TTL based Packet Marking for IP Traceback

Vamsi Paruchuri, Arjan Durresi and Sriram Chellappan

Abstract— Distributed Denial of Service Attacks continue to pose major threats to the Internet. In order to traceback attack sources (i.e., IP addresses), a well studied approach is Probabilistic Packet Marking (PPM), where each intermediate router of a packet marks it with a certain probability, enabling a victim host to traceback the attack source. In a recent study, we showed how attackers can take advantage of probabilistic nature of packet markings in existing PPM schemes to create spoofed marks, hence compromising traceback. In this paper, we propose a new PPM scheme called TTL-based PPM (TPM) scheme, where each packet is marked with a probability inversely proportional to the distance traversed by the packet so far. Thus, packets that have to traverse longer distances are marked with higher probability, compared to those that have to traverse shorter distances. This ensures that a packet is marked with much higher probability by intermediate routers than by traditional mechanisms, hence reducing the effectiveness of spoofed packets reaching victims. Using formal analysis and simulations using real Internet topology maps, we show how our TPM scheme can effectively trace DDoS attackers even in presence of spoofing when compared to existing schemes.

I. INTRODUCTION

Any attack on the Internet today can be highly devastating. Distributed Denial of Service (DDoS) attacks are among the most malicious Internet attacks, that overwhelm a victim system with data such that the victim response time is slowed or totally stopped. There have been many instances where DDoS attacks have caused damages worth billions of dollars [35, 36, 37]. Defending against DDoS attacks has hence become a major priority in the Internet community [1, 3-21, 24-25, 29-34].

Clearly, any defense against DDoS attacks is contingent on the ability of defenders to to identify the source of DDoS attacks. This process is known as Traceback. (i.e.tracing back the origin of attack traffic). To date, the best known approach for traceback is to place tracking information into rarely used header fields inside the IP packets as and when the traffic propagates through the Internet. Since, available space in IP header is limited, researchers have focused on routers probabilistically marking each packet with their IDs, along with their position in the routing path, called as Probabilistic Packet Marking (PPM) Scheme. For large number of DDoS packets, if each router probabilistically marks a packet, this approach is expected to provide enough router and path information at victim side in order to traceback the path and hence the source of attack traffic [6, 7, 8, 9, 13, 17, 31-34].

Despite, the large body of work, there are certain critical constraints for PPM schemes. Each router independently marks packets with a probability. The optimal marking probability to reduce the number of packets required to traceback is shown to be 1/d, where d is the length of the path taken by the packets [6, 7]. Since, most of the IP path lengths are less than 25 hops, the marking probability is set to 1/25. As we can see, the percentage of unmarked packets at a victim side is (1-1/d)^h, where h is the actual number of hops in the path. For an average case path length of around 16 hops [28], this means that more than half the packets reaching the victim are unmarked.

Attackers can take advantage of such unmarked packets by intelligently inserting their own packet markings. In a recent study [38], we demonstrated how attackers can exploit this and introduce a significant amount of anomaly at the victim preventing it from performing meaningful attack graph construction and traceback. Note that this shortcoming is there for all existing PPM schemes [38].

The main reason for this vulnerability is that current schemes require routers mark with a constant probability. In this paper, we propose a new Time-To-Live (TTL) based PPM scheme (TPM), where each router marks packets with dynamic probability. Specifically, each router marks a packet with a probability inversely proportional to the distance it has to travel. As such, a packet that has to traverse long distances is marked with higher probability, compared with a packet with shorter distances to traverse. This modification ensures that a packet is marked with much higher probability compared to existing mechanisms, which greatly reduces effectiveness of spoofed marks. We show that TPM can reduce the number of false positives by 90% when compared with PPM based techniques (e.g., FIT [9]), while requiring as few as one-third of the packets required by FIT for attack graph reconstruction.

The rest of the paper is organized as follows: We review traceback literature in Section II. Section III presents our TPM. In Section IV, we analyze and present results on effectiveness of TPM against attacker spoofing. Finally, Section V presents concluding remarks.

II. RELATED WORK

A. A General Background

Researchers have proposed various schemes to address the DDoS problem. The obvious countermeasure is ingress filtering [1] based on suspect source address. The next approach is victim pushback, where a site that believes to be under attack can send back messages installing filters at upstream routers [2, 3]. The IETF working group proposed...
that each router periodically selects a packet and “append” authenticated traceback information to this packet [4], by creating a second packet tailgating the original packet.

Snoeren et al. [5] propose storing a hash of each packet along with information about where it arrived from in a memory efficient fashion. A graph coloring approach to traceback employing packet marking is proposed in [16]. In approaches like [25], DDoS attacks are thwarted by offensive strategies where victim servers (under DDoS attacks) ask legitimate clients to increase their traffic.

