A

Dissertation

On

Prevention of DOS & DDOS Attack in Cloud Computing using Filtering Method

Submitted in partial fulfillment of the requirement of the

Award of the Degree of

Master of Technology

In

Computer Science & Engineering

Submitted By

DEVENDRA DADORIYA

University Roll No. 2K11/CSE/04

Under the esteemed guidance of

Mr. R. K. YADAV

Assistant Professor

Computer Engineering Department, DTU, Delhi

DELHI TECHNOLOGICAL UNIVERSITY

2011-2013

CERTIFICATE

This is to certify that the dissertation titled "**Prevention of DOS & DDOS Attack in Cloud Computing using Filtering Method**" is a bona fide record of work done at **Delhi Technological University** by **Devendra Dadoriya, Roll No. 2K11/CSE/04** for partial fulfilment of the requirements for degree of Master of Technology in Computer Science & Engineering. This project was carried out under my supervision and has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma to the best of our knowledge and belief.

Date: _____ **(Mr. R. K. Yadav)**

Project Guide

Department of Computer Engineering

Delhi Technological University

ACKNOWLEDGEMENT

First of all, let me thank the almighty god, my parents and my dear friends who are the most graceful and merciful for their blessing that contributed to the successful completion of this project.

I feel privileged to offer sincere thanks and deep sense of gratitude to **Mr. R. K. Yadav,** project guide for expressing his confidence in me by letting me work on a project of this magnitude and using the latest technologies and providing their support, help & encouragement in implementing this project.

I would like to take this opportunity to express the profound sense of gratitude and respect to all those who helped us throughout the duration of this project. **DELHI TECHNOLOGICAL UNIVERSITY**, in particular has been the source of inspiration, I acknowledge the effort of those who have contributed significantly to this project.

Devendra Dadoriya

University Roll no: 2K11/CSE/04

M. Tech (Computer Science & Engineering)

Department of Computer Engineering

Delhi Technological University

Delhi - 110042

ABSTRACT

Cloud Computing is an emerging technology. Cloud computing provides services to users on-demand and online. Users use services and according to service pay for that service. Cloud has more number of security issues. Availability is the one of the important security issue. Availability security issue are affecting by the DOS and DDOS attack.

DOS and DDOS are affecting victim. Victim does not provide services or deny the users for provide services. In the DOS and DDOS attack, attacker sends large number of packets to the victim and victim process useless packet. Attacker sends TCP SYN and TCP ACK packets. TCP SYN packet affects the memory exhaustion of the victim because attacker sends large number of the TCP SYN packet. TCP ACK packet affects the CPU exhaustion of the victim because victim does not response legitimate user and waste most of the time process useless packets.

Count based filtering method solve the DOS and DDOS attack. Count based filtering method extract the fields of the TCP frame format and count the packet which are the same fields that are extract and maintain count. According to count in the single timestamp packets will be accepted or reject.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AWS:	Amazon Web Services
CBF:	Confidence Based Filtering
CSP:	Cloud Service Provider
DDOS:	Distributed-Denial-of-Service
DOS:	Denial-of-Service
EC2:	Elastic Cloud Computing
GWT:	Google Web Toolkit
HPC:	High Performance Computing
IaaS:	Infrastructure-as-a-Service
IIS:	Internet Information Service
ISP:	Internet Service Provider
LAN:	Local Area Network
MPI:	Message Passing Interface
PaaS:	Platform-as-a-Service
QoS:	Quality-of-Service
S3:	Simple Storage Service
SaaS:	Software-as-a-Service
SLA:	Service Level Agreement
SOA:	Service Oriented Architecture
TTL:	Time-To-Live

VC: Virtual Computer

VM: Virtual Machine

WAP: Windows Azure Platform

CHAPTER 1

INTRODUCTION

Cloud computing is the one of the most leading technology in recent time and attains lots of attention of the organizations and researchers. Cloud computing has more advantages for customer applications. Its main feature is that information is available anywhere and unlimited. Cloud Computing has possibility for private information that cannot be shared to anyone and public information that can be shared to all. The data is available in the cloud, customer access by mobile, laptop or other devices. Now day's lots of people use mail online create picture album and upload their photos to the server. These are the applications that are running through the internet and store data remotely or in server. It means that data is not stored to the local machine.

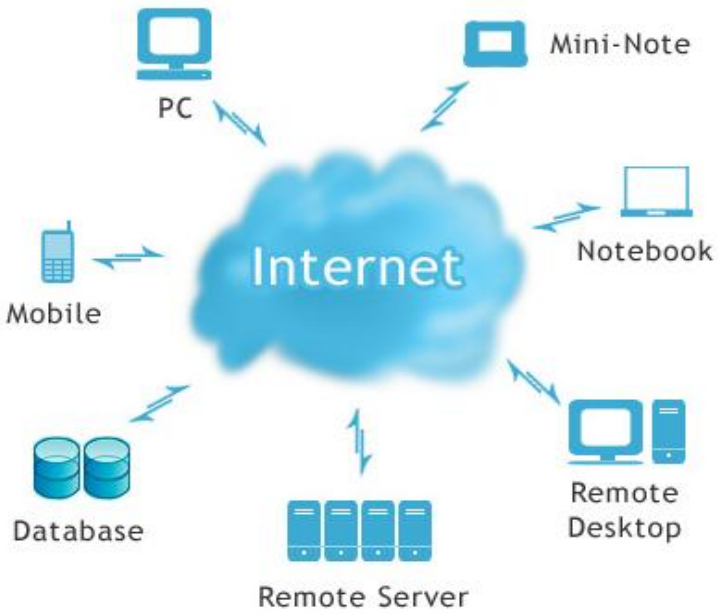

Figure 1.1 Cloud computing

These services can be provided free or by payment and that can be decided by which type of service you used. Virtual machine decide customer only those who use service that has been provided by cloud service provider (CSP). Cloud computing

provides services not only for customers but also for organizations. Organization use computational speed of the computers and some type of programming environment that are suitable for the organizations. All the resources that are provided by cloud is manage by CSP.

1.1 Research Question

Cloud computing is one of the most leading and important technology at present. In the cloud computing there are more security threats because of these threats cloud computing compromises its security. Availability is the most important security threats and this problem occurs because server is not available to respond the customers. Availability security problem occurs due to denial of service (DOS) attack and distributed denial of service (DDOS) attack.

How to Prevent DOS & DDOS Attack in Cloud Computing using Filtering Method

1.2 Purpose of Research

The purpose of the thesis is to find out the solution of prevention of server from DOS and DDOS attack. In this attack server leads to two types of problem CPU exhaustion and memory exhaustion. In the CPU exhaustion server show low computation capability and server do not respond to legitimate users and in memory exhaustion also servers do not respond to the legitimate users.

1.3 Organization of Thesis

In the chapter 1, we introduce the Introduction of the thesis.

In the Chapter 2, we will discuss the theoretical baselines which includes the basics of the cloud computing, comparisons of the cluster computing, grid computing, cloud computing and deployment model of the cloud, characteristic of cloud computing, security issues of the cloud computing.

In the Chapter 3, we are discussing the literature review. In that section we discuss about denial-of-service (DOS) attack and distributed-denial-of-service (DDOS) attack and its related work that is prevention method of the DOS and DDOS attack is included.

In the Chapter 4, we explain the proposed work, and discussed about the count based filtering methods that are the solution of the denial-of-service (DOS) attack and distributed-denial-of-service (DDOS) attack.

In the Chapter 5, we are discussing the performance evaluation and comparing results in the form of the graphs. These results are comparison for the DOS and DDOS attack.

In the Chapter 6, includes the conclusion and future work. We explain the future research area of the cloud computing that prevent the denial-of-service (DOS) attack and distributed-denial-of-service (DDOS) attack.

CHAPTER 2

THEORETICAL BASELINES

The aim of this part is to introduce the theoretical framework of my research area.

2.1 Evolution of Cloud Computing

In this part, we introduce how cloud computing is introduces and it become important part of the research recently.

- **Motivation**

Throughout the 60 years of history, computer system is evolved in the integration and distribution way. The 1970's introduces the centralized system, mainframe system and shared system. In the 1990's, decentralized system and private PC's came. In 2010's, shared machines invisible to users called cloud computing arise. The basic idea of cloud computing is to provide the computational resource of the system through internet.

The increasing availability of the internet and corporate IP connections enables the network-based services. While internet-based services manage the mail service that is stored for many years and offering of the storing data recently increases to combine network-based service and network-based computing. The new type of service introduces for organization and individual users are called "cloud computing" [1].

2.2 Definition of Cloud Computing

Cloud Computing is a recent technology which is widely used for storing data remotely. Cloud Computing provide services on-demand and user pay for that service [2]. Customer opening account in the cloud can use it only on internet. Customer willing to use application remotely can use services of cloud by paying for that according to its using services. Cloud computing is accounting model business. It

calculate to distribute the resources on the resource pool according to computer strength, storage space, and requirement of applications. This kind of resource pool is called “Cloud” [3]. The Cloud Computing is managing all computation resources. In the Cloud Computing virtualization introduces some techniques such as live-migration and pause-resume [2]. Live-migration service is provided when customer moves one place to other place geographically. In Pause-resume services customer can pause the application and after some time resume that application. Cloud service provider (CSP) satisfy cloud user requirement without affecting own utilization. Cloud is largely scalable so that provide resources dynamically or when user uses service. Cloud computing provides resources to the user through internet. Cloud computing charge for service and firstly infrastructure service released by Amazon and paved for many cloud applications, services, solutions etc [4]. Cloud computing use both the fields educationally and organizationally. In the educationally, cloud is the most research field area and number of university teaches cloud as a subject. In the organizationally, some companies has own cloud and works on it. Cloud has some types are follows as: public cloud, private cloud, community cloud, and hybrid cloud. Each type of cloud has its own uses.

Cloud computing are some key features: Access anywhere, On-demand, Elasticity [4].

(1) Cloud accesses anywhere and any location. Only you need internet connection and browser then you access cloud.

(2) Cloud computing provide services on-demand. Assigning and re-assigning the resources to the customer.

(3) The ability to scale up and available resources to assign the customer. Only customer pay for that how much service uses.

In the figure 2.1 [5] has six phase diagram. In the phase 1, has one or more user shared powerful mainframe using dummy terminal. In the phase 2, single PCs become powerful PCs to meet the users. In the phase 3, PCs, server, and users connect to the network and all are share resources and improve performance of the network.

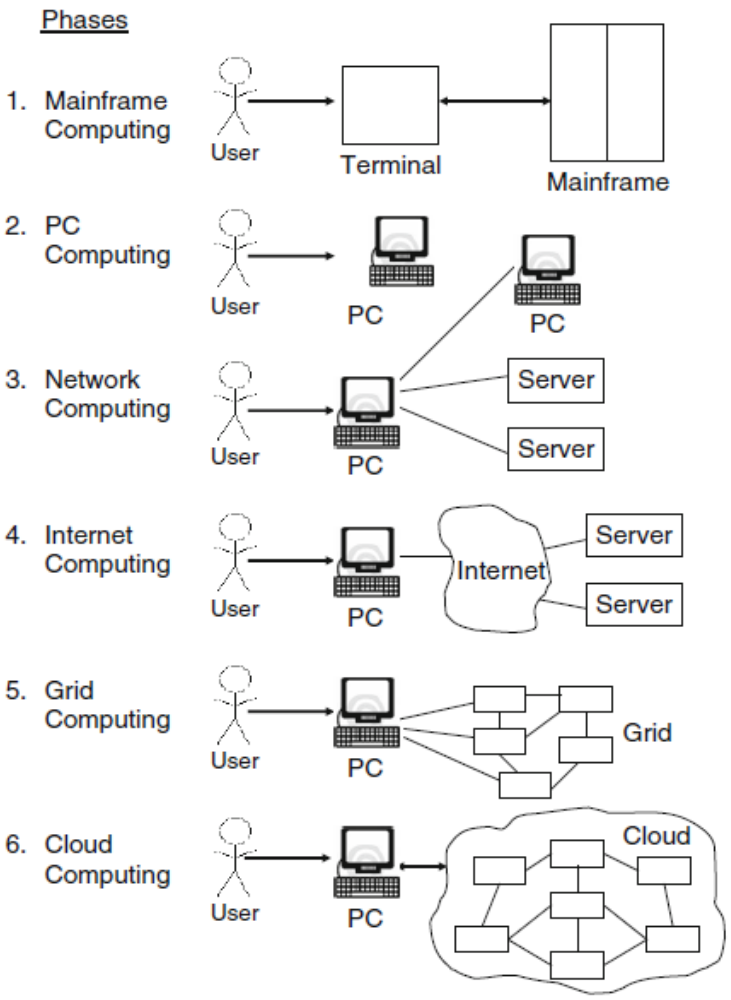

Figure 2.1 Six computing paradigms – from mainframe computing to Internet computing, to grid computing, and cloud computing [5]

In the phase 4, local network connect to the other local network and form the global network known as internet. Local network share the resource and use internet to utilize remote applications. In the phase 5, Grid computing provides shared computing power. Computation will be fast and storage space has available. Grid computing is distributed. In the phase 6, Cloud computing provide resources remotely and use resources are sharable and storage space will be available remotely. Cloud provides services and applications remotely.

2.3 Cluster Computing

Cluster computing uses commodity-off-the-shelf (COTS) hardware and commonly used [6]. Cluster computing has emerged of the several trends, high performance microprocessors and high speed network so that develop tool that enables the high performance distributed computing and high speed computation power for the applications. Cluster computing provides computing resources for the educational institutions. Colleges and institutions need not to invest much more money for the understanding of the concept of parallelism. They are using software that provides in the market or simulation tool. Cluster computing is using parallel computing that has more than one task done at a time. For example openMP, MPI and HADOOP are the tool for the parallel programming. Hadoop is provides programming environment in the cluster and understanding the how we develop and work on the single node.

A cluster is a type of parallel or distributed processing system that contains interconnected network, stand alone computers working as a single computer, and integrated computing resources [7].

