An IP Trace back System to Find the Real Source of Attacks

A.Parvathi and G.L.N.JayaPradha
M.Tech Student,Narasaraopeta Engg College, Narasaraopeta,Guntur(Dt),A.P.
Asso.Prof & HOD,Dept of I.T, ,Narasaraopeta Engg College, Guntur(Dt),A.P.

Abstract—Internet Protocol (IP) traceback is the enabling technology to control Internet crime. In this paper, we present a novel and practical IP traceback system called Flexible Deterministic Packet Marking (FDPM) which provides a defense system with the ability to find out the real sources of attacking packets that traverse through the network. While a number of other traceback schemes exist, FDPM provides innovative features to trace the source of IP packets and can obtain better tracing capability than others. In particular, FDPM adopts a flexible mark length strategy to make it compatible to different network environments; it also adaptively changes its marking rate according to the load of the participating router by a flexible flow-based marking scheme. Evaluations on both simulation and real system implementation demonstrate that FDPM requires a moderately small number of packets to complete the traceback process; add little additional load to routers and can trace a large number of sources in one traceback process with low false positive rates. The built-in overload prevention mechanism makes this system capable of achieving a satisfactory traceback result even when the router is heavily loaded. The motivation of this traceback system is from DDoS defense. It has been used to not only trace DDoS attacking packets but also enhance filtering attacking traffic. It has a wide array of applications for other security systems.

Keywords—DDoS attacks, IP trace back, performance evaluation, routers, security.

1 INTRODUCTION

Nowadays more and more critical infrastructures are increasingly reliant upon the internet operators Given the widespread use of automated attack tools, attacks against Internet-connected systems are now so common-place that Internet crime has become a ubiquitous phenomenon. Although a number of countermeasures and legislations against Internet crime have been proposed and developed, Internet crime is still on the rise. One critical reason is that researchers and law enforcement agencies still cannot answer a simple question easily: who or where is the real source of Internet attacks? Unless this question is fully addressed, effective defense systems and legislations against such crime would only be blusterous ornaments because knowing where the DDoS attacking packets come from, where a suspect intruder is located, where a malicious e-mail is originated, or where a terrorism website is hosted is the key to identify, track, report, arrest, and punish criminals.

The dynamic, stateless, and anonymous nature of the Internet makes it extremely difficult to trace the sources of Internet crime, since the attacker can forge the source address field in an Internet Protocol (IP) packet. To find the real source of Internet attacks, we must possess the capability of discovering the origin of IP packets without relying on the source IP address field. This capability is called IP trace back. IP trace back systems provide a means to identify true sources of IP packets without relying on the source IP address field of the packet header, and are the major technique to find the real attack sources [1], [2]. Although currently there have been many publications on IP trace back, some key issues that are essential to make an IP trace back scheme into a really usable trace back system were not solved, for the system is solved such as how many sources can be traced in one trace back process, how large is the false positive rate, how many packets are needed to trace one source, and how to alleviate the load of participating routers.

In this paper, a novel and practical IP traceback system, Flexible Deterministic Packet Marking (FDPM), is presented. FDPM belongs to the packet marking family of IP traceback systems. The novel characteristics of FDPM are in its flexibility: first, it can adjust the length of marking field according to the network protocols deployed (flexible mark length strategy); second, it can also adaptively change its marking rate according to the load of the participating router by a flexible flow-based marking scheme. These two novel characteristics of FDPM make it more practical than other current traceback systems in terms of compatibility and performance. Both simulation and real system implementation prove that FDPM can be used in real network environments to trace a large number of real sources, with low false positive rates, and with low resource requirement on routers.

The rest of this paper is organized as follows: Section 2 surveys previous work on IP trace back research. In Section 3, the system design of FDPM, including encoding scheme, reconstruction scheme, and flow-based marking scheme, is presented. Section 4 describes the simulation on how FDPM can effectively trace a large number of sources in a single trace back process with limited number of packets required. Section 5 describes the simulation on overload prevention of FDPM with its flow-based marking scheme.

