INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS IN ENGINEERING, TECHNOLOGY AND SCIENCES (IJ-CA-ETS)

AN ANOMALY DETECTION SYSTEM FOR DDOS ATTACK IN GRID COMPUTING

[1]Sumit kar, [2]Bibhudatta sahoo

[1,2] NIT Rourkela/CSE, Rourkela, India
sumitk-cs209@nitrkl.ac.in, bdsahu@nitrkl.ac.in

ABSTRACT:
Grid computing is rapidly emerging as a dominant field of wide area distributed computing. Grid computing is a collection of heterogeneous computers and resources across multiple organizations and delivers computing and resources as services to its users. The heterogeneity and scalability characteristics of Grid introduce potential security challenges. Distributed Denial of Service attack (DDoS) is one of the major threats to grid computing services. The perfect secure system for DDoS attack is based on the 3 steps: (i) Attack prevention, (ii) attack detection and recovery, and (iii) attack identification. This paper presents vulnerability of Grid computing in presence of DDoS attack. Our proposed method is based upon attack detection and recovery, and uses an Entropy based anomaly detection system to detect DDoS attack. A grid topology model is used to describe how to implement the entropy based anomaly detection system in grid environment.

Keywords: Grid computing, anomaly detection, entropy

1. INTRODUCTION

The Grid technology has emerged as a high throughput wide area distributed computing. Grid computing [1] [2] provides an infrastructure that supports the sharing and coordinated use of heterogeneous computers and resources spread across multiple administrative domains. Due to the dynamic and multi-organizational nature of grid, the issue of managing security of both users and resources are the most challenging. First generation grid was deployed in research labs, academic institutions and for military use. But now a day many enterprises are beginning to use grid technologies commercially as well [20], so maintaining the QoS and security are important to meet user's demand. Grid uses internet as an infrastructure to build communication and middleware like Globus are used that enable resource providers to make available their services for users. The fusion of web services and grid technologies further increases the concerns about security problem for their complex nature [17]. A good classification of possible threats in grid computing can be found in [6], which is based on threats on different users associated with Grid. From which Distributed Denial of Service attack (DDoS) is an immense threat to grid computing. For example Sun's new on-demand grid computing service becomes a victim with a denial-of-service (DOS) attack on its first day of operation [20]. The reminder of the paper is organized as follows section 2 discusses related work. section 3 presents vulnerability of DDoS attack in Grid computing and describe the 3 steps used in our proposed defense system to secure Grid. Section 4 outlines the need of an anomaly detection system in Grid and discussed how entropy based approach can be suitable for it. In section 5 we describe entropy based anomaly detection approach and our detection algorithm. In section 6 we have described using a grid topology model how it can implemented in Grid and conclusion in section 7.

2. RELATED WORK

The vulnerabilities of grid environment in the presence of DDoS have been presented in [10] and they have proposed a distributed defense system for Grid. Authors of [11] discussed the need for an intrusion detection system in grid environment. They have classified grid intrusions in to 4 types i.e. 1) Unauthorized access 2) Misuse 3) Grid exploit 4)Host or Network-specific attacks. They have proposed a model that is composed of high-level GIDS that utilizes functionality of lower-level HIDS (host intrusion detection system) and NIDS (network intrusion detection system) provided through standard inter-IDS communication. Different techniques and challenges involved in anomaly detection system can be found in [14]. Many articles like [13] use traffic volume [flow, packet, byte count] as the metric for anomaly detection system. Volume based detection scheme were proved as a good metrics, But like small DoS attacks that do not cause much changes in traffic can not be detected perfectly. Recently there has been use of entropy and traffic distribution for detecting DDoS attack anomalies. Author [3] uses entropy of distribution of source address for DDoS detection. In [4] included PCA framework with entropy based metrics and shown that it can detect anomalies more efficiently than before. In [5] suggested use of different information measures for detecting malicious activities. The authors of [7] use entropy rate to discriminate the DDoS attack from legitimate traffic. Our objective in this paper is to design an

anomaly detection system based on entropy and entropy rate to detect DDoS attack in Grid environment. We use normalized entropy which calculates the over all probability distribution in the captured flow in our algorithm to get more accurate result.

