Volume 4, Issue 3, March 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Survey on Grid Computing Environment

PrajnaParamita Ray
Asst.Professor,CA Dept.,GNIT,WBUT
India

Swati Sing, Riya Guchhayit
Trainee,CA Dept.,GNIT,WBUT
India

Abstract-Grid computation comes from the idea of super computation; it uses the internet to connect computers worldwide, to create a virtual supercomputer. Grid computation saves money and space, while a supercomputer is costly and wastes space. The main purpose of grid computing is to share computer resources online using internet connections. However, security is a concern because sharing resources always increases the risk of a computer virus or worm. The main contribution of this paper is to present a brief survey on some aspects of Grid computing environment including VO,TrustDomain,OGSA,GSI,Grid Middleware.

Keywords— VO, TRUST DOMAIN, OGSA, GSI, GRID MIDDLEWARE

I. INTRODUCTION

Grid computing provides an infrastructure that supports resource sharing across multiple administrative domains. Due to the dynamic and multi-organizational nature of grid, the issue of managing security of both users and resources are the most challenging thing. First generation grid was deployed in research labs, academic institutions and for military use. But now a day many enterprises are beginning to use grid technologies commercially as well , so maintaining the QoS and security are important to meet user's demand.The followings are some of the important aspects of grid Computing:

A. VO(Virtual Organization)

Grid Computation runs processes under an internet environment and therefore a specific domain need to be set up to restrict sharing of resources throughout the globe.This domain is known as VO(Virtual Organization).People who join a VO will be able to share some (but not all) of their computer's re-sources (storage drive, memory, or CPU…) with others in the same VO. Normally a web-based interface is designed for users to perform their requests.VO has a centralized authorized system which provides credentials to the users to access the resources. Examples of some VO Authorized Systems are:CAS(CommunityAuthorizationSystem),VOMS(VirtualOrganization Membership Service),EALS(Enterprise Authorization and Licensing System).

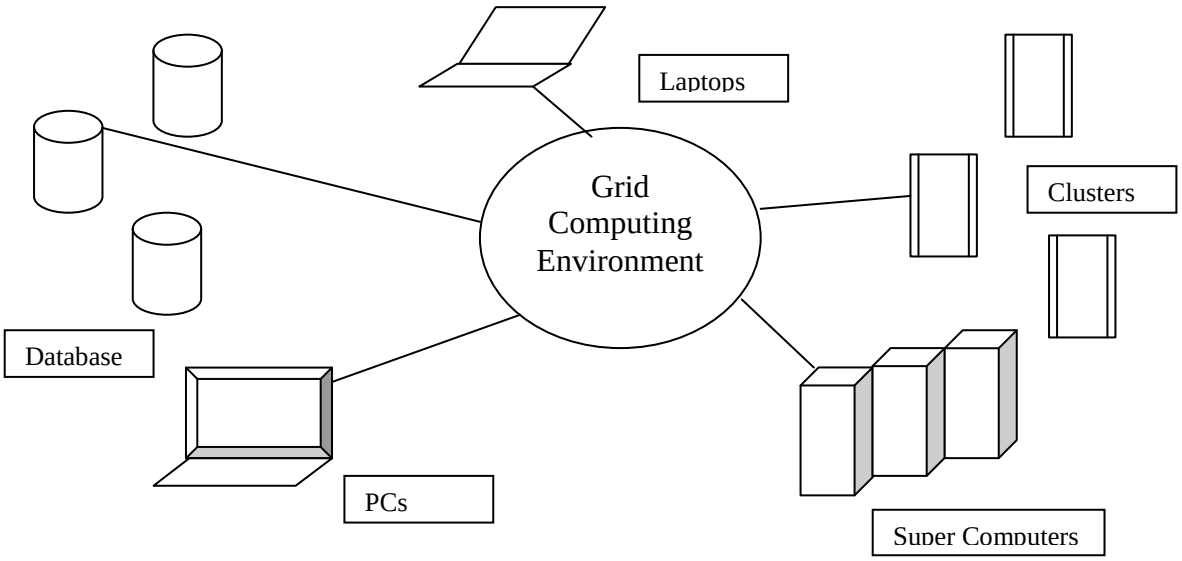

Fig 1: selecting, sharing,aggregating,hosting,offering service across the world

These virtual organizations require common solutions for resource management, data management and access,application development environments, and information services. In many ways, arguably, the most significant challenge for Grid computing is to develop a comprehensive set of mechanisms and policies for *securing* the Grid.Users need to know if they are interacting with the "right" piece of software or human, and that their messages will not be modified or stolen as they traverse the virtual organization.

