Efficient Traceback of DoS Attacks using Small Worlds in MANET

Yongjin Kim, Vishal Sankhla, Ahmed Helmy[1]
Department. of Electrical Engineering, University of Southern California, U.S.A
{yongjkim, sankhla, helmy}@ceng.usc.edu

Abstract

Mobile Ad hoc NETwork (MANET) is an increasingly promising area of research with many practical applications. However, MANET is vulnerable to a number of attacks including Denial-of-Service (DoS) attacks due to its autonomous nature. DoS attacker traceback is challenging issue in MANET since each node works as an autonomous terminal, acting as both host and router. Mobility of nodes in MANET makes problem even worse since it is hard to trace back attacker when they are moving around frequently changing network topology. We propose to use an efficient on-the-fly search technique to trace back DoS attackers. Our scheme is based on small world concept and effectively extends *Contacts* [3] for MANET utilizing location information. In addition, to deal with address spoofing problems in DoS attacks, we use *Traffic Patterns Matching (TPM)* [5] and propose to use *Traffic Volume Matching (TVM)* as matching-in-depth to identify an attacker. We also processing *in-network processing and directional expanded ring search* to reduce communication overhead in attacker traceback. We show that our scheme successfully trace back attacker using both TPM and TVM. In addition, we show our scheme incurs very low communication overhead.

1 INTRODUCTION

Denial-of-service (DoS) attacks consume the resources of a remote host or network, thereby denying or degrading service to legitimate users. Such attacks are among the most intricate security problems to address because they are easy to implement, difficult to prevent, and very difficult to trace. The most common DoS include attacks similar SYN Flood, Smurf, UDP Flood. Determining the source generating attack traffic is especially difficult when using stateless routing protocols (as in the Internet or geographic routing). Attackers routinely disguise their location using incorrect, or "spoofed", source address.

There are many IP traceback scheme proposed for the Internet such as link testing, packet marking, logging, ICMP traceback, etc [1],[5],[6],[7].

Such traceback schemes are not directly applicable to Mobile Ad Hoc NETwork (MANET) due to the following reasons.

- In MANET, there is no fixed infrastructure. Each node works as an autonomous terminal, acting as both host and a router.
- Each node moves in and out, frequently changing network topology.
- Network bandwidth and battery power are limited.

To perform efficient DoS attacker traceback under such a harsh environment in MANET, we propose to use an efficient on-the-fly search technique. For that, we propose to use the small world concept. Establishing a small world reduces the degrees of separation between the attacked node (victim) and the attacker and provides an efficient traceback mechanism. Helmy[1] et,.al [3] established the applicability of small world graphs to wireless networks. In this paper, we effectively extend *Contacts* [3] for MANET utilizing location information. By using location information, we can optimally select Contacts reducing coverage overlap and construct a small world to identify and trace attackers with reduced communication overhead.

In addition, to deal with address spoofing problems, we use *traffic patterns* [5] to identify an attacker. We also propose *traffic volume matching* to complement the traffic pattern matching. We call this *matching-in-depth*. A traffic pattern is defined by the sequence of number of packet in a time slot at each node.

We also use *in-network processing and directional expanded ring search* to reduce communication overhead. Our paper is organized as follows. In section 2, we provide related work on DoS attack traceback in the Internet. In section 3, we introduce our Contact-based DoS traceback architecture. We show simulation result in section 4. In section 5, we conclude our paper present future works.

[1] Ahmed Helmy was supported by NSF Career, Intel and Pratt & Whitney PWICE Institute

0-7803-8521-7/04/$20.00 © 2004 IEEE

2 RELATED WORKS

There are two existing approaches to the problem of determining the route of a packet flow in the Internet: one can audit [5],[6],[7] the flow as it traverses the network, or one can attempt to infer the route based upon its impact on the state of the network [1].

Route inference was pioneered by Burch and Cheswick who considered the restricted problem of large packet flows and proposed a novel technique that systematically floods candidate network links. By watching for variations in the received packet flow due to the restricted link bandwidth, they are able to infer the flow's route. This requires considerable knowledge of network topology and the ability to generate large packet floods on arbitrary network link. One can categorize auditing techniques into two classes according to the way in which they balance resource requirement across the network components. Some techniques require resources at both the end host and the routing infrastructure; others require resources only within the network itself. Of those that require only infrastructure support, some add packet processing to the forwarding engine of the routers while others offload the computation to the control path of the routers.

