

An Anomaly Detection Scheme for DDoS Attack in Grid Computing

*A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE AWARD OF DEGREE OF*

Master of Technology
in
Computer Science and Engineering
(Specialization: Information Security)

By
Sumit Kar



Department of Computer Science Engineering
NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA
ROURKELA-769008 (ORISSA)

May 2009

An Anomaly Detection Scheme for DDoS Attack in Grid computing

*A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE AWARD OF DEGREE OF*

Master of Technology

in

Computer Science and Engineering

(Specialization: Information Security)

By

Sumit Kar

Under the guidance of

Prof. Bibhudatta Sahoo



Department of Computer Science Engineering
NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA
ROURKELA-769008 (ORISSA)

May 2009

To my Parents



National Institute of Technology Rourkela
Rourkela-769008 (Orissa)

Certificate

This is to certify that the work in this Thesis Report entitled “*An Anomaly Detection Scheme for DDoS Attack in Grid computing*” by **Mr. Sumit Kar** has been carried out under my supervision in partial fulfillment of the requirements for the degree of **Master of Technology** in Computer Science and Engineering, Specialization: Information Security during session 2008-2009 in the Department of Computer Science and Engineering, National Institute of Technology, Rourkela. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Place: NIT Rourkela

Date:

Prof. Bibhudatta Sahoo

Senior Lecturer

Dept of Computer Science and Engineering

National Institute of Technology

Rourkela – 769008

Acknowledgements

It is not possible to complete a thesis without considerable support of many great people. First and foremost, I am greatly indebted to my advisor Prof. Bibudatta sahuo, Senior Lecturer in CSE department of NIT Rourkela for his guidance, encouragement, motivation, and continued support throughout my academic years at NIT. He has allowed me to pursue my research interests with sufficient freedom, while always being there to guide me. Working with him has been one of the most rewarding experiences of my professional life.

I am grateful to Prof. (Dr). B. Majhi, Head, CSE for his excellent support during my work. My sincere thanks go to Dr. A.K. Turuk, Dr. S. K. Rath, Dr. D.P. Mohapatra, Dr. P. M. Khilar, Dr. S.K jena for supporting my work.

I have been fortunate to have met great friends throughout my MTECH journey. I am forever grateful for their moral support, encouragement, and true friendship.

I am grateful to all the staffs of the computer science department for their generous help in various ways for the completion of this thesis. Last but not least, I am forever indebted to my parents, my sisters and the rest of my family. They have been a great source of inspiration to me. This would have not been possible without his love, support, and continuous encouragement.

Sumit Kar

M.Tech. (Comp. Sc.), 2008-2009

Abstract

The demand for computing power and storage is increasing continuously and there are applications like scientific research and industrial need, whose computational demand even exceeds the available fastest technologies. As a result it is an economically feasible mean to look into efficiently aggregate existing distributed resources. To achieving this goal makes it possible to build a shared large scale wide-area distributed computing infrastructure, a concept which has been named the Grid computing. The primary objective of Grid computing is to support the sharing of resources and service spanning across multiple administrative domains. Due to the inherently dynamic and multi organizational nature maintaining security of both users and resources is the challenging aspect of Grid. Grid uses internet as an infrastructure to build communication, with the fusion of web services and grid technologies further increases the security concerns for their complex nature.

This thesis takes a look at the vulnerability of Grid environment on denial of service attack. We found that deploying an efficient intrusion detection system to Grid can significantly improve its security and it can detect denial of service attack before it affects the victim. But due to the special characteristics and requirement of Grids, the existing traditional intrusion detection system can not work properly in that environment. The focus of this thesis is to investigate and design an anomaly detection system which can detect DoS and DDoS attack with high attack detection and low false alarm rate to achieve high performance. We have extensively surveyed the current literatures in this area; the main stress is put on feature selection for the Grid based anomaly detection system. An entropy based anomaly detection system has been proposed; also we have discussed the advantage of taking entropy as the metric. Finally the performance of the system has been analyzed using NS2 network simulator.

For shake of continuity each chapter has its relevant introduction and theory. The work is also supported by list of necessary references. Attempt is made to make the thesis self-content.

List of Acronyms

Acronyms	Description
ACK	Acknowledgement
ADS	Anomaly Detection System
CBR	Constant Bit Rate
CA	Certificate Authority
CPU	Central Processing Unit
CAS	Community Authorization Service
DoS	Denial of service Attack
DDoS	Distributed Denial of service Attack
GT	Globus Toolkit
GSI	Grid Security Infrastructure
GIDS	Grid intrusion Detection System
HIDS	Host Intrusion Detection System
IDS	Intrusion Detection System
IP	Internet Protocol
IEEE	Institute of Electrical and Electronics Engineers
IPV4	Internet Protocol Version 4

LAN	Local Area Networks
MLS	Message-level Security
NIDS	Network Intrusion Detection System
NS	Network Simulator
OS	Operating System
OGSA	Open Grid Services Architecture
PCA	Principal Component Analysis
PKI	Public Key Infrastructure
QoS	Quality of Service
SSL	Secure Socket Layer
TLS	Transport layer Security
VO	Virtual Organization
WAN	Wide Area Network
XML	Extensible Markup Language

Contents

Certificate	iii
Acknowledgement	iv
Abstract	v
List of Acronyms	vi
List of Figures	x
List of Tables	xi
1. Introduction	1
1.1 Motivation	3
1.2 Problem Statement and Objectives	4
1.3 Contributions	5
1.4 Organization of the Thesis	5
2. The Grid and Security Concerns	7
2.1 Need of Grid Computing	8
2.2 Types of Grid	9
2.3 Grid Applications	9
2.4 Grid Components	11
2.5 Steps of Task execution in Grid	12
2.6 Grid Architecture	13
2.7 Grid Security	14
2.7.1 Authentication	15
2.7.2 Authorization	16
2.7.3 Delegation and Single Sign on	16
2.7.4 Secure Communications	17

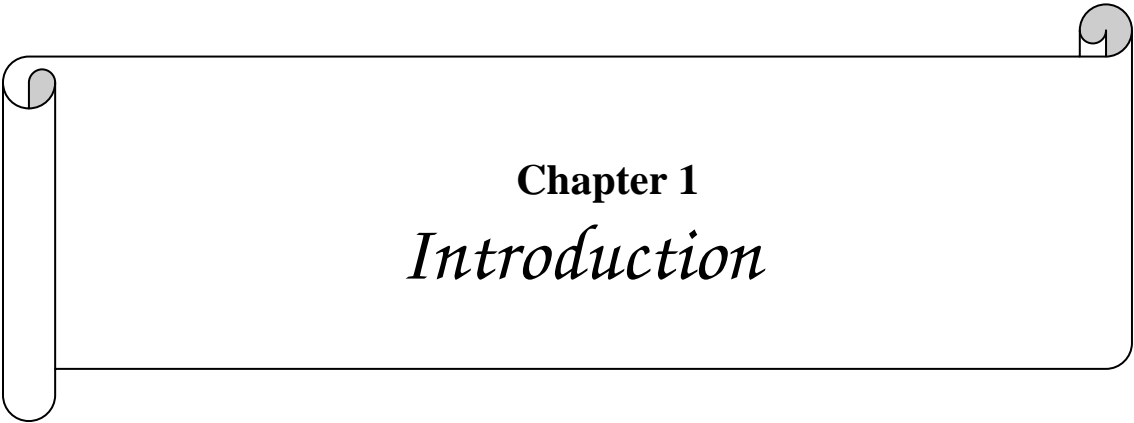
2.7.5 Accounting and Auditing	17
2.8 Security Threats Classifications in Grid	18
2.9 Conclusion	19
3. DDoS in Grid Environment	20
3.1 Related Work	21
3.2 DoS Attack	22
3.3 DDoS Attack	23
3.4 DDoS Attack Classifications	24
3.5 DDoS Defense System	25
3.5.1 Attack prevention	25
3.5.2 Attack Detection and Recovery	26
3.5.3 Attack source Identification	27
3.6 Conclusions	28
4. Proposed Entropy Based ADS	29
4.1 Entropy	30
4.2 Proposed ADS for Grid	30
4.3 Implementation Details	33
4.4 Conclusion	36
5. Simulation and Results	37
5.1 Experimental Setup	38
5.2 Performance Evaluation	39
5.3 Conclusion	42
6. Conclusion and further Enhancements	43
Bibliography.....	46
Dissemination	49

List of Figures

2.1 Task execution in Grid	12
2.2 Block Diagram of Computational Grid	13
2.3 Typical Grid Scenario	15
3.1 DoS and DDoS Attack Scenario	24
3.2 DDoS Attack Classification	25
4.1 DDoS Detection Algorithm	32
4.2 Grid Topology	33
5.1 Screen Shot of Setup	38
5.2 Effect of DDoS Attack	40
5.3 Screen Shot of Packet Drop	40
5.4 DDoS Detection Rate	41
5.5 False Positive Rate	42