B. Trust Domain

Before discussing trust domain we need to know some terminologies which are as follows:

- Subject: A subject is a User or a process operating on behalf of a user.
- Object: An object is resource protected by the security policy.
- Credential: A piece of information for authorization purpose of a user. Such as password.

VO must establish trust among not only the users(people) and resources (eg: computers) who join VO but also among the VO's resources so that they can be co-ordinated.These Trust Domains can span multiple organizations and must adapt dynamically as participants join, are created, or leave the VO.
In other words, a trust domain is a collection of both subjects and objects governed by single administration and a single security policy.

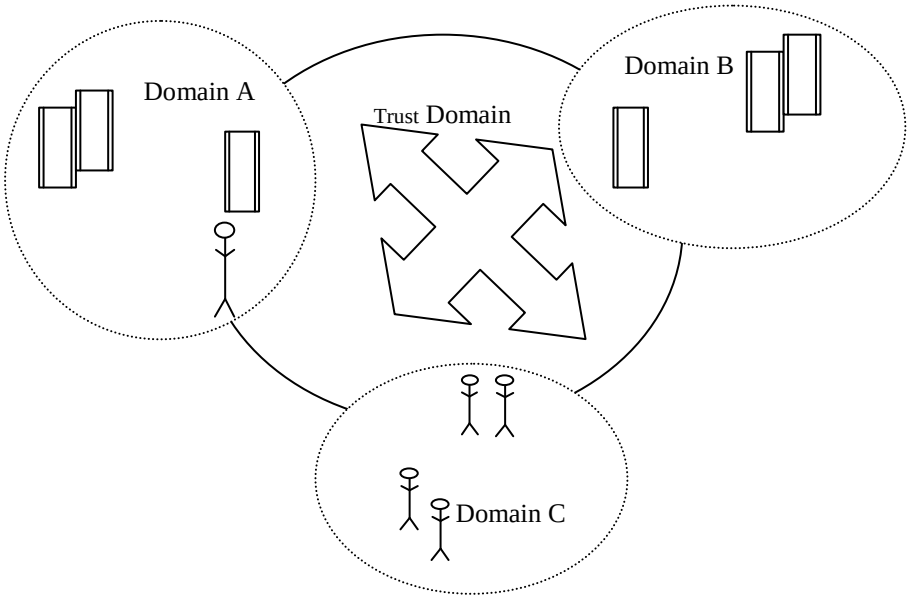

Fig 2: A trust Domain in VO shows how the participants from disparate domains can be pulled together into a common trust domain .

C. OGSA

In between 2002-2006 all the researchers and practitioners in the grid computing area opened a Global Grid Forum (GGF)which is now called as Open Grid Forum(OGF) have released a Open Grid Standards Architecture(OGSA).It is a layered architecture in which each layer defines comprehensive standards which governs Grid Computing.