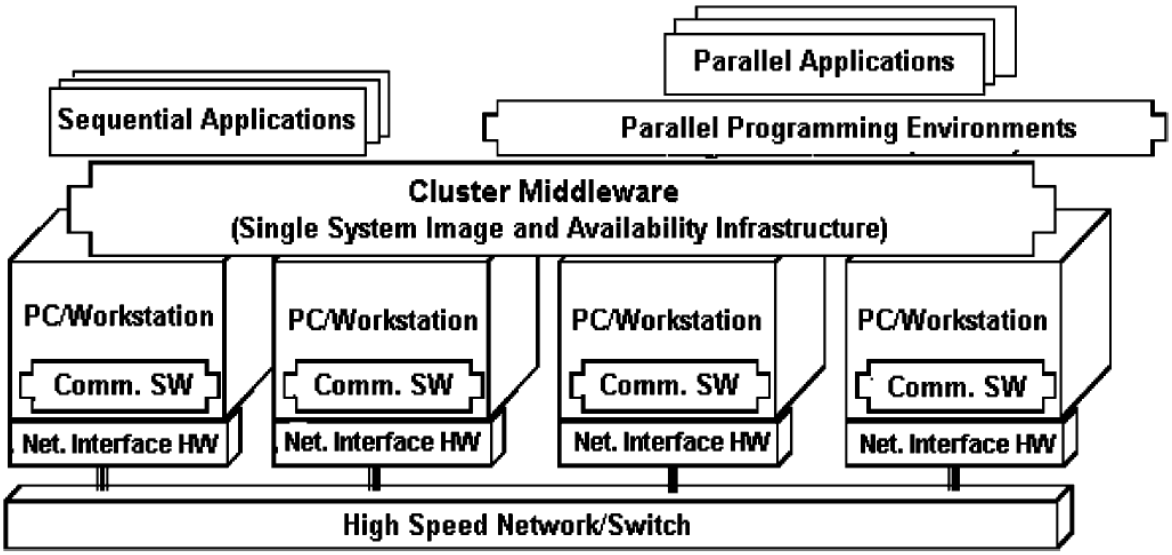

Figure2.2 Typical cluster architecture [7]

In the fig 2.2 any node has its own memory, operating system, input/output devices and each node can be single processor or multi-processor. In the cluster computing a system has a single node. A node has a connected via local area network (LAN). Also include for the communication between two nodes in the network are use transmission control protocol/Internet protocol (TCP/IP). Cluster middleware has responsible for offering the unified system image from the collection of independent and interconnected nodes [7]. Cluster middleware offers for the single system image (SSI) and high availability infrastructures for the processes, memory, networking, input/output and storage. The SSI can be implemented using software and hardware infrastructure. The modular architecture of the SSI allows use the services provided by the lower level layers for the used and implementation of the higher level layers.

Clusters are of many types [8]:

- Asymmetric cluster

In asymmetric cluster, a server exist which faces failure of some reasons. The node has some different capabilities. In this type of schema if one node has failure of some reasons then other node takes responsibility and available in the place of failure node. The standby server does some useful function and have same capacity as compare to primary server.

- Disadvantage
  - The standby nodes have the idle most of the times and replace only if server has failed.
  - It should be capable of replacing any node in asymmetric cluster and needs to qualify potential of various types of nodes in cluster.
- Symmetric cluster
  - In the symmetric cluster, each server works as a primary server for the particular set of applications. If one server failed then other server works remains working if server failed during process. In this schema server takes and does all the event of the failed server.
- Disadvantage

- Each node checks if it's corresponding right node working or not and this process is done for detecting failed node. If two successive nodes have failed then right node will be undetected.
- It used sender based message logging. If sender node is fail then logs will be lost and resulting in the system to rollback and affecting the efficiency of the system.

A cluster java virtual machine (JVM) views the cluster as a single computer, a single JVM that has implicitly spreads java threads to the cluster nodes [9]. Cluster nodes implement java memory model for support implicit data communication. The message passing interface (MPI) to java running own separate JVM, also we use Hyperion cluster JVM.

- Hyperion is a high performance implementation of the cluster JVM [9]. It provides the programmer that illusion of the single JVM but runs in the background multithreaded. Internally, it distributes application threads to the cluster machines to available the parallelism. Java threads can share objects, and developers. Hyperion implements a distributed shared memory (DSM) across the cluster's node.
- Cluster computing commonly uses MPI and well defined communication interface specification available in many platforms. mpijava is an object oriented java interface to MPI [9]. The reason for MPI's popularity is its performance, because MPI model application provides developers to manage all the data communication. It provides high performance parallelism application to understand the characteristic of the nodes, network and algorithm.

2.4 Grid Computing

There are three areas of major scientific research [10]:

1. Laboratory field, field experiment, and scientific experimentation.
2. Analytical and theoretical research.
3. The use of the computation to understand and solve complex problems.

Grid computing is very important area of the educational and scientific research point of view. Grid computing is the technology that allows user to access remotely supercomputer and does some computation perform to the supercomputer that is large scale data process. It is the combination of hardware and software resources that are access everywhere for the fastest computation of the world [10, 11].

In the grid computing security is effective problem. Grid computing provides security using secret key encryption of Kerberos, brokered authentication and authorization.

The growth of the internet is along with high availability of powerful computers, high speed networks, and low cost commodity component. These technologies are available of geographically distributed resources, storage, data resources, supercomputer, computation devices, special devices and services. Grid computing have challenges at all levels, conceptual and implementation models, application formulation and development, resource management, programming systems, infrastructures and services, application formulation and development, conceptual and implementation models, networking, security, and development of the global research community [11]. Parallel computers are the variety of architectures become commercially available and networking software and hardware are becoming widely deployed.

To program the parallel programming a long list of parallel programming languages and tool are develop and evaluate. This list includes Linda, BSP, Concurrent Prolog, Occam, Fortran-D, Programming Composition Notion, Compositional C++, pC++, Nexus, Mentat, lightweight threads, and Parallel Virtual Machine [11]. Developers use these new tools, it have obvious that computer networks allows machines to used one parallel code. Networks of workstations were use for parallel computation. Homogenous sets of machines, it is possible to use heterogeneous sets of machine. Network has rise the distributed computing. Systems using Distributed Computing Environment (DCE) are building the group of machines, well defined and closed configurations.

In the maintenance phase of the grid computing software maintenance is identified single greatest cost of computer software and systems. “The modification of a software product after delivery to correct faults, to improve performance or other

attributes or to adopt the product to a modified environment" [12]. The changing of grid flexibility, business needs, and user needs by continuous service, resource allocation, reconfiguration, and provisioning by executing organizing operation. Maintaining grid resources and services are important issues because [12]:

1. It can't be assumed that grid service always match all requirement of the grid users. For example limitation of deployment service, version issue and so on.
2. Software requirement and resource availability evolve rapidly and some mechanism needs to install and deploy these services for the applications of the large scale grid computing.
3. Grid is a dynamically changing capability and subject to many changes, some types of mechanism are needed to monitor and manage these changes.
4. Maintenance and performance of grid computing will be enhanced by maximizing the availability of the systems.

The daily maintenance of the grid systems includes [12]:

- **Service Provisioning:** It includes operations to deploy the new service components, upgrade the new or existing service components, and uninstall the deployed service components from the computational system dynamically upon quality of service (QoS) requirements.
- **Configuration Management:** This operation changes the capability, availability, and scale of the required resources for some applications in the grid computing.
- **Runtime Migration:** It includes the operation to clone a proper instance of some service components to a new computational node so that new computational node will do all the functionality of the previous node. This clone functionality is requiring for the fault tolerance and highly reliable systems.

For the maintenance problem the grid system is difficult while attempting to maintain grid computing systems, administrators want to [12]:

1. Schedule the maintenance time in advance
2. Find out the maintenance related dependency to improve the solution will be optimized
3. Includes the quality of both software and hardware resources of the grid system

4. Highly successful in propagating the maintenance instructions to all the nodes for the decrease any delay in the maintenance time.

From experience with ChinaGrid [12] and from logical deduction two main issues for the building maintenance of the grid computing are: scalability and complexity.

- **Scalability:** Grid computing is the large scale systems because it composes thousands even millions of networked nodes, so that maintaining is important and it has correctly propagated to all the nodes in the grid systems namely called scalability challenge.
- **Complexity:** Computing environment complexity could be characterized by the number and variety of components and their interactions, computation done with in time, computation capabilities, and rate of changes. Mapping this characteristics and grid characteristics we can identify the complexity as inconsistency challenge, dependency challenge, and dynamicity challenge.
    - **Inconsistency challenge:** Grid computing provides very diverse resource types including software and hardware (e.g., CPU power, radio telescopes, storage devices, files, sensors, computation resources, etc.) namely resource heterogeneity [12]. In the grid computing, resources belong to many different-2 organizations that allows other organizations to access them. Each and every organizations may establish different security and administrative policies and their own resources can be accessed namely policies heterogeneity. So that maintenance mechanism should be working with different-2 types of resources and different organization policies.
    - **Dependency challenge:** The need of dependable service is the fundamental service in the grid. Resources in the grid computing must be coordinated. Thus dependency recognition of both service and resources should be executing and maintaining task correctly.
    - **Dynamically challenge:** Grid computing environment is dynamic. In which any resources failure then grid node is migrating dynamically means leave failure resources and join some other resources. So that maintenance mechanism should deal with dynamicity [12].

The issue to be considered in the development of the grid architecture for the high performance computing may be summarized as follows [13]:

1. The decomposition of the computational problem to a set of atomic tasks and execution of the tasks to be distributed in a collection of the grid nodes.
2. The data flow density among processing latency, communication tasks, the network bandwidth, and communication.
3. The workflow management including the synchronization among running tasks.

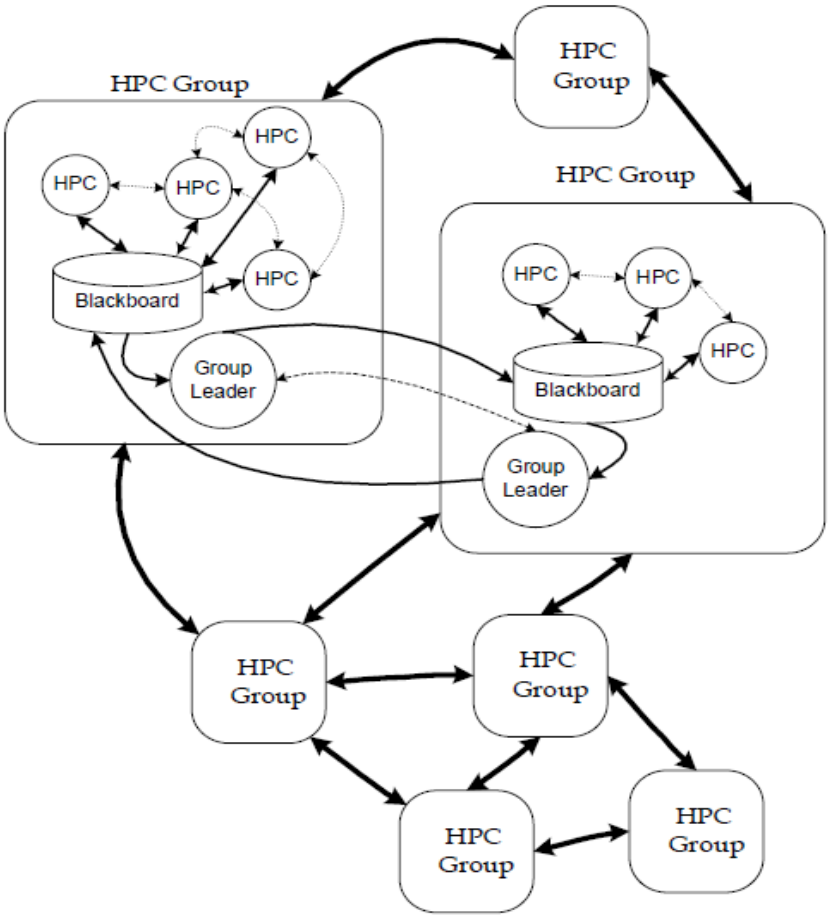

Figure 2.3 The Grid as a Federation of HPC Clusters [13]

In the above issues and the consideration grid environment conditions, the proposed architecture describes the following objectives:

1. Support a high level programming based on a web services so that grid applications should be run in a parallel environment.
2. Achieve scalability for the domain and multi-scale computational problem so that encountered in the CFD, material, and molecular simulations respectively.

In the fig 2.3 provides a high level grid topology for proposed architecture. The grid as a view of the network of high performance computing (HPC) groups, and each group is a collection of HPC clusters. In the grid computing interactions can be possible only neighboring groups, for which the communication path between neighboring HPC should be solid lines. Because of network latency, neighbors prefer geographically located closer network distance than non-neighbors. Each high performance computing (HPC) group has a blackboard; shared data stored use the members of the group to storing the output of their processing, and read the necessary input data that should be generated by other members of the group. It serves as the interface of the dataflow between the tasks executed by the group members. Reading and writing from the Blackboard is enabling by a web services. A light synchronization message is used by the HPC clusters to inform their neighbors within the same group exchange mechanism using TCP socket programming and data uploaded to the blackboard. In this topology node interacting only neighbors node and at the grid and group level motivated by the bandwidth consumption, message processing time, limit network congestion, which are all the contributors are the same latency.

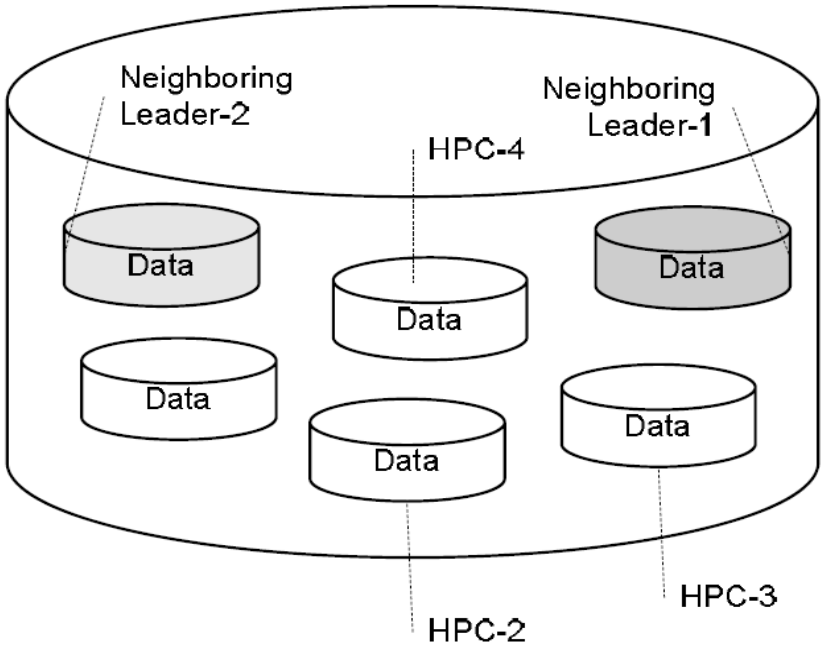

Figure 2.4 Blackboard partitions for an HPC group with two neighbors and four group members [13]

In the figure 2.4 shown that to avoid the bottleneck at the blackboard data stored and divided in the partitions and each partitions written by the single actor which could be group member or the leader of the neighboring group. All the group members and leader of the neighboring group may be read the partitions content [13]. High performance computing (HPC) group members, neighbor group leaders use light weight message for the TCP sockets share the group state and the state of the data uploaded neighboring blackboard. The neighborhood topology proposed architecture may provide scalability over centralized architectures.

2.5 Comparison b/w Cluster Computing, Grid Computing and Cloud Computing

In the cluster computing, all the nodes or system working with one standard operating system and cloud computing all the nodes working with hypervisor or virtual machine (VM). Virtual machine can be working with different-2 OS or environment.