List of Tables

1. Threats classification in Grid	19
2. Data for router 1	35
3. Data for router 3	35
4. Data for router 2	35
5. Traced Data	39
6. Normalized Router entropy Calculation	39



Chapter 1
Introduction

Chapter

1

Introduction

The availability of low cost powerful computers with the popularity of the internet and high speed networks have led the computing technology evolved from classical distributed system to Grid environment. The Grid computing is rapidly growing as a dominant field of high throughput wide area distributed computing. Its objective is to provide a service oriented infrastructure that supports the sharing and coordinated use of heterogeneous resources spread across multiple administrative domains that are not subject to any centralized control [1, 2]. Now a variety of middleware systems are available that enable resource providers to make available their resources for use by others in Grid. By combining distributed resources Grid forms an image of a single virtual system through which users from different places can access shared resources in an efficient manner. Unlike where internet focuses on communication among devices, Grid computing is seen as a network of computation. The term “Grid” was introduced in early 1998 with the launch of the book “The Grid: Blueprint for a new computing infrastructure” (Ian Foster, Carl Kesselman). This defined “Grid computing is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high end computational capabilities”. [1]

The name 'Grid' comes from analogy with the Electricity Grid. Users can obtain a resource such as electricity, or in this case computer processing from a variety of sources to supply their needs. The goal is to provide users with access to the resources they need and when they need them. Initially Grid systems were developed for supporting scientific computations. Today, many enterprises and researchers are looking to use the Grid approach to commercial uses and for applications in many different areas. Security in

Grid systems however has not been much addressed and yet is an important issue to make it usable in a variety of commercial applications.

1.1 Motivation

In Grid environments, where the computational, storage, and network resources are inherently heterogeneous, dynamic and multi organizational in nature, the issue of managing security of both resources and users are most challenging. 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 [11]. An intruder can explore security flaws in any of the other components like operating system (OS), network protocols and non grid application running in the same environment. Moreover grid cannot defend itself against stolen passwords and legitimated users who abuse their privileges to execute malicious activities and due to the huge resource capacity like computational and storage, grid may become a next platform for the attackers [24]. If an intruder got unauthorized access to grid, the grid resources can misused in different ways; like the huge computational power can be used for breaking passwords or security systems , the large storage capacity can be used to store illegal software , data and the huge bandwidth can be used for launching Distributed Denial of Service (DDoS) attack. Possible security threats associated in grid are discussed in section (2.8), out of which Denial of Service attack (DoS) and Distributed Denial of Service attacks (DDoS) [23] are the most common and deadly attack today.

Intrusion detection system (IDS) are widely used for detection of Denial of Service attack in a network or with in a system and it can be deployed in grid to complement existing measures like authentication, authorization, access control [23]. Along with those basic security measures Grid-specific Intrusion Detection System is capable of functioning in a real-life grid environment, which has the potential to detect before Denial of Service attack succeeds. The purpose of this research is to designing an anomaly detection system for detecting real-life grid misuse.

1.2 Problem Statement and objectives

As the users of the grid are from a number of different real organizations, each organization may have different access policies and security mechanism. These organizations share their resources collectively as grid which is in the form of a larger dynamic virtual organization (VO). Maintaining security in such an environment is a real challenge and most difficult part in security management of large high speed networks like grid is the detection of suspicious anomalies in network traffic patterns due to DoS and DDoS attacks.

To secure grid from DoS attack its must be detected before it affects the end user with high detection rate and low false alarm rate, So that attack traffic will be discarded, without affecting legitimate traffic. In case of DDoS attack, the attack packets comes from ten or thousand of sources and DoS defense system that is based upon monitoring the volume of packets coming from a single address or single network will fail since the attack comes from various sources. Intrusion detection systems [8] are widely used for DDoS attack detection. An intrusion detection system (IDS) inspects all inbound and outbound network and system activity and identifies possible security threats in a network or in a system. IDSs can be classified, based on their functionality, as misuse detectors and anomaly detectors [23]. Anomaly detection has an advantage over signature-based is that a new attack can be detected if it falls out of the normal traffic patterns. Due to the special characteristics and requirement of computational grids, detecting such difference in traffic patterns imposed some new unique challenges that did not exist in traditional intrusion detection system. One of the properties of Grid Security Infrastructure (GSI) [11] is the confidentiality of the data transferred over the network. For which the data transmitted over the grid must be encrypted and 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.

Accordingly we identify the objective of the thesis and list them as follows

- To study different possible security threats in grid.
- To investigate how DoS attack affect the grid performance.
- To investigate which metrics can be suitable for designing anomaly detection system for grid.
- To propose an anomaly detection algorithm for grid.
- To study the performance of the above algorithm through simulation.

1.3 Contribution

We found that DoS and DDoS attacks are major threats to the grid performance. The perfect secure system for DDoS attack can be divided in to 3 steps: (i) Attack prevention, (ii) attack detection and recovery, and (iii) attack identification. In thesis we emphasized to detect and stop DDoS attacks with in the intermediate network before affecting the victim. After surveying a lot of literature we found that entropy can be taken as a suitable metric for designing an anomaly detection system for detecting DDoS attack in grid environment. In section (4) we discussed our proposed entropy based Grid specific anomaly detection system. Finally using NS-2 network simulator [10] we evaluate the performance results.

1.4 Organization of the Thesis

The rest of the thesis divided into the following Chapters.

Chapter-2 It starts by giving an overview of grid and its architecture .It then describes basic security requirements and technology used in grid. Finally a threat classification model for grid has been presented.

Chapter-3 Discusses related work has been done to protect grid from DoS attack and intrusion detection system for grid. This chapter also explains in details the DoS and DDoS characteristics and the three steps used in DoS defense system.

Chapter-4 Explains the proposed entropy based anomaly detection system (ADS) for grid environment. Finally using a grid topology model the implementation details of the proposed algorithm has been presented.

Chapter -5 presents the experimental setup for validation of the proposed system using NS2 simulator and followed by simulation results.

Chapter -6 concludes this thesis with a discussion about future direction. Reference section includes detail list of necessary references used in this thesis.



Chapter 2

The Grid and Security Concerns

Chapter ---

2 The Grid and Security Concerns

This Chapter presents the concepts underlying this thesis and the necessary materials for understanding the rest of the thesis. The presentation here consists of three parts. The first part introduces the basic concept about Grid technology and its architecture. The second part discusses the security requirements and technology used in grid. Finally, the threats classifications in grid are discussed.

2.1 Need of grid Computing

The basic idea of grid computing is to provide an infrastructure for solving massive computational problems by using the idle resources (CPU cycles, disk storage, and network bandwidth) of large number of computers, servers embedded in a distributed infrastructure. The aim is getting computers to work together for solving a particular problem. In every organization, there are large amounts of underutilized computing resources. According to IBM survey Mainframes are idle 40% of the time, UNIX servers are actually "serving" something less than 10% of the time and most desktop machines are busy less than 5 percent of the time .[2]

Grid computing provides a framework for exploiting these underutilized resources to solve problems which are beyond the scope of single processor and also increases the efficiency of resource usage. [2] In addition to scientific experiment, industries such as bio-medical field, modeling, oil exploration, motion picture animation, weather prediction and many others needs massive computing power. In grid environment a single large job can be split into smaller pieces and run on several computers simultaneously which is too intensive for any stand alone machine.

2.2 Types of grids

Grid computing can be used in a variety of ways to address various kinds of application requirements. Often, grids are categorized by the type of solutions that they provide. The three primary types of grids are summarized below. [9]

Computational grid

A computational grid provides secure access to huge pool of shared processing power suitable for high throughput applications and computation intensive computing. It focused on specifically for computing power. Computational Grid also called Meta computer. Examples of computational Grid include NSF TeraGrid and SETI@home.

Data grid

Data grids provide an infrastructure to support data storage, data discovery, data handling, data publication, and data manipulation of large volumes of data stored in heterogeneous databases and file systems across multiple organizations. Different datasets stored in different locations create an illusion of mass storage. These are often, but not always, combined with computational grid systems.

Service grid

In service grid unused resources are exported to the users in the form of service. It creates a "grid" from the unused resources in a network of participants. With the evolve of open grid architecture (OGSA), open and standard protocols it provides services and supports dynamic creating, running, maintaining and cancellation of application. The QOS is an important parameter for a service grid system to meet the user demand.

2.3 Grid applications

Based on types of applications of grid it can be divided in to four types like on demand computing, high through computing, data intensive computing and collaborative computing.