3. DoS ATTACK IN GRID

DoS and DDoS attacks [9] are the most common and deadly attack today. A DoS attack involves sending large number of packets to a destination to prevent legitimate users from accessing information or services. DDoS use the computing power of thousands of compromised machines known as "zombies" to a target a victim. Zombies are gathered to send useless service requests, packets at the same time. DDoS attacks are not targeted at stealing, modifying or destroying information, but its aim is to prevent legitimate users from using a service. It is very difficult to detect a DoS attacker because they generally use spoofed IP address and it becomes more complicated in large distributed system like grid. Although Grid Security Infrastructure (GSI) of grid middleware provides several security features that required on grid environment; include X.509 certificates, authentication algorithm using Secure Socket Layer (SSL) protocol, authorization, delegation, auditing and single sign on [17]. Due to the scalability and dynamic nature of grid some security flaws are there and the huge resource capacity like computational and storage of grid computing may become a next platform for the attackers [10]. If an intruder got unauthorized access to grid, then the grid resources can misused in different ways; like the huge computational power can be used for breaking passwords or security systems and the large storage capacity can be used to store illegal software and data, and the huge bandwidth can be used for lunching DDoS attack. To secure grid from DDoS attack the defense system can be divided in to 3 steps. (i) Attack prevention (before attack), (ii) attack detection and recovery (during the attack), (iii) attack identification (after attack).

3.1. DDoS Attack prevention

The aim of attack prevention mechanisms is to take preventive measures which can not provide 100% security, but it will decrease the strength of DDoS attack. Based on the target of implementation of the mechanisms it can divide them in to system security and protocol security mechanisms. [9]

System security mechanisms

System security deals with those mechanisms which are implemented on the end host. In DDoS attack it is required thousands of compromised machines to target a victim, but if we will strengthen the overall security of each host of grid then it will difficult for an attacker to lunch an attack. Examples of system security mechanisms are firewall and micro firewall systems, anti virus systems, access control, packet filter and authorization systems.

Protocol security mechanisms

Protocol mechanisms increase the security by designing a safe protocol so that only resources are allocated to the clients after sufficient authentication and authorization are completed. For which resources will not waste time in attack like TCP SYN attack. Use of proxy sever has been proposed by authors [18].

3.2. Attack detection and recovery

The aim of attack detection and recovery is to detect DDoS attack before it affects the end user .Intrusion detection systems [11] are widely used for DDoS detection. An Intrusion detection system (IDS) is software and/or hardware which will monitor the network or a computer system for suspicious activity and alerts the system manager or network administrator. We can classify the IDS based the target of implementation as host based and network based. The technique adopted by IDS for intrusion detection classifies IDS in to two types signature based and anomaly based.

Signature based IDS

A signature based IDS will monitor packets on the network and compare them against a database maintained with known threats. If the signature of packets match with those known attacks it will marked as malicious. The advantage with signature based IDS is signatures are easy to develop. The disadvantage of is that they can only detect known attacks, for which a large up-to-date database of signature for every attack must be created,

Anomaly-based IDS

Anomaly-based IDS [14] creates the normal behavior of the users using the system or network to detect intrusions. If the deviation of user activity is outside a certain threshold value, it marked as malicious and a response is triggered. Anomaly detection has an advantage over signature-based in that a new attack can be detected if it falls out of the normal traffic patterns. Disadvantage of anomaly-detection system is the difficultly of defining rules. Setting of a perfect threshold is also very challenging because setting of a small threshold creates many false positives and setting of high threshold reduces the effectiveness of the IDS [14]. After detection of intrusion it's the work of response system to take action so that attack traffics will damaged with out affecting legitimate user. There are popularly two response mechanisms filtering and rate limiting algorithms are used against DoS attack.

3.3 Attack source identification

Another difficulty in defending to DDoS attack is to trace the source of the attacks, because the attackers are generally uses spoofed IP addresses in the IP packets. For which the attack identification mechanism should be flexible enough so that it can

trace the source of attack packets without depending on the source address field of the IP header. There are different mechanisms proposed by different authors like Advanced Marking Scheme and the Authenticated Marking Scheme [19], Probabilistic Packet Marking (PPM), Deterministic Packet Marking (DPM), has been proposed to trace back the source of spoofed IP packets. Also more efficient method like flexible Deterministic Packet Marking (FDPM) can be found [8].

