Step-tracking Algorithm of DDoS Attacks Based on Advanced Marking Scheme

YANG Xueqin[1, a], YANG Xuehui[2,b] and CHEN Chaobo[3,c]

[1]School of Telecommunication and Information Engineering Xi'an University of Posts and Telecommunications, No.563 Chang'an South Road, Xi'an, Shaanxi, P.R. China

[2]Shandong Institute of Aerospace Electronics Technology, Yantai, Shandong, P.R. China

[3]Xi'an Technological University, No.2 Xuefuzhong Avenue Xi'an, Shaanxi Province, P.R. China

[a]exceeding2@126.com, [b]exceeding3@126.com, [c]choby@xatu.edu.cn

Keywords: AMS algorithm; DDoS; the attack sources; AS.

Abstract. In this paper, a new scheme is presented. This scheme is based on Advanced Marking Scheme (AMS), and it works with Autonomous System (AS). In autonomous systems, routers are divided into two categories-borders- routers and internal routers. It can reduce the number of routers involved in marking and the number of marked packets, further improves the efficiency and reliability of the reconstruction of the path by different types of routers using different markers. Simulation results from the experiments show that obtained results are the same as theoretical results. Performance analysis and simulation experimental results shows, compared with the conventional methods, the new algorithm improves the performance of IP Traceback technique obviously.

Introduction

Distributed Denial of Service attacks (DDoS) have become major Internet attacks because of its destructiveness and difficult to trace back. IP Traceback is one main means of preventing such attacks. In its development, there are a variety of techniques. IP marking approach is one promising solution, because it has a low overhead for routers and the network and supports incremental deployment. Representatives of packet marking algorithm is the Advanced Marking Scheme (AMS) [1] proposed by Dawn Xiaodong Song and Adrian Perrig. Our approach is based on AMS and cooperates with Autonomous System (AS)[2] so that it has lower network and router overhead than AMS, yet our approach are much more efficient and accurate for the attacker path reconstruction.

Step-tracking Algorithm for DDoS Attack Sources

In this section, we describe our Step-tracking Scheme, in which we use new encoding schemes that are efficient and accurate even for DDoS attacks originating from over 1000 simultaneous attackers. This scheme is based on Advanced Marking Scheme (AMS), AMS algorithm is an edge marking algorithm, and it records each IP address' hash value instead of IP address of a router in order to reduce storage space. For our marking schemes, we assume the victim has a map of its upstream routers. We denote a directed acyclic graphs Gm. After dividing slice and recombination, through comparing Hash value and IP Hash value of address of router of network to reconstruct the attack path. Our scheme also is divided into two parts: marking procedure and reconstruction procedure. In this article, the marking procedure based on the AMS algorithm has been improved slightly.

Marking procedure. There is 16-bit identification field which is seldom used (only 0.25% packets is divided into fragments)[3,4,5], so that we can use this field to storage marking information. In this scheme, we divided the 16-bit IP Identification field into a 5-bit distance field and an 11-w bit edge field and w-bit FlagID field (Note that 5-bit can represent 32 hops which is sufficient for almost all Internet paths). We can encode the IP address of a router into 11-w bit, and the XOR of two neighbors to storage them in the field of 11-w bit field. There is also 3-bit flag field and only the first two bits are used, then we can use the third bit as label field. If the value of this label field is 0 which means this packet is marked by internal router(s)[6] and when a border router wants to mark a packet, it needs to

set the label field to 1.If the value of this label field is 1 which means this packet only can be marked by border routers and internal routers are not permitted to mark it. In order to reduce colliding, reduce false positive rate, we choose 2^w hash function. The encoding of IP head shows as Fig. 1.

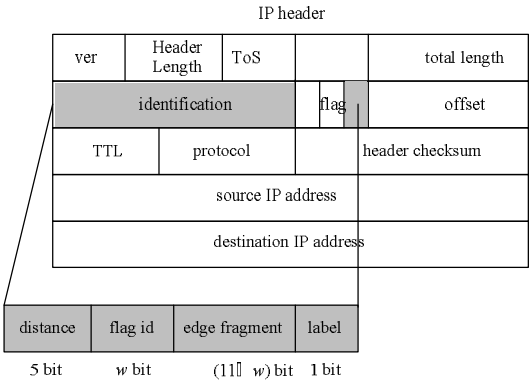

Fig. 1 Encoding into the IP header

Each router marks a packet with a probability q when forwarding the packet. If a router R_i decides to mark the packet P, it chooses a random number x of w bits and writes it into the FlagID field and use $g(<x,R_i>)$as its IP address encoding and writes it into Edge field, writes 0 into the distance field in packet P. Otherwise, if the distance field is 0 which implies its previous router has marked the packet, it chooses a random number y of w-bits and writes it into the FlagID field and XORs $g'(<y,R_i>)$with the edge field value and overwrites the edge field with the result of the XOR. At the same time if the router is an internal router which will set the label field to 0 or a border router will set the label field to 1.The router always increases the value of distance field if it decides not to mark the packet. The XOR of two neighbouring routers encode the edge between the two routers of the upstream router map. The edge field of the marking will contain the XOR result of two neighbouring routers, except for samples from routers one hop away from the victim. But when the value of label field is 1, the edge field of the marking will contain the XOR result of two neighbouring routers. Because $a \oplus b \oplus a = b$, we could start from markings from the routers one hop away from the victim, and then hop-by-hop, decode the previous routers. The reason to use two independent hash functions is to distinguish the order of the two routers in the XOR result.