2 PREVIOUS WORK ON IP TRACEBACK

2.1 Problem Description

Let A_i, i ε [0,n] be the attackers and V be the victim. The attackers and victim are linked by various routers R_j, j =2 ,1; m. The main objective of IP trace back problem is to identify the n routers directly connected to A_i. The key issue here is to completely identify the n routers with low false positive rates in a single trace back process (conducted by the same trace back point, e.g., V , for a certain period) because correlating the data in different traceback processes is not only extremely difficult but also meaningless for tracing a time-dependent event. In [3], it was stated that a practical IP traceback system should be able to identify a few hundred ð=Þ sources/routers out of 1 million routers. Some traceback schemes not only identify the n routers directly connected to A_i but also find the routes between the n routers to victim V. In this paper, we only deal with the problem of finding these n routers (not the routes). In fact, the packets starting from the same origin and arriving at the same destination still may take different routes because of the dynamic nature of the Internet. Therefore, considering routes may not have direct benefits to identify the real source of attacks.

2.2 Current IP Trace back Schemes

There are some survey papers discussing the tradeoffs of different IP trace back schemes, such as [4], [5], and [6]. Current IP trace back schemes can be classified into five categories: link testing, messaging, logging, packet marking, and hybrid schemes. The main idea of the link testing scheme is to start from the victim to trace the attack to upstream links, and then determine which one carries the attack traffic [7], [8]. It consumes huge amount of resources, introduces additional traffic, and possibly causes denial of service when the number of sources needed to be traced increases. Messaging schemes use routers to send ICMP messages from the participating routers to destinations. For a high volume flow, the victim will eventually receive ICMP packets from all the routers along the path back to the source, revealing its location [9], [10], [11]. The disadvantages of messaging schemes are that the additional ICMP traffic would possibly be filtered by some routers, and huge numbers of packets are required by the victim to identify the sources. Logging schemes include probabilistic sampling and storing transformed information. Logging schemes maintain a database for all the traffic at every router within the domain and to query the database to filter is used to reduce the data stored. The main disadvantage of logging schemes is that they heavily overload the participating routers by requiring them to log information about every packet passing by, although it is claimed that it needs only a single packet to find its origin . Identify the sources of an IP packet. Hash function or Bloom

Packet marking schemes insert traceback data into an IP packet header to mark the packet on its way through the various routers from the attack source to the destination; then the marks in the packets can be used to deduce the sources of packets or the paths of the traffic [16], [17], [18], [19], [20]. As this method overwrites some rarely used fields in IP header, it does not require modification of the current Internet infrastructure. This property makes it a promising traceback scheme to be part of DDoS defense systems [21]. However, the space in IP header that can be utilized is limited. Thus, the information that one packet can carry is also limited. Therefore, many challenges for this category of traceback schemes are raised. For example, the number of sources that can be traced could be limited, the number of packets required to find one source could be large, and the load of the traceback router could be heavy. In Sections 2.3 and 2.4, we detail current packet marking schemes and analyze their limitations.

Recently, there has been also some research on hybrid schemes. In a hybrid traceback scheme combining logging and packet marking is presented to achieve the small number of packets needed to trace a single source and the small amount of resources to be allocated to the participating routers. Although the hybrid schemes try to overcome the disadvantages of each traceback scheme, the complexity of such combination and the practicability of their implementation still need more research.

2.3 Probabilistic Packet Marking Schemes

Probabilistic Packet Marking (PPM) [16] is one stream of the packet marking methods. The assumption of PPM is that the attacking packets are much more frequent than the normal packets. It marks the packets with path information in a probabilistic manner and enables the victim to reconstruct the attack path by using the marked packets. PPM encodes the information in rarely used 16-bit Fragment ID field in the IP header. To reduce the data that is to be stored in 16 bits, the compressed edge fragment sampling algorithm is used.

Although PPM is simple and can support incremental deployment, it has many shortcomings that can seriously prevent it from being widely used. First, the path reconstruction process requires high computational work, especially when there are many sources. For example, a 25-source path reconstruction will take days, and thousands of false positives could happen [18]. Second, when there are a large number of attack sources, the possible rebuilt path branches are actually useless to the victim because of the high false positives. Therefore, the routers that are far away from the victim have a very low chance of passing their identification to the victim because the information has been lost due to overwriting by the intermediate routers.

advanced and authenticated PPM based on the assumption that the victim knows the mapping of the upstream routers. It not only reinforces the capability to trace more sources at one time but also solves the problem of spoofed marking. Another method to reduce the overhead of reconstruction was proposed in . It uses counters to complement the loss of marking information from upstream routers, in order to save computation time and reduce false positives. Adler analyzed the tradeoff between mark bits required in the IP header and the number of packets required to reconstruct the paths.

