IPv6 marked packets preprocessing scheme research based on Hash table

Ming-zhen Li, Fu-gui Luo,Jing Wei
Computer and information science department
Hechi University
Yizhou, China
e-mail: 583185636@qq.com

***Abstract*-Packet marking technology is one of IP traceback technology, which contains two parts of packet marking and path reconstruction. The efficiency of entire algorithm depends on both packet marking and path reconstruction. At present, the research on packet marking technology is mainly about packet marking algorithm, but is seldom about path reconstruction. It is impossible that improving the efficiency of the entire algorithm only by improving packet marking. Aiming at this problem, a scheme is proposed, which deals with marked packets before path reconstruction, using improved hash table to organize marked packets is proposed and making marked packets stored orderly, and then decreasing the time that spends on searching a specific marked packet.**

Keywords-Network Security; IPv6; packet mark-ing; path reconstruction; hash table

I. INTRODUCTION

Packet marking technology is one of IP traceback technology which stores routing information in some filed. When the victim detects attack, a traceback system will be started which utilizes the marked routing information to reconstruct attack path, and then some measures will be adopted to protect the host from being damaged.

Packet marking technology contains two parts that are packet marking and path reconstruction. The efficiency of the entire algorithm depends on the two parts. At present, almost packet marking technology researchers are about packet marking to improve the algorithm efficiency of packet marking technology by increasing the amount of marking information, and the researchers achieve rather great progress. There are several improved IPv4 and IPv6 packet marking algorithms. In IPv4, those are PPM[[1]], advanced and authenticated marking schemes for IP traceback[2], and adjusted probabilistic packet marking for IP traceback[3]. In IPv6, those are improved SPIE scheme which aims at tracing single packet in IPv6[4,5], an improved algorithm choosing the extension header Hop-by-Hop as marking area[6,7], etc. However, there is seldom research about path reconstruction, and only a random model about the quantitative relation between convergence time and packet marking probability [8] was improved, which provides a theoretical basis for further packet marking research except concrete proposal aiming at path reconstruction.

The limitation of IP header decides that the space of further improving the algorithm efficiency of packet is limitation. Thus, marked packets preprocessing scheme is proposed, which drops incomplete and repetitive packets before path reconstruction executed and makes marked packets orderly stored to reduce the time complexity of searching a specific marked packet during path reconstruction. The shape of attack path is inverted-tree, according to this characteristic, so the scheme uses improved hash table subchain to store and search a specific marked packet in order to achieve the goal that queries fast a specific marked packet.

II. PACKET MARKING TECHNOLOGY

A. Packet Marking Maintaining the Integrity of the Specifications

The core of packet marking algorithm is to improve the amount of marking information and reduce the amount of marked packets needed in path reconstruction and improve the accuracy rate of path reconstructioin as the starting point to choose appropriate marking area and marking probability and encoding method.

IPv4 address space will be exhausted in the near future, thus the application of IPv6 is an irresistible trend and will eventually replace IPv4. The emphasis of research should place on IPv6, so we choose an example of IPv6 packet marking algorithm (V6PPM) [7] to illustrate the proposed scheme in this paper. As shown in TABLE 1, V6PPM chooses extension header Hop-by-Hop as marking area, and with a probability r router marks packets that haven't been marked. If a packet has been marked, router marks the marked packet with a probability *r*/*stt*l. The information marked

*funded by Hechi University Youth Yopic :Based on Attack Source Traceback the Security Technology Research of Internet of things (2013A-N002)

into marking area is about an *edge* <*distance*, *ESAddr*, *EEAddr*>. *ESAddr* and *EEAddr* are the node IPv6 addresses of two adjacent routers and every IPv6 address (128bits) is directly marked into marking area without being compressed.

The amount of V6PPM marking information is very large about two IPv6 addresses; it reduces the probability of marking information being covered by determining the original marking probability and revising the marking probability r if a packet has been marked; encoding method is simple without route information being compressed. The overall performance of V6PPM is rather good.