2.3.1 On Demand Computing

On-demand computing is an increasingly popular enterprise model in which computing resources are made available to the user as needed [15]. Maintaining sufficient resources locally to meet peak requirements can be costly, On-demand applications use grid capabilities to meet short-term requirements in peak time. These resources may be computation, software, data repositories, and special scientific equipments. Computer Associates, HP, IBM, Microsoft, and Sun Microsystems are among the more prominent on-demand vendors. Sun launches Sun Grid, now anyone in the U.S. can get flexible, metered access to enormous computing power over the network for the affordable price of \$1/CPU-hr. [13]

2.3.1 High-Throughput Computing

In high-throughput computing, the grid is used to schedule large numbers of independent tasks, with the goal of utilizing unused processor cycles of ideal machines. Many scientific experiments are needed massive CPU power and they are strongly related to computing throughput so grid is used as high throughput computing in this case. For example The Condor [14] system from the University of Wisconsin-Madison is used to manage large collections of distributive owned heterogeneous computing resources around the world which is used for research purposes.

2.3.2 Data intensive computing

Data intensive computing is concerned with capturing, analyzing, managing large volume of data, which is maintained in geographically distributed repositories, digital libraries, and databases. Such data intensive applications exist on different areas like scientific research, cyber security, animation and business field. Also data-intensive applications require high degrees of fault-tolerance, reliability, and availability. Using data intensive computing rather only collect, analyze and store massive amounts of information, researchers are able to getting experimental results more rapidly and thoroughly .For example, Google runs an average of 100,000 Map Reduce jobs per day on its clusters, processing over 20 pet bytes data daily.

2.3.3 Collaborative Computing

It provides an ability to allow geographically distributed teams to develop, share and utilize common resources. Sharing is not limited to data, but also includes many other resources, such as scientific equipments, software, services, and licenses. These applications are often structured in terms of a virtual shared space to give a uniform interoperability among heterogeneous grid participants. With existing collaboration tools grids provide an infrastructure within which distributed users share computing tasks, analyses, and visualization results in the form of a collaborating computing. For example Persons from different companies in a virtual enterprise can work on different components of a CAD project without even disclosing their work using the grid.

2.4 Grid components

Grid basically contains five Components.

1. A portal
2. A service broker
3. Task scheduler
4. A task manager
5. A group of grid node.

The portal acts as a user interface, through which user can log in and use the grid. After having logged into the grid, a user can submit a task.

The service broker identifies the list of resources for handling a particular task submitted by the user and selects the optimal one which is available now.

A scheduler sets rules and priorities for scheduling jobs on a grid-based infrastructure. It is responsible for scheduling submitted tasks.

The task manager finally launches a submitted task.

The nodes can be desktops, servers, Workstations and clusters that belong to different LANs, WANs, or the Internet.

2.5 Steps of Task execution in grid

This grid VO model in figure 2.1 depicts typical steps of how grid jobs are submitted and executed.

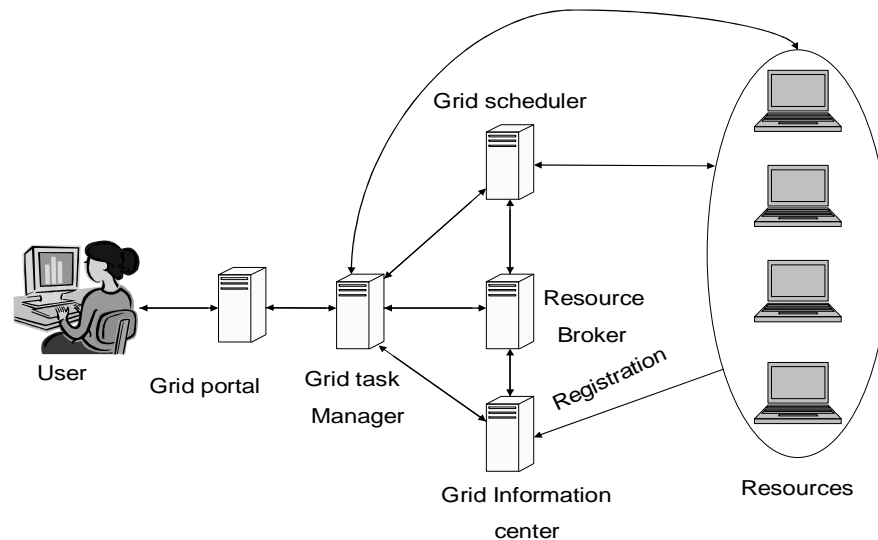


Figure 2.1: Task execution in Grid

- 1) A user log in to the grid through a portal, the portal acts as a user interface, through which user can log in and use the grid. After identification verification of the user he/she can submit its task to the task manager.
- 2) The task manager cooperates with information center, which maintains the list of available resources and resource broker, which identifies the list of resources which can satisfy a particular task submitted by the user.
- 3) Next the scheduler is responsible for scheduling submitted tasks on the resources identified by resource broker, Which Sets rules and priorities for scheduling task on a grid infrastructure.

4) Job manager supplies the user task, data to the selected resources and after execution of the task it returns the computed result to the user.

2.6 Grid architecture

Grid architecture identifies fundamental system components, specifies the purpose and function of these components, and indicates how these components interact with one another. [3] A computational grid can be modeled using 4-layer architecture as (1) Grid fabric, (2) Core Grid Middleware, (3) Grid Tools, and (4) Grid Applications. As shown in figure 2.2 [4]

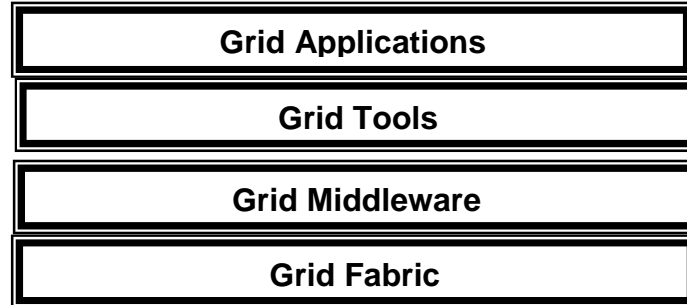


Figure 2.2: Block Diagram of Computational Grid

Grid Fabric consists of all the globally distributed resources that are accessible from anywhere on the Internet. The Fabric Layer includes the local protocols and interfaces for accessing and managing the local resources. Possible resources could be computers (such as PCs, clusters) with variety of operating systems (such as UNIX or Windows) as well as resource management systems such as LSF (Load Sharing Facility), Condor, PBS(Portable Batch System) or SGE (Sun Grid Engine), storage systems devices, databases, programs ,networks and special scientific instruments such as a radio telescope.

The **Core Grid Middleware** offers core services such as remote process management, co-allocation of resources, storage access, information registration and discovery, security and aspects of Quality of Service (QOS) such as resource reservation and

trading. These services abstract the complexity and heterogeneity method of accessing distributed resources.

The **Grid Tools (User-Level Grid Middleware)** includes application development environments, programming tools, and resource brokers for managing resources and scheduling application tasks for execution on global resources.

Application layer is the highest layer of the structure that grid users "see" and interact with. The application layer is often called serviceware. Grid applications are typically developed using grid-enabled languages and interfaces and brokering and scheduling services provided by grid tools or user level middleware. Applications may be in science, engineering, business, finance and more. Grid portals offer Web-enabled application services, where the users can submit and collect results for their jobs on remote resources through the Web.

2.7 Grid Security

Generally first generation grid was deployed across mutually trusting administrative domains like research laboratories, academic institutions and for military use, But OGSA (Open Grid Services Architecture) [5] as a standard for interoperable next-generation grids, many commercial enterprises are beginning to use grid technologies as well [13]. As grid is being used commercially, maintaining the QoS, level of security demand by users and resource security, resource integrity, confidentiality of communication, privacy of user information are prime concerns. Grid security itself imposes several unique security challenges, including managing user identities across local and global networks, trust relationships between entities, end-user key and credential management, and providing security to resources against malicious acts from grid users. [9] The fusion of web services and grid technologies further increases the concerns about security problem for their complex nature [16]. Security is a very important component of the Grid, as misuse of its vast resources can result in considerable damage.