B. Probabilistic marking Schemes:

The most studied and well accepted solution is to let routers probabilistically mark packets with partial path information during forwarding [6]. Song et al. [7] show that the approach in [6] has high computation overhead for the victim to reconstruct attack paths, and that the scheme is ineffective under distributed DoS attacks. Dean et al. [8] propose routers to algebraically encode the path or edge information iteratively using Horner’s rule. This scheme is susceptible to a GOSSIB attack [30]. Also, the number of packets required to reconstruct path is high.

Fast Internet Traceback [9] seems to be most efficient and scalable, requiring fewer packets to traceback and producing low false positives. FIT achieves these properties by using only a single bit to encode distance between a marking router and victim. Several other improvements have been proposed to improve the performance of PPM techniques [13, 17, 18, 31-34].

DPPM [12] lets routers dynamically set marking probabilities. The routers deduce how far a packet has traveled and then choose the marking probability as an inverse function of hop count traveled. For deducing the distance travelled, the fact that most initial TTL values fall in the set of $S = \{32, 64, 126, 255\}$ is used. However, attackers can spoof the TTL field and set it randomly or to other node(s). In a recent study in [38], we perform a detailed analysis on spoofing, and show that all existing PPM schemes suffer from the same problem. We show how an attacker could efficiently spoof a large number of packets in order to mislead the victim and hence hide its identity, even without any knowledge of network topology.

III. TTL BASED PACKET MARKING

In this section, we present TPM, a dynamic marking scheme based on the TTL field in order to minimize the number of spoofing packets that reach the victim. We first present the TPM marking mechanism and then analyze the number of packets needed to reconstruct the attack graph.

A. Proposed scheme

In simple terms, our goal is to design an approach which would significantly reduce, if not prevent, the number of unmarked packets reaching the destination. We propose not to mark all the packets with equal probability; instead each router would compute the marking probability for each packet based on the TTL field value, $t$. Each IP packet contains a Time-To-Live (TTL) field, which is decremented when a router processes the packet. If TTL reaches zero, the packet is discarded, preventing packets from running in circles forever and flooding a network.

The marking algorithm at a router in our scheme is shown in Figure 1, where $t$ is the TTL value of the packet being marked and $t_p$ is the maximum path length. To elaborate, a router marks a packet with probability $1/h$, where $h$ is the estimated maximum remaining distance that the packet would traverse. Thus, a packet traversing long distances is marked with high probability, while packets with short distances to traverse are marked with lower probabilities. This simple modification would ensure that a packet is marked with much higher probability than by the traditional mechanisms. For estimating the maximum remaining distance, we use the studies [28] that show that more than 99% of path lengths are less than 25 hops and more than 90% of path lengths are shorter than 20 hops. We choose a value of $t_p = 24$ corresponding to a maximum path length of 25 hops.

Our TPM scheme can be incorporated in all probabilistic marking techniques. However, for illustration purposes, we...

For each packet $Pkt$

$t \leftarrow t - 1$

if $t_p > t$

$h \leftarrow t_p - t$

else

$h \leftarrow 1$; $t \leftarrow t_p$

Let $r$ be a random number in $[0, 1)$

if $r \leq 1/h$

mark the packet

Fig 1. Proposed TPM based Packet Marking algorithm
assume a FIT [9] like mechanism mainly due to the following reason: FIT proposes a novel method that uses only 1-bit and the TTL field to compute the distance to the marking router. Thus, among all schemes, FIT allocates a maximum of 15-bits for encoding path information. This increase in number of encoding bits lowers false positives and enables faster reconstruction of the attack graph.

To reduce the number of false positives, schemes [7, 9] propose to split (encode) link (IP) information into multiple fragments. Then, the packets are probabilistically marked each time with one of these fragments and the corresponding fragment number. Let \( k \) be the total number of fragments the information is split into and the fragment size be \( b_{\text{frag}} \). For reconstruction, the victim needs to receive at least \( n_{\text{path}} \) distinct fragments from a router.