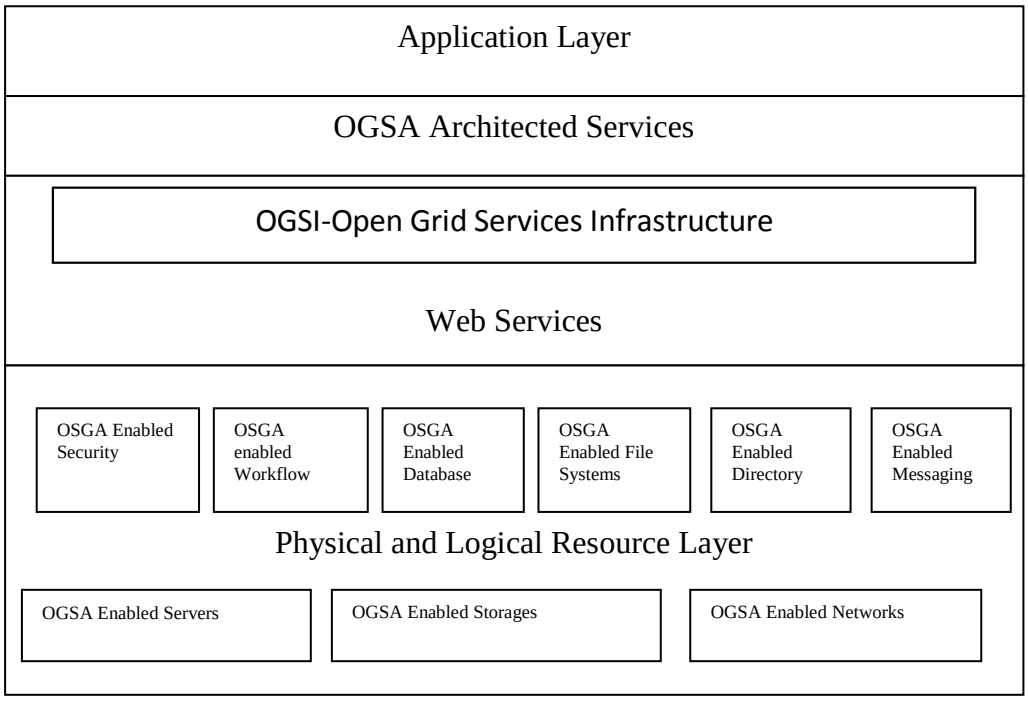

Fig 3:Layers of OGSA Architecture

The Layers of OGSA are as follows:

- **Physical and Logical Resource Layer:**
  Physical resources are Servers,Storages,Networks .Logical resources are Security,Workflow Managers,Database Managers.Logical resources manages Physical resources.
- **Web Services Layer:**
  Consists of Open Grid Services Infrastructure (OGSI)[2][6] sub-layer which specifies grid services and provide consistent way to interact with grid services.The sub-layer consists of 5 interfaces.5 interfaces are:

  - Factory : provide way for creation of new grid services
  - Life Cycle : Manages grid services life cycles
  - State Management :Manage grid service states
  - Service Groups :Collection of indexed grid services
  - Notifications :Manages notification between services and resources
- **OGSA Architected Services Layer:**
  Classified into 3 service categories:
  - Grid Core Services :Service Management: assist in installation, maintenance, troubleshooting tasks in grid system.It also provides policy services and security services.
  - Grid Program Execution Services : Supports unique grid systems in high performance computing,collaboration,parallelism..
  - Grid Data Services :Provide mechanism for access to distributed resources such as databases, files
- **Application Layer:**
  This layer comprise of applications that use the grid architected services.

D. GSI

The Grid Security Infrastructure (GSI),developed independently and later integrated as part of the OGSA Standards. It addresses all the information security challenges in Grid.GSI is based on some proven standards such like public key encryption,X.509 certificates,Single sign and delegation and enables authentication and communication over internet.The latest Version of GSI based on Globus Tool Kit 4.0(open source implementation of OGSA) also allows Web services based security.

- In Public Key Cryptography, the entities generate public/private key pairs based on some cryptographically secure mathematical function.A message is encrypted by a public key and can be decrypted using the corresponding private key.
- X.509[5] certificate is used for authentication purpose.This certificate contains information to identify and to authenticate the user and services
- By creating a proxy The GSI provides a single sign on and delegation procedure, which reduces the number of times the user must enter his or her password when multiple resources are used. Kerberos [6] is an authentication system designed to allow a single sign-on to many machines within an administrative domain. Only the Kerberos Key Distribution Center (*KDC*) needs to know everyone's password. It acts as a trusted third party between the user and a target host. Both the user and the target host establish an authenticated secure connection with the KDC. The KDC gives the user an encrypted token that he can present to a target server, who can then present it to the KDC to establish the identity of the requestor. In a Kerberos login the user does not send the password to the KDC, but uses it to encrypt a challenge phrase that Kerberos can decrypt using its copy of the password. Since the KDC stores all passwords, it becomes a single point of failure for the whole domain. Even though the KDC is expected to be operated in a very secure manner, the hosts that rely on a KDC tend to be limited to a single administrative domain.