TABLE I. TABLE TYPE STYLES

next header	*Exten header length*	*pad N option*	*opt data length =0*	*opt type=X*	*opt data length =34*	*sttl*	*distance*
ESAddr（128bit）							
EEAddr（128bit）							

B. Path Reconstruction

1) DDoS attack path[1]

As shown in figure 1, the attack path from victim to attacker is invert-tree structure. Victims are represented by A_i (i=1, 2, 3), routers are represented by R_j (j=1, 2, … ,7), and victim is represented by V. Dotted line is one of attack path.

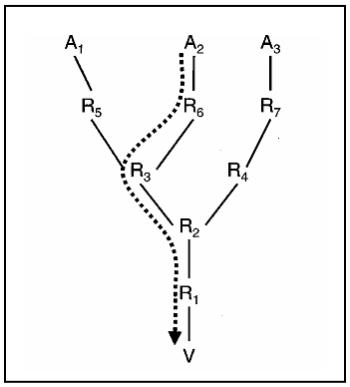

Figure 1. Attack path.

2) Path reconstruction[5]

Every marked packet contains an edge constituted by two IPv6 addresses. All marked information constitutes an edge set *E*. Path construction generates attack path topology which is the attack tree *T*. The specific process is as follows:

Step 1: *distance*=0, victim V is the root node of attack tree *T*;

Step 2: *distance*=1, find out all *edges* <1, R', V> from E that *EEAddr* is V. R' represents child node of V;

Step 3: *distance*=2, find out all *edges* <1, R'', R'> from E that *EEAddr* is R'. R'' represents child node of R';

Step 4: Repeat the process until *distance=d*, and find out all *edges* <1, R^d, R^{d-1}> from *E* that *EEAddr* is R^{d-1}. R^d is the leaf node of attack tree *T*.

Step 5: Get attack tree *T*, *T* is said all attack paths.

III. MARKED PACKETS PREPROCESSING

A. HashTable

The basic thought of hash table is establishing mapping between key key and the memory address of every element. We can determine memory address through hash value h(key). The time complexity of the insert, delete, and query for elements is O(1).

Hash value unavoidable has conflict, and the zipper method is one of the better in the existing conflict resolution methods. Hash table is composed of two parts, a basic region and a synonym child table. A synonym child table uses a linked list to represent. For example, Figure 1 corresponding hash table is s shown as figure 2. *distance* is represented by basic region and *edge* by subchain. The time complexity of the insert, delete, and query for elements of subchain is O(n).

B. Improved Hash Table Subchain Bases on Balanced Binary Tree (AVL tree)

Although hash value directly is mapped to element memory address, the time complexity of query achieves O(n) when the conflict of subchain is too much. The efficiency is rather low and the superiority of hash table doesn't be displayed. AVL tree realizes the rapid query and insert and delete operation by structuring an ordered tree. The time complexity is $O(\log_2 n)$. A particular structure of basic region being sequential storage structure and subchain being AVL tree is adopted. The improved hash table owns two superiorities of determining memory address through hash value h(key) and random access for sequential storage structure. At the same time, improved subchain own higher algorithm efficiency. Figure 1 corresponding improved hash table is s shown as figure 2.

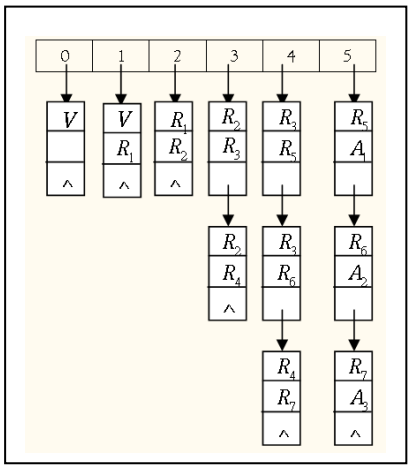

Figure 2. Hash table.