### B. Packets to Reconstruct:

In this section, we compute the number of packets needed to reconstruct the attack graph. Assume \( r_j \) routers are at distance \( d \) from the victim in the graph \( G \). Hash fragment size is \( b_{\text{frag}} \) and let \( r_n \) be the number of routers on the attack path at distance \( d \) from the victim in the attack graph \( G_d \). \( n_{\text{path}} \) is the number of distinct fragments needed to reconstruct an IP address. The probability of receiving \( j \) distinct hash fragments from a set of \( k \) total fragments after receiving \( y \) randomly selected fragments is [26]

\[
P_i[j,k,y] = \binom{k}{j} \left[ \frac{1}{2} \right]^y \binom{k}{j} \left( 1 - \frac{1}{2} \right)^{k-j}
\]

(1)

The probability that a packet carries a fragment from a router at distance \( i \) hops from the victim and the TTL value being \( t \) (\( i \geq 0 \)):

\[
P_m[i,t] = \left[ \frac{1}{24-t} \right] \left[ \frac{1}{24-(t-1)} \right] \left[ \frac{1}{24-(t-2)} \right] \cdots \left[ \frac{1}{24-(t-i)} \right]
\]

\[
= \left[ \frac{1}{24-t+i} \right]^{i+1}
\]

(2)

The first term on the RHS on line-1 of above Equation corresponds to probability that the router marks the packet. The second compound term is the probability that none of the latter routers in the route mark the packet.

Here we would like to direct the reader’s attention to the following trade off on the attacker’s side:

- Worst case probability \( P_m[i,t] = 1/24 \), occurs when \( t = i \).
- If the attacker marks the packet such that \( t < i \), the packet would get dropped.
- If the attacker marks the packet such that \( t > i \), the router markings would reach the victim faster since in this case \( P_m[i,t] > 1/24 \).

We note that for traditional PPM schemes (e.g., FIT), the probability of receiving a fragment from a router at distance \( i \) hops from the victim, given the marking probability is \( p_i \), is

\[
P_i = p_i (1 - p_i)^{-1}
\]

(3)

Table 1 shows the number of packets needed to reconstruct different path lengths with 50% and 95% probability. TPM-Best case correspond to the scenario when the attacker does not spoof the TTL field and \( t = 24 \). The worst case corresponds to the case when the attacker cleverly spoofs the TTL field such that \( t = i \). We note that, even in the worst case, TPM requires less than half the packets required by PPM based techniques. In the best, TPM requires less than one-third the packets required by PPM.

### IV. Performance Analysis

In this section, we first estimate the number of spoofed packets that reach the victim. We analyze the impact of these spoofed packets on the number of false positives with different spoofing strategies. We further complement the mathematical analysis with experimental results using representative Internet topologies. For illustration purposes, we consider Skitter data [28] – *cam* datasets that has an average path length of around 18. In the rest of this section, PPM refers to the specific case of FIT with \( k = 4 \) and \( n_{\text{path}} = 4 \). Thus, each router can mark a packet with any of four distinct marks and the victim has to receive all four of them to add it to the attack graph. Where needed, for simplicity and fair comparison, we assume that each attacker sends 300 packets as a part of the DDoS attack, so that the victim can reconstruct the attack graph.

#### A. Analysis of Spoofed Packets

Here, we state two lemmas on probability of a packet being marked by a router (thus overwriting the spoofed value) before it reaches the victim.

**Lemma 1:** All the legitimate packets would be marked at least once by an intermediate router before it reaches the destination (victim).

**Proof:** The proof directly follows from the algorithm.

**Lemma 2:** There is an upper bound on the probability that a spoofed (illegitimate) packet reaches the destination without being marked. This upper bound is a function of the distance between the sender (attacker) and the destination (victim).

**Proof:** The attacker’s goal is to bombard the victim with packets. As such the attacker will always set TTL to be as high as possible (i.e., \( t_o \)) in order to ensure that its attack packets reach the victim. As per our TPM protocol, the next router in this path will reset the TTL to 24, and for each subsequent hop, the TTL is decremented by one. In this case, each packet will traverse a maximum of 24 hops before reaching the victim. Assuming that the victim is \( i \) hops from the attacker, the probability that the victim receives a spoofed packet is:

\[
P[i,t] = \left[ \frac{1}{24-t+i} \right]^{i+1}
\]

(4)

For the worst case, where \( t = i \), the probability simplifies to

\[
P[i,i] = \frac{1}{24}
\]

(5)

Table 1 shows the number of packets needed to reconstruct certain path lengths.
hops away from the attacker, the probability that a packet is unmarked can be calculated as the probability that all intermediate routers in the $i$-hop path do not mark the packets. This is given by $P_{\text{un-marked}}$ as,

$$P_{\text{un-marked}} = \left( \frac{1}{24-i} \right)^i \left( \frac{1}{24-(i-1)} \right)^{i-1} \cdots \left( \frac{1}{24-(i-3)} \right)^3 \left( \frac{1}{24-(i-2)} \right)^2 \left( \frac{1}{24-(i-1)} \right)^{i-1}$$

$$= \frac{24-i}{24-i-1}$$

(4)

Since, each packet cannot traverse more than the TTL value, we have $P_{\text{un-marked}} = (24-i+1)/24$. As such $P_{\text{un-marked}}$ is bounded as a function of the hops distance between the sender (attacker) and the victim. To illustrate, we computed $P_{\text{un-marked}}$ for the cam dataset [28] using the hop-length distribution and we observed that $P_{\text{un-marked}} \approx 0.23$. Thus, on an average, less than one-fourth of the attacker packets can reach the victim unmarked. To compare, for FIT and other PPM based schemes more that 50% of the packets are unmarked. Figure 2 presents the maximum probability for a packet to reach the victim unmarked for different path lengths.