E. GRID MIDDLEWARE

It is a software stack enabling users to access multiple resources in a Grid.
Some name of Grid Middlewares are:

- Globus Toolkit[1]
- BOINC[1]
- gLite
- legion
- EMI
- DIET
- ARC
- Oracle Grid Engine
- Grid Way

Due to the decentralized approach and heterogeneous infrastructure of GRID, It is a tough challenge to the Grid Middleware to provide secure,stable,high quality of service to the users of GRID.Hence a Grid Middleware is divided into six modules as follows:

- Basic middleware
- Authentication and Authorization System
- Workload Management/Scheduling
- Data Management
- Fabric Management
- Information System

Basic Middleware: This module provides the basic abstraction layer from the system integrated in the grid and also the API for developing and running applications on the grid.
Authentication and Authorization System: This module is responsible for the authentication and authorization of users, virtual organizations, and processes accessing the grid.
Workload Management/Scheduling: This module manages the scheduling, distribution, and prioritization of jobs and processes running on the grid.
Data Management: This module manages the data storage and also the access to data on the grid.
Fabric Management: This module provides tools for installation management of grid applications and basic resource management, monitoring, and configuration.
Information System: This module collects available information about the grid like availability and status of resources, the job queue and the status of active jobs, information about users and virtual organizations, etc. These systems allow monitoring the grid and also providing tools allowing users to interact with their submitted jobs.

Globus Tool Kit[2] ,gLite [2]these are some grid middlewares those are used for authorization and authentication purpose .They both uses the PKI(public Key Encryption/Decryption)method and X.509 certificate for authentication and authorization.

BOINC[2] is a grid middleware which is used for scheduling purpose.BOINC, allows the owner of the resource decide at which project he wants to participate and all available resources are used. In contrast to Globus and gLite where the users are the ones accessing others' resources and thus have to trust the resources and the infrastructure, when participating at BOINC the users have to trust the software and data installed and transferred to them.

gLite[2] ,for example includes the web information and monitoring tool **GridICE[3]**,which can provide information about available memory, number of CPUs, storage size, etc. for computing and storage elements, which is useful information for administrators and users, but also possibly useful information for attackers.

II. BACKGROUND

The following are some of famous projects that have been designed for grid computation:

Human Genome Project (HGP)[1] The human genome is composed of 24 distinct chromosomes with about 3 billion DNA base pairs organized into 20,000~25,000 genes . To identify these genes and determine the sequences of 3 billion DNA base pairs, running a computer simulation would be expensive and time consuming. Coordinated by the U.S. Department of En-ergy (DOE) and the National Institute of Health (NIH), the Human Genome Project was completed in 2003, three years ahead of the target goal. The institutes involved in this research are Wellcome Trust, Sanger Centre, and HUGO Gene Nomenclature Committee. The Wellcome Trust Sanger Institute has accomplished almost 1/3 of the total work. The project identified about 20,000~25,000 genes in human DNA, deter-mined the sequences of the 3 billion chemical base pairs that make up human DNA, stored this information in databases, improved tools for data analysis, transferred re-lated technologies to the private sector, and addressed the ethical, legal, and social is-sues (ELSI) that may arise from the project .

Human Proteome Folding Project (HPF)
After the Human Genome Project [1]was completed, scientists wanted to understand the function of human proteins, which af-fect human health, to discover the cure for diseases such as AIDS and cancer. Today, only the function and struc-ture of 30 % of human proteins are known . To identify all human proteins could take up to 1,000,000 years using the most advanced personal computer to perform the analysis. Therefore, the Human Proteome Folding Project (HPF) was started and ran on two computational grids; World Computing Grid (WCG) and United Devices' grid.org , which tried to identify all human proteins' functions and structures in a short time. The institutes that participated in this research include the Institute for Systems Biology, the University of Washington, Seattle, and the IBM Corporation..

World Community Grid
The World Community Grid [1] is an experimental project led by IBM, which accepts volunteer members. It is a non-profit organization, which welcomes anyone in the world to donate some computing resources when staying online but

doing nothing. It supports all kinds of research that benefits humanity, at no cost. To participate in World Community Grid you can download the software from their website (http://www.worldcommunitygrid.org/) and install it. The software is free and secure. The current projects running are Help Defeat Cancer, FightAIDS@Home, and Human Proteome Folding - Phase 2 . You may also submit your project pro-posal to them by filling out the application form online.

Computational Chemistry

Chemical reactions or molecular behavior can be huge and complicated processes. Chemists have been trying to determine molecular structure, simulate molecular be-havior, and predict the reaction processes. Computational chemistry[1] has been opera-tional for a long time; however, some chemistry problems, like quantum mechanics, would take hundreds of years to simulate on a personal computer. Therefore, grid computation plays an important role in computational chemistry, which not only saves equipment costs but also processing time. Computational Chemistry Grid (CCG, https://www.gridchem.org/) is one of the most important virtual organizations, which provides all necessary software and resources for computational chemistry.