Both approaches are not feasible and inefficient in MANET since they consume significant bandwidth/power and each node moves around frequent changing network topology.

3 CONTACT-BASED TRACEBACK ARCHITECUTRE

3.1 *Design Requirements*

The design requirements for efficient traceback in MANET include:

(I) *Robustness to mobility*: The mechanism should be robust to handle frequent mobility. That is, we should be able to trace an attacker despite of frequent intermediate node mobility.

(II) *Robustness to address spoofing*: It is a common attacking technique to spoof addresses. We should be able to trace an attacker in spite of address spoofing.

(III) *Scalability*: Applications of large-scale ad hoc networks involve military and sensor network environments that may include thousands of nodes. Hence traceback mechanism should be scalable in term of communication overhead with increase in network size.

(IV) *Efficiency*: Ad hoc networks include portable devices with limited battery power. Traceback mechanism should be power-efficient.

(V) *Decentralized operation*: For the network to be rapidly deployable, it should not require any centralized control.

3.2 *Architecture and Definitions*

Each node maintains information only about its *Vicinity* using very limited broadcasts within square from the node. Unlike [3], a node does not need to maintain information about a set of nodes, called *Contacts*, beyond the vicinity. In our scheme, Contact is selected using location information, which further reduces communication overhead. Ideal location of Contacts is selected first. Then, nodes closest to the ideal Contacts are selected as Contacts. For instance, in fig.1, there are 8 ideal locations of Contacts and nodes which are the closest to the each ideal Contact locations in each rectangle are selected as actual Contacts. Contacts of a node are called level-1 Contacts. Contacts of the Contacts are called level-2 Contacts, and so on. During a search for the attack traffic pattern in the wireless network a node queries its Contacts, and their Contacts, so on, up to level-*D* Contacts. *D* is called the depth-of-search.

When each Contact performs limited broadcast, they send queries to neighbor nodes specifying its rectangle region. When, neighbor nodes receive the query, they check whether they are in the square region or they have already received the same query from other nodes. If they are outside the square region or they have already received the same query, they discard the query. Otherwise, they broadcast the query to their neighbors.

The attack *Traffic pattern* is defined by the variation of packet number over time. For instance, when the number of data is *m* for a time window, traffic pattern is expressed as $A= (A_1,A_2,A_3,\dots,A_m)$. In DoS attack, large amount of packets is generated towards the victim. For instance, 200-500 pps of SYN packets are generated [1]. However, in normal case, only one SYN packet is generated per connection. Accordingly, a large amount of SYN packets can be suspected as attack.

The queried nodes are asked to perform a TPM to determine the correlation coefficient between two traffic pattern (A,B). In case correlation coefficient of (A,B) is high (greater than 0.7), the traffic A is said to match traffic B (fig.2). For instance, when traffic pattern observed at node i is given as $L_i=(n_1,n_2,\dots,n_N)$, and traffic pattern observed at node j is $L_j=(m_1,m_2,\dots,m_N)$, correlation coefficient is obtained as follows.

$$r(A,B)=\frac{1}{nS_iS_j}\sum_{i=1}^{n}(L_i(k)-\overline{A})(L_j(k)-\overline{B})$$

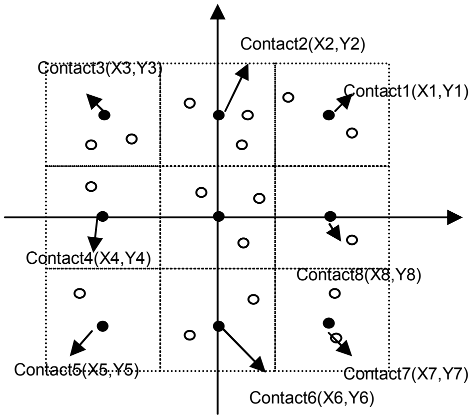
(Eq.1)

where,

$$S_i=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(L_i(k)-\overline{A})^2} \qquad \text{(Eq.2)}$$

$$S_j=\sqrt{\frac{1}{n}\sum_{j=1}^{n}(L_j(k)-\overline{B})^2} \qquad \text{(Eq.3)}$$

, and $\overline{A}$ & $\overline{B}$ is the average of data, L_i and L_j.