The attacker has to make a tradeoff between increasing the probability of having a spoofed packet go through and probability of misleading the victim by “mismarking” initial TTL field. There is a trade off – higher the initial TTL, higher is the probability that the packet gets marked.

### B. False Positives

We assume that the attacker randomly spoofs the packets i.e., the spoofed packets carry markings that were randomly generated. The reconstruction process would assume that a router at a distance $d$ has marked the packet. Under random spoofing, $d$ would be randomly distributed between 0 and 31. Thus, if $N\times Pkt$ were collected for reconstruction process, around $N\times Pkt \times P_{\text{cam}}$ packets would carry actual router markings while $N_{\text{spoof}} = N\times Pkt \times P_{\text{spoof}}/32$ packets would appear to carry markings from a router at a given distance $d$ ($0< d < 31$).

We consider FIT like marking scheme, where 15 bits are allocated for hash fragment and hash number. Thus, $M = 2^{15}$ markings are possible. For each $1 \leq j \leq M$, let $x_j$ be the random variable such that $x_j = 1$, if the $j^{th}$ marking is carried by at least one of the $N\times Pkt$ packets received by the victim. So, the total number of markings that are received at least once is $\sum_{i=1}^{M} x_i$ and we want to find its expected value.

For a given $x_i$, the event $x_i = 1$ could result due to either of following reasons:

- At least one of the $N_{\text{spoof}}$ packets carries $i^{th}$ marking originating at a given distance $d$. The probability of this event is

$$P_{\text{spoof}}(x_i, N_{\text{spoof}}) = 1 - \left( \frac{1}{M} \right)^{N_{\text{spoof}}}$$

(5)

- At least one of $R_{j}$ routers decides to mark with $i$. The probability of this event is

$$P_{\text{marked}}(x_i, R_{j}) = 1 - \left( \frac{1}{M} \right)^{j+R_{j}}$$

(6)

Finally, the marking $i$ is not received only if none of the spoofed packets carry $i$ and none of the routers mark with $i$. Thus, probability of the event $x_i = 1$ can be computed as

$$P(x_i = 1) = 1 - \left[ 1 - P_{\text{spoof}}(x_i, N_{\text{spoof}}) \right] \left[ 1 - P_{\text{marked}}(x_i, R_{j}) \right]$$

(7)

This is also the expected value of $x_i$. Since each $x_i$ has the same expected value over all $1 \leq j \leq M$, the expected number of distinct markings (Fig. 3) that appear to originate at distance $d$ is

$$M_d = M \times P(x_i = 1)$$

(8)

Let $M_d$ be the number of distinct markings received by the victim with the distance field $d$. The probability that a specific fragment of a router not on the attack matches that fragment of a router on the attack path can be computed similar to the SS scenario (eq. (4)) and can be expressed as

$$p_{\text{fm}} = 1 - \left( \frac{1}{2} \right)^{M_d}$$

(9)

Since, at least $n_{\text{path}}$ markings per router are required to add it to the attack path, the probability that a router is a false positive is

$$P_{\text{fp}} = \sum_{j=n_{\text{path}}}^{M} \binom{k}{j} \left( 1 - p_{\text{fm}} \right)^{k-j}$$

(10)

$P_{\text{fp}}$ is the expected number of false positive IP addresses per router to be reconstructed. Figures 4 and 5 show the estimated probability that a router is a false positive, and the number of false positives for the cam dataset.
V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new scheme for tracing back DDoS packets, where each router marks a packet in a manner such that packets traversing longer distances are marked with high probability, while packets with short distances to traverse are marked with lower probabilities. Such a type of marking ensures that each attack is marked with much higher probability by intermediate routers, which greatly reduces the impacts of spoofed marks. We demonstrated that our scheme can guarantee a much higher degree of traceback success with low false positives compared to existing schemes. Further, TPM requires as few as one-third of the packets required by other schemes for attack reconstruction. Our future work is to explore techniques like expected distribution of legitimate traffic vs. attack traffic to design more robust traceback and attack reconstruction approaches.

REFERENCES