Basic security requirement and technologies in grid

Any system's security goal is to provide easy access to legitimate users and to prevent users who don't have the proper privileges from accessing resources and information. This section presents the traditional security areas that play an important role in defining security for grid and the associated technologies. [12]

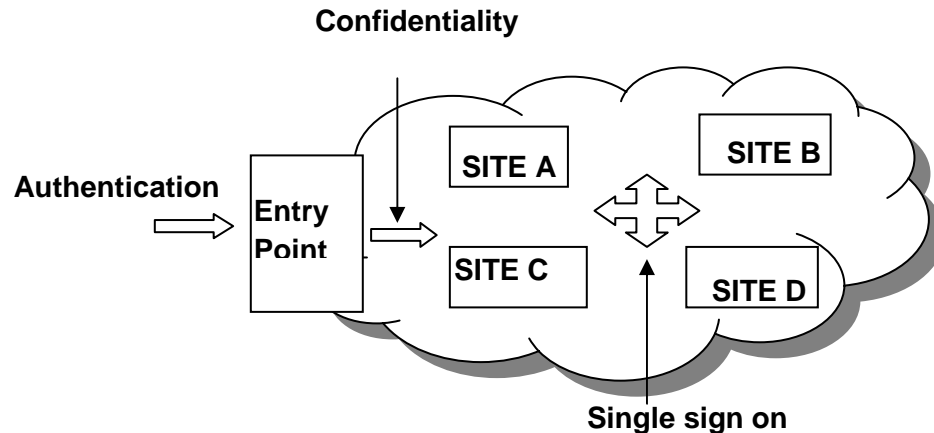


Figure 2.3: Typical grid Scenario

2.7.1 Authentication

Grid security requirements should contain authentication mechanisms at the entry points. Authentication deals with verification of the identity of an entity within a network. An entity may be a user, a resource provider or service provider. The main requirements for authentication include scalability, trust across different Certification Authorities and timeliness of revocation. It is possible to have different authentication mechanisms for different sites within a grid. Therefore, the security protocol should be flexible and scalable to handle all the different requirements and provide a seamless interface to the user.

The most prevalent mechanisms of authentication in a GSI (Grid Security Infrastructure) based grid is Public Key Infrastructure (PKI), which defines message formats and protocols that allow entities to securely communicate claims and statements. The most widely used certificate format in PKI is X.509. In this PKI, highly trusted

entities know as certificate authorities (CA) issue X.509 certificates where essentially a unique identity name and the public key of an entity are bound through the digital signature of that CA. In addition to certificate based mechanism, Kerberos and can either be a peer-to-peer relationship such as a password that only the client and the server know have also been implemented .

2.7.2 Authorization

After authentication, the second most fundamental challenge to security in a computational grid is authorization. Once the identity of a user can be established, it is then necessary to determine the permission of that user to use a particular resource. Authorization ensures that every authenticated user can only access the resources that he or she is allowed to. Thus authorization is closely related to access control trust.

There are several technologies used for handling authorization in grid environment. One of the widely used technologies was in the form of globus toolkit grid map file. [11] This is a text file that holds the list of the X.500 distinguished name of the grid users and the equivalent local user account names that they are to be mapped into. This must be installed and maintained on each participating host. The Community Authorization Service (CAS) was the next attempt by the Globus to improve the user authorization. CAS allows a resource owner to grant access to a portion of his/her resource to a VO and then let the community determine who can use this allocation. The US fusion grid project [8] used resource oriented authorization manager (ROAM) for simplifying the process of requesting, viewing and granting authorization.

2.7.3 Delegation and Single Sign-on

A single computation may require access to many resources and many times, but requiring a user to re authenticate on each occasion is impractical and generally unacceptable. Instead, a user should be able to authenticate once and then assign to the computation the right to operate on his or her behalf, typically for a specified period. This capability is achieved through the creation of a proxy credential [15] .The GSI provides single sign-on and delegation capabilities by creating a proxy consisting of a new

certificate (with a new public key in it) and a new private key. The new certificate contains the owner's identity, modified slightly to indicate that it's a proxy. The owner, rather than the CA, signs the new certificate, which also includes a time notation after which others should no longer accept the proxy. Another very important requirement for a grid-based security system is delegation. In a Grid, a user may need to ask an entity to perform an operation on his/her behalf. For example, a scheduler may need to query an information service to determine on which machines a user is allowed to and has enough allocation to execute a job. The scheduler does not have the privilege to ask the information service directly, because the information service protects its potentially proprietary information from unauthorized access. The information service would provide a user's information to a scheduler if it was acting on that user's behalf. That's the reason why delegation comes to the reality. Delegation is the technique that an entity could be asked to perform some operations on that user's behalf. To simplify the process even further, proxy certificates are also used for delegation. [15]

2.7.4 Secure Communication

Secure communication provides the ability for two or more entities to conduct a conversation with integrity, confidentiality and non-repudiation data communication. Depending upon on the layer where security is implemented there is two categories of security mechanism, transport level security and message level security. GT4 implements transport security using a Secure Socket Layer (SSL) implementation, which provides the security properties. As an alternative, GT4 uses message-level security (MLS) to implement encryption, authentication, and integrity mechanisms at the message layer, rather than at the transport layer, using Web services standards such as WS-Security and WS-Secure Conversation. [16]

2.7.5 Accounting and audit

Auditing allows resource providers and solution producers to see what actions were done by whom and when. While auditing is not directly a mechanism to secure a system, it is nevertheless a vital component in the overall security architecture. [15] It can also ensure

that all parties keep to their resource usage agreements. Accounting is a special form of auditing where resource and service usage is logged for billing purposes. Like auditing, accounting and billing is not a security service itself but a system which has to be secured. Accounting and audit play important roles in the grid environment.

2.8 Security Threats classification in Grid

Although grid security system provides fundamental security mechanisms like authentication, authorization, confidentiality also web services provides security systems like WS-security, WS-conversations, still there are many other possible security threats associated with grid environment. Below security threats classification given in table 1 are based on the four types of actors associated with grid. [6]

2.8.1 User Threats

When user credentials are not secure enough or the fundamental security mechanism like authentication or authorization are not carried out properly, there is a chance that different user threats will generate. [7]

2.8.2 Mediator Threats

During the communication between user and service or resource provider, mediator carries the corresponding data or message in XML format. So mediator threat consists of those threats originating from insecure service level communication. Insecure communication includes eavesdropping and intercepting service communication.

2.8.3 Service Provider Attack

The service provider takes the job submitted by the user, process them and send back the result to the user with quality-of-service. So service threats are those kinds of threats that composed of malicious input to get malicious goal like malicious code/ malware.

2.8.4 Resource Provider Attack

A resource provider provides supplies two types of resources including physical resources like CPU cycles, storage, bandwidth etc and software resources. So resource

provider threats composed of those kinds of threats that are unauthorized use of the physical resources or integrity threats or destroying the software resources.

Table 1: Threats classification in Grid

User Threats	Communication Threats	Service provider attack	Resource provider attack
1.User credential theft 2.User credential compromise 3.User impersonation	1.Network eavesdropping 2. Man in the middle attack 3. Replay/session hijack 4. Brute force attack 5. Denial of service attack 6. Sniffing/snooping 7. Mosque rending 8. Routing detours	1.malicious code / malware 2.SQL injection 3.Access control attack 4.Defeating auditing and accounting service 5.object reuse 6.virus and worms 7.dynamic xml 8. Improper error Handling	1.Resource and data attack 2.Improper privilege management 3.configuration vulnerability

2.9 Conclusion

From the possible vulnerability of grid as given in section 2.8, Denial of Service attack (DoS) and Distributed Denial of Service attack (DDoS) are immense threats to grid computing performance. 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]. In the next chapter we discussed how DoS and DDoS attack affects grid performance and literature survey of some related work to protect grid from denial of service attack.



Chapter 3

DDoS in Grid environment

Chapter**3****DDoS in Grid environment**

3.1 Related Work

Distributed Denial-of-Service (DDoS) and Denial-of-Service (DoS) are the most dreadful network threats in recent years. The vulnerabilities of grid environment in the presence of DDoS have been presented in [24] and they have proposed a distributed defense system for grid. Authors of [25] discussed the need for an intrusion detection system in grid environment. They have classified grid intrusions into 4 types i.e. 1) Unauthorized access 2) Misuse 3) Grid exploit 4) Host or Network-specific attacks. They proposed a model that is composed of high-level GIDS (Grid intrusion detection system) 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 [28]. Many articles like [27] use traffic volume [flow, packet, byte count] as the metric for anomaly detection system. Volume based detection schemes were proved as a good metric for anomaly detection system and it can detect anomalies that cause large traffic changes, but small DoS attacks that do not cause much changes in traffic can not be detected perfectly. 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) [11] 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. Recently there has been use of entropy and traffic distribution for detecting DDoS attack

anomalies. A number of articles suggested entropy as a metrics to summarizing traffic distribution for anomaly detection [18] [19] [20] [29]. Author [18] uses entropy of distribution of source (destination) address for DDoS detection. In [19] included PCA framework with entropy based metrics and shown that it can detect wide variety of types of anomalies that can not identified using volume based analysis only. In [20] Lee and Xiang suggested use of different information theoretic measures for detecting malicious activities. The authors of [21] use entropy rate to discriminate the DDoS attack from legitimate traffic. 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) [29]. The anomaly detection system discussed in this paper is based on by analyzing the change in entropy of above two traffic distributions.