2.5.1 Challenges

2.5.1.1 Cluster Computing

The research challenges of cluster computing are as follows [14]:

1. **Middleware:** To produce software environment for the illusion of the single system other than collection of computers.
2. **Program:** Applications run on the cluster which is explicitly written which division of tasks between nodes.
3. **Elasticity:** Elasticity is a real time response time when number of services changes dynamically.
4. **Scalability:** System should be scalable means dynamically increases the number of systems or resources and performance of the cluster should not decrease.

2.5.1.2 Grid computing

The research challenges faces in grids include [14]:

1. **Dynamically:** In the grid computing resources are own or manage by more than one organizations and any organizations may enter and leave the grid at any time. It is depends burden on grid [14].
2. **Administration:** A unified resource pool, heavy administration load is increase with other maintenance work and coordinate local administration policies with global ones.
3. **Development:** Problems are creates with writing software for the run on grid platforms, which includes distribute the processing elements or CPUs than assembling solutions.
4. **Accounting:** This challenge find the ways for the different accounting infrastructure, economic model and applications models that can cope with tasks that communicate frequently.
5. **Heterogeneity:** Finding ways, how to create a wide area data programming and scheduling framework in set of heterogeneous set of resources.
6. **Programming:** The grid computing is the distributed, so that it has the more complex for the programming in applications in grid manner.

2.5.1.3 Cloud Computing

The challenges of cloud computing includes the following [14]:

1. **Dynamic scalability:** The nodes are scaling up and down in the dynamic manner by the application according to the response time. The scheduling delays involved are concern which leads need of effective and dynamic load management system.
2. **Multi-tenancy:** when more than one application run and compute on the same node and increases application on the node then bandwidth of each application will be reduce according to application and performance will be degrade.
3. **Querying and access:** Secure access of information and scalable querying are open problems for both cloud computing and grid computing.

4. **Standardization:** Every organization has its own protocol and APIs. User makes the users data so that interoperability and integration of all the services and application is a challenge.
5. **Reliability and fault-tolerance:** The organizations develop tools for calculating the occurrences of the fault, so that develop the reliable system which is less failure and fault tolerable.
6. **Debugging and profiling:** The remote user debugging and parallel debugging has the problem in the cloud computing because it creates the problem for developing high performance computing (HPC) programs and its issue occurs in the cloud computing also.
7. **Security and privacy:** In the cloud computing user has no idea about data. Where data stored and which geographically located and user has no idea about data, how many cloud users use the same data or data will be public or private.
8. **Power:** Cloud computing automatic energy aware resource management. It is highly useful and required. Because of cloud computing provides many types of services.

Table 2.1 comparison of Cluster Computing, Grid Computing and Cloud Computing [14]

	Clusters	Grids	Clouds
Service level agreement (SLA)	Limited	Yes	Yes
Allocation	Centralized	Decentralized	Both
Resource Handling	Centralized	Decentralized	Both
Loose coupling	No	Both	Yes
Protocols/API	MPI, Parallel Virtual	MPI, MPICH-G, GIS, GRAM	TCP/IP, SOAP, REST, AJAX
Reliability	No	Half	Full
Security	Yes	Half	No
User friendliness	No	Half	Yes
Virtualization	Half	Half	Yes

Interoperability	Yes	Yes	Half
Standardized	Yes	Yes	No
Business Model	No	No	Yes
Task Size	Single large	Single large	Small & medium
SOA	No	Yes	Yes
Multi-tenancy	No	Yes	Yes
System performance	Improves	Improves	Improves
Self service	No	Yes	Yes
Computation service	Computing	Max. Computing	On demand
Heterogeneity	No	Yes	Yes
Scalable	No	Half	Yes
Inexpensive	No	No	Yes
Data Locality exploited	No	No	Yes
Application	HPC,HTC	HPC,HTC, Batch	SME interactive apps.
Switching cost	Low	Low	High
Value Added Service	No	Half	Yes

2.5.2 Projects and Applications

2.5.2.1 Cluster Computing

There are some projects on the field of cluster computing.

1. **Condor** [15]: Our goal is to develop, implement, and evaluate mechanisms that support high throughput computing (HTC) and all the work done on large resources that collected on distributed manner. The challenges by both the technological and sociological of such computing environment, the high throughput computing (HTC) at UV-Madison has building the open source HTCondor distributed computing software and enable engineers and scientists to improve the throughput.

2. **ShaRCS (Shared Research Computing Services)** [16]: This project name is ShaRCS and it is pronounced as sharks and its acronym for Shared Research Computing Services pilot. The ShaRCs project designed to define and demonstrate how memory and resources shared and cluster residing in the regional data centers. The sharks want to show that research computing services for better capability can be developed, overall cost will be reduced to UC and retain low barrier to access service.

The applications of the cluster computing are as follows:

1. The Weather Research and Forest (WRF) [14] is project that is the collaboration of some institutions and organizations to develop next regional forecast model and previous data for operational numerical weather prediction and atmospheric research.
2. Hadoop [17] is an open source framework for the running data applications. The Apache Hadoop project develops open source software for scalable, reliable and distributed computing. The apache hadoop software library is the framework for the distributed computing for the large amount of data across clusters using simple programming model. It is work for the scale up of the single machine to thousands of machines and each work as locally store the data and computation done.
3. Cluster computing is always used for solving challenging applications which is automobile crash simulations, weather modeling, computational fluid dynamics, life sciences, electromagnetic, image processing, nuclear simulations, aerodynamics, astrophysics and data mining [14]. Cluster computing uses both data mining applications which involve both data intensive operations and compute. They are also use for commercial applications such as banking which required to high availability and backup.

2.5.2.2 Grid Computing

Some of the projects of the grid computing are as follows:

1. **Globus:** It is an open source software and it has most challenging problem in the distributed resource sharing. The open grid service architecture (OGSA) represents a grid system architecture based on web services technologies and

concepts [18]. The Globus toolkit is use for the development of environment for producing grid services that follow open grid services architecture principles. Member of the OGSA alliance have some contribution to the development of OGSA. Globus provision is a tool for the deploying fully configured globus systems on the Amazon EC2 [19]. The globus provision will take care of creating certificates, user accounts, setting up NFS/NIS, etc. Globus provision currently support the following software is GridFTP 5.1, MyProxy 5.2 and GRAM (coming soon).

2. **EGI-InSPIRE** project (Integrated Sustainable Pan-European Infrastructure for research in Europe) [20]: Started on 1st May 2010 and is co-funded by the European Commission for four years. EGI-InSPIRE project is divided into seven work packages (WP) are as follows:
    - WP1: management
    - WP2: Community Engagement
    - WP 4: Operations
    - WP 5: Provisioning the Software Infrastructure
    - WP 6: Services for the Heavy User Community
    - WP 7: Operational Tools
3. Some other grid projects are NASAs Information Power Grid, GriPhyN, Particle Physics Data Grid, NSFs National Technology Grid, NEESgrid and the European Data Grid [14].

The grid applications are for oil reservoir simulation, High Energy Nuclear Physics (HENP), advanced manufacturing, weather modeling, numerical wind tunnel, particle physics research, popular science web services, and terrain analysis of nature observation, bio-informatics and scientific database [14].

1. **MammoGrid** [21]: MammoGrid is service oriented architecture and it is based on a medical grid application. The aim of the MamoGrid is to deliver the set of evolutionary prototypes to demonstrate the specialist radiologists working in breast cancer screening, mammogram analyst can use grid information infrastructure to solve the common image analysis problems.
2. **DDGrid** (Drug Discovery Grid) [22]: The aim of the project is to build a collection of platforms that are collaborating for the drug discovery using the peer

to peer and grid computing technology. DDGrid project solves large scale computation and the data intensive scientific applications in the field of molecular biology and medicine chemistry with the help of grid middleware.

2.5.2.3 Cloud Computing

Some of the cloud projects that are the initiatives:

1. **CERN:** It is a Nuclear Research organization located in Europe [14]. This organization is developing for the mega cloud computing to distribute data to scientist around the world as part of the LHC project.
2. **Unified Cloud Interface (UCI):** This project is proposed by Cloud Computing Interoperability Forum (CCIF) and working on the development of the standard APIs. This APIs is used by all the cloud service providers to reduce the interoperability issues.
3. **Cloud-Enabled Space Weather Platform (CESWP):** The purpose of this platform is to provide power and flexibility to cloud computing for the space weather physicist. The goal is to lower the barrier for physicists to conduct their science i.e. develop space weather models, produce visualizations, collaborate with other scientists, enable provenance and run simulations.
4. **OpenNebula:** This project is most advanced and its aim is to develop highly scalable, adaptable software toolkit for cloud computing management. It is the plan for the operation network to simplify the management of OpenNebula cloud instance, enhanced management of images and templates, fault tolerance functionality to maximize uptime in the cloud, enhanced support for the federation of data centers, new security functionality and support for multi-tier architectures [14].
5. **TClouds** [23]: The TClouds is an advanced cloud infrastructure that can be used privately, for storage and resilient computing. It is scalable, simple, cost efficient and deliver a new level of security and for this we need outsourcing critical system for the clouds, scientists will be building an advanced cloud of clouds framework for the project.

Some of the applications of the cloud computing are multi-tenant service, web hosting, HPC, media hosting, multi-enterprise integration, distributed storage, etc.

1. The public cloud established in many university libraries, it may conserve library resources and improve its user satisfaction. There are Inter-library loan (ILL) and Online Public Access Catalog (OPAC) services; access to the shared resources by a uniform access platform is difficult.
2. RoboEarth is a European project which is led by the Eindhoven University of Technology, a giant database where robots can share information about objects, Netherland, environments, to develop a WWW for robots and tasks. A-Star Robotics Laboratory, Singapore (ASORO) researchers have built a cloud computing infrastructures that are allows robots to create 3-D maps of their environment much faster than their onboard computers [14].
3. Cloudo is free computer that are lives in the internet and working on the web browser. It allows to access photos, documents, movies, music and all other files any computer and any mobile phones or devices.
4. Panda Cloud antivirus [24], it is the first free antivirus from the cloud. In the panda antivirus are regular updates are not a problem and uses very little system resources, simple interface, uses collective intelligence servers for fast detection and protests PC offline.

2.5.3 Tools and Simulation Environment

2.5.3.1 Cluster Computing

1. Nimrod is a tool for the cluster computing. By using nimrod it is easy to create plan for parametric computing and runtime system submit, compile, run and collect the results from multiple cluster nodes [14].
2. PARMON [25] is a tool for cluster computing and that allows monitoring the system resources and its activities at the each level: component, node and system. PARMON is used to control C-DAC PARAM 10000 supercomputer and the cluster of the 48 Ultra-4 workstations is powered by sun Solaris [14].
3. MPI and OpenMP: Message passing libraries provides high level interface and data between process executions on distributed memory systems. It wants to achieve high performance computing of the each and every cluster nodes.

2.5.3.2 Grid Computing

1. Paradyn is a performance analysis tool for grid computing and it supports performance experiment management through techniques for numerically comparing more than one experiment and performance evaluation based on the dynamic instrument. This tool experiment is done manually, whereas performance evaluation is done automatically.
2. Nimrod-G is a grid computing tool and it uses Globus middleware services for the dynamic resource recovery and sends jobs for the computational grids. It allows engineers and scientists to model transparently and experiments stage, program and data at the remote sites and execute the program on each element of the different machine and finally collect the results on the remote sites and on the user machine.
3. Condor-G [15] toolkit represents the work of Globus and Condor project. In this project G stands for the Globus. This project enables the utilization of the large number of resources that are located on multiple geographical areas. The Globus uses protocols for the secure inter domain communications and standardized access to remote batch systems. Condor consists some functions are as follows: job submission, job allocation, creation of friendly execution environment and error recovery.
4. GRID computing and BUSiness (Gridbus) [26] toolkit project is the design for the development of the cluster and grid technologies for the service oriented computing. Gridbus provides end to end services to aggregate and lease services of the distributed resources depending on the capability, cost, availability, performance and quality of service requirements.
5. Legion [27] is an object oriented based meta-system. It supports transparent data management, site autonomy, core scheduling, fault tolerance and a middleware with a wide range of security options.

2.5.3.3 Cloud Computing

There are various tools and products for the development of the cloud computing:

1. Zenoss [14] is the cloud product which is integrated that manage the entire IT infrastructure, which is deployed (virtual, physical, or in cloud). It manages the servers, networks, storages, virtual devices, and cloud deployments.
2. Spring Roo is an advanced generation application development tool for the cloud computing. It is combines with the power of Google web toolkit (GWT) that enables developer to build the rich browser applications in the enterprise production environments.
3. CloudSim and CloudAnalyst [14] is the most important tool for the developer in cloud computing to evaluate the requirements of large scale cloud applications in the sense of geographically distributed of both the computing servers and user. The former was developed for the purpose of studying the behavior of the applications under various development configurations.
4. Cloudera [28] is an open source Hadoop software framework. It is using in the cloud computing deployments due to its data intensive queries, flexibility with cluster based and other tasks. Cloudera will allow exploring complex, non-relational data in the native form.

2.6 Layered Architecture

Cloud computing is a collection of the services and it is presented on the layered architecture. These services are offered by the IT service through cloud remotely. It offers the services of the three major types: software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). They create hierarchy of the each layer. We will be discussing each layer architecture of the cloud computing. Each layer has its own work and functionality.

Software as a Service (SaaS) layer provides resources and applications remotely and works on applications. Platform as a service (PaaS) layer has provided platform for organizations and institutions. This platform provides applications remotely for the institutions and organizations that pay for the service.

Infrastructure as a service (IaaS) layer works on the infrastructure of the network and organizations.

2.6.1 Software-as-a-Service (SaaS)

Software-as-a-service (SAAS) provides applications. Its applications implement various functions and offers users as a service. They are using cloud infrastructures or platforms services for the cloud based services [4].

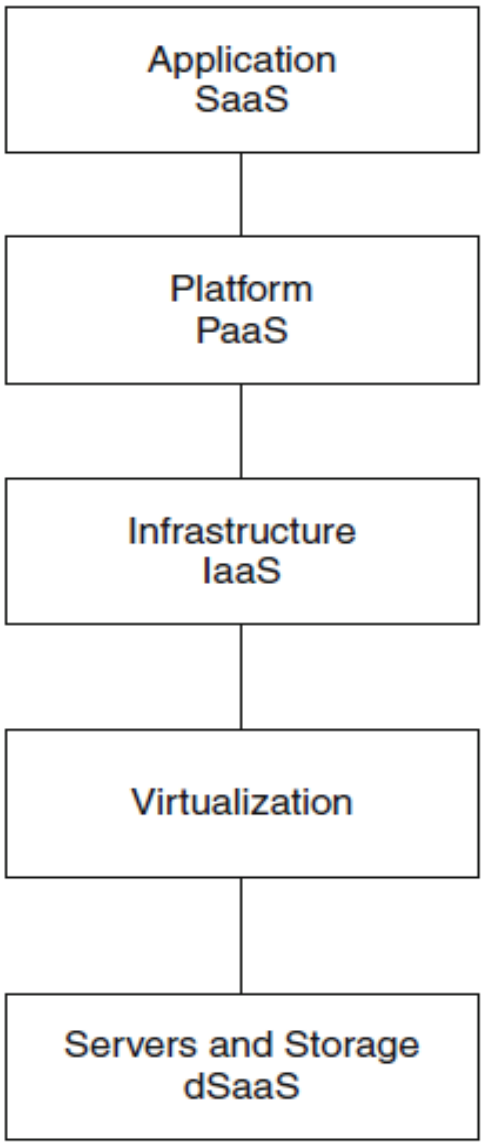

Figure 2.5 Layered architecture of Cloud Computing [5]

This service refers to the prebuilt and vertically integrated applications (e.g. a human resource management, database processing, payroll processing, email system and other application processes) and these are delivered to customers as a service [29].