C. Marked Packets Preprocessing.

The packets that victim has received include common packets of not being marked and marked packets. The marked packets include marking information incomplete or repeated marked packets. Those received packets are so huge and desultory that the time complexity of querying a specific packet increases, moreover, the path reconstruction efficiency is been greatly reduced. In order to solve this problem, a scheme of marked packets preprocessing is proposed, which removes invalid data packets and adopts improved hash table to store valid marked packets to make marked packets well-organized. The specific marked packets preprocessing process and path recon-struction process scheme of is as follows:

Marked packets preprocessing process:

Step 1: If *ESAddr* and *EEAddr* are both null, the packet is common and don't do any process;

Step 2: If *ESAddr* is null, drop the packet. If *EESddr* is not null, check up the value of distance. If *distance*=1, the packet is valid and store it to hash table as Step 3; If *distance*>1, The packet is invalid and drop it;

Step 3: If *ESAddr* and *EEAddr* are both non-null, compared the marking information with the information of hash table having stored. If they are different, insert the marking information into hash table subchain; if not, don't do any process;

Step 4: Repeat the steps 1 to step 3 until form a stable hash table.

Path reconstruction process (choose the attack path of dot line representing in figure 1 as an example):

Step 1: Set V as the root node of attack tree and query hash table. When *distance*=1, query the subchain of the element of hash value h(key)=1 and add *edge* <1, R_1, V> into set *T*;

Step 2: When *distance*=2, query the subchain of the element of hash value h(key)=2 and add *edge* <1, R_2, R_1> into set *T*;

Step 3: Repeat the steps 1 to step 3 until *distance*= *sttl*=5. Get set *T*

$$V \rightarrow R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_6 \rightarrow A_2.$$

Set *T* is the reconstructed attack path set.

D. Extending

The format of IPv6 routing information stored in hash table which is created on marked packets preprocessing stage still can be applied to IPv4 marking algorithm. We only need to do corresponding change about marking information format aiming at the structure of subchain node.

The proposed scheme is universal applicability and can apply to any packet marking algorithm in IPv4 and IPv6 network environment.

IV. Analysis of Experimental Results

A. Algorithm Analysis

The purpose of algorithm improvement is dropping the invalid packets, making marked packets be orderly stored, and then reducing the time of querying a specific packet. The victim needs to maintain a hash table. The size of hash table is related to the length of attack path. Generally speaking, $d \le 16$ [9] hop and the number of adjacent node for every node is 3.15 [10] (here *d*=4). So the size of hash table is

$$S(d,s) = s \times \sum_{i=0}^{d-1} 3^i .$$

d represents the length of attack path, *i* represents the hops between the node marking packets and victim. When the number of nodes achieve $\sum_{i=0}^{15} 3^i$, the hash table tends towards stability. The size of hash table is

$$S(d,s) \le s \times \sum_{i=0}^{15} 3^i .$$

On path reconstruction stage, only need to query the subchain of h(key) corresponding. The time complexity for query is $O(\log_2 n)$. However, the time complexity is O(n) before algorithm improvement.

B. Analysis of Experimental Results

To prove the performance of the proposed scheme in this paper, we use NS2 network simulator to experiment in the Linux system. The experiment respectively realizes the scheme of packets preprocessing and having no preprocessing before path construction. The value for the length of attack path *d* is from 1 to 16, and $d \le 16$. The quantity relation between P which represents the number of

packets needed to query in order to query a specific packet and *d* which is the length of attack path is as shown in figure 3.