2.4 Deterministic Packet Marking Schemes

Another stream of packet marking methods, which does not use the above probabilistic assumption and stores the source address in the marking field, is in the category known as the deterministic approaches, such as Deterministic Packet Marking (DPM) our FDPM (the first version of FDPM was published in), and Deterministic Bit Marking . Recently in the DPM scheme was modified to reduce false positive rates by adding redundant information into the marking fields. Unlike PPM, deterministic approaches only keep the first ingress edge router's information in the marks (but not the whole path). Moreover, they record marks in a deterministic manner (but not a probabilistic manner as in PPM). This category of schemes has many advantages over others, including simple implementation, no additional bandwidth requirement, and less computation overhead. However, enough packets must be collected to reconstruct the attack path (e.g., in the best case, at least two packets are required to trace one IP source with any of the above schemes). Importantly, all previous works neither perform well in terms of, nor have addressed the problems of, the maximum number of sources that the traceback system can trace in a single traceback process, the number of packets needed to trace one source, and the overload prevention on participating routers.

3 Flexible Deterministic Packet Marking Scheme

3.1 System Overview

The FDPM scheme utilizes various bits (called marks) in the IP header. The mark has flexible lengths depending on the network protocols used, which is called flexible mark length strategy. When an IP packet enters the protected network, it is marked by the interface close to the source of the packet on an edge ingress router. The source IP addresses are stored in the marking fields. The mark will not be overwritten by intermediate routers when the packet traverses the network. At any point within the network, e.g., the victim host, the source IP addresses can be reconstructed when required.

Processing packets consume resources such as memory and CPU time of a participating router. Therefore, it is possible for a router to be overloaded when there are a large number of arrival packets waiting for FDPM to mark them. A question that has been raised is how much computing power is needed by the marking process of FDPM and is it worth selectively reducing the marking rate? According to the research in , the complexity of the forwarding process in a typical router is low (e.g., 2.1 instructions executed per byte of data in a packet) but other processing applications such as data encryption or data compression impose much more complexity (e.g., 10^2 instructions executed per byte of data in a packet). Packet marking requires a router to generate marks including different parts by certain computation methods such as hashing and random number generating. The complexity of packet marking is not measured in this paper; however, it must be more than the forwarding process (as it will be proven in Section 6.3). The flow-based marking scheme is proposed to solve the overload problem. When the load of a router exceeds a threshold, the router will discern the most possible attacking packets from other packets then selectively mark these packets. The aim is to alleviate the load of the router while still maintaining the marking function.

The flexibility of FDPM is twofold. First, it can use flexible mark length according to the network protocols that are used in the network. This characteristic of FDPM gives it much adaptability to current heterogeneous networks. Second, FDPM can adaptively adjust its marking process to obtain a flexible marking rate. This characteristic prevents a traceback router from the overload problems.

The complexity of packet marking schemes can be expressed by the number of packets needed to reconstruct one source. Let b be the number of bits allocated to traceback, and let n_s be the length of the description of the source, e.g., 32 for one source IP address. Because of the deterministic feature of FDPM, it requires only $O(n)$ packets to reconstruct one source. However, all the probabilistic schemes require a greater number of packets. For example, an improved PPM [25] requires $O(bn^2_s 2^b (2 + e)^{4n=2_b})$ packets, for any constant $\varepsilon > 0$, to reconstruct the source with probability greater than 1/2. Section 4 will give the estimated number of packets needed to reconstruct one source and the experiment results.

3.2 Utilization of IP Header

FDPM is based on IPv4. Possible IPv6 implementation of FDPM will involve adding an extension header in IPv6 packets, which is different with the IPv4 design. The necessity of FDPM IPv6 implementation needs more research because IPv6 has built-in security mechanisms such as authentication headers to provide origin authentication.

Three fields in the IP header are used for marking; they are Type of Service (TOS), Fragment ID, and Reserved Flag. The TOS field is an 8-bit field that provides an indication of the abstract parameters of the quality of service desired. The details of handling TOS and specification of TOS values can be found in [32]. The TOS has been rarely supported by most routers in the past. Some proposed standards such as Differentiated Services in TOS [33], used to indicate particular Quality-of-Service needs from the network, are still under development. Therefore, in FDPM, the TOS field will be used to store the mark if the underlying network protocol does not use the TOS filed.

Fragment ID and Reserved Flag are also exploited. Given that less than 0.25 percent of all Internet traffic are fragments [34], Fragment ID can be safely overloaded without causing serious compatibility problems.