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.

3.2 DoS Attack

A Denial of Service (DOS) is an attack with the intent of compromising availability. A DoS attack involves sending large number of unused packets to a particular destination to prevent legitimate users from accessing information or services. The attack is aimed at shutting down a computer or a network, by exhausting a computer's resource or by exhausting bandwidth limits of a network.

Examples of DoS Attack

- Smurf attack.

- Ping flood.
- SYN flood.
- ICMP flood.
- UDP flood.

3.3 DDoS Attack

The most common implementation of DoS attack is DDoS attack. The attack is distributed because it is a coordinate attack on the availability of given target system or network that use the computing power of thousands of compromised machines known as “zombies” to target a victim. A DDoS attack is carried out in several phases. To launch a DDoS attack malicious user first searches for some compromised hosts. This process is usually performed automatically through scanning of remote machines, looking for security holes. Subsequently the malicious user installs attack tools on the compromised host (Zombies). In a typical DDoS attack, Zombies are gathered to send useless service requests, packets at the same time. There are four popular known DDoS attack tools:

- Trinoo
- Tribe Flood Network (TFN)
- Tfn2k)
- Stacheldraht (German for barbed wire)

DoS or 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. Due to the scalability and dynamic nature of grid some security flaws are there and the huge resource capacity of grid computing may become a next platform for the attackers [24]. If an intruder got unauthorized access to grid, then it may pose as an actively participating node in the grid. Through message request in the grid application and by using the huge bandwidth the

malicious node may use grid network as a massive traffic generator for launching DDoS attack.

As a secure framework for 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). DoS and DDoS scenario has been shown in the figure 3.1.

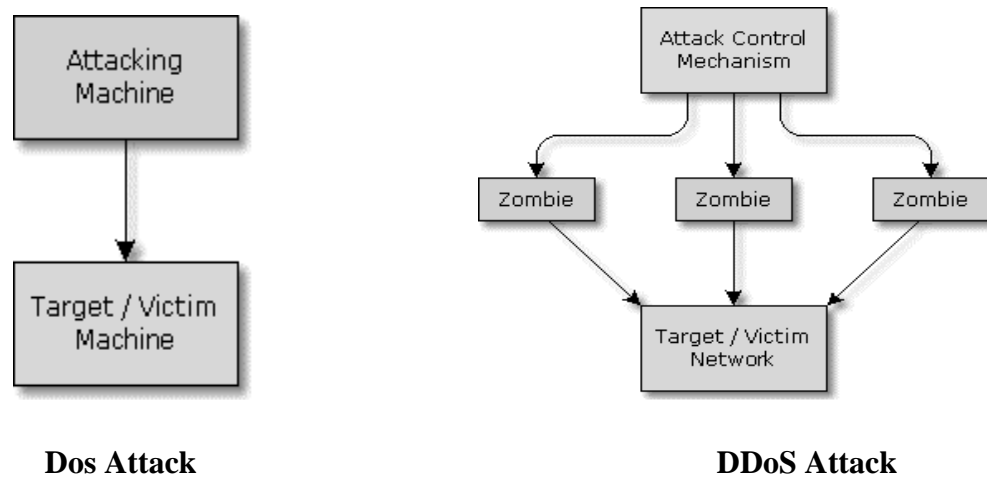


Figure 3.1: DoS and DDoS Attack Scenario

3.4 DDoS Attack Classifications

We can classify of DDoS attacks into bandwidth depletion and resource depletion attacks as shown in the figure 3.2. A bandwidth depletion attack is designed to flood the victim network with unwanted traffic so that legitimate traffic will discard. A resource depletion attack is designed to tie up the resources of a victim system. This type of attack targets a server or process on the victim system making it unable to process legitimate requests for service. Flooding attacks which comes under bandwidth depletion is the most possible attack in grid computing. Flooding is an attack that is designed to bring a network or service down by flooding it with large amounts of traffic.

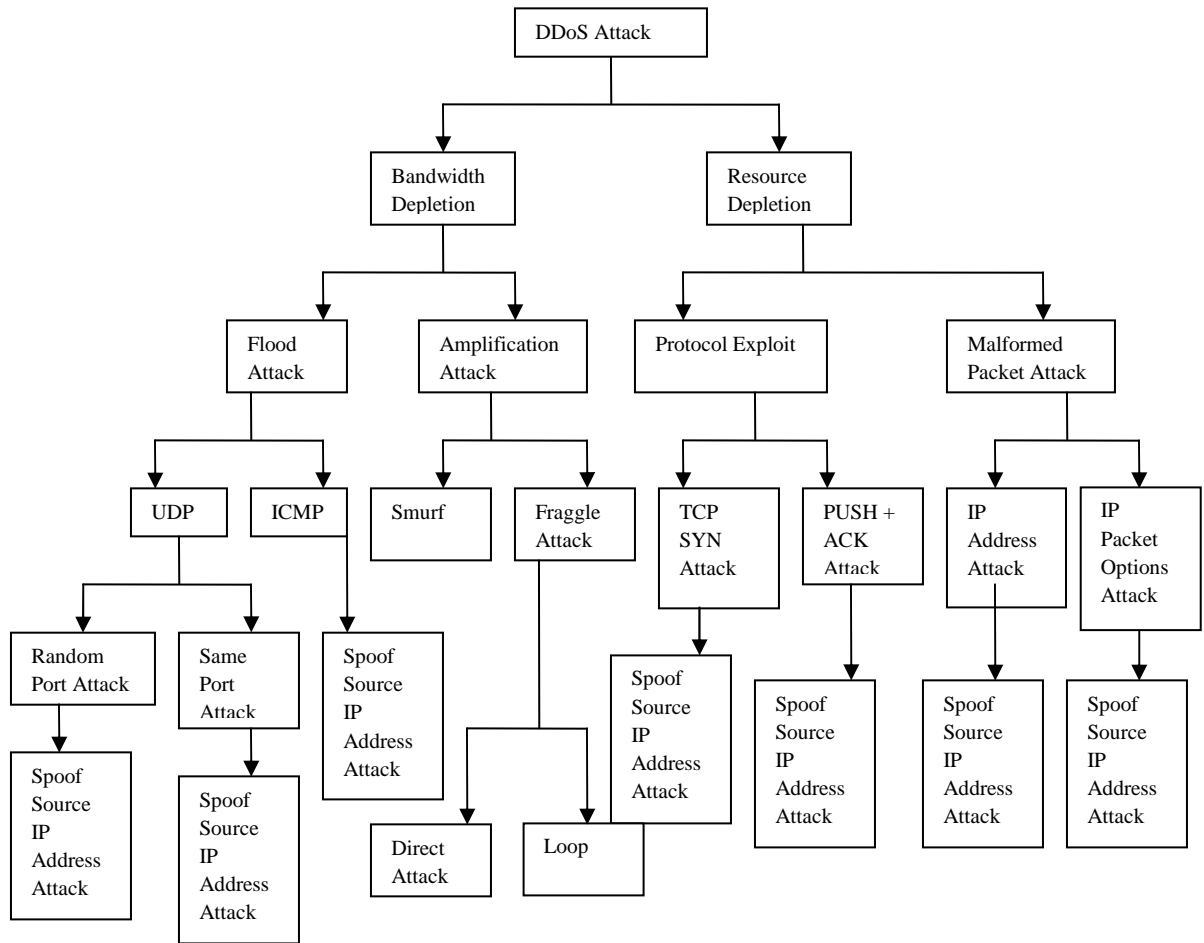


Figure 3.2: DDoS Attack classification

3.5 DDoS Defense Systems

Denial-Of-Service attacks have become an increasingly prevalent form of security threat, and the problem appears to be quite difficult to solve. Several countermeasures have been proposed. Based on the DDoS attack detection and defense strategies, we have divided complete defense system for DDoS attack into three 3 steps: (i) Attack prevention, (ii) attack detection and recovery, and (iii) attack identification.

3.5.1 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 [23].

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 [31].

3.5.2 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 [25] 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 [28] 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 difficulty 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 [28]. 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.

3.5.3 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 [17], 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. [22]

3.6 Conclusion

As pointed in this chapter the performance and QOS (Quality of Service) can be adversely if the arising security issues due to DDoS and DDoS attack are not addressed properly in time and accurately. In the next chapter we discussed our proposed anomaly detection system based on entropy and entropy which can detect Denial of service attack with in the network before it affects the victim.