4. ANOMALY DETECTION SYSTEM FOR GRID

Previously, much work has been done in volume (no. of bytes, packets, flows) as a principal metric for anomaly detection system [13]. Volume based detection scheme method can detect anomalies that causes large traffic changes ,But anomalies like small DoS attacks which do not cause much changes in traffic volume can not be detected well. The attack discussed above can be better detected by analyzing distribution of traffic features. A traffic feature is a field in the header of the packet. One of the properties of Grid Security Infrastructure (GSI) [17] is confidentiality of the data transferred over the network. For which the data transmitted over the grid must be encrypted. So the system could not see the data payload portion of the packet because of encryption. Analysis would be based only on the low level information, which can be extracted from the packet header. The Next problem is to find a metric that can extract distribution of traffic features that can be used in anomaly detection system. A number of articles suggested entropy as a metrics to summarizing traffic distribution for anomaly detection [3] [4] [5] [15]. The use of entropy for analyze changes in traffic distribution has two benefit. i) Using entropy for anomaly detection increases the detection capability than volume based methods. ii) It provides additional information to classify among different types anomaly (worms, DoS attack. Port scanning) .We considers two classes of distribution i) flow header features (IP address, ports, and flow sizes) ii) behavioral features (the number of distinct destination / source address that a host communicates with) [15]. The anomaly detection system discussed in this paper is based on by analyzing the change in entropy of above two traffic distributions.

5. ENTROPY BASED APPROACH

Entropy [16] is a measure of the uncertainty or randomness associated with a random variable or in this case data coming over the network. The more random it is, the more entropy it contains. The value of sample entropy lies in range [0, logn].The entropy shows its minimum value 0 when all the items (IP address or port) are same and its maximum value logn when all the items are different. The entropy of a random variable X with possible values $\{x_1, x_2 \ldots, x_n\}$ can be calculated as

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \qquad (1)$$

If we are interested in measuring the entropy of packets over unique source or destination address then maximum value of n is 2^{32} for ipv4 address. If we want to calculate entropy over various applications port then n is the maximum number of ports. Here $p(x_i)$ where $x_i \in X$ is the probability that X takes the value x_i. Suppose we randomly observe X for a fixed time window w, then $p(x_i) = m_i/m$, where m_i is the frequency or number of times we observe X taking the value x_i I.e. $m = \sum_{i=1}^{n} mi$.

$$H(X) = -\sum_{i=1}^{n} (m_i/m) \log (m_i/m) \qquad (2)$$

H (X) = Entropy.

If we want calculate probability of any source (destination) address then,

m_i = number of packets with x_i as source (Destination) address and

M = total number of packets

$$P(x_i) = \frac{\text{Number of packets with } x_i \text{ as source (destination) address}}{\text{Total number of packets}}$$

Here total number of packets is the number of packets seen for a time window *T*. Similarly we can calculate probability for each source (destination) port as

$$P(x_i) = \frac{\text{Number of packets with } x_i \text{ as source (destination) port}}{\text{Total number of packets}}$$

Normalized entropy calculates the over all probability distribution in the captured flow for the time window *T*.

$$\text{Normalized entropy} = (H / \log n_0) \qquad (3)$$

Where n_0 is the number of distinct x_i values in the given time window.

In a DDoS attack from the captured traffic in time window *t*, the attack flow dominates the whole traffic, as a result the over all entropy of the traffic decreased in a detectable manner. But it is also possible in a case of massive legitimate network accessing. To confirm the attack we have to again calculate the entropy rate. Here flow is packages which share the same destination address/port. In this mechanism we have taken one assumption that the attacker uses same function to generate attack packets at "zombies", and it is a stationary stochastic process. According to [16] for a stochastic processes the entropy rate $H(\chi)$ of two random processes are same.

$$H(\chi) = \lim_{n \to \infty} \frac{1}{n} H(x_1, x_2 \ldots x_n) \qquad (4)$$

The steps in DDoS detection algorithm are described in figure 1.

Algorithm 1 : DDoS detection algorithm

1. Collect sample flows for a time window *T* on the edge routers
2. Calculate router entropy $H(x) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$
3. Calculate NE = (H / log n_0) where, NE = normalized Router entropy
4. If NE < threshold (δ_1), identify the suspected attack flow
5. Calculate the entropy rate $H(\chi) = \lim_{n\to\infty} \frac{1}{n} H(x_1, x_2 \ldots x_n)$ of the suspected flow in that router and the routers in down stream
6. Compare $H_i(\chi)\ \forall i \varepsilon$ entropy rates on routers
8. If $H_i(\chi) \leq$ threshold (δ_2), it is a DDoS attack
   Else legitimate traffics
9. Discard the attack flow.

Figure 1: DDoS detection algorithm

Definition 1:
A stochastic process {X (t), t ∈ T} is a collection of collection of random variables. For each t ∈ T, X (t) is a random variable. We refer X (t) as the state of the process at time t. The set T is called the index set of process.