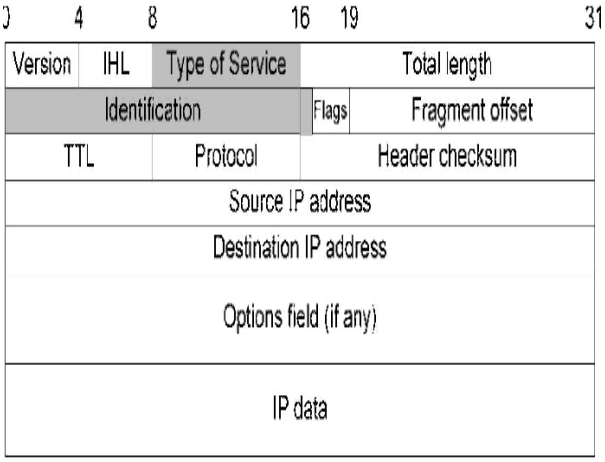


Fig. 1. The IP header fields (darkened) utilized in FDPM.

available for the storage of mark information if the protected network allows overwriting on TOS. When considering the possibility that the TOS field may be unavailable partly or totally, the minimum number of the bits in the IP header is 16 (excluding the 1-bit Reserved Flag). The Reserved Flag is not considered into the marking fields because it is used as the control bit to indicate whether or not the TOS field is being used, which will be discussed later. FDPM can adjust the mark length according to the protocols of the network in which FDPM is deployed. Therefore, even when FDPM is deployed among networks with different protocols, it can still work well because FDPM can differentiate the networks by the control bits.

Because the maximum length of the available mark is 25 bits, more than one packet is needed to carry a 32-bit source IP address. This is the reason why a segment number is needed to reconstruct an IP address into its original order. Each packet holding the mark will be used to reconstruct the source IP address at any point within the network. After all the segments corresponding to the same ingress address have arrived at the reconstruction point, the source IP address of the packets can be reconstructed. In order to keep track of the set of IP packets that are used for reconstruction, the identities showing the packets coming from the same source must be included; therefore, a hash of the ingress address is kept in the mark, known as the digest. This digest always remains the same for an FDPM interface from which the packets enter the network. It provides, on the victim's end, the ability to recognize which packets being analyzed are from a same source, although the digest itself cannot tell the real address.

Even if the participating router is compromised by attackers , this scheme will not be affected because the packets with irrelevant digest will be discarded during the reconstruction process. In essence, this will not introduce false positives, but will result in requiring more packets to reconstruct the sources. In this paper, we have the assumption that no participating router is compromised.

3.3 Encoding Scheme

Before the FDPM mark can be generated, the length of the mark must be determined based on the network protocols deployed within the network to be protected. According to

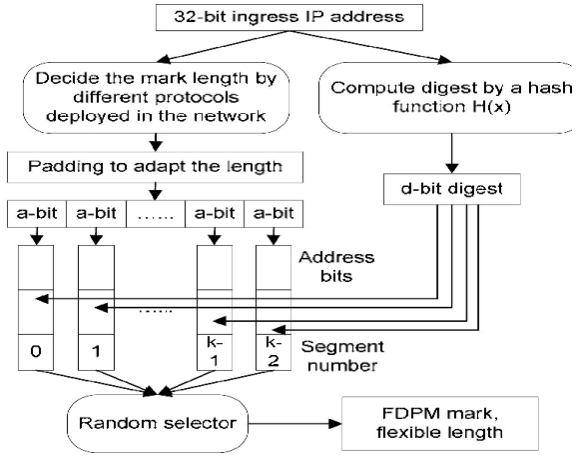


Fig. 2. FDPM encoding scheme.

Encoding scheme, which are named as FDPM-24, FDPM-19, and FDPM-16 in the rest of this paper. FDPM encoding scheme is shown in Fig. 2. The ingress IP address is divided into k segments and stored into k IP packets. The padding is used to divide the source IP address evenly into k parts. For example, if k= 6, the source address is padded with 4 bits of 0, making it 36 bits long, then each segment will be 6 bits long.