Marking procedure at route R_i:
for each packet P
If(P.label==0)
{ if this router is Internal Router
Call Marking procedure
Else this router is Area Border Router
{Call Marking procedure
P.label ß 1}}
Else(P.label==1)
{ if this router is Area Border Router
Call Marking procedure
Else do nothing}
Marking procedure at route R_i:
let u be a random number from[0,1]
if($u \le q$) then
let x be a random number from[0,7)
P.fID ß x
P.distance ß 0
P.edge ß $g(<l,R_i>)$
Else
if(P.distance==0)then
P.edge ß P.edge $\oplus g'(<P.fID,R_i>)$

P.distance ß P.distance+1

Reconstruction procedure. In this procedure, the marked packets are divided into two types: Label =0(the value of label field of this kind of packets is 0) and Label=1 (the value of label field of this kind of packets is 1).We can use the Label=1 to trace back the autonomous systems which attackers belong to and use the Label=0 to determine the exact location of attackers.

Two-dimensional threshold reconstruction algorithm[7] is implemented in the reconstruction procedure of AMS algorithm. For a $m_{f,d}$ - threshold scheme, a node u in G_m will only be considered as on an attack path if more than $m_{f,d}$ of its hash values from the 2^w hash functions match the right markings in the attack packets.

Reconstruction procedure is completed at the victim. Intuitively, to reconstruct the attack paths, the victim uses the upstream router map G_m as a road-map and performs a breadth-first search from the root. Let's denote the set of edge fields marked with a distance, FlagID l($l\in(0,2^w-1)$ as$\psi_{d,l}$ (do not include duplicates).At distance 0, the victim enumerates all the routers one hop away from itself in G_m and regarding the value of l($l\in(0,2^w-1)$ checks which routers have the hash value of their IP addresses, $g(<x, R_i>)$, matched with the edge fields in$\psi_{0,l}$, and denotes the set of matched IP addresses as. Therefore S_0 is the set of routers one hop away from the victim in the reconstructed attack graph. S_d denotes the set of routers at distance d to the victim in the reconstructed attack graph. Then for each edge x in$\psi_{d+1,l}$, and for each element y in S_d, the victim computes $z=x\oplus g(<l,y>)$.For all values of l the victim then checks whether any child R_i of y in G_m has the hash value of its IP address,$g(<l,R_i>)$,equal to z. If the victim finds a matched IP address R_u, then it adds R_u to the set S_{d+1} (initially S_{d+1} is empty). The victim repeats the steps until it reaches the maximal distance marked in the packets, denoted as $maxd$. Thus, the victim reconstructs the attack graph.

Reconstruction procedure at victim v:

```
D=0
for each packet p whose P.label==1
{Call Reconstruction procedure
If Distance>D
D=Distance}
for each packet p whoseP.label==0
{if Distance >=D
Call Reconstruction procedure}
```

Reconstruction procedure at victim v :

let S_d be empty for $0\le d\le maxd$

```
for each child R of v in G_m
let count=0
  for l:=0 to 2^w-1
    if g(<l,R>)∈ψ_{0,l} then
      count=count+1
  if count>m_f, 0 then
    insert R into S_0
for d: =0 to maxd-1
  for each y in S_d
      for each child u of y in G_m
      let count=0
      for l: =0 to2^w-1
```

for each x in $\psi_{d+1,l}$

$z=x\oplus g(<l,y>)$

if g(<l,u>)=z then

count=count+1 ; break

if count>$m_{f,d}$ then

insert u into S_d+1

output S_d for 0≤d≤$maxd$

Algorithm performance analysis and simulation

Performance Analysis. Compared with the AMS algorithm, victim host will receive two types of marked packets: Label=0 this kind of packets is marked by internal routers; and Label=1 which is marked by border routers. We all know that the number of border routers is only a small fraction of the number of all routers in the whole attack path, so that the number of routers involved in marking of our scheme is much less than AMS algorithm, which can greatly reduce not only the workload of each router but also the marked packets. For the victims, they can use Label=1 packet to trace back the autonomous system which attackers belong to and then determine the exact position by Label=0 packets, which can shorten the path reconstruction time, improve the tracking efficiency.

Simulation Result. To verify the feasibility of the step-tracking scheme for DDoS attack sources, we build an experimental environment to simulate attacks using programs written with C++ language. We assume that there are six autonomous systems in the network and there are six routers in every autonomous system including border routers and internal routers, topology shown in Fig.2.In the same circumstances, compare the simulation results of the two schemes, AMS algorithm and the step-tracking scheme. During the simulation, the marking probability q is equal to 4%, the w is equal to 3 and m is equal to 6.In all the tests, we use only one victim. We simulate the process that routers mark the attack packets, and the process that the victim reconstructs the attack graph using the markings in the packets.

