An Investigation of 2^n Direction Geographical Traceback Using Direction Ratio Sampling Algorithm (DRSA) & IP Traceback Strategies

S.Karthik
Research Scholar, Professor and Head, Department of Computer Science Engineering
SNS College of Technology, Sathy Main Road,
Coimbatore-641035, Tamil Nadu, India.

Dr.V.P.Arunachalam
Principal and Research Supervisor, SNS College of Technology, Sathy Main Road,
Coimbatore-641035.

Dr.T.Ravichandran
Principal, Hindustan Institute of Technology,
Pollachi Main Road, Coimbatore-641032.

ABSTRACT:
DoS / DDoS(Distributed Denial of Service) attacks deny regular, internet services accessed by legitimate users, either by blocking the services completely, or by disturbing it completely, so as to cause customer baulking. Several traceback schemes are available to mitigate these attacks. The simulation approach also can be used to test the performing effects of different marking schemes in large-scale DDoS attacks. Based on the simulation and evaluation results, more efficient and effective algorithms, techniques and procedures to combat these attacks may be developed. DGT8, directional geographical trackback scheme, with 8 directions is one of them. Having a limited set of 8 directions, DGT8 may not work for routers with more than 8 interfaces. In this paper, we propose M-DGT i.e DGT 16, a 16 directional geographical traceback scheme having all the advantages of DGT. The 16 directions, though not having exactly equal interface, have nearly equal measures, and are identified using a novel scheme of Segment Direction Ratios (SDR). The SDR concept and the associated marking scheme allow the victim to defend against DDoS attacks independent of its ISP and also the generalization to DGT2n, having 2n directions (n>4).

Keywords:
DoS, DDoS, DGT (Directed Geographical traceback), IP traceback, SDR (Segment Direction Ratio)

1. INTRODUCTION
A denial of services attack (DoS) is an attempt to prevent legitimate users of a service, from using that service. DoS attacks are essentially, resource overloading attacks and either crash the communication system of the host with the rest of the Network or degrade the host's service rendering it unavailable for legitimate users. A DDoS attack, in general, consumes the target's resources, so that it cannot provide service. The resource is either an internal host resource on the target system or data transmission capacity in the local network.

IP traceback is the process of identifying the actual sources of attack packets. This has the benefit of holding attacker accountable for abusing the internet. It helps in mitigating DoS attacks by isolating identified attack sources. To abort these attacks, many IP traceback schemes [12] have been advocated.
Broadly they can be categorized into 3 groups: those which reconstruct the entire attack path the attack packets have traversed[1],[8],[9] such as Probability Packet Marking (PPM); those which focus only on the sources of attack packets, irrespective of the path taken[3] such as Deterministic Packet Marking (DPM);and the third is the Directed Geographical traceback (DGT) and geographical mapping techniques [11], [12]

The DGT Scheme of [12] possesses many desirable features such as fast convergence, light weight, good scalability and attack mitigation capability.

The DGT Scheme of [12] considers only 8 directions and may not work well for Routers that have more than 8 interfaces. In this paper, we are generalizing the DGT scheme to 16 interfaces of nearly equal measures.

By the novel scheme of Segment Direction Ratios(SDR), the 16 directions are identified by their SDR and every Router need know only the SDR of its immediate neighbors.

The rest of this paper is organized as follows. The traceback mechanisms are discussed in IP Traceback. The concept of Segment Direction Ratios (SDR) is introduced in Concept SDR Section-2. The SDR of scheme DGT 16 are presented together with the assumptions of DGT in DGT16 procedure is explained IN Section -3. Storage formalities are discussed in section-4 for Encoding Requirements. Qualitative comparison with other schemes and the limitations of DGT 16 constitute section-5

Comparison of DGT16 with other traceback schemes. Generalization to DGT 2n is discussed in Results & discussion followed by the Conclusion.

2. Basic Marking Algorithms

Node Append

Simplest marking algorithm is used to append each nodes address to the end of the packet as it travels through the network from the attacker to the victim. Every packet received by the victim will have the complete path traversed. As Shown in figure1 Algorithm is robust and extremely quick to converge.