The encoding algorithm is shown in algorithm as Follows

1.Marking process at router R,edge A in network N
2.Set the bit array Digest and Mark to 0
3.if N Does not utilize TOS
4.Reserved_Flag:=0
5.7^{th} and 8^{th} bit of TOS:+0
6.Length_of_Mark:=24
7.else
8.Reserved_Flag:=1
9.if Nutilizes Differentitiated Services Field or
10 .N Supports Procedence and Priority
11.7^{th} and 8^{th} bit of TOS:=1
12.Lenth_of_Mark:=24
13.else if N Supports Precedence and Priority
14. 7^{th} bit of TOS:=1
15.8^{th} bit of TOS:=1
16. Lenth_of_Mark:=19
17.else if N Support Priority but not Precedence
18. 7^{th} bit of TOS:=0
19. 8^{h} bit of TOS:=1
20. Lenth_of_Mark:=19
21.Decide the lengths of each part in the mark
22.Digest:=Hash(A)
23.for i=0 to k=1
24.Mark[i].Digest=Digest
25. Mark[i].Segment=Segment
26. Mark[i].Address_bit:=A[i]
27.for each incoming packet p passing the encoding router
28.i:=random integer from 0 to k=1
29.write mark[i] into p,Mark

Fig No : 3. Algorithm of Encoding

The other situations, the length of mark will be 19 or 16, with relevant bit(s) in TOS marked. If the network supports TOS Precedence but not TOS Priority, fourth to sixth bits of TOS are utilized for marking; and if the network supports TOS Priority but not TOS Precedence, first to third bits of TOS are utilized for marking.

3.4 Reconstruction Scheme

The reconstruction process includes two steps: mark recognition and address recovery. When each packet arrives at the point that requires reconstruction, it is first put into a cache because, in some cases, the reconstruction processing speed is slower than the Reconstruction algorithm as follows

1. Reconstruction at Victim V, in network N
2. for each coming packet p passing the reconstruction point
3. mark recognition
4. if all fields in one entry are filled
5. output the source IP
6. delete the entry
7. else
8. if same digest and segment number exist
9. create new entry
10. fill address bits into entry
11. else
12. fill the address bits into entry

Fig No : 4. Algorithm of FDPM reconstruction scheme

The first step, address recovery, analyzes the mark and stores it in a recovery table. It is a linked-list table; the number of rows is a variable, and the number of columns in the table is k number is used to correlate the data into the correct order. The row of the table means the entry; usually each digest owns one entry (source IP address). However, different source IP addresses may have the same digest because the digest is a hash of the source IP address, and is shorter than an IP address. In this case, hash collision is unavoidable. When the hash collision occurs, more than one entry may be created in order to keep as much information as possible. The advantage of this design is that it can reconstruct all possible sources but the disadvantage is it also brings possible irrelevant information. Compared with DPM in [27], our reconstruction process is compatible with different protocols and will not lose any sources even when hash collision occurs. More details about the benefits of this design can be found in next Sections.

3.5 Flow-Based Marking Scheme

The possibility of the overload problem always exists because the resources of a router are always limited. If the router is overloaded, the marking scheme can be totally ineffective. All packet marking traceback schemes consume the computing power and storage capacity of routers as they need to overwrite many bits in the IP header. Therefore, overload prevention is important to all packet marking traceback schemes. There are many methods to lighten the burden of a router. One is to increase the computing capability of the router, for example, by

using Multi core based architecture [36]. Another is to apply an adaptive algorithm to reduce the load of processing of packets when the load of the router exceeds a threshold, which is our novel approach, flexible flow-based marking scheme.

The idea of flow-based marking is to selectively mark the packets according to the flow information when the router is under a high load. Therefore, it can reduce the packet.

Flow 1	Destination IP Address 1	Number of packets in flow 1:$npkts_1$
Flow 2	Destination IP Address 2	Number of packets in flow 1:$npkts_2$
.....	.....	.....
.....	.....	.....
Flow n	Destination IP Address n	Number of packets in flow 1:$npkts_n$

Fig. No 5. Dynamic flow table T and FIFO queue Q in FDPM flow-based marking scheme

The goal of flow-based marking is to mark the mostPossible DDoS attacking packets (from the same sources but not necessarily with same source IP addresses and to the same destination), then let the reconstruction process in the victim end reconstruct the source by using a minimum number of packets. Ideally, the flow-based marking scheme should be able to keep a separate state for every flow thatthe router needs to forward, regardless of whether the flow contains large or small number of packets. In our flowbased marking scheme, we aim at reducing complexity and increasing efficiency. It
does not keep the state for each flow, but simply uses a single first-in, first-out (FIFO) queue which can be shared by all flows. The advantage of this is that it can be easily implemented in current router architecture, with little impact on the router's packet processing capability. This process is similar to some congestion control schemes such as the Random Early Detection (RED) [37], which isolates the flows that have an unfair share of bandwidth and drops the packets in those flows. The flow-based marking scheme needs to isolate and mark the
flows that occupy more bandwidth containing most possible DDoS attacking packets. It can mark packets with a certain probability from each flow, in proportion to the amount of bandwidth the flow uses.

