Volume 4, Issue 5, May 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Security Analysis for Data Grid Middle wares

Raafiya Gulmeher
Research Scholar
Computer Science & Engineering
J.J.T University, jhunjhunu, Rajasthan, India

Dr. Mohammed Abdul Waheed
Associate Professor
Computer Science & Engineering
VTU Regional office, Gulbarga, Karnataka, India

Abstract—Grid computing is believed to be ultimate solution for meeting the increasing computation needs of organizations. At present the major focus is on load balancing in Grid computing in order to improve the performance of grid. However, the user running an application on a remote machine in the grid-computing network requires assurance about privacy and integrity of his data. Similarly the local host requires a similar assurance regarding the client data and processes that run on the host. So the focus must be on the security of data and applications along with the high performance of the grid. The purpose of this paper is to explore the security problems in grid computing and the steps that can be taken to solve them

General Terms: Performance, Reliability, Security

Keywords: Data grid, Replication, Resources, Security, Interoperability, Grid workflows.

I. INTRODUCTION

A Grid is a collection of heterogeneous computers and resources spread across multiple administrative domains with the intent of providing users easy access to these resources. In simple term, it is the pooling of computing resources. However, the inherent scale, heterogeneity, dynamism and non-determinism of grids and grid applications have resulted in complexities that are quickly breaking current paradigms, making both the infrastructure and the applications insecure. So there is a need for a fundamental change in how grids and grid applications are developed and managed.

The major aim of this paper is to study about data grid security issues and provide a solution to guard data or information in Grid Services that are appeared while operating in data storage systems
and we present a cryptographic & fragment based scheme to accomplish the server protection requirements associated with a standard Data Grid environments. The Data Grid is a kind of distributed structure in which mutual assets (CPU or storage space) are offered in a cooperative style by the peers. These surroundings likely to present productive assets not only for processor-based jobs, but as well for programs which need major sum of primary memory, physical memory space and network performance. Data Grids are based on a number of extensively spreader and distrusted data storage peer, hence doesn't assures accessibility or safety to the preserved data.

In order to guarantee the privacy, reliability, as well as ease of use of significant data in dispersed storage systems, furtive distribution also erasure coding-based plans have been exercised. For achieving recital objectives in information entrances, we use information crumbling approach which is collective along by means of dynamic duplication. Usually erasure coding systems are utilized to examine the issue of most favorable distribution of responsive information stuffs. We putrefy the distributed facsimile share issue into 2 associated issues: The OIRSP is used to determine which clusters share replicas. We also use Optimal Intra need cluster Share Allocation Problem (OISAP) where the quantity of distribute copies required in a group and its locations is determined. We propose two heuristic algorithms in favour of the two associated issues. From the Experimental studies it is determined that heuristic algorithms help in achieving high-quality recital in decreasing communiqué outlay.

II. MIDDLEWARE

Middleware provides all the services and applications necessary for efficient management of datasets and files within the data grid while providing users quick access to the datasets and files Data access services work hand in hand with the data transfer service to provide security, access controls and management of any data transfers within the data grid. Security services provide mechanisms for authentication of users to ensure they are properly identified. Common forms of security for authentication can include the use of passwords. Authorization services are the mechanisms that control what the user is able to access after being identified through authentication.

III. GRID MIDDLEWARE TECHNOLOGIES

A high-level view of activities involved within a seamless, integrated computational and collaborative Grid environment is shown in Figure1.. The end users interact with the Grid resource broker that performs resource discovery, scheduling, and the processing of application jobs on the distributed Grid resources. In order to provide users with a seamless computing environment, the Grid middleware systems need to solve several challenges originating from the

inherent features of the Grid. One of the main challenges is the heterogeneity in grid environments, which results from the multiplicity of heterogeneous resources and the vast range of technologies encompassed by the Grid. Another challenge involves the multiple administrative domains and autonomy issues because of geographically distributed grid resources across multiple administrative domains and owned by different organizations. Other challenges include scalability (problem of performance degradation as the size of Grids increases) and dynamicity/ adaptability (problem of resource failing is high). Middleware systems must tailor their behavior dynamically and use the available resources and services efficiently and effectively.

