is M.

(3) There are K nodes at distance d+1 in the REAL attack tree (no false positive among the K nodes).

6.1 Number of packets needed

The same number of packets is needed to get attack paths of the same length for Basic PPM and Advanced

PPM[8]. We simulated Basic PPM and our method 10 000 times for each path length from 2 to 30. Figure 8 is about the average numbers of packets required in reconstructing paths of varying lengths for these methods.

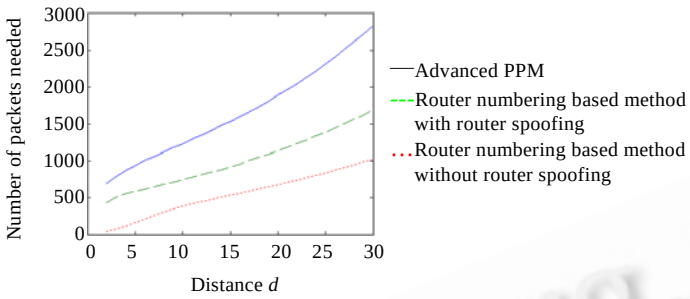

Fig.8 Packets needed for different methods

In our experiments, without spoofing of routers, namely, with all the routers following the rule, the number of packets needed for our method is about 30.39% of that for Basic PPM in average. Even with some routers spoofing, the number of packets needed for our method is about 60.5% of that for Basic PPM in average.

6.2 False positives

Before comparing our method with others, we first show that Numbering scheme II is inferior to Numbering scheme III in terms of false positive. For Numbering scheme II, suppose that the victim receives $k_0^{\#}$ and $k_1^{\#}$ different fragments with distance d and offset 0 and 1 respectively, then the number of false positives at distance* d+1 is about $2^{-20}kM\cdot(k_0^{\#}\cdot k_1^{\#}-K)$ in average. Let $V(N,n)$ denote the number of distinct values we get while sampling n times repeatedly, equi-probably, and with replacement from the population D={1,2,...,N}. Then $k_j^{\#}$ has the same distribution as $V(2^{10},k)$. The number of false positives for Numbering scheme III in the same situation would be $2^{-36}kM\cdot(\prod_{j=0}^{3}k_j^{*}-K)\approx 2^{-36}kM\cdot\prod_{j=0}^{3}k_j^{*}$ in average, with k_j^{*} has the same distribution as $V(2^9,k)$. Since

$$\frac{2^{-20}kM\cdot(\prod_{j=0}^{1}k_j^{\#}-K)}{2^{-36}kM\cdot(\prod_{j=0}^{3}k_j^{*}-K)}\approx\frac{2^{-20}kM\cdot\prod_{j=0}^{1}k_j^{\#}}{2^{-36}kM\cdot\prod_{j=0}^{3}k_j^{*}}=2^{16}\frac{\prod_{j=0}^{1}k_j^{\#}}{\prod_{j=0}^{3}k_j^{*}}.$$

Generally, this formula is larger than 1, That is to say, the numbering scheme II is inferior to numbering scheme III in terms of false positive rate.

The number of false positives for the Basic PPM in the same situation would be $2^{-32}k(\prod_{j=0}^{7}k_j-K)\approx 2^{-32}k\prod_{j=0}^{7}k_j$ in average, for Advanced PPM it is $2^{-64}(kM-K)\prod_{j=0}^{7}k_j'\approx 2^{-64}kK\prod_{j=0}^{7}k_j'$ in average, and for our method it is $2^{-36}kM\prod_{j=0}^{3}k_j^{*}$. Here** k_j and k_j' have the same distribution as $V(2^8,k)$ while k_j^{*} has the same distribution as $V(2^9,k)$. Obviously $M<2^{32}$, so the number of false positives for Basic PPM is larger than that for Advanced PPM. The ratio between the number of false positives for Advanced PPM and that for our method is

$$ratio=\frac{2^{-64}kM\prod_{j=0}^{7}k_j'}{2^{-36}kM\prod_{j=0}^{3}k_j^{*}}\approx\frac{2^{-28}\prod_{j=0}^{7}k_j'}{\prod_{j=0}^{3}k_j^{*}}.$$

While k is small, the false positives for both Advanced PPM and our method are few and that for our method

* At the victim, if the distance field in a packet is k, the router that marks the packet is k+1 hops away.

** For simplicity, we suppose that IP addresses are randomly distributed.

are a little more than that for Advanced PPM. With k grows, the false positives for our method grow slower than that for Advanced PPM, and finally, the false positives for our method will be fewer than that for Advanced PPM. The larger the k is, the more superior our method is compared with Advanced PPM in terms of false positives. Figure 9 depicts the ratio as a function of k.