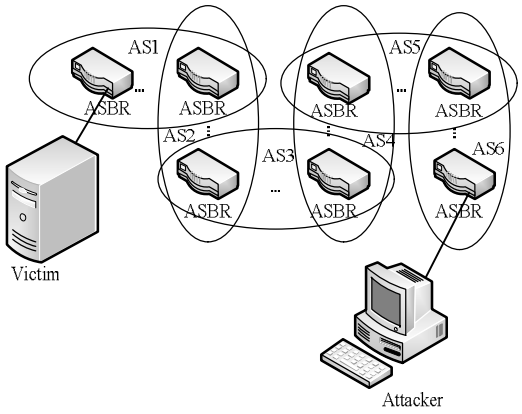

Fig. 2 network topology

Firstly, we test the number of packets required to reconstruct the attack graph. Fig.3 shows the simulation results of the number of packets required reconstructing the paths of various lengths with 95% probability f in presence of only one attacker or the Advance Marking Scheme and our step-tracking scheme. Each point sends attack packets at a certain distance from the victim.

It can be seen from the Fig.3 that under the same marking situation the step-tracking scheme requires greatly fewer packets for the reconstruction especially for long distance.

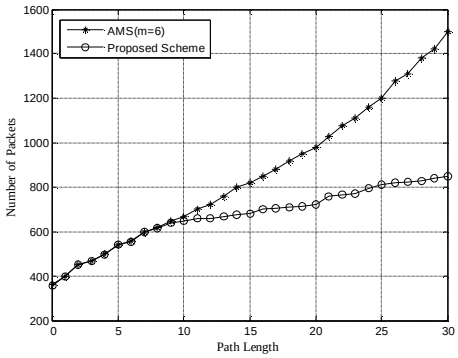

Fig. 3 number of packets required for reconstruction (q=4%)

We also test the number of false positives of the Advanced Marking Scheme and step-tracking scheme. The network topology consists of routers that are 4-29 hops away from the victim, and there are 60 attackers which assure more than 95% accuracy of reconstruction. The simulation results are shown in Fig.4.

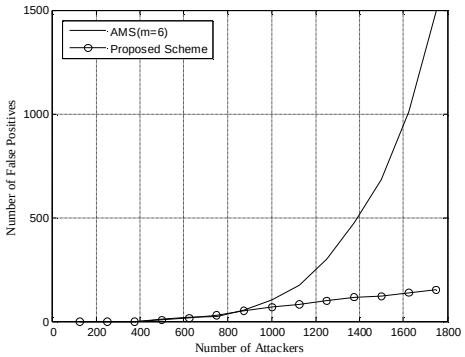

Fig. 4 False Positives

We can also find that under the same situation the number of the false positives of the step-tracking scheme is less than the Advanced Marking Scheme. We know that both marking packets and reconstructing the attack path consume network resources. Compared with all routers mark the packet with a probability, a part of routers can reduce a lot of unnecessary overhead and improve the efficiency of scheme.

Conclusions

The step-tracking scheme presented in this paper reduces overhead, improves the stability and accuracy of reconstruction. In contrast to Advanced Marking Scheme, our algorithm has significantly higher precision (lower false positive rate) and lower complexity in reconstructing the attack paths under large scale distributed Denial-of-Services attacks.

Acknowledgements

This work was financially supported by the Shaanxi Province Industry Science and technology Plan Project Fund (15K06-42) and Xi'an University of Posts and Telecommunications Principal Fund (205010321).

References

[1] D.X. Song, A. Perrig. Advanced and Authenticated Marking Schemes for IP Traceback[C]//INFOCOM 2001: Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol.2 (2001), p. 878

[2] D.Magoni, J Pansiot. Analysis of the autonomous system network topology: ACM Computer Communication Review, Vol.31 (2001), p. 26

[3] R.L. Carter and M.E. Crovella. Server Selection Ssing Dynamic Path Characterization in Wide-area Networks[C]//INFOCOM'97: Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution., Proceedings IEEE, Vol.3 (1997), p. 1014
[4] W. Theilmann, K. Rothermel. Dynamic distance maps of the Internet[C]//INFOCOM 2000: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol.1 (2000), p. 275
[5] Cooperative Association for Internet Data analysis on http://www.caida.org
[6] A. Durresi, V. Paruchuri and L.Barolli. Fast Autonomous System Traceback: Journal of Network and Computer Applications, Vol.32 (2009), p. 448
[7] X. Yang, C. Pei, C. Zhu and Y. Li. AMS Based Reconstruction Algorithm with Two-dimensional Threshold for IP Traceback[C]//Parallel and Distributed Computing: Applications and Technologies, 2005. PDCAT 2005. Sixth International Conference on. IEEE, (2005), p. 781