Definition 2:
A stochastic process is said to be stationary if the joint distribution of any subset of random variables is invariant with respect to shifts in the time index i.e.
Pr{X1 = x1, X2 = x2. . . Xn = xn}
= Pr{X1+l = x1, X2+l = x2. . . Xn+l = xn}

Definition 3:
The entropy rate is the rate of growth of entropy of a random process. If we have a sequence of n random variables, then the entropy rate of a stochastic process { x_i } is defined by

$$H(\chi) = \lim_{n\to\infty} \frac{1}{n} H(x_1, x_2 \ldots x_n)$$

6. IMPLEMENTATION IN GRID

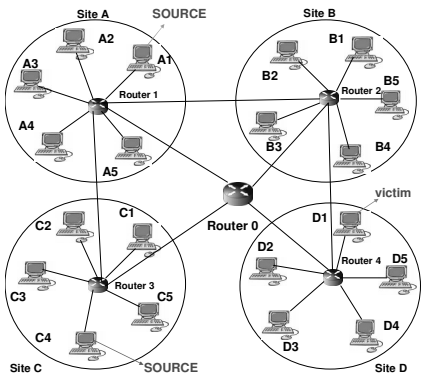

Figure 2. Grid topology

Grid computing can be thought as a virtual organization which is a collection of some real organizations or sites [12]. In the figure 2 we have shown a Grid topology model which is a collection of four sites i.e. site A, site B, site C and site D and they are connected by 5 routers. We employ our proposed anomaly detection system in each router of the grid infrastructure. Edge routers near the source of traffic will capture flows for a predefined time window *T* and calculate the router entropy and normalized router entropy. If the normalized router entropy is less than certain threshold δ_1 identify the suspected DDoS attack flow from the traffic. But it is also possible in a case of massive legitimate network accessing. To confirm the attack the entropy rate of the suspected flow is calculated in that router according To "Eq. (4)".

Based on the destination address on the IP header of the packets and routing table it discovers the downstream routers and sends security alarm to those routers to calculate entropy rate of the suspected flow. As discussed above, the entropy rates of attack flows at different routers in the network are same. If the calculated entropy rates on routers are same or very near, the attack is confirmed. No real Grid environment is available for testing the performance. So the real life experiments could not be performed. We have considered two examples using figure (2) how the detection scheme works. Suppose From figure (2) node A1 and C4 are attack sources and D1 is the victim. Based on the DDoS detection algorithm flows coming A1 will first captured by router 1 and flows coming from C4 will be captured by router 3.

Suppose at router 1, router 2 and router 3 we have captured flows as given in table 1, table 3 and in table 2 respectively in a fixed time window *T*. The entropy of the flows are calculated according to "Eq. (2)"and "Eq. (3)".For easy understanding we have assigned IP addresses to each host.

Table I. Data for router 1

Source node	Destination address	No of packets	entropy
A1	D1[134.11.78.56]	6	0.47
A5	B5[192.168.1.121]	2	0.44
A2	B2[192.168.1.118]	3	0.51

Here Router entropy = 0.47 + 0.44 + 0.51 = 1.42
n_0 = 3
Normalized Router entropy NE = 1.42/ $\log_2$ 3 = 0.89

Table II. Data for router 3

Source node	Destination address	No of packets	entropy
C1	B4[192.168.1.122]	2	0.48
C2	D3[134.11.78.54]	2	0.48
C4	D1[134.11.78.56]	5	0.47

Router entropy =0.47+0.48+0.48 = 1.43
Here n_0 = 3
Normalized Router entropy NE = 1.43/ $\log_2$ 3 = 0.90

Table III. Data for router 2

Source node	Destination address	No of packets	entropy
B1	C2[192.168.213.109]	2	0.52
B2	C3[192.168.213.108]	2	0.52
B3	C1[192.168.213.110]	2	0.52

Router entropy=0.52 + 0.52 +0.52 = 1.56

$n_0 = 3$

Normalized Router entropy NE = 1.56/ $\log_2 3$ = 0.98

Although the data are taken manually we can see that for router 1 and router 3 the normalized router entropy is less than the router 2. In the first two cases one flow dominates the whole traffic as a result the normalized entropy decreases. If the threshold δ_1 is perfect, suppose 0.94 for the above example, it will treat flow coming from node A1 and C4 as suspected flows. After which the entropy rate is being calculated. In figure 2, for router 1 the entropy rate of suspected flow is calculated and compared in router1 and router 0. Similarly for router 3 the entropy rate of those flows will be calculated and compared both in router 3 and router 0. While the entropy rates of different routers are same or less than δ_2, the attack is confirmed and attack flow is discarded.