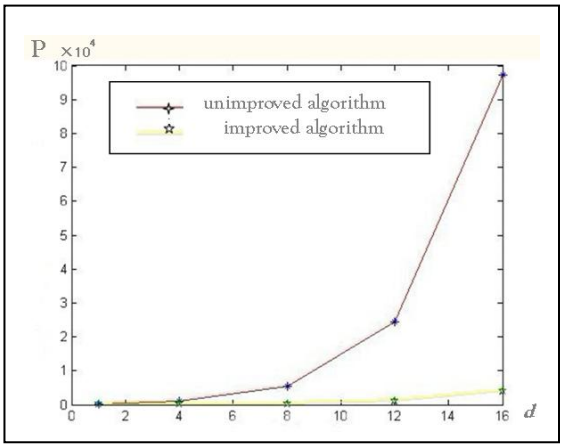

Figure 3. The quantity relation between P which represents the number of packets needed to query in order to query a specific packet and *d* which is the length of attack path

With the grown of *d*, P is rapidly growth in no marked packets preprocessing scheme. However, it is relatively stable in the proposed scheme. Experimental Results show that the improved algorithm can effectively reduce the number of packets querying a specific packet.

V. CONCLUSIONS

The proposed scheme preprocesses the marked packets before path renconstruction, makes valid marked packets orderly stored, and thus improves the algorithm efficiency of path reconstruction. The experiment proves that the proposed scheme has obvious superiority and wide range of application. However, it is difficult to maintain a real-time and accurate because it will expend more memory and CPU resource. And the algorithm improvement is only about packets preprocessing and isn't real path reconstruction algorithm improvement. In future, the research emphasis is real path reconstruction algorithm improvement and how to combine marked packets preprocessing with path reconstruction algorithm improvement to improve the entire stage algorithm efficiency.

REFERENCES

[1] SAVAGE S, WETHERALL D, KARLIN A, et a1. Practical network support for IP traceback. [C]. Proceedings of the 2000 ACM SIGCOMM Conference. New York, USA: ACM Press, 2000:295-306.

[2] SONG D, PERRIG A. Advanced and authenticated marking schemes for IP traceback [C] Proc. of IEEE INFOCOM 2001. Alaska, USA:IEEE Press, 2001: 878-886.

[3] PENG T, LECKI C, RAMAMOHANROA K. Adjusted probabilistic packet marking for IP traceback [C]. Proceedings of Networking 2002. Pisa, Italy: IFIP Press, May 2002: 697-708.

[4] TIMOTHY STRAYER W, et al. SPIE-IPv6: Single IPv6 Packet Traceback, Local Computer Networks, 2004. 29th Annual IEEE International Conference on 16-18 Nov. 2004.

[5] ZHAN Yong-jun , XIE Dong-Qing , ZHOU Zai-hong. An IP Traceback Scheme Based on the Improved SPIE in IPv6[J]. Computer Engineering & Science, 2007,29(4):11-13.

[6] OBAID CHOONG S, SEON HONG C. On IPv6 Traceback[C]. Proc. of ICACT2006, Phoenix Park, Republic of Korea, Feb. 20-22, 2006: 2139- 2143.

[7] YANG Jun, WANG Zhen-xing, GUO Hao-ran. IPv6 Attack Source Traceback Scheme Based on Extension Header Probabilistic Marking.[J]. Application Research of Computers, 2010, 27(6):2335-2340.

[8] WANG Xiao-jing, XIAO You-lin, WEI Shen-jun. A Stochastic Model of Attack Path ReconstructionProblem for IP Traceback.[J]. Transactions of Beijing Institute of Technology, 2011, 31(2):168-172.

[9] W. Theilmann and K. Rothermel. Dynamic distance maps of the Internet. In In Proceedings of the 19th Annua lConference of IEEE Communications and Computer Societies (INFOCOM 2000), pages: 275-285. IEEE Communications and Computer Societies, March 2000.

[10] C. R. Palmer, G. Siganos, M. Faloutsos, C. Faloutsos, and P. B. Gibbons. The connectivity and fault-tolerance of the Internet topology. The 2001 Workshop on Network-Related Data Management; in cooperation with ACM Special Interest Group on Management of Data/Principles of Database Systems, May 2001.