Chapter 4

Entropy Based ADS

Chapter

4

Entropy Based ADS

4.1 Entropy

Entropy or Shannon-Wiener index is an important concept of information theory, which 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 $[0, \log n]$. The entropy value is smaller when the class distribution is pure i.e. belongs to one class. The entropy value is larger when the class distribution is impure i.e. class distribution is more even. Hence comparing the value of entropy of some sample of packet header fields to that of another sample of packet header fields provides a mechanism for detecting changes in the randomness.

The entropy $H(X)$ of a random variable X with possible values $\{x_1, x_2, \dots, x_n\}$ and distribution of probabilities $P = \{p_1, p_2, \dots, p_n\}$ with n elements, where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$ can be calculated as

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

4.2 Proposed ADS for Grid

In our proposed anomaly detection algorithm we use entropy as a principal matrix. We use change of entropy of traffic distributions (IP address, port) for DDoS detection. 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. $m = \sum_{i=1}^n m_i$.

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

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 .

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

Here 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 normalized 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 [30] for a stochastic processes the entropy rate $H(\chi)$ of two random processes are same.

$$H(\chi) = \lim_{n \rightarrow \infty} \frac{1}{n} H(x_1, x_2, \dots, x_n) \quad (4)$$

The steps in our proposed DDoS detection algorithm are described in figure 4.1.

<i>Algorithm 1 : DDoS detection algorithm</i>	
Step i:	Collect sample flows for a time window T on the edge routers.
Step ii:	Calculate router entropy $H(x) = - \sum_{i=1}^n P(x_i) \log P(x_i)$.
Step iii:	Calculate $NE = (H / \log n_0)$ where, $NE =$ normalized router entropy.
Step iv:	If $NE < \text{threshold} (\delta_1)$, identify the suspected attack flow.
Step v:	Calculate the entropy rate $H(\chi) = \lim_{n \rightarrow \infty} \frac{1}{n} H(x_1, x_2, \dots, x_n)$ of the suspected flow in that router and the routers on down stream.
Step vi:	Compare $H_i(\chi) \forall i \in \mathcal{E}$ entropy rates on routers.
Step vii:	If $H_i(\chi) \leq \text{threshold} (\delta_2)$, it is a DDoS attack. Else legitimate traffics.
Step viii:	Discard the attack flow.

Figure 4.1: DDoS detection algorithm

Definition 1

A stochastic process $\{X(t), t \in T\}$ is a collection of collection of random variables. For each $t \in T$, $X(t)$ 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.

$$\Pr\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

$$= \Pr\{X_{1+l} = x_1, X_{2+l} = x_2, \dots, X_{n+l} = x_n\}$$

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

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(x_1, x_2, \dots, x_n)$$

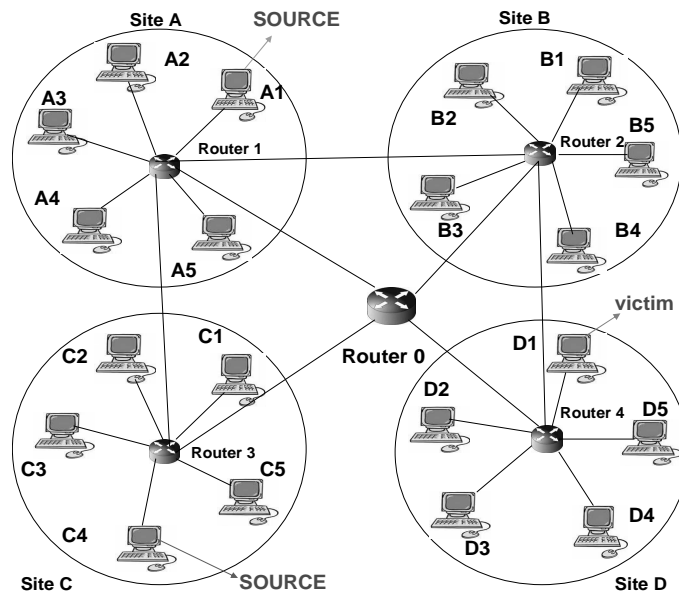
4.3 Implementation Details

Figure 4.2: Grid Topology

In this section it is described using figure 4.2 how the proposed anomaly detection system can be implemented in grid infrastructure and how routers will communicate with each other to detect the attack. Grid computing can be thought as a virtual organization which is a collection of some real organizations or sites [26]. In the figure 4.2 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.

We have taken two examples using figure 4.2 how the detection scheme works. Suppose 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 3 and router 2 we have captured flows as given in table 2, table 3 and in table 4 respectively in a fixed time window T . The router entropy is calculated according to “Eq. (2)” and the normalized router entropy is being calculated using “Eq. (3)”.

Table 2. Data for router 1

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

Here router entropy = $0.47 + 0.44 + 0.51 = 1.42$ and $n_0 = 3$

Normalized router entropy NE = $1.42 / \log_2 3 = 0.89$

Table 3. Data for router 3

Source node	Destination node	No of packets	entropy
C1	B4	2	0.48
C2	D3	2	0.48
C4	D1	5	0.47

Here Router entropy = $0.47 + 0.48 + 0.48 = 1.43$ and $n_0 = 3$

Normalized Router entropy NE = $1.43 / \log_2 3 = 0.90$

Table 4. Data for router 2

Source node	Destination node	No of packets	entropy
B1	C2	2	0.52
B2	C3	2	0.52
B3	C1	2	0.52

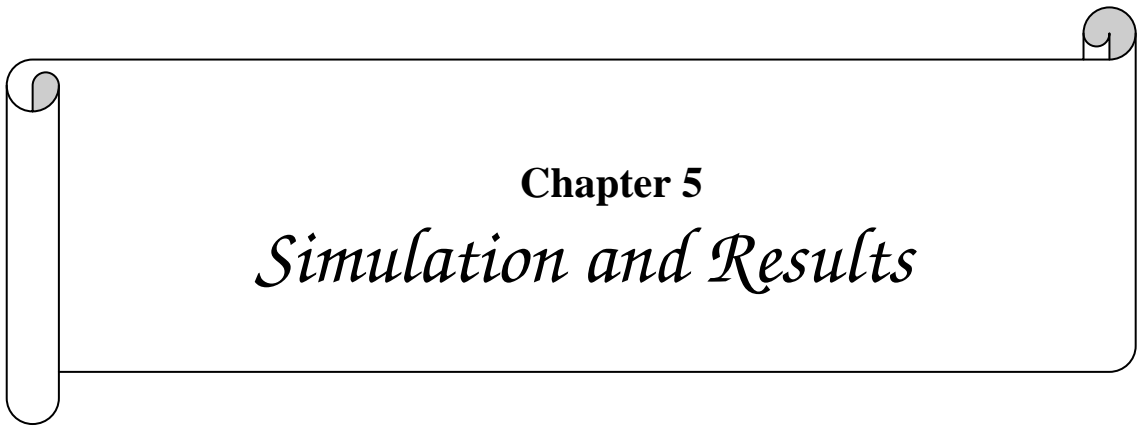
Here Router entropy = $0.52 + 0.52 + 0.52 = 1.56$ and $n_0 = 3$

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 comparatively 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 4.2, 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 above calculations are based on \log_2 .

4.4 Conclusion

In this chapter we have described about our proposed entropy based anomaly detection algorithm. In the next chapter we are going to evaluate the performance of this algorithm using NS2 network simulator



Chapter 5
Simulation and Results

Chapter

5

Simulation and Results

The simulation was done using NS-2 simulator [10] to evaluate the performance of our DDoS detection algorithm with results from the experiment. We tested our anomaly detection algorithm on a 3.0GHz processor, in linux (Ubuntu 8.04) environment. This section introduces the experimental setup and reports performance results.

5.1 Experimental setup

Our simulation includes 3 source, 2 intermediate routers and 2 destination nodes as shown in figure 5.1. Out of which 3 source nodes 2 nodes are attackers and 1 node is a legitimate user. The bandwidth of legitimate traffic is set constant and the simulation of attack traffic is achieved by randomly generating many pairs of Constant Bit Rate (CBR) UDP flows in NS2. The legitimate user to send packets from the time of .20 second and the attacker starts sending attack traffic at .30 seconds. The experiment lasts at 3 second. We traced no of packets received in every 0.5 second interval. In table 5 the traced data are shown.