All the calculations are based on $\log_2$.

7. CONCLUSION

A DDoS attack is a major and complex threat for Grid computing. The aim of this study was to investigate how DDoS attack affects the grid performance and designing an anomaly detection system. The attack must be detected and blocked before reaching the victim and with high detection rate and low false alarm rate. In this paper we have used information theoretic parameters entropy and entropy rate to model the anomaly detection system for Grid. We have implemented anomaly detection system in each router of the grid environment and the router will cooperate with each other to detect anomaly. The main advantage of the above proposed method is the attack is detected and blocked before reaching the victim and with high detection rate. But the challenge lies in this approach if the attacker will use different packet generation functions in an attack and setting a good threshold. The effectiveness of the proposed method has been proved theoretically. In the future work we will simulate the proposed algorithm and analyze the results and we will provide a road map for further research in this area.

8. REFERENCES

1. *I. Foster, C. Kesselman, "The Grid: Blueprint for a new computing infrastructure".Morgan Kaufmann publishers, 1999.*
2. *R. Buyya, S. Venugopal,"A Gentle Introduction to Grid computing and Technologies ", CSI Communications, 19 July 2005.*
3. *L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred. "Statistical approaches to DDoS attack detection and response". In Proc of DARPA Information Survivability Conference and Exposition, 2003.*
4. *[4] A. Lakhina , M. Crovella and C. Diot , " Mining anomalies using traffic feature Distributions ". In proc. of ACM SIGCOMM, 2005.*
5. *W. Lee, D. Xiang, "Information-theoretic measures for anomaly Detection", In Proc. of IEEE Symposium on Security and Privacy, 2001.*
6. *N. Jiancheng, L. Zhishu, G. Zhonghe,S. Jirong, " Threat analysis and Prevention for grid and web security services", SNPD, pp. 526-531, 2007.*
7. *S. Yu , W.Zhou,"Entropy-based Collaborative Detection of DDoS attacks on Community Networks", In Proc. of IEEE international conference on pervasive computing and Communications, 2008.*
8. *Y. Xiang and W. Zhou, "A Defense System against DDoS Attacks by Large- Scale IP Traceback", In Proc. of Third International Conference on Information Technology and Applications (ICITA'05).*
9. *J. Mirkovic, J. Martin and P. Reiher, "A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms,"ACM Computer Communications Review, Vol.34, No. 2, April 2004.*
10. *Y. Xiang , W. Zhou, " Protect Grids from DDoS attack" , In proc. of third International conference on grid and cooperative computing", vol. 3251 , pp. 309-316, 2004*
11. *A. Schulter, J. A. Reis, F. Koch, C. B. Westphall "A Grid-based Intrusion Detection System", In Proc. of ICNICONSMCL'06.*
12. *T. Znati, J. Amadei, Daniel R. Pazehoski , S. Sweeny "Design and Analysis of an Adaptive, Global Strategy for Detecting and Mitigating Distributed DoS Attacks in GRID Environments "In Proc. of the 39th Annual Simulation Symposium (ANSS'06), 2006.*
13. *A. Lakhina, M. Crovella, and C. Diot. Diagnosing Network-Wide Traffic Anomalies. In ACM SIGCOMM, Portland, August 2004.*
14. *P. G. Teodoro, J. D.Verdejo, G. M.Fernandez, E.Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges", computer & security, 2008.*
15. *G. Nychis, V. Sekar, D. G Andersen, H. Kim and H.Zhang, "An Empirical Evaluation of Entropy-Based Traffic Anomaly Detection". Tech. Rep. CMU-CS-08-145, Computer Science Department, Carnegie Mellon University, 2008.*
16. *Thomas M. Cover and Joy A. Thomas, "Elements of Information Theory", second edition, 2007.*
17. *S. Shirasuna, A. Slominski L. Fang and L. Gannon, "Performance comparison of security mechanisms for grid services", Proc. of the fifth IEEE/ACM International Workshop on Grid Computing, pp.360-364, 2004.*
18. *J. Wang, X. Liu and A. Chien, "Empirical Study of Tolerating Denial-of-Service Attacks with a Proxy Network", In Proc. Of* the USENIX Security Symposium*, 2005.*
19. *D.Xiaodong Song and A. Perrig "Advanced and Authenticated Marking Schemes for IP Traceback "IEEE INFOCOM, 2001.*
20. *http://www.sun.com/service/sungrid/index.jsp*