Limitations

1. Router overload – infeasible
2. Length of path cannot be predetermined – space constraints in packets
3. Length of packet increases – fragmentation

Marking procedure at router R:
for each packet w, append R to w

Path reconstruction procedure at victim v:
for any packet w from attacker extract path (Ri..Rj) from the suffix of w

Figure 1: Node Append Algorithm.

Node Sampling

The attack path is sampled one node at a time and avoids recording the entire path. When a router receives a packet it chooses to write its information (address) in the node field based on some probability p. when the victim receives the marked packets it will have atleast one sample for every router in the attack path and the path can be converged as Shown in figure2.

Advantages

1. Requires only the addition of a write and checksum update
2. Currently high speed routers are available which can handle the marking efficiently.

Limitations

3. Inferring the total router order is a slow process
4. Routers far away from the victim contribute lower number of samples and cam lead to disordering. (requires more samples to avoid this i.e. probability of marking in these routers must be higher)
5. If multiple attackers are present, then multiple routers may be present at the same distance and hence will be sampled at same probability. Hence technique not robust against multiple attackers.

Marking procedure at router R:
for each packet w
let x be a random number from [0..1]
if x < p then,
write R into w.node

Path reconstruction procedure at victim v:
Let NodeTbl be a table of tuples (node,count)
for each packet w from attacker
z:= lookup w.node in NodeTbl

if zz!= NIL then
increment z.count
else
insert tuple (w.node,1) in NodeTbl
sort NodeTbl l by count
extract path (R_i..R_j) from ordered node fields in NodeTbl

Figure 2: Node Sampling Algorithm

Edge Sampling

As Shown in figure3 Instead of encoding individual node information in the packet encode the edge information. This includes the start and end nodes of the link and a distance field. When a router wants to mark the packet it enters its own address as the start information and sets the distance field to zero. If the distance field is already zero indicates that the packet was marked by previous router. In this case the router adds its information to the end field and increments the distance by 1. Even if the router does not mark a packet it has to increment the distance field by

1.

Marking procedure at router R:
for each packet w
let x be a random number from [0..1]
if x < p then,
write R into w.start and 0 into w.distance
else
if w.distance = 0 then
write R into w.end
increment w.distance

Path reconstruction procedure at victim v:

Let G be a tree with root v

Let edges in G be tuples (start, end, distance)

For each packet w from attacker

if w.distance = 0 then

insert edge (w.start, v, 0) into G

else

insert edge (w.start, w,end, w.distance) into G

remove any edge (x, y, d) with d ≠ distance from x to v in G

extract path (R_i..R_j) from ordered node fields in NodeTbl

Figure 3: Edge Sampling Algorithm

3. Simulation of IP Traceback Methods

Ns2 was used as our simulative tool. The network topology was constructed as a three layers tree with victim to be the root. As Shown in figure4 the basic assumptions made are that

1. The attacker may generate any number of packets and the packets may be lost or reordered during transit.
2. Multiple attackers may be involved and attackers may or may not be aware that they are being traced.
3. The path between attacker and victim is fairly stable.
4. Routers have limited CPU and memory constraints and are not widely compromised.

Assuming a marking probability p, set to 1/25, the experimental results for number of packets needed to reconstruct paths of varying lengths is as shown in figure 8.

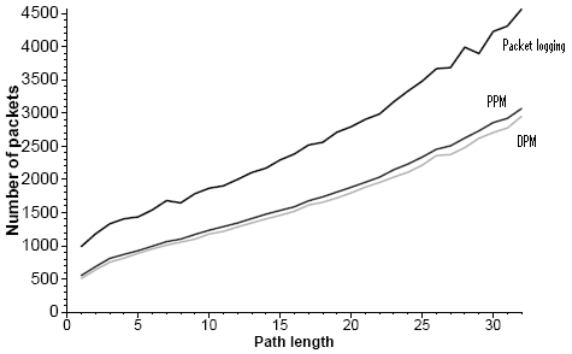

Figure 4: Experimental results for number of packets required for path reconstruction with marking probability set at 1/25.

