International Journal On Engineering Technology and Sciences – IJETS™
ISSN (P): 2349-3968, ISSN (O): 2349-3976
Volume 2 Issue 4, April -2015

Multivariate Correlation Analysis based detection of DOS with Tracebacking

Jasheeda
P Student
Department of CSE
Kathir College of Engineering
Coimbatore
jashi108@gmail.com

T.K.P.Rajagopal
Associate Professor
Department of CSE
Kathir College of Engineering
Coimbatore
tkprgrg@gmail.com

R.Subathra
Head /Department of CSE
Kathir College of Engineering, Coimbatore
rsubatra2000@yahoo.co.in

Abstract—**Denial of service (DoS) attack is a serious threat to the security of cyberspace. It typically exhausts bandwidth, processing capacity, or memory of a targeted machine or network. This paper proposes a novel trace back method for DoS attacks that is based on entropy variations between normal and DoS attack traffic, which is fundamentally different from commonly used packet marking techniques. DoS attacks are detected using multivariate correlation analysis. Using traceback method it analyses the source of the attack. This strategy is efficiently scalable, robust against packet pollution, and independent of attack traffic patterns. Once the DoS attack detection algorithms raise the alarm of a potential attack, it starts to calculate the distance among the different suspicious flows in the community network.**

***Index Terms*—Denial of Service, Traceback, Multivariate correlation analysis**

I. INTRODUCTION

DoS attacks attempt to exhaust the victim's resources. These resources can be network bandwidth, computing power, or operating system data structures. To launch a DoS attack, malicious users first build a network of computers that they will use to produce the volume of traffic needed to deny services to computer users. To create this attack network, attackers discover vulnerable sites or hosts on the network. Vulnerable hosts are usually those that are either running no antivirus software or out-of-date antivirus software, or those that have not been properly patched. Vulnerable hosts are then exploited by attackers who use their vulnerability to gain access to these hosts. The next step for the intruder is to install new programs (known as attack tools) on the compromised hosts of the attack network. The hosts that are running these attack tools are known as zombies, and they can carry out any attack under the control of the attacker. Many zombies together form what we call an army.

DoS attack detection is essential to the protection of online services. Network-based detection mechanisms are widely used. Network–based detection systems[1] are classified into misuse-based detection systems and anomaly-based detection systems[2].Due to various drawbacks of misuse-based detection systems, anomaly based detection systems are widely used. Since spoofed packets are used for DoS attack, it is difficult to find out the route of attack. An effective method for tracebacking is also necessary.

II. RELATED WORKS

Even though various DoS attack detection mechanisms are available, one with effective tracebacking is necessary. Anomaly based network intrusion detection techniques are a valuable technology to protect target systems and networks against malicious activities [2]. Anomaly-based detectors attempt to estimate the ‘‘normal’’ behavior of the system to be protected, and generate an anomaly alarm whenever the deviation between a given observation at an instant and the normal behavior exceeds a predefined threshold. Another possibility is to model the ‘‘abnormal’’ behavior of the system and to raise an alarm when the difference between the observed behavior and the expected one falls below a given limit.

Tan et al. [3] proposed a system which applies the idea of Multivariate Correlation Analysis (MCA) to network traffic characterization and employs the principal of anomaly-based detection in attack recognition. This makes the solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. Furthermore, a triangle area technique is proposed to enhance and speed up the process of MCA. Traffics are monitored at destination. Anomaly based detectors, sample by sample detection, multivariate correlation based method along with Triangle Area Map generation are used to recognize the malicious users.

Security community does not have effective and efficient trace back methods to locate attackers as it is easy for attackers to disguise themselves by taking advantages of the vulnerabilities of the World Wide Web, such as the dynamic, stateless, and anonymous nature of the Internet. The memory less feature of the Internet routing mechanisms makes it extremely hard to trace back to the source of these attacks. As a result, there is no effective and efficient method to deal with this issue so far. Number distributions of packet flows, which will be out of control of attackers once the attack is launched, and found that the similarity of attack flows are much higher than the similarity among legitimate flows. An approach, based on *ICMP messaging* [4], is to have each router X decide, with some probability q (typically = 1/20000 is mentioned), for each packet P to send an additional ICMP packet to the destination, which identifies X and some content of P. The main idea of this approach is that during a DDOS, a sufficient amount of attack packets will trigger ICMP messages from the routers in the attack tree T so that the victim can identify the leaves of T from these messages. The main drawback of this approach is that it causes additional network traffic even when no DDOS is present. Also it is not efficient, for identifying the n leaf nodes in the attack tree T.

Some researchers have advocated a logging approach to the IP traceback problem. In a logging solution, we either ask routers to log the packets they process or we augment the data packets themselves to contain a full log of all the routers they have encountered on their way to their destinations. Stone [5] and Baba and Matsuda [6] advocate logging of packet information at the routers, and Snoeren *et al.* [7] propose the logging of message digests of packets at the routers. The drawback with these approaches is that they require additional storage at the routers.