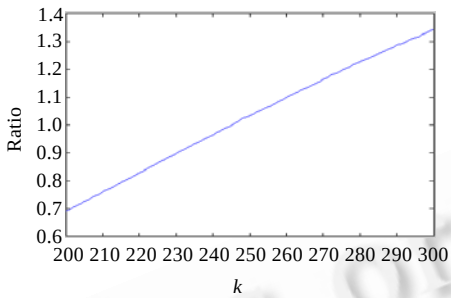

Fig.9 The false positive ratio between advanced PPM and router numbering based packet marking

6.3 Workload

In Advanced PPM, if the victim gets k and k' nodes at distance d and d+1 respectively (some are false positives), the victim must have done more than $(kM–k')+8k'=kM+7k'$ hash operations in average while getting the k' nodes. In our method, the major operations in path reconstruction are combination and comparison, and both are much more efficient than hash and with limited numbers. It's safe to say that the workload for the victim in our method is much less than that in Advanced PPM. Since workload of Advanced PPM is much smaller than that of Basic PPM, the workload of our method is much smaller than that of Basic PPM too.

7 Conclusion

In this paper, a router numbering based packet marking scheme for traceback is given, which is superior to the previous ones in that it needs fewer packets and less workload in path reconstruction and has fewer false positives which result in a better scalability. This method also limits attackers' room for trace message spoofing. Furthermore, the Adaptive marking policy developed in this paper could be used to enhance other similar marking methods.

References:

[1] CERT. CERT statistics. http://www.cert.org/stats/#incidents

[2] Park K, Lee H. A proactive approach to distributed DoS attack prevention using route-based packet filtering. Technical Report, CSD00-017, Purdue University, 2000.

[3] Savage S, Wetherall D, Karlin A, Anderson T. Practical network support for IP traceback. In: Proc. of the 2000 ACM SIGCOMM Conf. Stockholm, 2000. 295−306.

[4] McGuire D, Krebs B. Attack on Internet called largest ever. 2002. http://www.Washingtonpost.com/ac2/wp-dyn/A828-2002Oct22?

[5] Lemos R. Attack targets .info domain system. ZDNet News, 2002. http://news.zdnet.co.uk/internet/0,39020369,2126521,00.htm

[6] CERT. Overview of attack trends, 2002. http://www.cert.org/archive/pdf/attack_trends.pdf

[7] Ferguson P, Senie D. Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing. rfc2827, 2000.

[8] Song DX, Perrig A. Advanced and authenticated marking schemes for IP traceback. In: Proc. of the IEEE INFOCOM 2001. 2001.

[9] Park K, Lee H. On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack. In: Proc. of the IEEE INFOCOM 2001. 2001. 338–347.

[10] Snoeren AC, *et al*. Hash-Based IP traceback. In: Proc. of the ACM SIGCOMM 2001. 2001. 3–14.

[11] Burch H, Cheswic B. Tracing anonymous packets to their approximate source. Usenix LISA, 2000. 313−321.

[12] CISCO. Characterizing and tracing packet floods using Cisco routers. http://www.cisco.com/warp/public/707/22.html

[13] Bellovin S, Leech M, Taylor T. ICMP traceback messages. Work in Progress, Internet Draft, draft-ietf-itrace-02.txt, 2001.

[14] Stone R. Centertrack: An IP overlay network for tracking DoS floods. In: Proc. of the 9th USENIX Security Symp. 2000.

[15] Stoica I, Zhang H. Providing guaranteed services without per flow management. In: Proc. of the '99 ACM SIGCOMM Conf. Boston, 1999. 81−94.

[16] Houle KJ, Weaver GM, Long N, Thomas R. Trends in Denial of Service Attack Technology. CERT® Coordination Center. 2002

[17] Li DQ, Su PR, Feng DG. Notes on packet marking for IP traceback. Journal of Software, 2004,15(2):250−258 (in English with Chinese abstract). http://www.jos.org.cn/1000-9825/15/250.htm

[18] Li DQ, Xu YD, Su PR, Feng DG. Adaptive packet marking for IP traceback. Acta Electronica Sinica, 2004,32(8):1334−1337 (in Chinese with English abstract).

[19] Li DQ. Denial of Service Attack. Beijing: Publishing House of Electronic Industry, 2007 (in Chinese).

附中文参考文献:

[17] 李德全,苏璞睿,冯登国.用于 IP 跟踪的包标记的注记.软件学报,2004,15(2):250−258. http://www.jos.org.cn/1000-9825/15/250.htm

[18] 李德全,徐一丁,苏璞睿,冯登国.IP 追踪中的自适应包标记.电子学报,2004,32(8):1334−1337.

[19] 李德全.拒绝服务攻击.北京:电子工业出版社,2007.

LI De-Quan was born in 1969. He is with the State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences. He got a Ph.D. degree from the graduate school of Chinese Academy of Sciences in computer science. His research area is network security.

SU Pu-Rui was born in 1976. He is a Ph.D. candidate at the Institute of Software, the Chinese Academy of Sciences. His research area is network security.

WEI Dong-Mei was born in 1974. She holds a master degree in electronics. She is a lecturer in College of Information & Control Engineering, Southwest University of Science and Technology. Her research area is information security.

FENG Deng-Guo was born in 1965. He is a professor and doctoral supervisor at the Institute of Software, the Chinese Academy of Sciences and a CCF senior member. His current research areas are information security and network security.