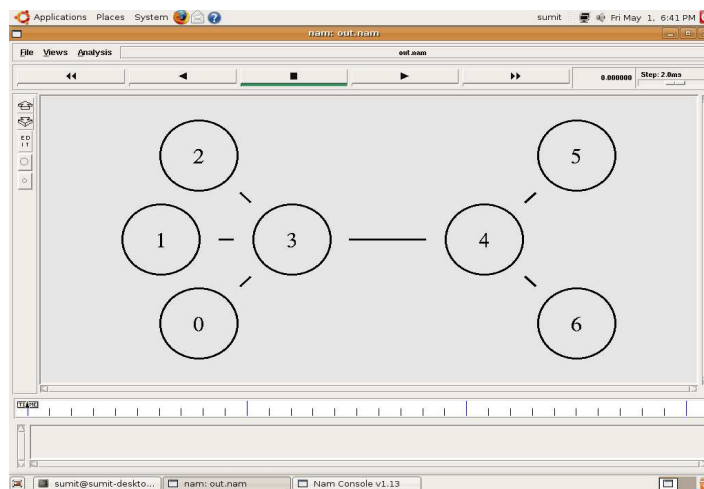


Figure 5.1: Screen shot of Setup

Table 5: Traced Data

Time in second	Number of attack packet received	Number of legitimate packet received
0-0.5	106	47
0.5-1.0	268	25
1.0-1.5	266	49
1.5-2.0	266	31
2.0-2.5	268	134
2.5-3.0	266	133

Table 6: Normalized Router Entropy Calculations

Time in Second	Normalized Router Entropy
0-0.5	0.8899
0.5-1.0	0.4207
1.0-1.5	0.6235
1.5-2.0	0.4825
2.0-2.5	0.9183
2.5-3.0	0.9183

5.2 Performance Evaluation

We consider 2 situations in our simulation to evaluate the performance of our proposed algorithm. In the first situation, we start the attack without using any defense mechanism and study how the system performance is being degraded. In this case the router at the victim simply drops packets including legitimate packets which it can not handle. In the

second situation, we deploy the entropy based anomaly detection system in the simulation network and examine how the system performance increases significantly.

The graph in Figure 5.2 depicts the effect of DDoS attack without any defense system. It shows numbers of attack packets as well as legitimate packets with respect to time. We can mark that with in same time different zombies combined to send attack packets as a result the number of attack packets increases significantly. For which the packet drop rate of legitimate users increases dramatically as shown in the Figure 5.3.

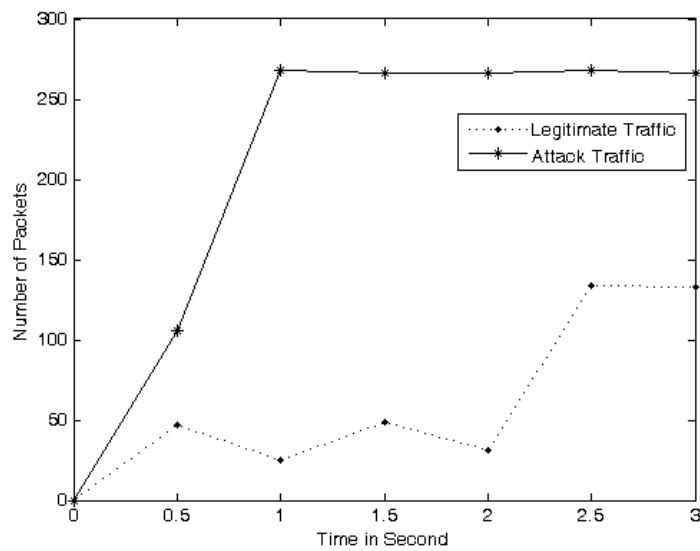


Figure 5.2: Effect of DDoS Attack

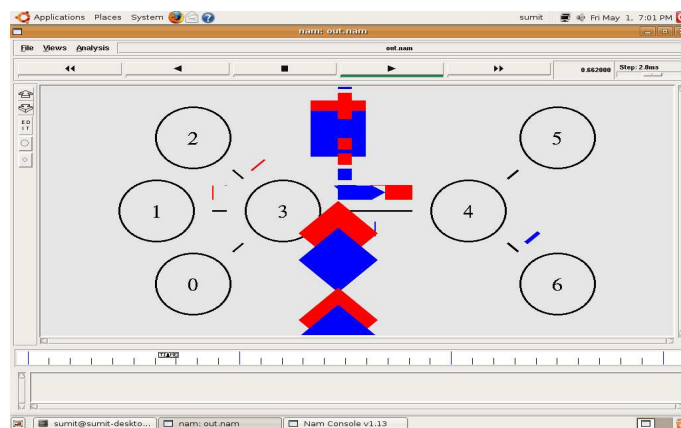


Figure 5.3: Screen shot of packet drop

We use 2 metrics attack detection rate and false-positive rate as performance evaluation metrics.

1) Detection Rate $R_d = d / n$

d = number of attack packet detected in the simulation experiments.

n = total number of attack packet generated.

The figure 5.4 shows attack detection rate of our proposed algorithm with respect to detection threshold. We found that the attack detection rate detection is almost 100% when the detection threshold δ_1 i.e normalized entropy is more than 0.92. But after calculating the false positive rate we will finalize the threshold δ_1 in this experiment.

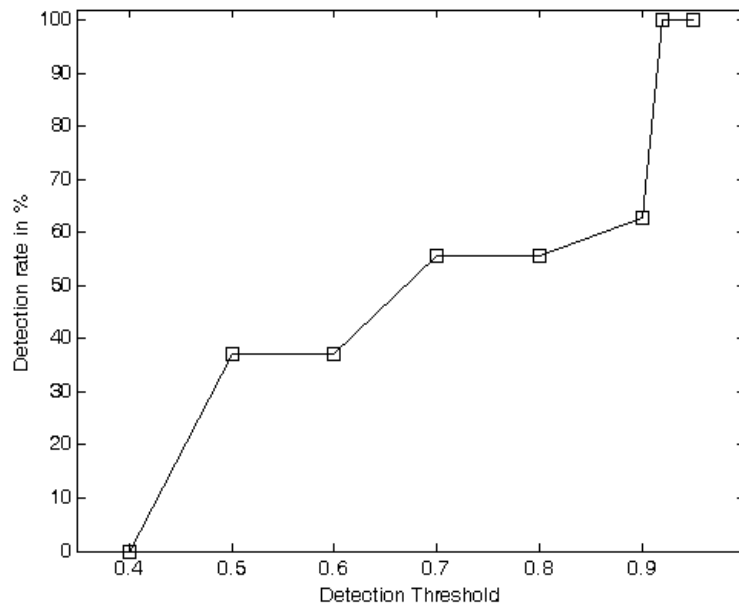


Figure 5.4: DDoS attack Detection Rate

2) False positive Rate $R_{fp} = p / m$

P = number of false positive raised

m = number of legitimate traffic flow checked by the simulator

When calculating false positive rate we found that up to setting detection threshold 0.94 the false positive rate is 0, but putting threshold 0.95-0.99 increases false positive rate 67.26 and after then false positive rate 100%.as shown in figure 12.

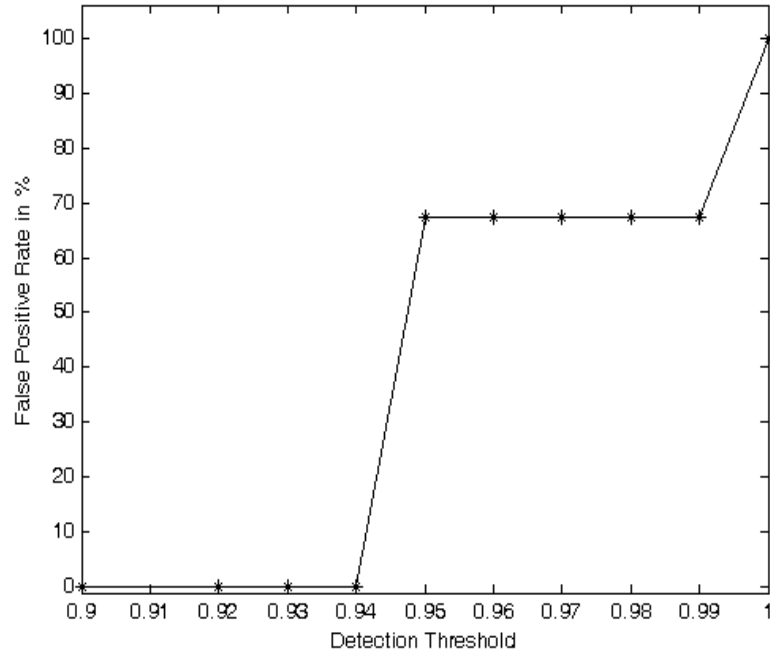
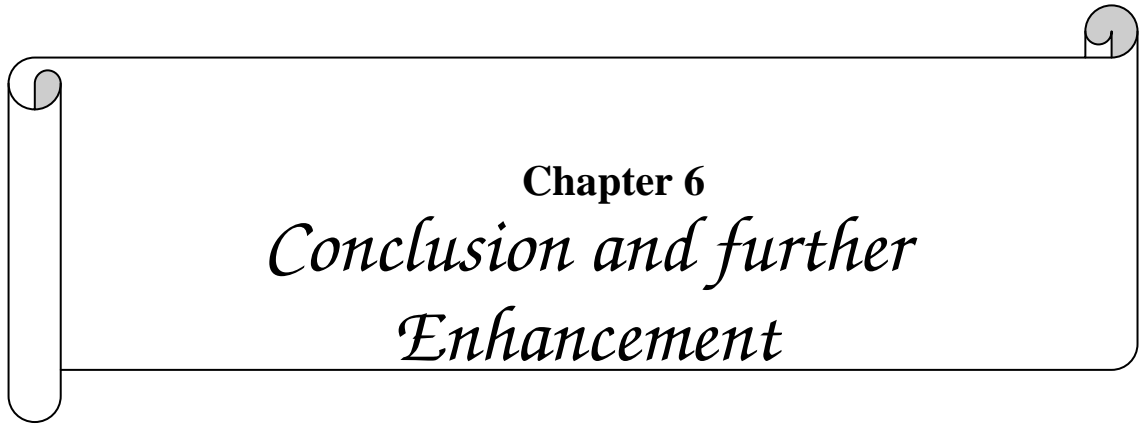