[Figure.1] Location-based Contact selection

We propose to use TVM to complement the traffic pattern matching. We define that traffic volume is matching between two points, when L_i and L_j shows similar traffic volume size. Mathematically, we use the following equation (least-squares method) to know the matching level.

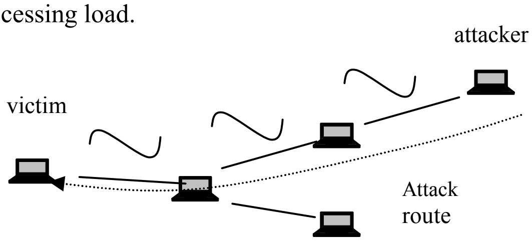
$$a_{ij}=\frac{\sum_{k=1}^{N}L_i(k)L_j(k)}{\sum_{k=1}^{N}L_i(k)^2}$$

(Eq.4)

When, the a_{ij} is close to 1, the traffic volume is matching. Traffic volume matching is necessary for correct traceback in MANET since other node can show high correlation coefficient under heavy background traffic. By checking TVM level as well as TPM level (we call this matching-in-depth), we can reduce false positives in our trace back. Note that mere traffic volume matching is also not enough since traffic volume can fluctuate showing different traffic volume in each node depending on background traffic.

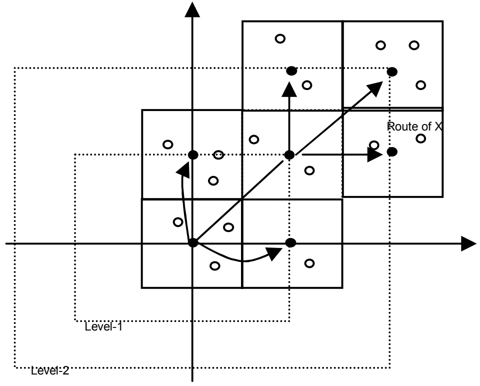
3.3 *Mechanism Description*

Each node monitors the traffic pattern/volume for a certain time window. The time varies based on the attack type. A node keeps only the variance of packet number over time, which reduces the processing load.



[Figure.2] Traffic pattern based traceback

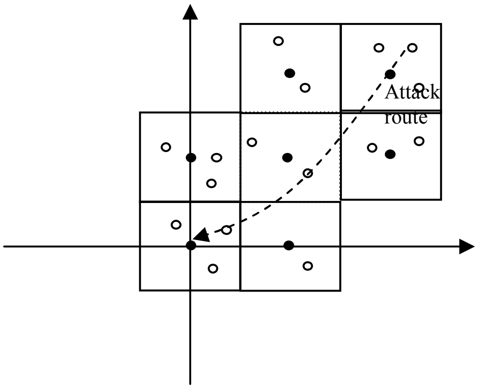
We describe the traceback scheme as follows: (1) When a victim node, *s*, detects attack such as SYN flooding in application level, it sends query to nodes within vicinity and level-1 Contact specifying depth of search (*D*) which is large enough to detect an attacker. We use greedy forwarding to send a query to Contact. In case of local maximum, perimeter mode [3] is used to take a detour. (2) In case a traffic pattern/volume matching report is observed by victim and other nodes, first step of trace is competed. For instance, we send query to 3 level-1 Contacts around the victim. (Fig.3.) Then, one level-1 Contact reports that some of its vicinity nodes observed matching traffic pattern/volume.



[Figure.3] Queries to level-1 Contacts

(3) Next, only the Contacts that observe matching traffic patterns in their vicinity send next level query to level-2 Contacts with the path from victim after reducing *D* by 1. Other Contacts stop forwarding the query *(In-network processing).* In doing so, we can perform *directional expanded ring search.* (4) When there are no more Contact reports, last Contact report to the victim the complete attack route (Fig.4.).

Our scheme is based on majority node report. That is, even if some nodes move out from the attack route, we can still find an attack route. Response after tracing back the attacker may include filtering, rate limiting, re-organizing to preclude the compromised node, or blacklisting.