A tremendous amount of effort has been spent in the design and implementation of middleware software for enabling computational Grids. Several of these software packages have been successfully deployed and it is now possible to build Grids beyond the boundaries of a single local area network. Examples of Grid middleware are UNICORE (UNiform Interface to COmputing REsources) , Globus, Legion and Gridbus . These middleware systems aim to provide a grid-computing infrastructure where users link to computer resources, without knowing where the computing cycles are generated.

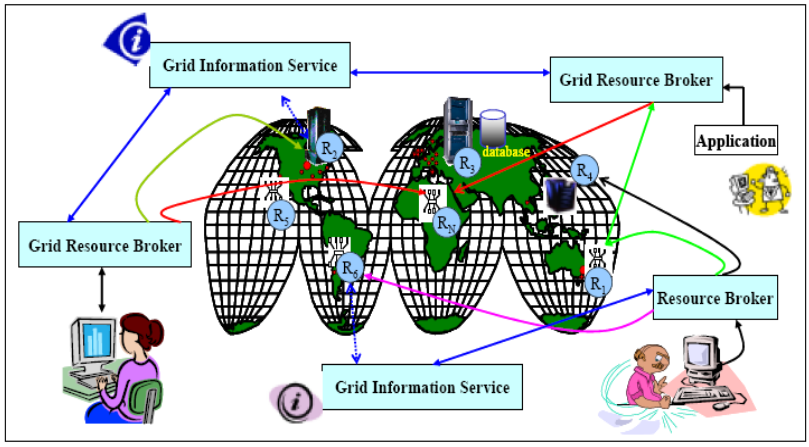

Figure 1: A world-wide Grid computing environment.

IV. GRID SECURITY ISSUES

Security is one of the main issues that usually arise when considering a grid computing environment. While the safeguards of a traditional system aim at protecting the system and data from its users, the security orientation of grid systems need to go a step ahead and also protect applications and data from the system where the computation takes place. Therefore, some unique requirements and challenges are to be counted while adopting a security infrastructure in a grid computing system. In this section, we point out the Technical and Non-technical security requirements that should be in place in grid computing systems .

Technical Issues

a)Logging information: There must be a proper record of log-in and log-out information of every event on grid.

b) Single sign-on: A user should be able to authenticate once and initiate computations without further authentication of the user.

c) Protection of credentials: User credentials such as passwords, private keys etc. must be protected.

d) Uniform credentials/certification infrastructure: There should be a standard such as X.509v3 for encoding credentials for security principals.

e) Delegation of access rights: Providing mechanisms to allow delegation of access rights from requesters to services while ensuring that the access rights delegated are restricted to the tasks intended to be performed within policy restrictions.

f) Message integrity: Ensuring that unauthorized changes made to message content or data can be detected at the recipient end.

g) Privacy: Allowing both a service requester and a service provider to define and enforce privacy policies.

h)Interoperability with local security solutions: Security policy for inter-domain access must be in accordance with security policy for local resources. There should not be any need to modify local security policy according to inter-domain security policy.

i) Support for secure group communication: A computation can comprise a number of processes that will need to coordinate their activities as a group. The composition of a process group can and will change during the lifetime of a computation. Therefore, support is needed for secure authenticated communication for dynamic groups.

j) Support for multiple implementations: The security policy should not dictate a specific implementation technology rather it should be possible to implement the security policy with a range of security technologies.

Non-Technical Issues

There are also many non-technical issues related to the adoption of grid computing. Corporate culture of many organizations may be opposed to it. The grid computing is based on the optimum utilization of resources and this optimism is achieved not only on the basis of computer resources connected within an organization. The resources are

shared in multi-administrative domains. The purpose is that if at any time an organization needs more computing resources than its capacity, then there is no need to arrange these rather it can use the underutilized resources of other organizations sharing their computing resources with it. Similarly there can be a reverse situation. However, the data access and applications processing policies of these organizations may not be in accordance to one another. So grid computing may require the redefining of these policies and the redefinition of ownership, copyright and licensing. As grid computing progresses such cultural, legal and economic issues will have to be resolved by adjusting these in accordance with what the technology will provide.

V. DATA GRID ORGANIZATION

Figure 2 shows a taxonomy based on the various organizational characteristics of Data Grid projects. These characteristics are central to any Data Grid and are manifest in different ways in different systems.