This layer user are purchasing for the functionality. Applications are designed for the ease to use on the based of proven business model. This is a user level layer and it is classified into two separate layers:

- Services (which is stand alone application e.g. billing service)
- Applications (which is a combinations of functionalities)

SaaS is providing broad area services, it can be anything from web based email to inventory control, database processing and even in some cases online banking services. Hotmail, IBM WebSphere, Gmail, Boomi, Quicken Online and SalesForce are some of the well known SaaS providers and products [29].

2.6.2 Platform-as-a-Service (PaaS)

Platform-as-a-Service (PAAS) is providing computational resources that provide API to access its direct execution of the tasks and capabilities [4]. The solution is implemented by the platform and it can be specific for any type of application. It is used for Google App Engine for designing the web applications. This layer works on the software and product development tools (e.g. database servers, middleware, application servers, portal servers, etc) which users purchase the facility so that they can deploy and build their own applications. It can increase as much as flexibility and control of the consumer [29]. There are offers which includes runtime environment for the cloud services, application code, storage, compute power and networking infrastructure. In this layer of services may be regard developer level layer.

2.6.3 Infrastructure-as-a-Service (IaaS)

Infrastructure-as-a-Service (IaaS) layer provide virtualized resource to the user as service. The cloud storage provides data that are remotely stored in the data centers. There are different flavors of IaaS [4]:

- Non-relational data is stored in the optimized manner, e.g. Amazon SimpleDB
- Basic unstructured data is remotely stored, e.g. Amazon S3
- There are cloud based relational databases, e.g. Microsoft SQL Azure

There are integration services that arise from need of the cloud computing applications for interactions to others [4].

- Publisher subscribe message delivery, e.g. Amazon Simple Notification Service (SNS)
- Message exchange via queues, e.g. Amazon Simple Queue Service (SQS)

Cloud computing is providing access to computational resources. There are computational resources providing virtual machines multiplexed by the hypervisors on the power hardware [4].

The following table shows different types of cloud technologies used in different types of services like:

Table 2.2 IaaS, PaaS and SaaS [5]

Service type	IaaS	PaaS	SaaS
Service category	VM Rental, Online Storage	Online Operating Environment, Online Database, Online Message Queue	Application and Software Rental
Service Customization	Server Template	Logic Resource Template	Application Template
Service Provisioning	Automation	Automation	Automation
Service accessing and Using	Remote Console, Web 2.0	Online Development and Debugging,	Web 2.0

		Integration of Offline Development Tools and Cloud	
Service monitoring	Physical Resource Monitoring	Logic Resource Monitoring	Application Monitoring
Service level management	Dynamic Orchestration of Physical Resources	Dynamic Orchestration of Logic Resources	Dynamic Orchestration of Application
Service resource optimization	Network Virtualization, Server Virtualization, Storage Virtualization	Large-scale Distributed File System, Database, Middleware etc	Multi-tenancy
Service measurement	Physical Resource Metering	Logic Resource Usage Metering	Business Resource Usage Metering
Service integration and combination	Load Balance	SOA	SOA, Mashup
Service security	Storage Encryption and Isolation, VM Isolation, VLAN, SSL/SSH	Data Isolation, Operating Environment Isolation, SSL	Data Isolation, Operating Environment Isolation, SSL, Web Authentication and Authorization

Table 2.2 Explain the comparison of the Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) on the different-2 services.

2.6.4 Virtualization

The advantage of cloud computing is the ability that virtualizes and share resources on the different-different applications for the better utilization of the server. In the figure 2.6 non-cloud computing has three independent platforms that exists for the three different applications running on the own server. In the cloud, server may be virtualized or shared for operating systems and applications. The virtualization

technology considers virtual machine technologies like a VMware and Xen, VPN and virtual networks [5].

In the cloud computing, there are different types of network resources that virtualized-as-a-services which combine with the cloud service provided to the end users [30]. Network virtualization techniques offer simple ways for users. Network virtualization technique can enhance the diversity, manageability and flexibility of the network services so that it improves the performance of the network requirements in the cloud computing. Virtualization has the main challenge in the cloud computing for modeling and optimization of cloud service provisioning based on network virtualization. This virtualization technique is used because more than one cloud services are used by single user [30]. Network performance has the significant impact on the cloud computing service provisioning in more than one network cases become the bottleneck that limits cloud computing supported by the high performance computing applications.

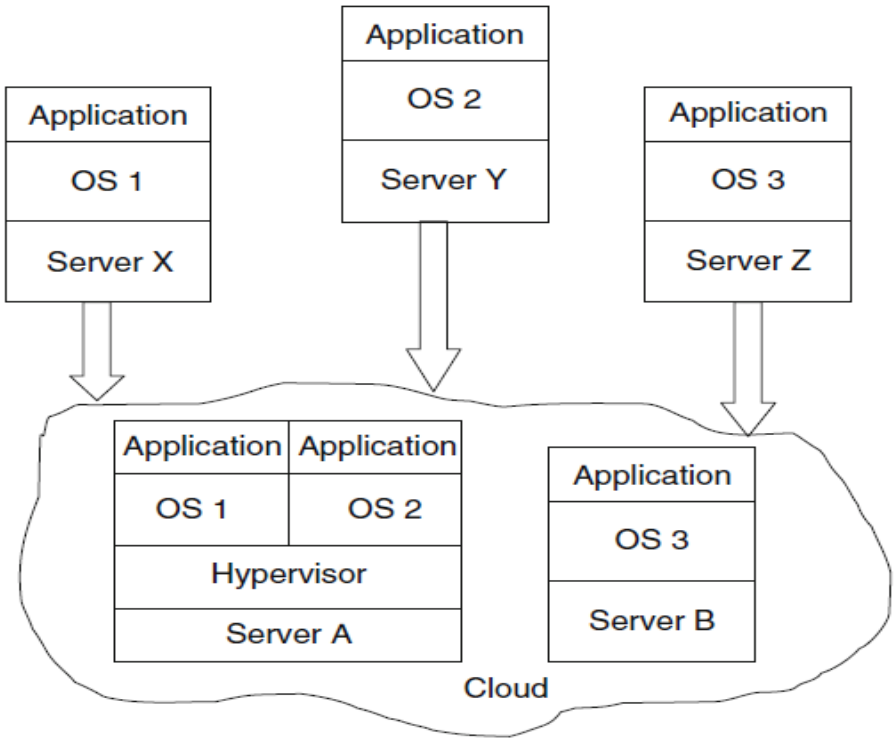

Figure 2.6 An example of virtualization

We cannot apply directly for the virtualization based cloud computing for Quality of Service (QoS) that provides guaranteed service composition. Network virtualization places new challenges may be actually large scale of heterogeneous networks with divert network resources in different domain. Network virtualization in cloud computing has participated in the three major contributions:

- We are presenting a model of the system with cloud services including the network virtualization technology. Using this model, we become aware of problem of the Quality of Service (QoS) service composition in the cloud computing in the different form of the MCOP [30]. The first model of QoS aware its service composition on the basis of virtualization in the network in cloud computing.
- We are proposing the algorithm to solve the QoS to aware service composition problem on the basis of our presenting model. We study the theoretical properties of the proposed algorithm [30].
- We are comparing the proposed algorithm and the best modified algorithm of the QoS algorithm through numerical experiments.

2.6.5 Other Services

The software-as-a-service (SaaS), platform-as-a-service (PaaS) and infrastructure-as-a-service (IaaS) are the basic three services in the cloud computing. But it not clear and there is considerable amount of overlap layers. Some researchers presented some combined services model as follows: SaaS & IaaS, SaaS & PaaS, PaaS & IaaS, and SaaS & PaaS & IaaS. Other services are as follows:

- Storage-as-a-Service: This service provides storage of the data remotely.
- Database-as-a-Service: This service manages different type of database structure.
- Security-as-a-Service: In this service cloud computing provides security protocols and security software remotely so that data, connection, communication, user information will be secured.
- Communication-as-a-Service
- Management/Governance-as-a-Service
- Integration-as-a-Service

- Testing-as-a-Service
- Business Process-as-a-Service

There are some architectures that provides more than one services to the single architectures in the cloud computing. In the Table 2.3 define these relations between architectures and cloud services:

Table 2.3 mapping of cloud provision to a generic EIA [29]

Architectures	Cloud computing Services
Business architecture	Process-as-a-Service, Management-as-a-Service, Governance- as-a-Service, …..
Application architecture	Software- as-a-Service, Integration- as-a-Service, Testing- as-a-Service, …..
Data architecture	Information- as-a-Service, Database- as-a-Service, Security- as-a-Service, …..
Technology architecture	Storage- as-a-Service, Infrastructure- as-a-Service, Platform- as-a-Service, …..

2.7 Cloud Service Applications

The cloud computing mainly divided applications into three basic layers that are as follows as: SaaS, PaaS and IaaS.

2.7.1 SaaS Application

The software-as-a-service layer is deciding that how types of applications provides this layer and that applications are as follows:

2.7.1.1 Desktop as a service

Desktop as a service is a special types of software-as-a-service. Desktop as a service provides virtualized desktop for the personal usage and sends its images to the user real desktop. The user can access own desktop on the cloud from the geographically different-different places and use benefits of the SaaS service at the same time. The Global Hosted Operating SysTem is a complete and free internet based virtual computer (VC) service includes files, applications and personal system. The Global Hosted operating SysTem applications are hosting by the Amazon Web Services (AWS) so that users can utilize the Elastic Cloud Computing (EC2) and Simple Storage Service (S3) resources through Global Hosted Operating SysTem desktop [5].

2.7.1.2 Google Apps

Google Apps is the generalized Software-as-a-Service implementation. Google Apps applications provides much more web applications with the same functionality and it also enables communicate too users, create and collaborate efficiently and easily. All the applications are available in the internet and it is accessed through web browser. Users may access all the applications using internet and it is not need to want anything extra locality [31].

Google Apps has the several components. The communication component contains Google Talk and Google mail which is allows for the communication through voice calls, instant messaging and email [31]. The office components are including spreadsheets and docs through users create document online and also facilitate collaboration and searching.

2.7.1.3 Salesforce

Salesforce is business software-as-a-service platform that are provides mostly Customer Relationship Management (CRM) services, customizable applications, to consumers. There are mainly two major products presenting by Salesforce are as follows [32]:

- Sales Cloud is a group of applications for improve the convenience and efficiency of the business activities.
- Service Cloud is providing to integrate social network applications like twitter and facebook, to constructing the customer service community.

The shared application model could interference between users; the Salesforce SaaS has these advantages [5]:

- The service provider can be develop only one version of the applications and do not need to worry about heterogeneous environments of the execution..
- Sharing of the operating system, physical computing resources and runtime environment reduces the cost of the applications.
- Users are free for choose the applications and version and customize it to fit their business.

2.7.2 PaaS Application

Platform-as-a-service (PaaS) is a cloud system provides software execution environment and that environment application can run. The environment is not a previously installed, it is installed or integrated with the programming language level platform that provides remotely using internet and this platform using for develop and build the applications.

2.7.2.1 Google App Engine

The main goal of the Google App Engine is to efficiently run users on the web application. The front-end of the Google App Engine is HTTP request with load balancing and routing strategies based on the contents [34]. The applications can

combine data services and other Google App Services, such as image storage, email and so on through API provides by the Google App Engine.

2.7.2.2 Microsoft Azure

Microsoft cloud technology creates cloud platform that are users may move their applications in the seamless way, to ensure its managed resources both are accessible cloud services and on-premises applications. Microsoft introduces Windows Azure Platform (WAP) which is combined of the cloud operating systems with Windows Azure. Microsoft Windows Azure is the main component of the WAP. Virtual machine is a runtime environment. Applications in the Microsoft cloud offering are divides into two types [34]:

- Web role instance: It is serve web request through the internet information service (IIS)
- Worker role instance: It is receives messages from other web role instances or on-premises applications.

2.7.2.3 Force.com

Force.com is the enterprise edition of the cloud computing platform offered by Salesforce. It involves vendor develop and deliver stable, scalable and secure applications. There are two key technologies of Force.com are multi-tenancy and metadata [35].

- The multi-tenancy approach is used for different users to share applications templates on the physical computing resource pool and the instance of the application is independent from each other.
- Metadata architecture that are generates application components according to own description has proposed.

2.7.3 IaaS Application

2.7.3.1 Amazon Elastic Compute Cloud (EC2)

Amazon has provides the universal solution of the cloud computing called as the Amazon Elastic Compute Cloud (EC2). The EC2 is providing the complete control over the customers computing resources so that their capacity may be scaled quickly through the simple web services. Amazon is the biggest online cloud service provider. Amazon calculates cloud platform for independent software development personal as well as the developer. Amazon calculates their cloud platform to be called as Elastic Compute Cloud (EC2). It is providing long distance cloud to calculating the platform service the company. Elastic Compute Cloud (EC2) comes by name Amazon Web Services exist platform development [3]. Amazon Web Services (AWS) is mainly composed by four core services:

- Simple Storage Services (S3)
- Elastic Compute Cloud (EC2)
- Simple Queuing Service
- SimpleDB

2.7.3.2 Amazon Simple Storage Service (S3)

Amazon Simple Storage Services (S3) provides data storage and the gain web service connection. The data is any type and any size of the data will be stored on the data centers and it access any where through internet. The object size is in the memory from 1 byte to 5 GB [3]. The Amazon Simple Storage Services (S3) is accessible to users through web services, by involving a SOAP interface, REST-style HTTP interfaces. Amazon S3 are storing data space into many more buckets and each bucket has its own unique name globally to help locate data addresses, identify users account for the payments and gathering information. Simple Storage Services (S3) deals with all types of data as objects and storing their metadata into bucket chose by the data owner.

2.8 Deployment Model

Cloud computing deployment model are mainly define in four types: public cloud, private cloud, hybrid csloud, and community cloud.