The simple data structures include a dynamic flow table T and a FIFO queue Q, as shown in Fig. 6. Each record in T stands for a flow. Here, the flow means the group of packets that have defined specific subsets of identifiers and are in the Q at a certain time. In DDoS scenarios, attacking packets areclassified into different flows according to the destination IP address in the IP header because the aggregation effect is the major feature in DDoS attack traffic. The flow records in T are the destination IP addresses and the number of packets from this flow in the queueQ, denoted as npkts.

The algorithm of flow-based marking is

1. if (load of router R>threshold Lmax)
2. do not mark any packets
3. turn on congestion control mechanisms
4. else if (load of router R>threshold Lmin)
5. turn on flow-based marking at R,edge interface A, in network N
6. for each incoming packet p
7. check npkts with same destination address of p from T
8. if(npkts==0),means no such flow in T)
9. add a new entry in T,set in pkts =1
10. else
11. npkts++
12. insert packets p into Q
13. calculate marking probability p_a
14. with probability p_a mark the packet
15. if Q is full
16. dequeue
17. else
18. mark all the packets at R, edge interfaceA, in network N

There are two load thresholds Lmax and Lmin for the trace back router. Lmax is the threshold that controls the whole packet marking process, which means the router will not mark any packet if its load exceeds this value .Congestion control mechanisms can be turned on in order to guarantee a best effort service [38] for the router. The load threshold Lmin means that if the load exceeds this value, the router can still work, but it must reduce its marking load. If the load stays below Lmin, then the router will just mark all the incoming packets because the router can process all packets without having performance penalty. These two thresholds should be set according to real situations in routers.When flow-based marking is turned on, the probability of marking an incoming packet from a particular flow is roughly proportional to the flow's share of bandwidth through the router. We define this probability where npkts is the number of packets in the flow containing current incoming packet, L is the current load of the router. This definition has Pa =0; 1_. When the current load of the router L reaches Lmax, Pa becomes 0, which means nomarking is performed. we apply a low-pass filter with exponentially weighted moving averages (EWMA), which is a fast and practical approach. CUSUM and related algorithms are not used because, here, the detection rate is not the major concern be considered. but keeping low complexity is.

Therefore, when calculating the marking probability Pa, we use the EWMA n pkts which is defined as where p is the filter constant, which dictates the degree of filtering, e.g., how strong the filtering action will be. By using this low-pass filter, the historical effect of npkts can be implemented.

4 Real System Implementation

4.1 Evaluation Measurements: Number of Packets

Needed to Trace One Source and Maximum Forwarding Rate Currently, most existing works on IP traceback are based on simulation or theoretical analysis. Few traceback schemes have been implemented and tested by real system imple-mentation. It is very difficult to test the real performance of a traceback scheme if only simulation is conducted. The motivation of real system implementation of FDPM is that we want to know how well it can perform under real environments. The main evaluation measurements we used are the marked rate ɜ, the number of packets needed to trace one source N_N , and the maximum forwarding rate $ɜ_{max}$. Maximum forwarding rate is the rate at which an FDPM-enabled router can forward 64-byte packets over a range of input rates. It is difficult to be measured in simulation, but it can be measured in real system implementation. The maximum forwarding rate can be plotted as the line in input rate and forwarding rate coordinates. Ideally, if a router has unlimited computing power and storage, and if the inter faces' bandwidth is unlimited, it would forward every input packet regardless of input rate, corresponding to the line y = x.