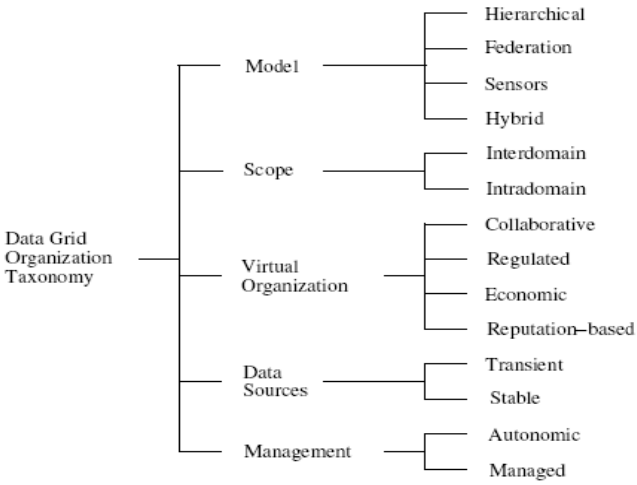

Figure 2: Data Grid Organization Taxonomy.

Data Transport:

The data transport mechanism is one of the fundamental technologies underlying a Data Grid. Data transport involves not just movement of bits across resources but also other aspects of data access such as security, access controls and management of data transfers. A taxonomy for data transport mechanisms within Data Grids is shown in Figure 3.

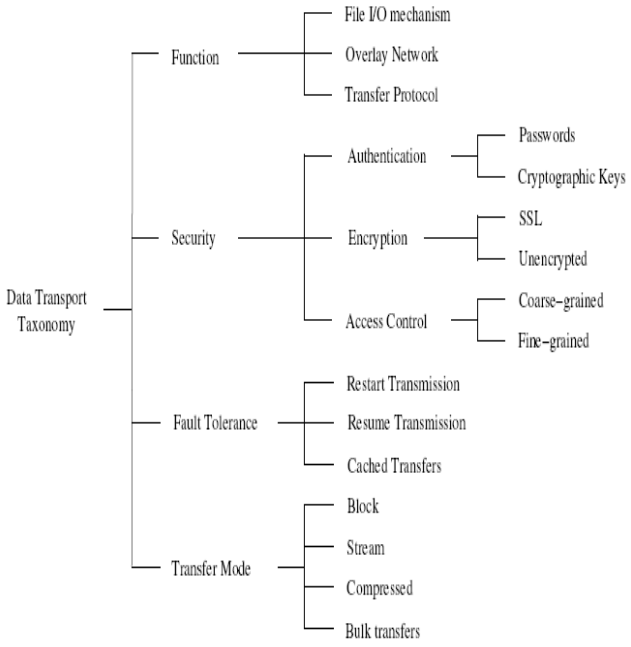

Figure 3: Data Transport Taxonomy

Data Replication and Storage

The important elements of a replication mechanism are therefore the architecture of the system and the strategy followed for replication. The rst categorization of Data Grid replication is therefore, based on these properties as is shown in Figure 4. The architecture of a replication mechanism can be further subdivided into the categories shown in Figure 5.

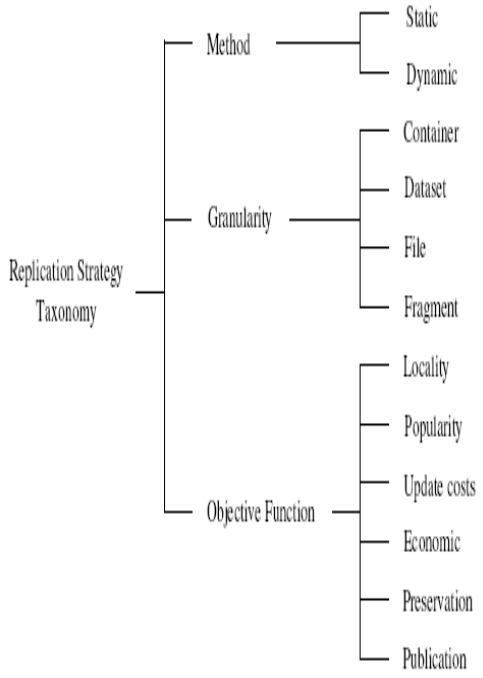

Figure 4: Replication Strategy Taxonomy.