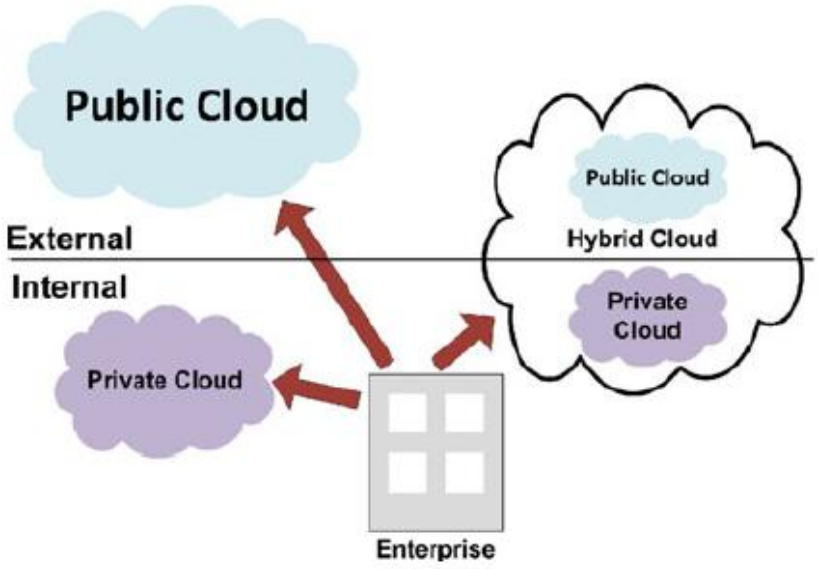

Figure 2.7 Cloud types: public, private and hybrid clouds [5]

2.8.1 Public Cloud

In the public clouds, any user or customers are shares the same resources that are provided by single cloud service provider. Customer are access the resources and pay for that resources. Public cloud are less secure than the other cloud models because it open and accessible to all the customers any-time and any-where. Advantages of public clouds are regulatory, compliances, exist the hidden danger of security and quality of service (QoS) [36].

2.8.2 Private Cloud

In the private cloud, the computing resources are used and controlled by the enterprises privately. Private clouds are managing by the internal cloud service provider and its user or customers are the internal person of the organization. Private

clouds are used by within the organizations. Advantages of the private clouds are security, QoS and compliance [36].

2.8.3 Hybrid Cloud

In the third type of cloud, hybrid cloud is the combination of the public cloud and private cloud. Hybrid cloud is the linked to one or more external services, provisioned as a single unit, centrally managed. Hybrid cloud is provide more secure control of the data applications and that are allowed to various parties to access information over the internet [5].

2.8.4 Community Cloud

In the community cloud, several organizations are jointly created and share the single cloud are called the community cloud. Community clouds are managed by one of the organizations that are the share the cloud. The community cloud creates the degree of economic, democratic equilibrium and scalability [36]. The cloud infrastructure can be hosted by the third party or within one of the organizations in the community.

2.9 Cloud Computing Features

In this section, we will define the features of cloud computing. Cloud computing has new features to other computing paradigms. There are described in this section.

- **On-Demand Services**

Cloud computing provides resources and service for the users that is on-demand. The resources are the scalable over the several data centers [5].

- **User-Centric Interface**

Cloud computing is location independent that means cloud can be access anywhere. It is accessible form anytime and anywhere through internet with web browsers [5].

- **Guaranteed Quality of Service (QOS)**

Cloud computing can be guarantee quality of service for the users in the terms of hardware or CPU performance, memory capacity and bandwidth.

- **Autonomous System**

In the cloud computing, autonomous systems are managing transparently to the users. So that software and data inside clouds can be automatically configured again (reconfigured) and consolidated to a simple platform depending on users need.

- **Pricing**

Cloud computing is not require up from investment. No hardware or resource required. Users pay for services and capacity as they need them or as they use them. Users do not require high speed networks or computation. Users have only available internet connections [5].

- **Elasticity**

In the elasticity, cloud computing has ability to scale up suddenly to cover the peak loads of your service requests [4].

- **Resource Pooling**

Resources pooling is the main feature of cloud computing. In the resource pooling allows combining computing resources (e.g. software, hardware, network bandwidth and processing) to present multiple customers or consumers and resources are assigned dynamically [29].

- **Scalability**

Scalability of cloud computing is large. Cloud of the Google has more than one million servers. Amazon, IBM, Microsoft and yahoo they have more than thousands server. Cloud computing is working with large computing power [37].

- **Virtualization**

In the virtualization, cloud computing make sure users get service anywhere through any kind of terminal. Resource is required come from instead of visible entity. You can complete all the work you want through internet service using the notebook, PC, mobile phones. Users can complete the task that can't complete in a single computer [37].

- **High Reliability**

In the reliability, cloud uses data multi transcript fault tolerance, computation node is isomorphism exchangeable and so on ensure the high reliability of the service. Using the cloud computing is more reliable than the local computer.

- **Versatility**

In this section, cloud computing does not aim at the certain special applications. It can produce various applications that supported by the cloud. One cloud can be support different-2 applications and running it at the same time [37].

- **High Extensibility**

The scale of cloud computing be extended dynamically so that meet the increasingly requirement.

- **Service Oriented**

The service oriented concept is the similar but more practical than the concept of service oriented architecture (SOA) in grid computing. Accessibility and abstraction are two keys to achieve the service oriented conception. Using virtualization and other technologies the architecture is abstracted without exposing much to user. Abstraction is reduces both the need for cloud user to learn the cloud details of the cloud architectures and threshold of the application development. Cloud users can consume all the capacity easily by exploring system parameters such as storage capacity and processing performance [38]. The type of providing capability, the services of cloud computing are mainly divided into three categories:

infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and Software-as-a-service (SaaS).

- **Loose Coupling**

In the loose coupling, it is technical fundament of cloud computing and goes beyond the loose coupling method of application interaction. Using virtualization or other technologies, the infrastructures are the separate in the physic or logic. The behavior of the one part hardly affects the other part. The platform is an abstract layer which is isolates different applications running on it [38].

- **Strong Fault Tolerant**

In this section, many fault tolerant method in parallel computing. At low level, there is always contains some fault correction mechanism with specific hardware. At middle level, checking point is the one of the most effective method. At high level, many specific applications are studying with methods aiming at algorithms. In the large scale parallel computer systems, interval of the two failures can be shorter than applications execution time. There are mainly four places where faults may be occurring in cloud computing: user-across, provider-across, provider-inner and provider-user [38].

- **Ease Use**

In the ease use, User experiences which is belonging to the subject of human computer interactions is an important criteria when evaluating whether applications is successful or not. In the cloud computing, users' experiences improve lot than its ancestors like grid computing. The good service should be easily accessed for cloud user. The base of user experiences is achieving ease to use. Ease to use is not only simple but also elegant [38].

2.10 Security Issues of Cloud Computing

Recently, Cloud computing uses education and organizational areas. In the organization, cloud computing widely used for storing data on data centers and

providing on-demand services. Data are storing globally then it has some security flaws. Users used services of cloud computing via internet through login id and password.

- **Security**

In the security issue, where is data more secure in the local computer or on the high security server in the cloud data centers? If local system is not connect to the internet then data is more secure in the local computer or if local computer is connect to the internet then it more secure to the cloud data centers. In the cloud, data will be stored in the distributive manner [36].

- **Privacy**

In the privacy is different from the traditional cloud computing model. Cloud computing uses the virtual computing technology. User's personal data may be stored in the various virtual data centers that are located in the various countries. If data will store in the different country with respect to users country, so that data will be face the controversy of different legal systems.

- **Reliability**

Cloud servers have the same problem as your own resident server. The cloud server sometimes slow down, the difference is that users have a higher dependent on cloud service provider (CSP) in the model of cloud computing [36]. If one system has down in the 1 minute but more frequently down and second server down for 10 minutes but less frequent then second server will be more reliable with compare to first server.

- **Legal Issues**

In this issue, the efforts into the lawful situation, such as Amazon Web Services are provide to major market by developing restricted rail and road networks and users to select availability zones [36].

- **Open Standard**

Open standards are the critical growth of the cloud computing. Some vendor has adopted others APIs and there are number of open standards under development including the open cloud computing interface. Open Cloud Consortium (OCC) is the working to develop consensus on early cloud computing standard and practices.

- **Compliance**

In the compliance security issue, cloud has some regulations that pertain to the storage and use of data require and regular reporting the data and regular audit data. In the cloud computing provider must be enable to customers approximately these regulations. In the cloud computing requirements which customers are subject, data centers are maintaining by the cloud service provider (CSP) may also be subject to compliance requirements [39].

- **Freedom**

Cloud computing does not allow to the users to physically possess the storage of the data control in the hands of cloud provider and leaving the data storage. Cloud computing users do not allows to store data in the geographically location to own choice. Cloud users only choice to store data, retrieve the data, update data and remove or delete the data to your own choice. Cloud users access data at anytime and anywhere through internet [40].

- **Long-Term Viability**

In this security issue, cloud service provider be sure that the data you put in the cloud will never become invalid even your cloud computing provider goes down or failure or swallowed and acquired up by the larger company [41].

- **Recovery**

In this cloud security issue, if you do not know where your data is located then cloud service provider to ask what will happen to your data and service in the case of disaster. If any offering that does not replicate the data and application across multiple sites is the vulnerable to the total failure [39].

- **Data Segregation**

In the data segregation, data in the cloud data centers is typically in a shared environment along with data from other customers. Encryption is the effective way to secure the data but it is not cure the all type of data. Encryption and decryption is the traditional way to secure the data but therefore it could not ensure to provide perfect solution for it [42].

- **Policy Integration**

In the integration, cloud computing is the heterogeneous which means that cloud servers may be different mechanism to ensure the clients data security so that policy integration is one of the concerns [43].

- **Authentication & Identification**

In the cloud computing, multi-tenancy is one of the major features of the cloud computing. This feature allows one single instance of software to serve various or more than one clients at a time. Due to multi-tenancy different-2 users may be use different identity tokens and communication protocols which will effects the interoperability defects [43].

- **Non-Repudiation**

In the non-repudiation cloud issue can be obtained by the applying the traditional e-commerce security protocols and the token provisioning to the data transmission within the cloud computing applications such as digital timestamp, signatures and confirmation receipts services [44].

- **Availability**

In this issue, availability is the one of the most critical information security requirements in the cloud computing because it is the key decision factor when decides among public, private or hybrid cloud vendors in the delivery model. Service level agreement is the most important document which is the highlights trepidation of availability in the cloud computing services and the resources between the client and cloud provider [44].

CHAPTER 3

LITERATURE REVIEW

In the literature review section, we will discuss the types of DOS and DDOS attack. DOS and DDOS attack, types of attack and related work to prevent the cloud server and then attack will be remove.

3.1 DOS & DDOS Attack

In the denial-of-service (DOS) attack, this attack in which one or more machines target to the victim and want to prevent the victim from doing useful work. The victim can be a client, router, network server, an individual internet user or a company doing business using the internet, a network link or a entire network, internet service provider (ISP), or any combination of the variant of these. The denial-of-service (DOS) attacks may be involve gaining unauthorized access to the network or computing resources but for the most part in this document we will focus on the cases where the denial-of-service (DOS) attack itself do not involve a compromise of the victim computing facilities.

Because of closed context of the original ARPANET, no other types of considerations were given by denial-of-service attacks in the original internet architectures. Almost all the internet services are vulnerable to the denial-of-service (DOS) attacks of the sufficient scale. In the most of the cases sufficient scale can be achieved by compromising enough end-systems or routers, and using those compromised hosts to perpetrate the attacks.

This document is serving the several purposes [45]:

1) To highlight possible for attack and by the so doing encourages protocol designers and network engineers towards designs that are the more robust.
2) To discuss the partial solutions that reduces the effectiveness of the attacks.
3) To highlight how some partial solutions can be taken advantages of by the attackers to perpetrate alternative attacks.

We note that the in principle it is not a possible to distinguish between a sufficiently DOS attack and a flash crowd that is unexpected heavy but non-malicious traffic has the same traffic as a DOS attack. Such as malicious attacks are usually more expensive and to launch than many of the rush attacks that has been seen to date. Thus, prevent against DOS is not about preventing all the possible attacks but rather than is largely questions of the raising the sufficiently high for the malicious attacks [45].

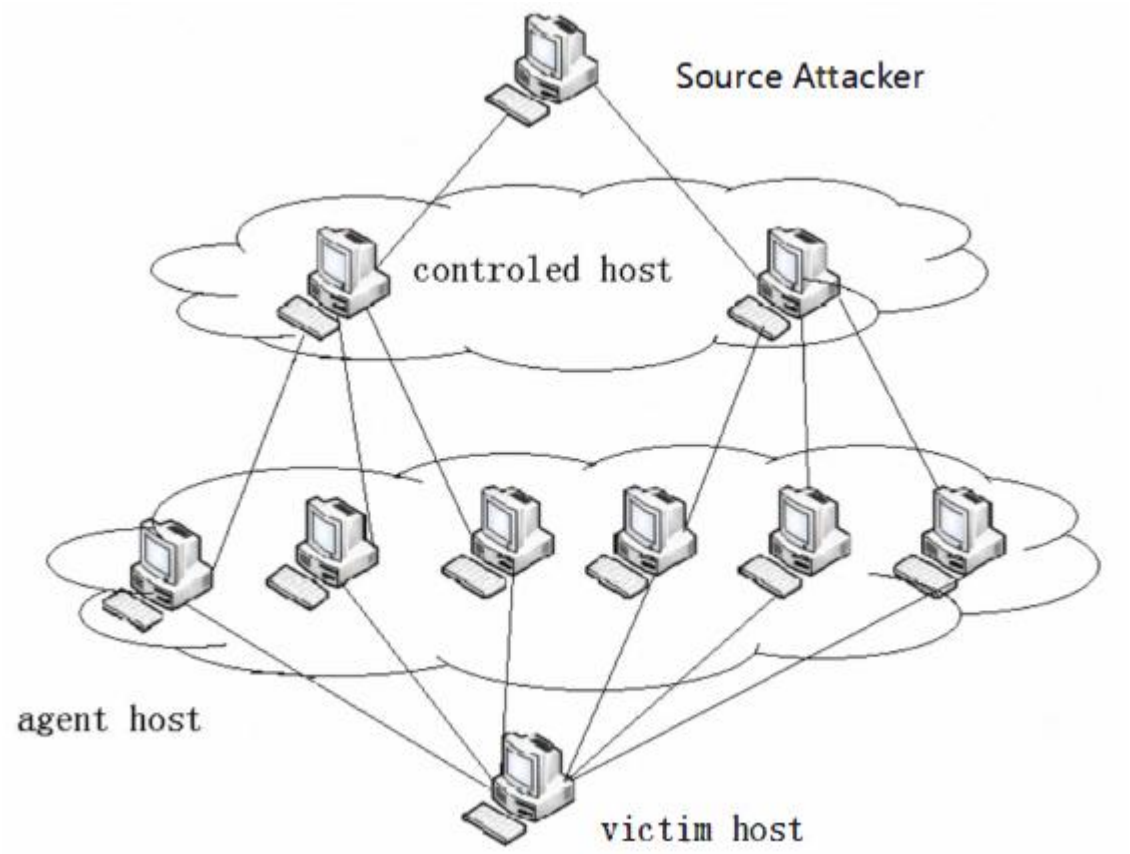

Figure 3.1 A complete system of DDOS attacks [46]

DDOS attack in understanding before a look of its predecessor DOS attack. In the DOS attack attacker sends large number of packets for the service requests over the network, occupy and beyond the processing capacity of the attacked host, to stop legitimate users to provide normal network services, consuming system resources or network bandwidth, leading to overload the network or systems [46].