Figure 12: False Positive Rate

5.3 Conclusion

According to the obtained simulation results, if we will take the detection threshold δ_1 as 0.94 then, the proposed anomaly detection system can detect DDoS attack traffics with 100% detection rate and with out any false positive. If the normalized router entropy in any case will be less than 0.94 it starts detecting the suspected flow. Again entropy rate is being calculated according to the DDoS detection algorithm given in section 4.2 to confirm the attack.



Chapter 6
*Conclusion and further
Enhancement*

Conclusion

This thesis takes a look at different vulnerability associated with grid. Out of which Distributed Denial of Service is an immense threat to grid services. We found that applying IDS to it can significantly increase its security which can deploy along with existing security measures. But existing IDSs are not designed to be grid specific. The objective of this study was to investigate how DDoS attack affects the grid performance and designing a grid specific anomaly detection system. DDoS attack is a major threat that can not be addressed well if various defense systems do not organize in to framework exchanging information and services. For which we have divided DDoS defense system in to 3 steps (i) Attack prevention (before attack), (ii) attack detection and recovery (during the attack), (iii) attack identification (after attack).

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.

Our experiment results also shows that if the threshold setting is perfect our proposed entropy based ADS can efficiently detect DoS and DDoS attack with a high detection and low false positive rate.

Further Enhancements

The main challenge lies in our proposed anomaly detection system are:

- 1) Discriminating legitimate traffics when there is a large number of legitimate accessing to a particular server.
- 2) Generally attacker uses same attack generation function to create attack packets at compromise hosts or zombies, but if attacker starts generating attack packets using different attack packet generation function.
- 3) Setting up a good threshold Value is very difficult and that should be dynamic depending up on the current situation.

As a next step we are going to work to eliminate those above issues and apart from DDoS attack, there are many other security threats possible on grid and in near future we will work more security issues.

Bibliography

- [1] I. Foster, C. Kesselman, "The Grid: Blueprint for a new computing Infrastructure", Morgan Kaufmann publishers, San Francisco, USA, 1999.
- [2] "Introduction to Grid Computing with Globus", sep 2003, <http://www.ibm.com/redbooks> .
- [3] I. Foster, C. Kesselman & S.Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations ", international journal of high Performance computing applications, vol. 15, pp. 200-222, sage publishers, London, UK 2001.
- [4] J. Joseph, M. Ernest, C. Fellenstein, "Evolution of grid computing architecture and grid adoption models ", IBM systems journal, vol 43, no 4, 2004.
- [5] Ian Foster, Carl Kesselman, Jeffrey M. Nick and Steven Tuecke, "The Physiology of Grid: An Open Grid Services Architecture for Distributed System Integration", Tech. Report, Globus Project, June, 2002, <http://www.globus.org/research/papers/ogsa.pdf>.
- [6] N. Jiancheng, L. Zhishu, G. Zhonghe, S. Jirong, " Threat analysis and Prevention for grid and web security services" , In proc. of Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, pp. 526-531, 2007.
- [7] Y. Demchenko, L. Gommans. C.D Laat., B.Oudenaarde "Web services and grid Security vulnerabilities and threats analysis and model", in proc. of the 6th IEEE/ACM International Workshop on Grid Computing, pp. 262-267,2005.
- [8] J.R Burruss, T.W. Fredian, M.R Thompson," Security on the US fusion grid", Fusion Engineering and Design, 81 (15), pp. 1949-1955, Jul 2006.
- [9] E. Cody, R. Sharman, R.H. Rao, S. Upadhyaya, "Security in grid computing: A review and synthesis", decision support systems, volume 44, issue 4, pp.-749-764, March 2008.
- [10] McCanne S. and S. Floyd, NS-2 Network Simulator, <http://www.isi.edu/nsnam/ns>, 1997.
- [11] The Globus Project, the Globus Toolkit 4, 2004, <http://www.globus.org/toolkit/>

- [12] A. Chakrabarti, "Grid computing security", New York, springer, 2007.
- [13] Sun Grid, <http://www.sun.com/2006-0320/feature/index.jsp>
- [14] Michael Litzkow, M. Livny, and M. Mutka." Condor - A hunter of idle workstations. In Proceedings of the 8th International Conference of Distributed Computing Systems, pp. 104-111, June 1988.
- [15] M. Smith, T. Friese, M. Engel, B. Freisleben, "countering security threats in service oriented on-demand grid computing using sandboxing and trusted computing techniques", journal of parallel and distributed computing, pp. 1189-1204, 2006.
- [16] S. Shirasuna, A. Slominski, and D.Gannon, "Performance comparison of security mechanisms for grid services", Proceedings in 5th IEEE/ACM International Workshop on Grid Computing, pp. 360-364, 2004.
- [17] D. Song, A. Perrig, "Advanced and authenticated marking schemes for IP traceback", in Proceedings of IEEE INFOCOM, 2001, pp. 878-886.
- [18] 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 (DISCEX 2003), pp. 303-314, April 2003.
- [19] A. Lakhina, M. Crovella and C. Diot, "Mining anomalies using traffic feature Distributions ". In proc. of ACM SIGCOMM, pp. 217-228. August 2005.
- [20] W. Lee, D. Xiang, "Information-theoretic measures for anomaly Detection", In Proc. of IEEE Symposium on Security and Privacy, (Oakland, CA), pp. 130-143, 2001.
- [21] S. Yu , W. Zhou, "Entropy-based Collaborative Detection of DDoS attacks on Community Networks", In Proc. of 6th IEEE international conference on pervasive computing and Communications, pp. 566-571, 2008.
- [22] 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), pp.431-436, July 2005.
- [23] J. Mirkovic, J. Martin and P. Reiher, "A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms," ACM SIGCOMM Computer Communications Review, Vol.34, No. 2, pp. 39-53, April 2004.

- [24] 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.
- [25] A. Schulter, J. A. Reis, F. Koch, C. B. Westphall “A Grid-based Intrusion Detection System”, In Proc. of International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies ICNICONSMCL’06, pp.187.2006.
- [26] 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), April 2006.
- [27] A. Lakhina, M. Crovella, and C. Diot. ” Diagnosing Network-Wide Traffic Anomalies”, In ACM SIGCOMM Computer Communication Review, Portland, pp. 219 - 230, October 2004.
- [28] P. G. Teodoro, J. D.Verdejo, G. M.Fernandez, E.Vazquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges”, computer & security, Volume 28, Issues 1-2, pp.18-28, 2008,
- [29] 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.
- [30] Thomas M. Cover and Joy A. Thomas, “Elements of Information Theory”, second edition, John Wiley & Sons, New York, 2007.
- [31] J. Wang, X. Liu and A. Chien, “Empirical Study of Tolerating Denial-of-Service Attacks with a Proxy Network”, In Proc. 14th conference Of the USENIX Security Symposium, 2005.

Dissemination

- [1] Sumit Kar, BibhuDatta Sahoo, “*An Anomaly Detection System for DDOS Attack in Grid Computing*”, International Journal of computer Applications in Engineering Technology and Sciences (IJ-CA-ETS), Volume 1, Issue 2, pp. 553-557, April 09.

- [2] Sumit Kar, Bibhudatta Sahoo, “*Securing Grid against Distributed Denial of Service attack using an Entropy based Anomaly Detection System*”, National Conference on Research Trends in Computer Applications (NCRCTCA- 09), pp. 47-51, March 2009.

- [3] Sumit Kar, Bibhudatta Sahoo,” *Security Issues and Preventive Model for Grid computing*”, National Conference on Modern Trends of Operating Systems, (MTOS- 09), pp.35-38, March, 2009.