We used the Click modular router [47] to implement our FDPM on PC-based router (Intel Pentium 4 Processors 2 GHz, DRAM 1 Gbyte, double D-Link network 100-Mbps adapters). Click router is a software architecture running on PCs for building flexible and configurable routers, which is assembled from packet processing modules called elements. The FDPM Encoding element, Reconstruction element, Flow-based Marking control element, and other associated measuring elements were added to this architecture. Turn-ing on added elements reduces the forwarding capability of the router. The tradeoffs of packet marking schemes will be discussed in Previous Section

4.2 Number of Packets for Reconstruction

The Above shows the relationship between the number of packets needed to trace one source N_N and the marked rate ɜ for flow-based marking scheme and random marking scheme in Click router implementation. The condition of Figure is that the router uses two packets to carry a source IP address (k=2) and the percentage of attacking packets ɜ=0.1. The condition of Figure is that the router uses eight packets to carry a source IP address (k =8) and the percentage of attacking packets ɜ = 0:5. From the comparison of Figures we can see that the simulation and real system implementation show the same trend. This clearly demonstrates the capability of the

FDPM to selectively mark the most likely DDoS packets in case of high load on routers

4.3 Maximum Forwarding Rate

This section evaluates FDPM-enabled router's performance of forwarding IP packets under different conditions. Below Figure shows the maximum forwarding rate ₃max for the raw Click router without any packet marking function. This figure can be used as the baseline to compare with FDPM-enabled router's maximum forwarding rate. In our experiments, the maximum forwarding rate ₃max of the Click router is 69,000 packets per second. When the input rate exceeds this rate, the router will discard received packets due to the bottleneck of the router's computing power.

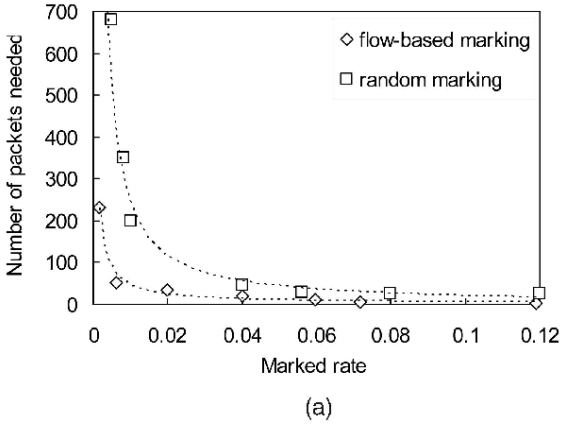
5 Conclusion

FDPM is suitable for not only tracing sources of DDoS attacks but also DDoS detection. The main characteristic of DDoS is to use multiple attacking sources to attack a single victim (the aggregation characteristic). Therefore, at any point in the network, if there is a sudden surge in the number of packets with the same destination address and with the same group of digest marks, it can be a sign of a DDoS attack. More details can be found in [48]. In FDPM, the marks in packets do not increase their size; therefore, no additional bandwidth is consumed. Moreover, with the overload prevention capability, FDPM can maintain the traceback process when the router is heavily Loaded, whereas most current traceback schemes do not have this overload prevention capability. Compared with other schemes, FDPM only needs 102 packets to trace up to 105 sources, so the sources/packets ratio is the highest. FDPM requires little computing power and adaptively keeps the load of routers in a low degree. Where Compatibility is concerned, FDPM does not need to know the network topology, and it can be implemented gradually because it has the control bits to differentiate different network protocols used.

An effective traceback system is essential to control Internet crime. While some research has been done, to the best of our knowledge, none of the previous work has fully solved problems such as the maximum number of sources that a traceback system can trace in one traceback process, and the possible overload problem of participating router. We are among the first to examine overload prevention in traceback systems. Compared with other IP traceback schemes, FDPM provides more flexible features to trace IP packets than other packet marking schemes, and can obtain better tracing capacity. To summarize this paper, we list our major contributions here:

1. A novel and practical packet marking traceback system, incorporating a flexible mark length strategy and flexible flow-based marking scheme, is proposed.

2. Simulation and real system implementation show FDPM produces better performance than any other current traceback scheme in terms of false positive rates, the number of packets needed to reconstruct one source, the maximum number of sources that can be traced in one traceback process, and the maximum forwarding rate of traceback enabled routers.