Michael T. Goodrich [8] proposed IP traceback based on the probabilistic packet marking paradigm called *randomize-and-link*, uses large checksum *cords* to "link" message fragments in a way that is highly scalable, for the checksums serve both as associative addresses and data integrity verifiers.
The DPM method [9] requires all the internet routers to be updated for packet marking. The DPM mechanism poses an extra ordinary challenge on storage for packet logging for routers. DPM require update on the existing routing software which is extremely hard to achieve on the internet. The DPM tries to spare space of a packet with the packet's initial router information. Therefore the receiver can identify the source location of the packets once it has sufficient information of the marks. The major problem of DPM is that it involves modification of the current routing software and it may require large amount of marks for packet reconstruction.

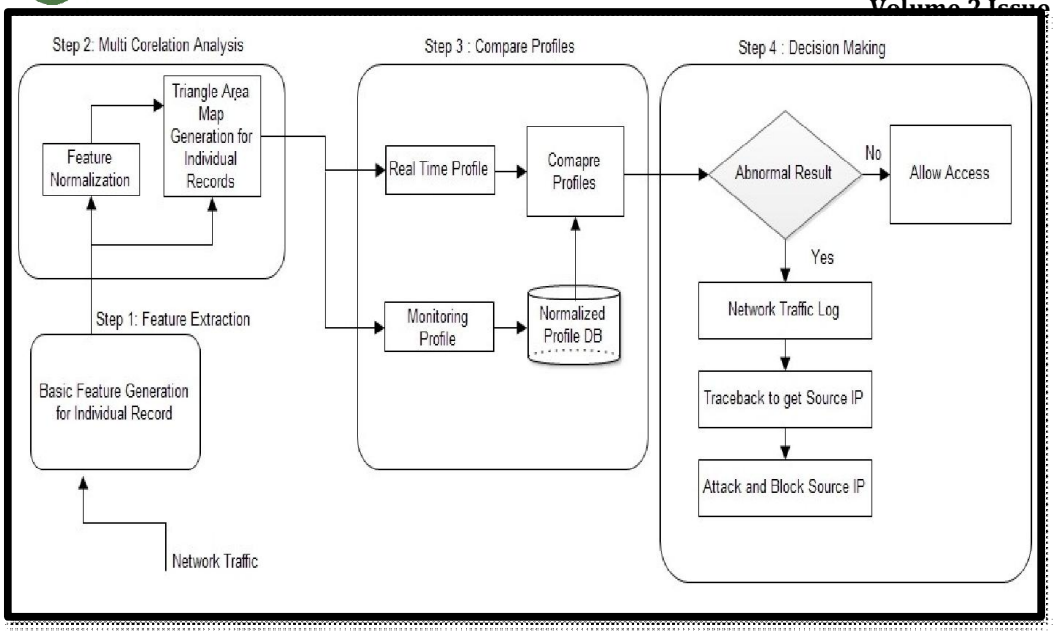

Fig.1 Architecture for DOS detection and tracebacking

III. PROPOSED SYSTEM

Architecture for proposed denial of service (DOS) attack detection and tracebacking mechanism is shown in fig.1. This system consists for four major steps. Sample by sample detection mechanism is involved in it.

At the first step basic features are generated from ingress network traffic to the internal network where protected servers reside in and are used to form traffic records for a well defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic.

Step2 provides multivariate correlation analysis, in which the "triangle area map generation" module is used to extract the correlations between two distinct features within each traffic record coming from first step or the traffic record normalized by the "feature normalization" module in this step 2. The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify intrusive activities. All the extracted correlations, called, triangle areas stored in triangle area maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to differentiate between legitimate and illegitimate traffic records.

Step 3 provides a comparison among real time profile and monitored profile based on anomaly based detection mechanism [2] to identify the anomalies among the records. Step 4 provides a decision based on step3.If there is no anomalies between real time profile and monitored profile then the user is legitimate one and access is provided. If there are anomalies greater than a specific threshold then the user is identified as a DoS attacker.

If the user is an attacker it is obvious that his identity is spoofed. To get his identity and to block the particular user, certain measures are necessary. An efficient IPtracebacking algorithm is installed at routers. The proposed detection system has detected attacks in routers and then the proposed trace back algorithm calculates information distances based on difference of their local traffic and the forward traffic from their immediate upstream routers, and also will find that there are no attacks in LAN and subsequent LANs. Therefore on routers, the proposed algorithm calculates continually information distances based on variations of their local traffic and the forward traffic from their immediate upstream routers. This can find there is an attack (zombie) in LAN, so the router will stop forwarding the traffic from the zombie immediately.