While IP-level traceback algorithm could be an important part of the solution for stopping denial-of-service attacks, it is by no means a complete solution. These algorithms attempt to determine the approximate origin of attack traffic – in particular, the earliest traceback-capable router involved in forwarding attack traffic from the source that directly generated it. Finally, traceback is only effective at finding the source of attack traffic, not necessarily the attacker themselves. Stopping an attack may be sufficient to eliminate an immediate problem, but long term disincentives may require a legal remedy and therefore the forensic means to determine an attacker's identity (http://www.cisco.com/warp/public/707/newsflash.html). Even with perfect traceback support, unambiguously identifying a sufficiently skilled and paranoid attacker is likely to require cooperation from law enforcement and telecommunications organizations.

4. The Concept of SDR

We assume a two dimensional square grid with Routers at selected grid points [12]. The edge between 2 routers is thus a line in two dimensions whose directions are specified by its direction cosines (Cosα, Cosβ), where α,β are the angles made by the edge with positive E and N directions (refer fig.5). Direction cosines satisfy $\cos^2 \alpha + \cos^2 \beta = 1$, always.

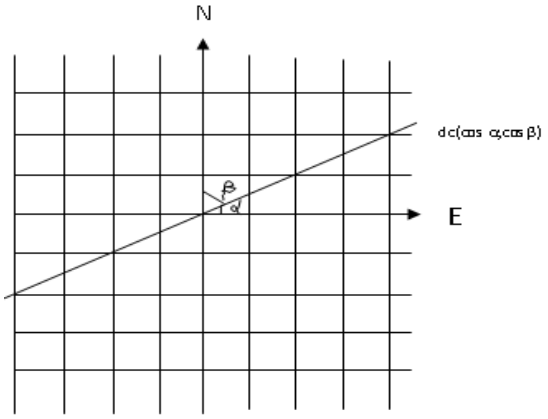

Fig 5: Square grid where an edge line has d.c (Cos α,Cos β).

Since most Cosθ values are cumbersome rationals and irrationals in [-1, 1], the concept of direction ratios (d.r) was introduced. Direction ratios (d.r) are proportional quantities to Direction cosines (d.c); are integers, denoted by (a,b) where in general $a2 + b2 \neq 1$. From direction ratio (a, b) we can get the directional cosine (cosα, cosβ) as (a/r, b/r) where r = $\sqrt{a^2+b^2}$. In fig1, the direction ratios of the line are (2, 1), from which we can recover the dc as (2/√5, 1/√5). By segment, we mean the edge between 2 adjacent routers, with coordinates (x1, y1),(x2,y2) with suitable origin O, and OE, ON as axes of reference. The coordinates are in units of the grid size. If AB is the edge joining 2 routers A, B with coordinates of A (x1, y1) and B(x2, y2) then SDR (Segment Direction Ratio) of AB are defined as (x2 – x1, y2-y1) where $|x2-x1|, |y2-y1| \leq 2$ and co primes (refer fig.6). In general for DGT of 2n directions we

handle SDR with $|x2-x1|,|y2-y1| \leq (n-2)$, and co primes for $n \geq 3$.

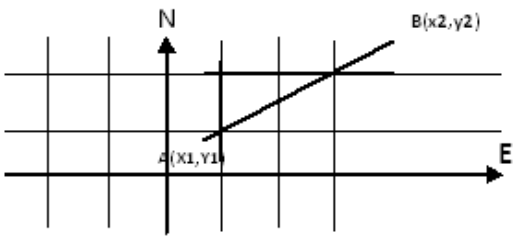

Fig 6: For edge AB between routers at A, B with SDR (x2-x1,y2-y1)=(2,1)

It is easy to see that (x2-x1),(y2 – y1) are only the grid steps to be taken in ± OE, ± ON directions (depending on the sign of SDR), to reach B from A. They are the projections of the edge AB on OE, ON with appropriate sign attached.