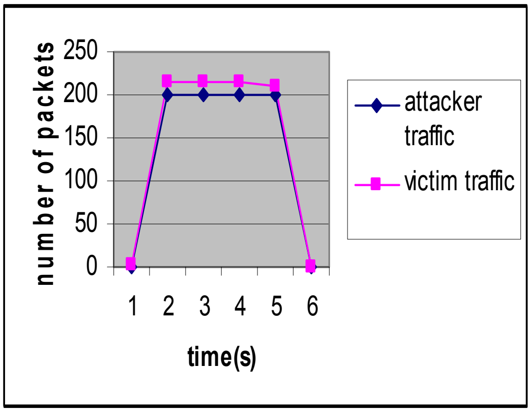


[Figure.4] Attack route

4 SIMULATION RESULTS

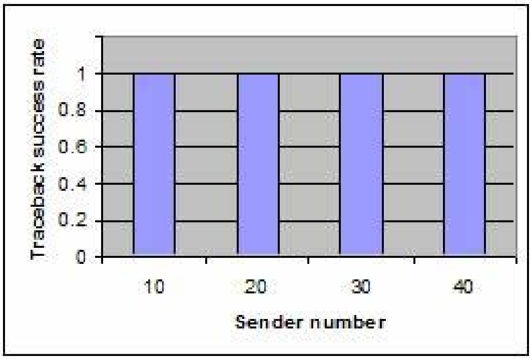
We performed simulation to investigate the design space parameters and evaluate the performance of our protocol. We put 1000 nodes in 1000m x 1000m areas and transmission range is taken as 250m. We used greedy forwarding as a routing protocol. Note that our scheme is generally applicable to other ad-hoc routing protocols (e.g., DSR, AODV). As attack traffic, we used SYN packets and 200 pps traffic was generated from attacker to victim. Geographic locations of all nodes are randomly chosen inside the region. Background traffic is generated randomly among [0,10] pps. Since background traffic can impact on the correct traceback of attacker, we varied the number of senders that generate SYN packet in a given time window and evaluated the impact of background traffic on correct traceback.

Figure 5 shows the traffic pattern taken at the attacker and victim node. Random number of background SYN traffic is generated by randomly chosen 50% of total nodes (i.e, 500 nodes) at every second. Traffic sample is taken every second. At 2 second, we can observe sudden SYN packet increase. We have sampled the traffic pattern when it goes up more than 100 packets (Th_{up}) per second. When, the traffic goes down below ($Th_{up}/2$), we stopped sampling traffic pattern. In figure 4, we can observe very similar traffic pattern between victim node and attacker node.

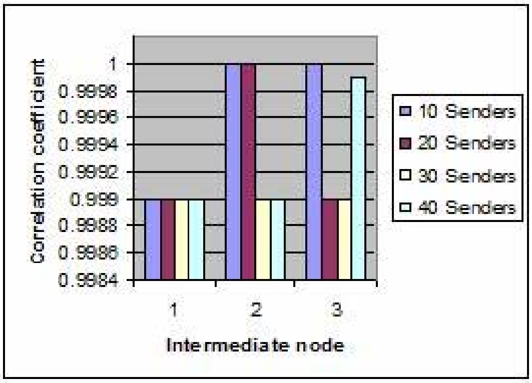


[Figure.5] Sample traffic pattern comparison between attacker and victim

Figure 6 shows successful traceback rate with small background traffic (e.g., up to 4% of nodes generates background traffic). It shows 100% traceback success rate with only TPM method. The correlation coefficient of intermediate nodes located between attacker and victim node shows high value over 0.9 (fig.7.). In this case, background traffic volume is very low, so we could obtain high correlation among victim node and intermediate node. In addition, no other node in vicinities showed high correlation coefficient (greater than 0.7) except the nodes which attacker's packet have traversed.