DDOS attack is the further evolution of the DOS attack. A simple DOS attack is an attack that a target source is one-to-one mapping between source attacker and the victim. The DDOS attack is the introduction of the clients/server system to enhance the introduction of the client/server system to enhance the concept of distributed. It is the concept of many-to-one mapping between source attacker and the victim. DDOS attack is more powerful than the DOS attack and more destructive, and damage.

In DDOS attack, Source attacker wants to DDOS attack to victim host. Source attacker control some specific n number (n = positive integer) of hosts and all n controlled hosts attack separately. Controlled hosts broadcast the packet for attack to victim host. Non-controlled hosts (m = positive integer) have not necessary all the hosts send a packet.

We are firstly discussing on end-systems. The end-systems in this context are typically network server pr PC, but it can also include any communication end points. A router is also an end-system from the point of view of terminating TCP connections for BGP.

- **Exploiting poor software quality**

The simplest denial-of-service (DOS) attack on the end-systems that are exploit the poor software quality on the end-systems, and cause that software to simply crash. For example buffer overflow attacks might be used to compromise the end-systems but even if buffer overflow can't be used to gain access, it will be usually possible to overwrite memory and cause the software to crash. Such as vulnerabilities can be in principle affect any software that is uses data supplied from the network. Software will be crashed because of the poor coding affect not only application software but also operating system kernel itself. The classic example is called "ping of death", which is become widely known in 1996. This exploit may caused many popular operating systems to crash when send a single fragmented ICMP echo request packet whose fragments totaled more than the 65535 bytes allowed in an IPV4 packet.

- **Application Resource Exhaustion**

Network applications may exist in a context that has finite number of resources. In the processing network traffic, such an application uses these resources to do its

intended task. An attacker may be able to prevent the applications from performing its task by causing the applications to exhaust the finite supply of a specific resource.

The obvious resources that might be exhausted include:

1. The CPU cycles available.
2. Available memory
3. The disk space is available for the applications.
4. The number of threads or processes or both the applications is permitted to use.
5. The configured maximum number of simultaneous connections the applications is permitted.

The lists are clearly not exhaustive but it illustrates a number of points.

Some resources are self renewing: CPU cycle will fall in this category. If the attack removed then more CPU cycles become available.

Some resources are constrained by configurations: the maximum number of simultaneous connections and the maximum number of processes are not normally hard limits, but rather are configured limits. Such type of limits is clearly to allow the machines to perform the other tasks in the events of the applications misbehaves.

For example, if a machine single task is to be an FTP server, then setting the maximum number of simultaneous connections significantly less than the machine can service makes the attacker's job easier.

- **Operating System Resources Exhaustion**

Operating system resource exhaustion and the application resource exhaustion are very similar. In the case of application resource exhaustion then the operating system may be able to protect other tasks from the affected by the DOS attack. DOS attack on an operating system is the TCP SYN-flood, which is called memory exhaustion. Then the attacker send a flood TCP SYN packets to the victim and requesting connections setup then does not complete the connections setup. Victim has initiated to handle the incoming connections [45].

CPU exhaustions attacks may be exacerbated by poor operating systems handling of incoming network traffic. An ideal operating system should behave as follows:

1. As incoming traffic increases the useful work done by the operating system should increase until some resources that are saturated.
2. From this point, as incoming traffic continues to increases the useful work done, it should be constant.

3.1.1 Methods of DOS and DDOS Attack

In the DOS and DDOS attack, we discuss method of attack. How attack is done to end-systems. This attack is use flood in the network. In the flooding, source broadcast the packet in the network and destination address in the packet is victim address. Due to flooding victim receives multiple copies of the packet and source target victims packet processing capability [47]. Methods of DOS and DDOS attack are as follows:

3.1.1.1 TCP SYN flood

In this attack, Attacker sends a large number of TCP SYN packets to victim. Victim receives the packet and reply TCP SYN+ACK packet to attacker. Attacker does not response to victim TCP SYN packet and victim is on TCP half-open state [45]. After timeout victim close the half-open connection. In this attack attacker target the memory exhaustion of the victim [45]. If attacker creates more TCP half-open connection before closing previous half-open connection by victim then memory exhaustion problem occur.

3.1.1.2 TCP ACK flood

In this type of attack, an attacker sends large number of TCP ACK packets to victim. Victim will process all the packets. Victim does not process legitimate packets. Most of the time victim will process attacked or spoofed packets. This attack affects on CPU exhaustion [45]. Victim process large number of packet and CPU processing is slow. In this attack, Victim suffers from livelock [45]. Because

incoming number of packets are increases and CPU does not do useful work due to more number of interrupts.

3.1.1.3 TCP Connections

In this attack, Zombie hosts creates the more number of three way handshake TCP connection to victim. These connections are creating until memory or resources are not exhausted [46]. An attacker will create the TCP connections between attacker and victim. Attacker sends TCP SYN packet to the victim and victim reply the piggybacked packet TCP SYN + TCP ACK packets. Attacker sends the TCP ACK packet and TCP connection creates successfully.

3.2 Related Work

In recent years, DOS and DDOS attack is a wide area of the research. This area have existed many successful solution. The DDOS problem has been attracted much attention from the research community. From observation there are three major branches for the research in DDOS attack are as follows:

- **Attack detection:** by monitoring protocol behavior
- **Attack traceback:** by packet marking
- **Attack traffic filtering:** research in the traffic filtering can be categorized into three main areas based on the point of protection.

3.2.1 Different type attack prevent

Filtering of the packets can be done on different-2 types like: victim-initiated, path-initiated, and source-initiated [48].

3.2.1.1 Source initiated

In the source-initiated, source are responsible for the guaranteeing that the outgoing packets are attack free. For example include disabling ICMP, ingress filtering, removing unused services to the prevent computers from becoming attack agents or filtering unused traffic from the source. The viability of the approaches on voluntary cooperation among majority of the ingress network administrator's internet-

wide, these approaches are impractical given the scale and the uncontrollability of the internet.

3.2.1.2 Path-initiated

Path-initiated drop the packet if packet is not coming from appropriate router. If the packet follow the correct path are allowed otherwise packet has be discarded. If any packets coming with a wrong source IP for the particular router port then is considered as a spoofed packet and dropped the packet. This method drops up to 88% of the spoofed packets. These approaches are considered practical they have high probability of false negative that is falsely accepting attack packets [48].

3.2.1.3 Victim-initiated

Victim-initiated reduce the incoming traffic of the packet. For example the pushback scheme, the victim starts reducing excessive incoming traffic and then requests the upstream router to perform rate reduction as well. There are the other methods based on a packet marking, TCP flow filtering, overlay network and statistical processing.

In the figure 3.2, the agents and the servers are many-to-many mapping. If a proxy server will bring much impact. Until all the agents are compromised, there is always one proxy server can also be provide the service, speed will affected in some. The impact depends on the number of compromised agents. These could be effectively solves the flooding attacks. Because of only large number of packets or large flooding attacks can be capture all the agents, ISP network operators can be change the paradigm of the wavelet to detect the occurrences of attacks, this massive arrival of the attack packets, to filter by the fingerprint recognition.

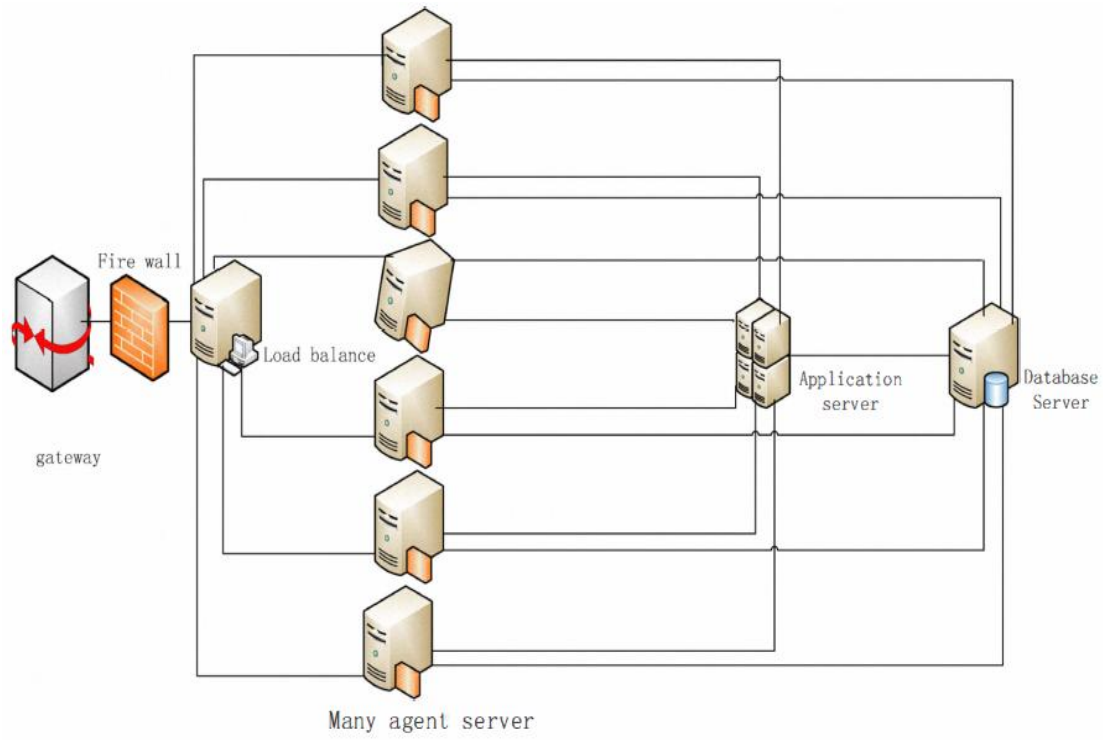

Figure 3.2 The relationship between the server and agent [46]

All services in the central server services agent on the uniform distribution. All the service does not place on the each proxy. Each agent is installed on the some of the services, reducing agency costs and increase operating efficiency. When the database services save the virtual table, do not delete the memory, both fast responses, and reducing the overhead caused by access to the database. DDOS attack can be protected application layer.

Confidence based filtering (CBF) [49] is other method for prevent DDOS attack. In this method find the correlation pattern of the incoming packets. In order to attack packets from legitimate ones the method proposed in this utilizes correlation patterns. The concept of the correlation refers to the situation that is some inner characteristic takes places at the same time in the packet flows. So that the basics assumptions of this method there are some unique correlation patterns in the legitimate packet flows.

Table 3.1 Key Terms used in this work [49]

Terms	Description
N	The number of the attributes under consideration in the method
A_i	The i^{th} attribute in the packet, $(1 \leq i \leq n)$
m_i	The number of values which attribute *Ai* can have
$a_{i,j}$	The *j*-th value of attribute A_i, $(1 \leq j \leq m_i)$
T	A time interval in packet flows
N_n	The total number of packets in the packet flow in one time interval *t*
$N (A_i = a_{i,j})$	The number of packets whose attribute A_i has value $a_{i,j}$ in this packet flow in one time interval *t*
$N (A_r = a_{r,x}, A_s = a_{s,y})$	The number of packets whose attribute A_r has value $a_{r,x}$, attribute *As* has value $a_{s,y}$ in this packet flow in one time interval *t*
P	A packet in the packet flows
P(i)	The value of attribute A_i in packet *p*

In the confidence based filtering (CBF) method, we have focused on the transport layer and network layers. The correlation patterns are the two layers co-appearances between attributes in the IP header and TCP header. The pairs of the attributes are distinctive patterns because some characteristic of the operating systems and network structure. It is the main thing that the key point of its success is utilizing the correlation pattern between time-to-live (TTL) and source IP address. So that generalizes the idea to all correlation patterns between attributes in the IP header and TCP header.

The concept of the correlation is reflects how much trust between correlation pattern and attribute pairs.

Definition1 (Confidence): confidence is the frequency of appearances of the attributes in the packet flows. Confidence for the single attributes and pairs of the attributes are calculated as equation (1) and (2).

$$conf\left(A_i = a_{i,j}\right) = \frac{N\left(A_i = a_{i,j}\right)}{N_n} \quad (1)$$

Where i= 1,2,3,…,n_i, j= 1,2,3,4,…….,m_i

Confidence for the attribute pairs:

$$conf\left(A_{i_1} = a_{i1,j1},\ A_{i_2} = a_{i2,j2}\right) = \frac{N\left(A_{i_1} = a_{i1,j1}, A_{i_2} = a_{i2,j2}\right)}{N_n} \quad (2)$$

Where i_1 = 1,2,3,…….,n, i_2 = 1,2,3,…….,n, j_1 = 1,2,3,…….,m_1, j_2 = 1,2,3,…….,m_2

Meaning of the variables in the equations is listed in the above table. The more times any attribute pair appears in the legitimate packet flow the higher confidence value of the pairs we can get. The concept of the confidence is the basis of the calculations of the confidence based filtering (CBF) score and the complete filtering process.

Definition2 (CBF Score): CBF score of a packet is weighted average of the confidence of the attribute pairs of the value in it. The CBF score for the packet p is calculated as (3):

$$score(p) = \frac{\sum_{k=1}^{d} W\left(A_{k_1}, A_{k_2}\right) conf\ \left(A_{k_1} = p(k_1), A_{k_2} = p(k_2)\right)}{\sum_{k=1}^{d} W\left(A_{k_1}, A_{k_2}\right)} \quad (3)$$

In this definition, d is the total number of attribute pairs involved in the calculations of the score. A_{k_1} and A_{k_2} are the two attributes in the k^{th} attribute pairs. $W\left(A_{k_1}, A_{k_2}\right)$ is the weight of the k^{th} attribute pairs. And consider the range of the each confidence value is in [0,1]. The range of the score (p) is also the [0,1].

We also calculate the confidence of all the pairs of the each attributes value in legitimate packet beforehand. In our method, we calculate these confidence values also called as nominal profile. The attribute pairs which may not be easily copied by the attackers will be given as a high weight. So that if any packet calculate packet

score is high it means its appearance frequency is high and difficulty copied correlation pattern. So that we can choose the discarding threshold value between [0,1] to make the judgment of filtering.

Definition3 (CBF Legitimate Packet): The legitimate packet in the CBF is one who's CBF score is the above discarding threshold. So that those packets with scores lower than the discarding threshold are regarded attack ones. The nominal profile structure which is contains two attributes pair (TTL, packet size) and (TTL, Source IP address).