As explained the proposed system detects the DOS attack efficiently and traceback the source of

attack. Multivariate Correlation Analysis (MCA) based on Triangle Area Map[1][3][4] is used for generating the normal profile and a detection algorithm based on Mahalanobis distance[1] is used to detect the attack. After finding the attack a tracebacking algorithm will run at routers to find the source of attack.

A. Algorithm for Normal profile generation based on triangle-area-map

Require: with g elements

1. ←
2. Generate covariance matrix *Cov* for
3. for i=*1* to *g* do
4. ← MD (, .)
5. end for
6. μ← – normal, i
7. σ← ——— 2
8. *Pro*← (N (μ, $\sigma^{2)}$, , *Cov*)
9. return *Pro*

Where generated lower triangles of the TAMs of the set of *g* legitimate training traffic records, are the legitimate training traffic records is the expectation of g legitimate training traffic records, MD is the Mahalanobis distance which is adopted to measure the dissimilarity between traffic records and described through mean μ and the standard deviation σ , *Pro* is the normal profile generated which contains the obtained distribution *N (μ, $\sigma^{2)}$* of the normal training traffic records, and *Cov*.

B. Algorithm for attack detection based on Mahalanobis distance

Require: Observed traffic record ,normal profile *Pro* :(N (μ, σ^2), , *Cov*) and parameter α

1. Generate for the observed traffic record
2. ← MD (,)
3. if μ σ α ≤ μ σ α then
4. return Normal
5. else
6. return Attack
7. end if

A specific Threshold which is μ σ α is used to differentiate attack from the legitimate one. For a normal distribution α is usually ranged from 1 to 3.If the MD between an observed traffic records and the respective normal profile is greater than the threshold, it will be considered as an attack.

C. IP Traceback algorithm in DoS attack detection

The below algorithm helps us to trace the source of attack.

```
IP_Traceback_Algorithm ()
{
  while(true)
   call Check_ForwardTraffic(0)//Checks attack
   on router R0(or victim)
}
Check_ForwardTraffic
{
   Calculate information distance
   if        > σ
   Call Check_LocalTraffic
   for j= 1 to n
      k= the ID of the jth immediate upstream
      router of
      call Check_ForwardTraffic
   end for
   end if
}
Check_LocalTraffic
{
   Calculate information distance
   if        σ
     stop forwarding the attack traffic to
     downstream routers(or destination),label
     the zombie
   end if
}
```

The proposed traceback algorithm calculates information distances based on variations of its local traffic and the forward traffic from its immediate upstream routers. If the information distance based on its local traffic is more than the

specific detection threshold σ , the proposed detection system detects an attack in that LAN; this means that the detected attack is an internal attack. The proposed algorithm calculates continually information distances based on variations of their local traffic and the forward traffic from their immediate upstream routers, and then can find there is an attack (zombie) in LAN so the router will stop forwarding the traffic from the zombie immediately.

IV CONCLUSION

This paper has presented an efficient and effective method to detect DoS attack based on Multivariate Correlation analysis along with proper tracebacking to find the source of attack. This system is able to distinguish both known and unknown DoS attacks from legitimate network traffic. The proposed IP traceback scheme based on information metrics can effectively trace all attacks until their own LANs (zombies).

REFERENCES

[1] Zhiyuan Tan, Aruna Jamdagni, Xiangjian He, Priyadarsi Nanda, Member, and Ren Ping Liu, "A System for Denial of Service Attack Detection based on Multivariate CorrelationAnalysis"*IEEE transactions on parallel and distributed systems*,vol.25, no.2, 2014

[2] P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez,and E.Vzquez, "Anomaly-Based Network Intrusion Detection: Techniques,Systems and Challenges," Computers and Security, vol. 28, pp. 18-28, 2009.

[3] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R.P. Liu, "Triangle-Area-Based Multivariate Correlation Analysis for Effective Denial-of-Service Attack Detection," Proc. IEEE 11th Int'l Conf. Trust,Security and Privacy in Computing and Comm., pp. 33-40, 2012.

[4] S. M. Bellovin. ICMP traceback messages. In *Work in Progress, Internet Draft draft-bellovin-itrace-00.txt*, 2000

[5] R. Stone. Centertrack: An IP overlay network for tracking DoS floods.In *Proc.of 9th USENIX Security Symposium*,August 2000.**Michael**

[6] T. Baba and S. Matsuda. Tracing network attacks to their sources. *IEEEInternet Computing*, 6(2):20–26, 2002.

[7] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio,S. T. Kent, and W. T. Strayer. Hash-based IP traceback. In *Proc. Of the ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, 2001

[8] Michael T. Goodrich: Probabilistic Packet Marking for Large- Scale IP Traceback, IEEE/ACM transactions on networking,vol. x, no.x,2007