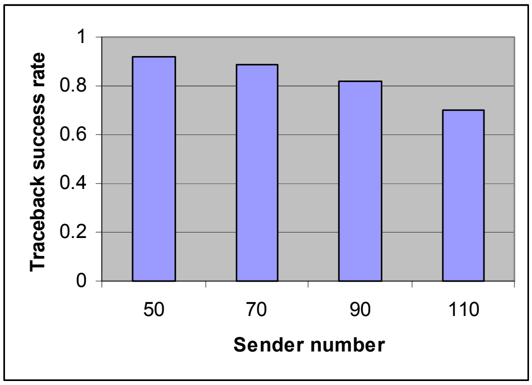


[Figure.6] Success rate with low background traffic



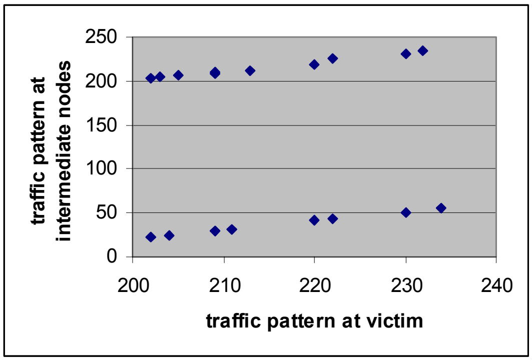
[Figure.7] Correlation coefficient with low background traffic

Figure 8 shows traceback success rate when varying rather high volume of background traffic. In this case, traffic volume matching becomes necessary since other nodes show high correlation due to heavy background traffic. In case of using only TPM method, traceback success rate goes down as background traffic increase.



[Figure.8] Success rate with high background traffic

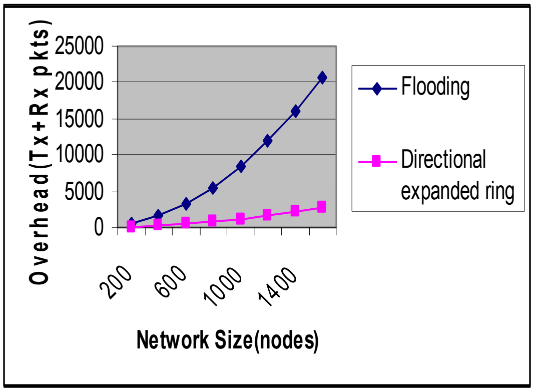
It is because of *clustering effect* as shown in figure 9. Both clustering show high correlation coefficient.



[Figure.9] Traffic volume correlation

We can separate the clustering by traffic volume matching using equation (4). We set 0.5<r<1.5 (Proper value of low bound and high bound depends on background traffic volume). Note that this is much computationally lighter compared to clustering algorithms such as K-means method [2]. Our simulation showed that by using both TPM and TVM, traceback success rate becomes 100%.

We compared our proposed scheme to flooding in terms of query overhead. Figure 10 shows query traffic generated. 1000m x 1000m area, 1000 nodes, and 80 meter transmission range was used in the simulation. Overhead includes transmission as well as reception packet number. As we can expect, our query scheme incurs much less overhead since our scheme performs directional expanded ring search. As network size becomes bigger, the difference becomes significant.



[Figure.10] Overhead comparison

5 CONCLUSIONS AND FUTURE WORK

Our Contact-based DoS attacker traceback mechanism in MANET has the following advantages: (I) By using Contacts/directional expanded-ring-search/in-network processing, we can effectively reduce communication overhead to trace an attacker. (II) Using the traffic pattern enables us to find attack routes efficiently with reduced processing load even if the node address is spoofed. (III) Even under mobility of intermediate nodes, we can trace back by utilizing less mobile nodes along the attack route. In the future, we will perform simulation with different mobility model to verify the efficiency of our scheme under dynamic topology change.

[REFERENCES]

[1] H. Burch, et al, “Tracing Anonymous Packets to Their Approximate Source”, Proc. 2000 USENIX LISA Conf., pp.319-327, Dec. 2000

[2] V. Guralnik and G. Karypis, Workshop on Data Mining in Bioinformatics (2001) 73-80

[3] A.Helmy, et al, “A Contact-based Architecture for Resource Discovery in Ad Hoc Networks”, ACM Baltzer MONET Journal, 2004

[4] B. Karp, T. Kung, “GPSR: Greedy Perimeter Stateless Routing for Wireless Networks” ACM Mobicom, Aug. 2000

[5] G.Mansfield, et al., “Towards trapping wily intruders in the large”, Computer Networks, Vol.34, pp.650-670, 2000

[6] Alex C. Snoeren, et al, “Hash-Based IP Traceback”, ACM SIGCOMM, 2001

[7] Stefan Savage, et al., “Practical Network Support for IP Traceback” , ACM SIGCOMM, 2000