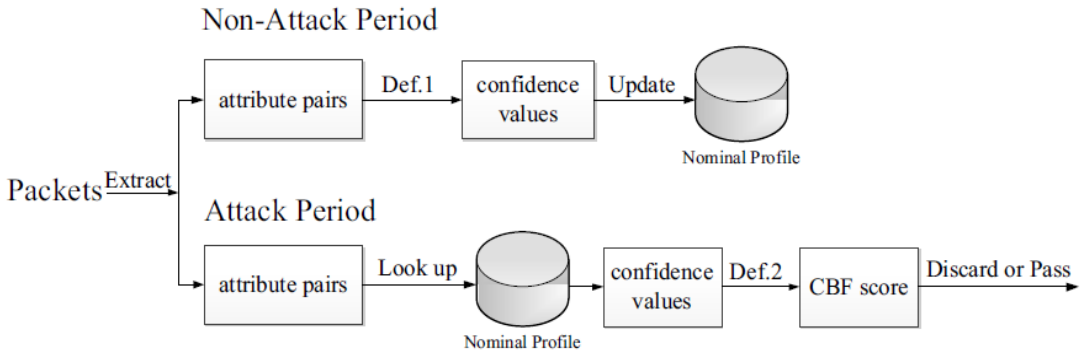

Figure 3.3 Outline of Confidence Based Filtering [50]

CBF method working on two periods: attacking period, Non-attacking period. If non-attack period, Nominal profile generate based on the fields of the packets and calculate confidence. If attack period, Nominal profile generate based on the fields of the occurrences of the packet and calculate confidence. According to confidence packet will be accept or discarded.

We will be introduces the structure of the nominal profile. We select six candidate single attributes. We combine the every two of the six attributes and then get 15 pairs (not the same) of the attributes. After combination the values of the attributes will have 32 bit size so that the 6 single attributes all have the size of the no more than 16 bit.

The overall construct the nominal profile is divided into small intervals which is called windows. The size of the windows can be fixed or the variable and changed to dynamically. In the each time of intervals t, our methods CBF count the number of the values appearances of these 15 pairs of attributes and then calculate the confidence of the each of the 15 pairs. After that end of the each window the new calculated confidence values are used to update the nominal profile. So that minimize the false negative rate, highest value of the confidence of an attribute value pairs in the nominal profile is stored which means that the updating only takes place when the new calculated confidence value is higher than the one stored in the nominal profile. We need only store the confidence value of the attribute value pairs which is higher than the predetermined threshold minconf (minimum confidence) e.g. 0.001, it means that the minimum confidence value in the nominal profile.

CHAPTER 4

PROPOSED WORK

Our idea is to prevent the cloud server due to denial-of-service (DOS) attack and distributed-denial-of-service (DDOS) attack that affects the availability security issue. Because due to DOS and DDOS attack server will goes down and server not available for the service and availability security issue will goes.

Our approach is working with transport layer. Because of transport layer is the end-to-end connections established. In the TCP layer, client can't communicate with server without establishing three way handshakes. Header of the TCP layer will not modify in the intermediate hops. All the information is available about packet in the TCP header filled by the source of the packet. So that if source is broadcast the any packet then receiver will receives the multiple copies of the same packets and all the multiple copies of the packet IP header information is different because these packet comes from the different-2 hops but information of the TCP header is same because TCP layer is working on the end-to-end. So that we are analyze the TCP header of the packet.

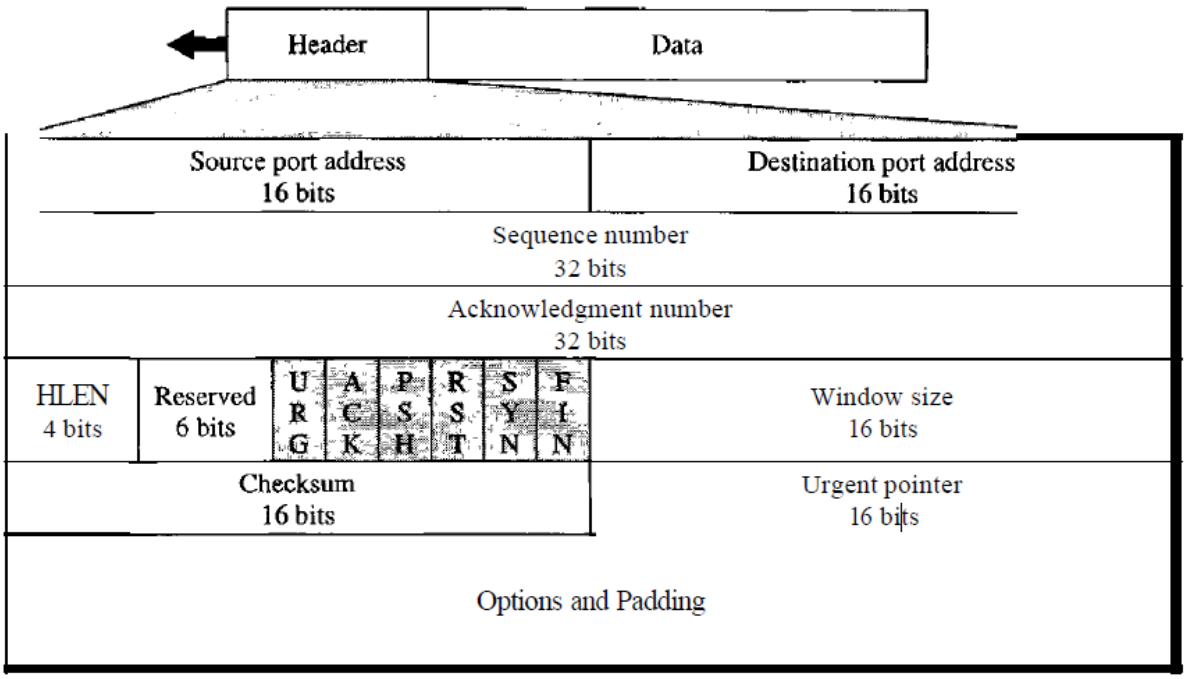

Figure 4.1 TCP frame format [50]

TCP frame header contains 192 bits. In this TCP header contain many types of fields. The fields of the TCP header are each field has its own functions.

Table 4.1 TCP header fields description

Field	Number of bits	Description
Source port number	16	The source port number
Destination port number	16	The destination port number
Sequence number	32	It represents the sequence number. If SYN bit present then it initialize by 1.
Acknowledge number	32	If ACK control bit is set this field contain the value of the next sequence number.
HLEN	4	The number of 32 bit words in the TCP header. It indicated where data begins.

Reserved	6	For the future use.
URG	1	Urgent pointer field significant.
ACK	1	Acknowledgement field significant.
PSH	1	Push function.
RST	1	Reset the connections
SYN	1	Synchronize sequence number.
FIN	1	For finish the connections. No more data from sender.
Window size	16	This field indicates the acknowledgement field which the sender of this segment is willing to accept.
Checksum	16	Checksum field is the 16 bit one's complement.
Urgent pointer	16	Urgent pointer communicates the current value of the urgent pointer as a positive offset.

We are using for our approach three fields in the TCP header:

- Source port number
- Destination port number
- Sequence number

Using these three fields, we are identifying the uniqueness of the packet. We are maintaining four tables are table_r, table_q, table_p and rejection_table. All the tables are maintains the four attributes are as follows:

< Source port number, Destination port number, Sequence number, Count >

Field Attribute (FA): < source port number, destination port number, sequence number >

It is a combination of these attributes.

$Field_{(i)\_table\_name}$: where i= tuple number of the specified table

$Count_{(i)\_table\_name}$: where i= tuple number of the specified table that contain count value

Then all the tables are update after every completion of the timestamp and table_r is update after every incoming of the packets.

Timestamp: Timestamp is a positive integer value. It is a time in millisecond. In single timestamp, we are updating one table_r. Timestamp value do not fix too small or vary large. If timestamp value is very small then most of the time of CPU taken computation work and it increases the number of updates in the rejection table. If timestamp value is large then update in the rejection_table is too late and all the tables size is increases and searching any attribute in the table taking more time on an average. At the i^{th} timestamp three table exist are i^{th} time (table_r), $(i-1)^{th}$ (table_q) and $(i-2)^{th}$ table_p.

$$Timestamp <= \frac{cycle\ time}{3} \quad (1)$$

Cycle time: After this time sequence number of the packet can be repeat

After first timestamp table_q and table_r are empty and table_r move the entry to table_q. After second timestamp table_p is empty and field entry move from table_q to table_p and table_r to table_q. In the third timestamps the entire entry delete from table_p then entry move from table_q to table_p and entry move from table_r to table_q then delete from table_r.

Life cycle of particular field's entry of the packet:

- Packet comes and entry will be updated in the table_r.
- After completion of timestamp entry will be updated from rejection_table and entry move from table_r to table_q.
- After again completion of timestamp entry will be move from table_q to table_p.

- After again completion of timestamp entry will be deleting from rejection_table and entry also delete from table_p.

Table_r, table_q and table_p: Attributes of the table is source port number, destination port number, sequence number and count. Count is doing number of packets that contain remain three attributes are same.

Rejection_table: This table contains attributes are source port number, destination port number, sequence number and count. This table contains all the entries of the table_q and table_p. If source port number, destination port number, sequence number are same in table_q and table_p then rejection_table contain addition of the count of the both table.

This method we are counting the number of packets per timestamp. According to its count packets will be accepted or rejected. rejection_table will update after each timestamp.

This algorithm works as follow:

```
1.  if ( ack_bit == 1 )
2.  {
3.          update ( Field_packet )  //update algorithm written in below
4.          if ( check( Field_packet ) ) //check algorithm written in below
5.          {
6.                  Process packet
7.          }
8.          else
9.          {
10.                 Discard packet
11.         }
12. }
13. else if ( syn_bit == 1 )
14. {
15.         update ( Field_packet )//update algorithm written in below
16.         if ( check( Field_packet ) ) //check algorithm written in below
```

```
17.         {
18.                 Process packet
19.         }
20.         else
21.         {
22.                 Discard packet
23.         }
24. }
25. else
26. {
27.         process packet
28. }
```

After this algorithm, we are discussing update ($Field_{packet}$) algorithm. This algorithm updates the table_r.

```
1. if ( Field_packet == Field_(i)_table_r )
2. {
3.         count_(i)_table_r = count_(i)_table_r + 1 ( corresponding entry in the Field_table_r
   )
4. }
5. else
6. {
7.         add new entry in the table_r
8.         Count = 1    // corresponding entry
9. }
```

Now, we are discussing algorithm for check($Field_{packet}$). This algorithm returns true if packet is processed otherwise return false.

```
1. if ( isEmpty ( rejection_table ) )  // return true if rejection_table is empty
2. {
3.         if ( ( Field_packet == Field_(i)_table_r ) && ( count_(i)_table_r >1 ) )
```

4. {
5. return false
6. }
7. else
8. {
9. return true
10. }
11. }
12. else
13. {
14. if ($Field_{packet}$ == $Field_{(i)\_rejection\_table}$)
15. {
16. if ($count_{(i)\_rejection\_table}$ >1)
17. return false
18. else
19. return true
20. }
21. else
22. {
23. if (($Field_{packet}$ == $Field_{(i)\_table\_r}$) && ($count_{(i)\_table\_r}$ >1))
24. return false
25. else
26. return true
27. }
28. }

Finally, we are discussing algorithm for update rejection_table. This table update after completion of each timestamp. All the update in the rejection_table according to algorithm is as follows:

Algorithm for modify rejection_table.

1. if (isEmpty (table_p)) //table_p is empty
2. {

```
3.      if ( isEmpty ( table_q ) )  //table_q is empty
4.          table_q = table_r
5.      else                    //table_q is not empty
6.      {
7.          table_p = table_q
8.          table_q = table_r
9.      }
10.     while ( ! isEmpty ( table_r ) ) // copy all the entry of table_r to
    rejection_table
11.     {
12.         if ( Field(i)_table_r == Field(j)_rejection_table )
13.         {
14.             count(j)_rejection_table = count(j)_rejection_table + count(i)_table_r
15.         }
16.         else
17.         {
18.             Add Field(i)_table_r and count(i)_table_r in the rejection_table
19.         }
20.         delete i^th tuple in the table_r
21.     }
22. }
23. else        //table_p is not empty
24. {
25.     while ( ! isEmpty ( table_p ) ) // delete all the entry in the
    rejection_table which is in the table_p
26.     {
27.         if ( Field(i)_table_p == Field(j)_rejection_table )
28.         {
29.             count(j)_rejection_table = count(j)_rejection_table - count(i)_table_p
30.             if (count(j)_rejection_table == 0 )
31.                 delete j^th tuple in the rejection_table
32.         }
33.         delete Field(i)_table_p in the table_p
34.     }
```

35. table_p = table_q
36. table_q = table_r
37. while (! isEmpty (table_r)) // copy all the entry of table_r to rejection_table
38. {
39. if ($Field_{(i)\_table\_r}$ == $Field_{(j)\_rejection\_table}$)
40. $count_{(j)\_rejection\_table}$ = $count_{(j)\_rejection\_table}$ + $count_{(i)\_table\_r}$
41. else
42. Add $Field_{(i)\_table\_r}$ and $count_{(i)\_table\_r}$ in the rejection_table
43. delete i^{th} tuple in the table_r
44. }
45. }

This method work good if we choose the value of the timestamp is appropriate.

CHAPTER 5

PERFORMANCE EVALUATION

In this section, we are using some statistics to test this method results. We are testing this method in the environment of 2.66 GHz Intel core i5 processor and 4GB memory. We are implementing this method in the cloud computing environment using cloudsim3.0.2 simulator. The cloudsim3.0.2 is a java library. For using this java library we must have to install jdk6.0 or above version and eclipse SDK. We are checking this method for denial of service attack and distributed denial of service attack results calculated and compare these results with confidence based filtering (CBF) method.

5.1 Simulation Conditions

The average rate of arrival packets are 6000 to 7000 packets per second and arrival of packets rate is 22.33Mbps [49]. Sequence number contain 32 bit so that sequence space is 2^32. According to [51] if data rate is 100Mbps then cycle time is 5.4 minutes and 22.33Mbps is much smaller than 100Mbps. We are using 100ms value of timestamp. It means after every 100ms rejection_table will be updated and table_r will get empty. Timestamp value should not be large or small. If it is large then table size will be large and if it is small then it increases CPU computation. Initially table_p, table_q, table_r and rejection_table are empty. In this simulation probability of arrival of attacking packets are 0.7 to 0.8. The nature of attack that table 5.1 contain are due to denial of service attack and source attacker is one.

Table 5.1 Nature of the DOS attack

Attack Intensity	Attack Packets	Legitimate Packets	Total no of Packets
1x	5,326	1,674	7,000
5x	25,326	9,674	35,000
10x	51,345	18,655	70,000
20x	1,02,929	37,071	1,40,000

Then the next Table 5.2 that is for the nature of the attack for the three source of attacker or distributed denial of service attack (DDOS). Then the Table 5.2 attacking with the probability of the attacking packet is 0.7 to 0.8.