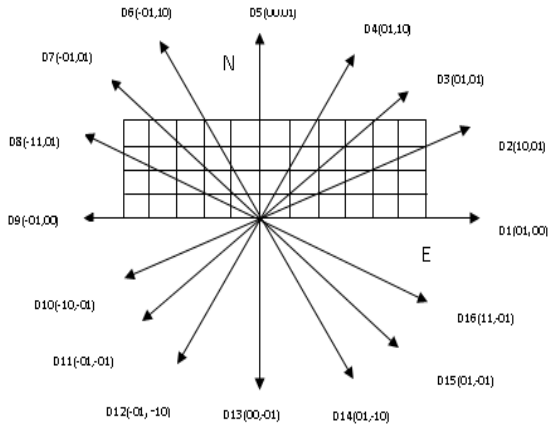

Fig:7 DGT 16 SDR

Fig 7, gives the 16 directions D1, to D16 (where D1 = OE, D5:= ON directions) with their SDR in bits.

The SDR of DGT 16 are given as ordered 2 bits with appropriate sign. It is easily verified that for such SDR (a,b); (a,-b), (-a, b), (-a,-b) are also SDR.

The assumptions of DGT2n for $n \geq 4$ are the same as in DGT8. The following basic assumptions are standard.

a. Any number of packets can be generated by an attacker.
b. Attackers are aware of trace attempts on them.
c. The routing behavior may be unstable.
d. Circuits routing is not there.
e. A router knows the SDR of its neighboring routers in one of the 2n directions ($n \geq 4$). Specifically for n=4, in the 16 directions D1 to D16.

Most of these assumptions are common to traceback schemes of one type or the other.

5. DGT16 Procedure With Encoding Requirements

When a packet arrives at router Ri and is destined for router Rj where the direction Dij, is one of D1 to D16 the only task that Ri, has to perform is to add the ordered SDR values of Dij, to the corresponding ordered subfields in the IP header and subtract 1 from the TTL value. Thus for the implementation of DGT16, we require 2 subfields in the IP header, to keep track of the cumulative grid step movements, from router to router, through their SDR.

In this way, when a packet arrives at the victim, the geographical location of the attack router can be obtained from the data in the SDR subfields, regardless of the source IP address which may be incorrect or compromised.

Assuming that the lengths of internet paths seldom exceed 32 hops, the cumulative SDR value cannot exceed in magnitude, the integer 64, for DGT16. Hence 2 (1+7) = 16 bits are needed in the IP header for the CSDR totals. To calculate the total number of hops between the attack router and the victim router, as the difference of initial TTL value and the final TTL value, we need to store the initial TTL value in the IP header. Assuming that the IP header has (16+8 +1) 25 bits, for DGT 16, we use the 8 bit segment for storage of initial TTL value. Location of the attacker and the hop count enables the victim to process the traceback.

6. Comparison of DGT16 with other Traceback Schemes

Comparison with DGT 8 : DGT16 and DGT8 being like schemes, offer equivalent advantages with respect to computational burden, scalability and mitigation capability of the attack, except for the fact that 16 directions are available now, with nil or negligible additional computations.

Qualitative comparison with other schemes like PPM and SPIE : DGT, PPM and SPIE being different types of trackback schemes only qualitative comparison is possible The inferences are same as those reported in [12] with respect to computational, scalability and capability parameters.

Limitations of DGT16 : A limitation of DGT16 is the inequality (though marginal) among the interfaces. This is the cost we have to pay to satisfy the integer requirements of the SDR and generalization to DGT2n.

7. Traceback Procedure:

We require an address field R, a direction ratio field DR, and a distance field S, in the packet header to implement this algorithm.

Assuming that the IP header has (16 + 8 + 1) = 25 bits, for DRSA, we can allot 10 bits each. For the address field, and DR Field and 5 bits for the distance field. This is acceptable since, routers are numbered serially; the 10 digit field can accommodate the last 3 digits of the serial number and is sufficient for R mod (1000). Since a 9 bit field is enough for the 4, 9 direction set of DR (2), 10 bits aare sufficient for the DR field. Since any IP path never exceeds 32 hops, a 5 bit distance field is taken at in Fig 7.