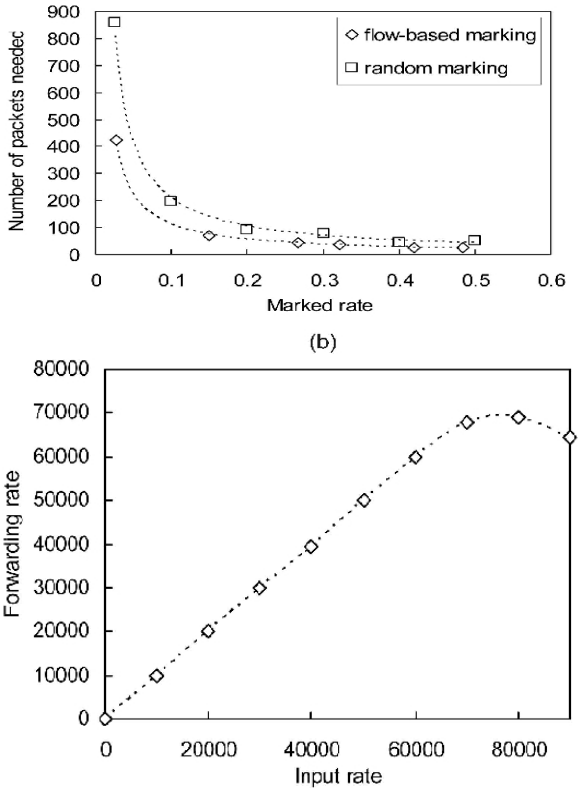


Fig. 6. The relationship between the number of packets needed to trace one source N_N and the marked rate ₃ for the flow-based marking scheme and the random marking scheme in real system implementation

(a) k =2, ₃ = 0:1. (b) k ¼=8, ₃ = 0:5.

REFERENCES

[1] H. Farhat, "Protecting TCP Services from Denial of Service Attacks," Proc. ACM SIGCOMM Workshop Large-Scale Attack Defense (LSAD '06), pp. 155-160, 2006.

[2] H. Wang, C. Jin, and K.G. Shin, "Defense against Spoofed IP Traffic Using Hop-Count Filtering," IEEE/ACM Trans. Networking,vol. 15, no. 1, pp. 40-53, 2007.

[3] M.T. Goodrich, "Efficient Packet Marking for Large-Scale IP Traceback," Proc. Ninth ACM Conf. Computer and Comm. Security (CCS '02), pp. 117-126, 2002.

[4] H. Aljifri, "IP Traceback: A New Denial-of-Service Deterrent," IEEE Security and Privacy, vol. 1, no. 3, pp. 24-31, 2003.

[5] A. Belenky and N. Ansari, "On IP Traceback," IEEE Comm., vol. 41, no. 7, pp. 142-153, 2003.

[6] Z. Gao and N. Ansari, "Tracing Cyber Attacks from the Practical Perspective," IEEE Comm., vol. 43, no. 5, pp. 123-131, 2005.

[7] H. Burch and B. Cheswick, "Tracing Anonymous Packets to Their Approximate Source," Proc. 14th Systems Administration Conf. (LISA '00), pp. 319-327, 2000.

[8] R. Stone, "CenterTrack: An IP Overlay Network for Tracking DoS Floods," Proc. Ninth USENIX Security Symp.(Security), pp. 199-212, 2000.

[9] S.M. Bellovin, ICMP Traceback Messages—Internet Draft, Network Working Group, 2000.

[10] A. Mankin et al., "On Design and Evaluation of Intention-DrivenICMP Traceback," Proc. 10th Int'l Conf. Computer Comm. And Networks (ICCCN '01), pp. 159-165, 2001.

[11] C. Jin, H. Wang, and K.G. Shin, "Hop-Count Filtering: An Effective Defense against Spoofed DDoS Traffic," Proc. 10th ACM Conf. Computer and Comm. Security (CCS '03), pp. 30-41, 2003.

[12] N.G. Duffield and M. Grossglauser, "Trajectory Sampling for Direct Traffic Observation," Proc. ACM SIGCOMM '00, pp. 271-282, 2000.

[13] A.C. Snoeren et al., "Single-Packet IP Traceback," IEEE/ACM Trans. Networking, vol. 10, no. 6, pp. 721-734, 2002.

[14] T. Baba and S. Matsuda, "Tracing Network Attacks to Their Sources," IEEE Internet Computing, vol. 6, no. 3, 20-26, 2002.

[15] J. Li et al., "Large-Scale IP Traceback in High-Speed Internet: Practical Techniques and Theoretical Foundation," Proc. IEEE Symp. Security and Privacy (S&P '04), pp. 115-129, 2004.

[16] S. Savage et al., "Network Support for IP Traceback," ACM/ IEEE Trans. Networking, vol. 9, no. 3, pp. 226-237, 2001.

[17] K. Park and H. Lee, "On the Effectiveness of Route-Based Packet Filtering for Distributed DoS Attack Prevention in Power-Law Internet," Proc. ACM SIGCOMM '01, pp. 15-26, 2001.