Business Computation

Grid computation is not only used in science, but also in business computation, where all corporate resources can be pooled so they can be processed efficiently in parallel, according to the business demand. Based on this design, enterprise level business-to-business (B2B) collaborations will be the virtual organization, which handles re-sources management .
Oracle has developed the most famous database management system in the world, and most enterprises like its reliable data management ability and powerful data query process. The Oracle 10g, in which g stands for "grid," has become the first database management system for grid computation. The Oracle 10g runs all database systems in a virtual environment (grid) where all systems are considered a resource pool, us-ing resources efficiently and dynamically for business needs .

SETI@Home Project

Searching for extraterrestrial intelligence (SETI)[1], is a compelling scientific research topic. SETI@Home, directed by UC Berkeley, utilizes grid computation technology to analyze space-based radio signals collected from a radio telescope, at Arecibo, Puerto Rico. This project uses a new platform, Berkeley Open Infrastructure for Net-work Computing (BOINC), to support the research. This platform will automatically update without having to download new versions. Sun Microsystems Inc. has made a great contribution to the SETI@Home project by providing UC Berkeley with required equipment and software. By downloading a screensaver program bundled with required software, almost a half million personal computers have been connected, through this technology, to perform grid computation for the SETI@Home project.

Other Applications

Grid computation can also be used in financial modeling, earthquake simulation, and climate/weather modeling, which are complex processes requiring an intricate infra-structure. A dynamic grid environment, which can perform parallel processing under a collaborative network, must be created to deliver the information

Some ongoing International projects are as follows:

D-Grid(Germany)

The first D-Grid projects started in September 2005 with the goal of developing a distributed, integrated resource platform for high-performance computing and related services to enable the processing of large amounts of scientific data and information.

Dutch Grid(The Netherlands)

DutchGrid[4] is the platform for grid computing and technology in the Netherlands. Open to all institutions for research and test-bed activities, DutchGrid aims to coordinate various grid deployment efforts and to offer a forum for the exchange of experiences on grid technologies.

ENAGRID(Italy)

ENEAGRID[4] makes use of GRID technologies to provide an integrated production environment including all the High Performance and High Throughput computational resources available in ENEA, the Italian National Agency for New Technologies, Energy and the Environment. Interoperability with other grid infrastructures is currently in operation.

Grid-Ireland(Ireland)

Grid-Ireland [4]fosters and promotes grid activities in Ireland, involving partners across the country.

Hungrid (Hungary)

Hungrid [4]was the first official Hungarian Virtual Organisation of EGEE and has become a national alliance from which NGI_HU has emerged. Hungrid joins the grid resources of all national education and/or research institutes of the Hungarian Grid Competence Centre. The aim of Hungrid VO is to allow grid users of Hungarian academic and

educational institutes to perform computing activities relevant for their research. Thus the VO functions as a catch-all for all the Hungarian participants that do not (yet) have an established VO in their respective field of research.

III. CONCLUSIONS

In this paper We have discussed on various important aspects of Grid Computing Environment.We have discussed about Virtual Organization, Trust Domains,OGSA Architecture. We have discussed about the modules of Grid Middleware as well as how security challenges are handled in theGSI layer .We have also given some project's names those are either completed or is running or will start based on Grid Computing.Some of them are national projects and some are international.We have also given names and purposes of some Grid Computing Softwares also in our Paper.

FUTURE SCOPE

In future we will survey on different security mechanisms, current security models and their advantages and drawbacks. We will also try to design a new security model for Grid Computing Environment in future.

REFERENCES

[1] Grid Computation – The Fastest Supercomputer in the World By Chao-Hsu Yao
[2] International Paper Security for Grids: Marty Humphrey, Mary R. Thompson and Keith R. Jackson
[3] Security in Grid Computing :Muhammad Asif Habib, Michael Thomas Krieger, Johannes Kepler University, A-4040 Linz, Austria
[4] http://www.gridcafe.org/EN/national-grid-initiatives.html
[5] CCITT Recommendation X.509: The Directory –Authentication Framework. 1988.
[6] International paper by IEEE on Mobile OGSI.NET: Grid Computing on Mobile Devices by-David C. Chu EECS Department University of California, Berkeley , CA 94720-1776