10	10	5
R mod (1000)	(a, b, c)	s
R Field (RF)	DR Field (DRF)	Distance Field (SF)

Fig 7: IP Header format for DRSA

Here is Ri: router at (xi, yi, zi) with a given serial number Dj = (aj, bj, cj) = an element of DR (2) indicating the direction ratio of the next router Rj (from Ri). Note that Ri (Rj) = Rj (the router from Ri in the direction Dj is the unique Rj since Dj is in (1 – 1) correspondence with Rj from a given Ri)

8. DRSA (Direction Ratio Sampling Algorithm)

The marking procedure at a router Ri of every packet w from the attacker is as follows:

Let x be a random number in (0, 1) and p is a chosen probability level. If x < p, then if the packet is unmarked, then write Ri mod (1000) in RF, Dj in DRF, 0 in SF. Otherwise (if the packet is already marked) or (x ≥ p) then only increment the distance field SF.

After sufficient number of samples are dream, then using the property Ri (Dj) = Rj and the distance field count, the attack path can be reconstructed. The victim uses the DR (along with R) sampled in these packets to create a graph leading back to the source (s) of attack.

9. Results and Discussions:

If we constrain p to be identical at each router, then the probability of receiving a marked packet from a router d hops array is $p (1-p)^{d-1}$ and this function is monotonic in the distance from the victim. Because the probability of receiving a sample is geometrically smaller, the further away it is from the victim, the time for this algorithm to converge is dominated by the time to receive a sample from the further router.

We conservatively assume that samples from all of the d routers (in the path from A toV) appear with the same likelihood as the furthest router. Since these probabilities are disjoint, the probability that a given packet will deliver a sample from some router is at least $dp (1-p)^{d-1}$ by addition law for disjoint events. As per the well known Coupon Collector problem [3], the number of trials required to select one of each of d equiprobable items .

From (6.1) we can show that E(X) is optimal if p = 1/d (ie dE / dp = 0, $d^2E / d^2p > 0$ for p = 1/d).

n	2n	SDR bit length	Max step moves	Max CSDR value	IP Header CSDR Length
3	8	1	1	32	2 (1+6)
4	16	2	2	64	2 (1+7)
5	32	2	3	96	2 (1+7)
6	64	4	4	128	2 (1+8)

For example, if p=1/d, where d= attack path length, then the victim can typically reconstruct the path after receiving

$E(x) = d^d \ln d / (d-1)^{d-1}$ packets for d=10; E(x)≤75 and hence a victim can typically reconstruct the path after receiving 75 packets from the attacker.

The concept of SDR allows us to extend the DGT 16 to DGT2n Table 1: DGT 2n Specifications nt manner.

The only additional requirement that arises is the increased CSDR upper limits and consequently more bits in the IP header, for the 2 subfields, are needed.

Specifically DGT2n restricts SDR of segment joining grid points A (x1,y1) and B (x2,y2) to the constraint of |x2-x1|,|y2-y1| being co primes and satisfying.

|x2-x1|,|y2 – y1 | ≤ n -2, (n ≥ 3), and imparts a corresponding increased requirement for the two CSDR maximum totals for an optimal 32 hop situation.

The SDR of the DGT32 scheme are given below IN Fig 8. These SDR with first or second or both components changed in sign give the SDR of the remaining directions, in Quadrants II, IV and III respectively.

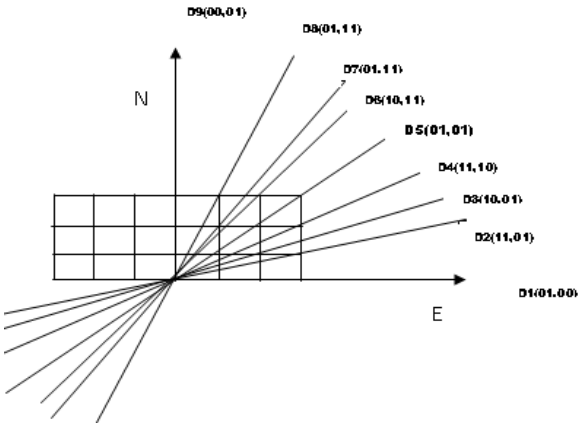

Fig 8 : DGT32 SDR in the directions D1 to D9 in quadrant I

Ultimately the number n of scheme DGT2n, depends solely on the IP header bit capacity as is evident from the following table.