Table 5.2 Nature of the DDOS attack

Attack Intensity	Attack Packets	Legitimate Packets	Total no of Packets
1x	5,122	1,878	7,000
5x	26,305	8,695	35,000
10x	53,149	16,851	70,000
20x	1,04,907	35,093	1,40,000

5.2 Simulation results

In this section, we compare the count based filtering method and CBF method. Timestamp value is 100ms. Table2 denotes the comparison of both methods.

Table 5.3 Comparison of count based filtering and CBF method

Attack Intensity	Process Time in Seconds		
	Count based filtering for DOS	Count based filtering for DDOS	CBF
1x	0.263	0.293	0.332
5x	0.758	0.730	1.073
10x	1.623	1.543	1.919
20x	3.179	3.212	3.661

We are improve the results because of count based filtering method does less computation work compare to CBF method.

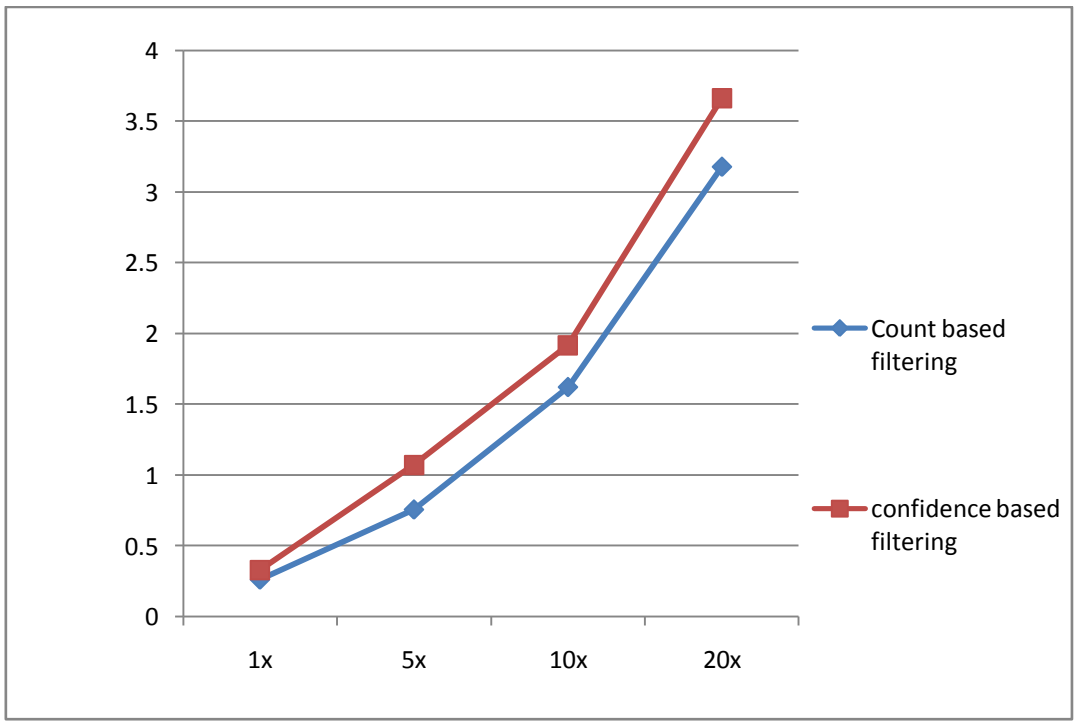

Figure 5.1 Comparison between confidence based filtering and count based filtering method for DOS attack

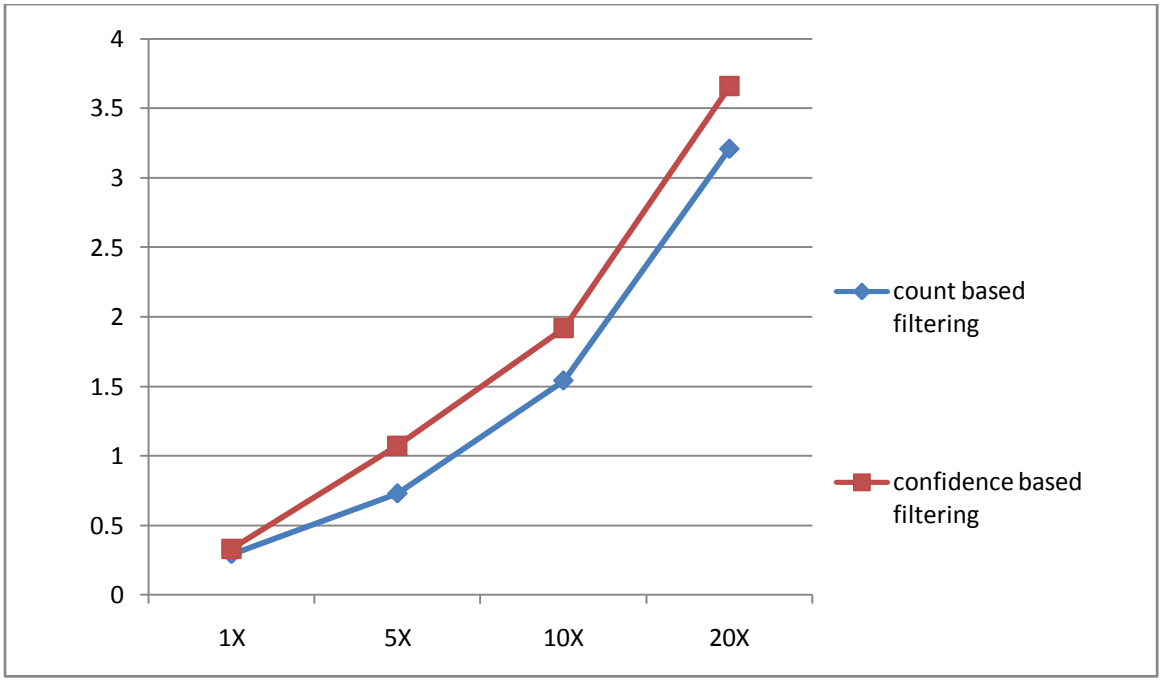

Figure 5.2 Comparison between confidence based filtering and count based filtering method for DDOS attack

This method gives performance same if attack is denial of service attack (DOS) or distributed denial of service attack (DDOS). Result of count based filtering method may vary according to change in the value of timestamp.

Since discarding or accepting a packet of CBF method first calculate confidence of packet and according to minconf packet it will be accepted or rejected. Count based filtering method calculates count of packet and according to count; packet will be accepted or rejected.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Cloud computing is the recent technology that provides services remotely and users are paying for the service. Availability is a most important security problem. Denial-of-service (DOS) attack and Distributed-denial-of-service (DDOS) attacks are threats to the availability security issue. TCP SYN flood attack effect the memory exhaustion and TCP ACK flood attack effect the CPU exhaustion.

We also discussed about count based filtering method. This method creates tables that are table_r, table_q, table_p and rejection_table and all the tables will update with the completion of every timestamp. According to count of the same field packet every packet will be processed or discarded. After completion of every timestamp all the tables will be modified. Performance of this method depends on the value of timestamp. This method will work well if we choose the nominal value of the timestamp. Denial-of-service (DOS) attack and Distributed-denial-of-service (DDOS) attack both are different and Distributed-denial-of-service (DDOS) attack is more dangerous as compare to Denial-of-service (DOS) attack.

Denial-of-service (DOS) attack and Distributed-denial-of-service (DDOS) is most important and most famous attack. Count based filtering method give the good performance and using nominal value of the timestamp this method performance is more enhanced.

REFERENCES

[1] Jayant Baliga, Robert W. A. Ayre, Kerry Hinton, and Rodney S. Tucker, "Green Cloud Computing Balancing Energy in Processing, Storage, and Transport," in *IEEE*, Vol. 99, No. 1, pp. 149-167, Jan. 2011.

[2] Dong Xu, "Cloud Computing: an Emerging Technology," in International Conference on Computer Design And Applications (ICCDA 2010) *IEEE*, vol. 1, pp. 100-104, 2010.

[3] Shufen Zhang, Shuai Zhang, Xubin Chen, and Shangzhou Wu, "Analysis and Research of Cloud Computing System Instance," in International Conference on Future Networks (ICFN 2010) *IEEE*, pp. 88-92, 2010.

[4] Eleonora Mocanu, Mugurel Ionut Andreica, and Nicolae Tapus, "Current Cloud Technologies Overview," *in International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, IEEE,* pp. 289-294, 2011.

[5] Borko Furht, and Armando Escalante, "Handbook of Cloud Computing," in *Springer,* 2010.

[6] Amy Apon, Rajkumar Buyya, Hai Jin, and Jens Mache, "Cluster Computing in the Classroom: Topics, Guidelines, and Experiences," in IEEE, pp. 476-483, 2001.

[7] Amy Apon, Jens Mache, Rajkumar Buyya, and Hai Jin, "Cluster Computing in the Classroom and Integration With Computiing Curricula 2001," *in IEEE,* Vol. 47, No. 2, pp.188-195, May. 2004.

[8] Ashwini Patil, Ankit Shah, Sheetal Gaikwad, Akassh A Mishra, Simranjit Singh Kohli, and Sudhir Dhage, "Fault Tolerance in Cluster Computing System," *in International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, IEEE*, pp. 408-412, 2011.

[9] Philip Hatcher, and Mathew, "Cluster Computing with Java," in *University of New Hampshire, Durham,* pp. 34-39, Mar.-Apr. 2005.

[10] Jens Mache, and Amy Apon, "Teaching Grid Computing: Topics, Exercises, and Experiences," *in IEEE,* Vol. 50, No. 1, pp. 3-9, Feb. 2007.

[11] Manish Parashar, and Craig A. Lee, "Scanning the Issue: Special issue on Grid Computing," *in IEEE,* Vol. 93, No. 3, pp. 479-484, Mar. 2005.

[12] Shadi Ibrahim, Hai Jin, Li Qi, and Chunqiang Zeng, "Grid Maintenance: Challenges and Existing Models," *in National Science Foundation of China.*

[13] Youcef Derbal, "Grid Architecture for High Performance Computing," *in IEEE,* pp. 514-517, 2007.

[14] Naidila Sadashiv, and S. M. Dilip Kumar, "Cluster, Grid and Cloud Computing: A Detailed Comparison," *in 6th International Conference on Computer Science & Education (ICCSE 2011), IEEE,* pp. 477-482, Aug. 3-5, 2011.

[15] "Condor toolkit", *http://www.cs.wisc.edu/condor/condorg*.

[16] "ShaRCS", *http://srcs.ucop.edu/pilot.php*.

[17] "Hadoop", *http://hadoop.apache.org*.

[18] "Globus", *http://www.globus.org/ogsa/*.

[19] "Globus project", http://www.globus.org/demogrid/.

[20] "EGI-InSPIRE", http://www.egi.eu/projects/egi-inspire/.

[21] "MammoGrid", *http://mammogrid.vitamib.com*.

[22] "DDGrid", *http://www.ddgrid.ac.cn*.

[23] "Cloud Security Alliance (CSA)", *http://www.cloudsecurityalliance.org/trustedcloud.html*.

[24] "Panda Cloud", *www.cloudantivirus.com/*

[25] "PARMON", *http://www.cloudbus.org/course/parmon.php*.

[26] "Gridbus toolkit", *http://www.gridbus.org*.

[27] "Legion toolkit", *http://www.legion.virginia.edu*.

[28] "Cloudera", *http://www.cloudera.com*.

[29] Zaigham Mahmood, "Cloud Computing: Characteristics and Deployment Approaches," *in International Conference on Computer and Information Technology, IEEE,* pp. 121-126, 2011.

[30] Jun Huang, Yanbing Liu, and Qiang Duan, "Service Provisioning in Virtualization-based Cloud Computing: Moodeling and Optimization," *in Communications QoS, Reliability and Modelling Symposium, IEEE,* pp. 1710-1715, 2012.

[31] "Google App", http://www.google.com/apps/intl/en/business/index.html.

[32] "Salesforce Homepage", http://www.salesforce.com/crm/.

[33] "Google App Engine", http://code.google.com/appengine/.

[34] "Windows Azure platform", http://www.microsoft.com/windowsazure/.

[35] "Force.com Whitepaper", *The Force.com multitenant architecture*. http://www.apexdevnet.com/media/ForcedotcomBookLibrary/Force.com_Multitenancy_WP_101508.pdf.

[36] Jianfeng Yang, and Zhibin Chen, "Cloud Computing Research and Security Issues," *in IEEE,* 2010.

[37] Shuai Zhang, Shufen Zhang, Xuebin Chen, and Xiuzhen Huo, "Cloud Computing Research and development Trend," *in International Conference on Future Networks, IEEE,* pp. 93-97, 2010.

[38] Chunye Gong, Jie Liu, Qiang Zhang, Haitao Chen, and Zhenghu Gong, "The Characteristics of Cloud Computing," *in International Conference on Parallel Processing Workshops, IEEE,* pp. 275-279, 2010.

[39] Krešimir Popović, and Željko Hocenski, "Cloud Computing security issues and challenges," *IEEE,* pp. 344-349, May 24-28, 2010.

[40] Jack Schofield. June 2013,http://www.guardian.co.uk/technology/2013/jun/cloud-computingjack-schofield.

[41] Gartner, "Cloud-computing security risks", http://www.infoworld.com June, 2013.

[42] Farzad Sabahi, "Cloud Computing Security and Responses," *in IEEE,* pp. 245-249, 2011.

[43] Ziyuan Wang, "Security and privacy issue within the Cloud Computing," *in International Conference on Computational and Information Sciences, IEEE,* pp. 175-178, 2011.

[44] Ramgovind S, Eloff MM, and Smith E, "The Management of Security in Cloud Computing," *IEEE,* 2010.

[45] M. Handley, and E. Rescorla, "Internet Denial-of-Service Considerations," IETF, RFC 4732, http://tools.ietf.org/html/rfc4732.

[46] An Lei, and Zhu Youchen, "The Solution of DDOS attack based on Multi-agent," *in International Conference on Educational and Information Technology (ICEIT 2010) IEEE,* Vol. 2, pp. 530-532, 2010.

[47] W. Eddy, "TCP SYN Flooding Attacks and Common Mitigations," IETF, RFC 4987, http://tools.ietf.org/html/rfc4987.

[48] Yoohwan Kim, Wing Cheong Lau, Mooi Choo Chuah, and Jonathan Chao, "PacketScore: A Statistics-Based Packet Filtering Scheme against Distributed

Denial-of-Service Attack," in Transactions on dependable and secure computing *IEEE*, Vol. 3, No. 2, pp. 141-155, April-June 2006.

[49] Qi Chen, Wenmin Lin, Wanchun Dou, and Shui Yu, "CBF: A Packet Filtering Method for DDoS Attack Defense in Cloud Environment," in Ninth International Conference on Dependable, Autonomic and Secure Computing, *IEEE*, pp. 427-434, 2011.

[50] Behrouz A. Forouzan, and Sophia Chung Fegan "Data Communications and Networking," in *Tata McGraw Hill, Fourth Edition,* 2007.

[51] Jon Postel, "TRANSMISSION CONTROL PROTOCOL," IETF, RFC 793, http://tools.ietf.org/html/rfc793.