8. CONCLUSION

The Internet has transformed from an information repository to a vital channel for conducting business. Unfortunately, with this positive change has come an increased frequency in malicious attacks (http://portal.acm.org/citation.cfm?id=1092845). All the proposed traceback schemes have their own specific advantages and disadvantages. Currently, no single solution could fulfill all the requirements outlined for an effective trace-back method [4]. For any of these IP traceback solutions to be effective, they would need to be deployed across corporate and administrative boundaries in a substantial portion of the Internet infrastructure. The Internet has transformed from an information repository to a vital channel for conducting business. Unfortunately, with this positive change has come an increased frequency in malicious attacks ([4]. All the proposed traceback schemes have their own specific advantages and disadvantages. Currently, no single solution could fulfill all the requirements outlined for an effective trace-back method [9]. For any of these IP traceback solutions to be effective, they would need to be deployed across corporate and administrative boundaries in a substantial portion of the Internet infrastructure. This in itself seems to be one of the biggest obstacles to a unified approach to IP traceback. Also, some measures are ineffective against DDoS attacks, are resource intensive, cause network overhead, and cannot be used for post-attack analysis. One conclusion we can draw from this is that unless IP traceback measures are deployed all over the Internet, they are only effective for controlled networks than for the Internet.

This same algorithm can efficiently discern multiple attacks. When attackers from different sources produce disjoint edges in the tree structure of reconstruction. The number of packets needed to reconstruct each path is independent of other paths.

The limitations imposed by restricting the number of DR to /DR (2)/=49 at every stage and using R (mod 1000) instead of the full serial number of router R are marginal in nature . We need more space in the packet header to use elements of DR (3) and the full representation of the R serial number. In conclusion DRSA is a robust scheme of 3 dimensional, multi-directional, geographical IP trace back

The authors are working towards to extend this multidirectional geometrical two dimensional traceback scheme to three dimensions.

REFERENCES

[1]. Yaar etc., FTT : Fast Internet Trackback., IEEE INFOCOM'05, Miami, Florida, Mar. 2005.

[2]. Al – Duwairi B., etc., Topology Based Packet Marking, IEEE int. Conf. Computer comm.. and Networks (ICCN) Oct. 2004.

[3]. Basheer Al-Duwairi etc., Novel Hybrid Schemes Employing Packet Marking and bagging for IP Traceback, IEEE Transactions on Parallel and Distribution Systems, Vol 17. No5. Pp 403 – 418, May 2006.

[4]. Meadows. "A formal framework and evaluation method for network denial of service." In Proceedings of the 12th IEEE Computer Security Foundations Workshop, June 1999. (8 and 10)

[5]. Cisco. "Strategies to Protect Against Distributed Denial of Service Attacks." http://www.cisco.com/en/US/tech/tk59/technologies_white_paper09186a0080174a5b.shtml

[6]. Cisco. "Strategies to Protect Against Distributed Denial of Service Attacks." http://www.cisco.com/warp/public/707/newsflash.html

[7]. Computer Incident Advisory Capability. *Network Intrusion Detector Overview*. http://ciac.llnl.gov/cstc/nid/intro.html

[8]. D.X. Song, and A. Perrig, Advanced and Authenticated Marking Schemes for IP Traceback, IEEE INFOCOM'01 Anchorage AK, AP 2001, pp 878 – 886.

[9]. S. Savage, D. Wetherall, A. Karlin, and T. Anderson. Practical Network Support for IP Traceback. In Proceedings of ACM SIGCOMM 2000, August 2000.

[10]. S. Savage, D. Wetherall, etc., Practical Network Support for IP Traceback IEEE / ACM transactions. Networking Vol 9 – pp 226 – 237, Jun 2001.

[11]. V. Padmanaban and L.Subramanian., An Investigation of Geographic Mapping Technologies for Internet Hosts, ACM SIGCOMM01. San Diego., 2001, pp. 173 – 185.

[12]. Zhiqiang Gao and Nirwan Ansari. "Directed Geographical Traceback", IEEE, transactions. IEEE paper 221-224, 2005.