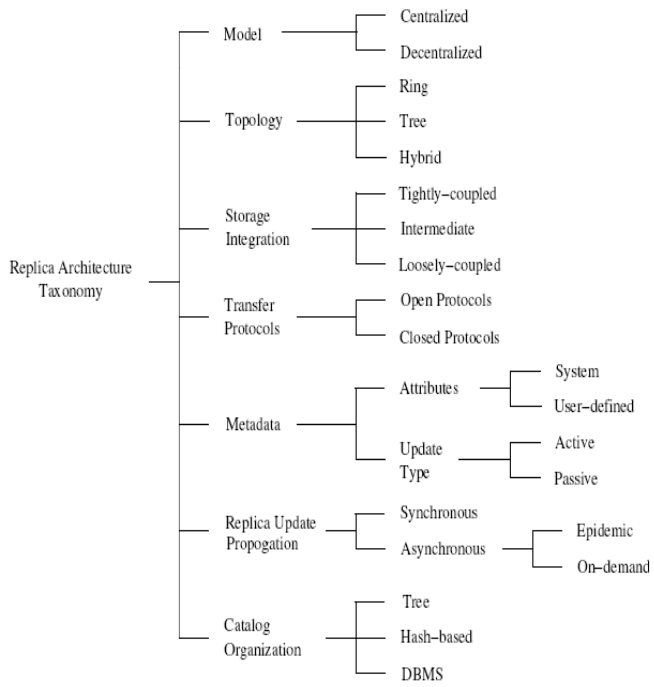

Figure 5: Replica Architecture Taxonomy.

Resource Allocation and Scheduling

The requirements for large datasets and the presence of multiple replicas of these datasets scattered at geographically-distributed locations makes scheduling of data-intensive jobs different from that of computational jobs. Schedulers have to take into account the bandwidth availability and the latency of transfer between a computational node to which a job is going to be submitted and the storage resource(s) from which the data required is to be retrieved. Therefore, the scheduler needs to be aware of any replicas close to the point of computation and if the replication is coupled to the scheduling, then create a new copy of the data. A taxonomy for scheduling of data-intensive applications is shown in Figure 6.

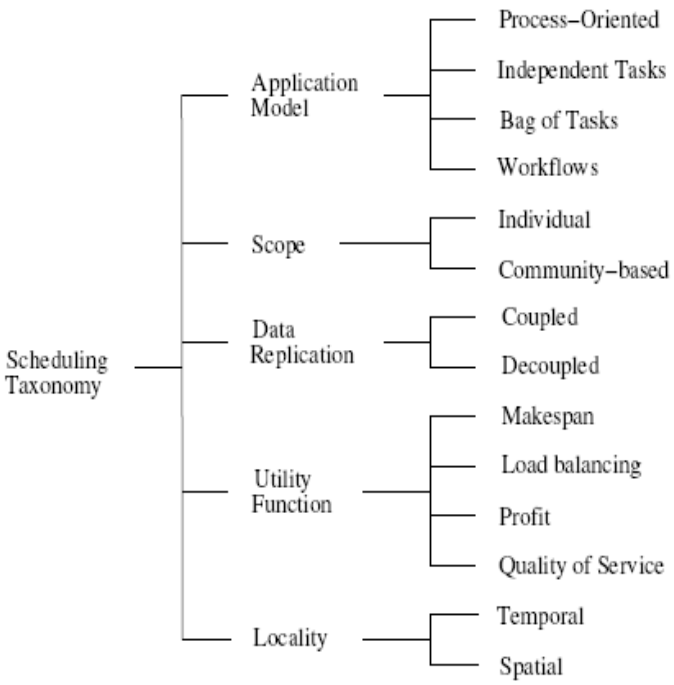

Figure 6: Data Grid Scheduling Taxonomy.

Dual Layer Mechanism for providing security:

Layer 1: Application Layer (GUI) → User Id / password Authentication

User Can View Only his Own Uploaded Files

Layer 2: Data Grid Level Security → Cryptography Techniques for Data Security. Automatic Encryption / Decryption

This dual layer is maintained by the application itself without user interference. As a preliminary measure to restrict unauthorized access we propose layer-1 and for data security on grids we propose encryption of data. Layer-1 is proposed when user starts the system and tries to log in and layer-2 is proposed when user uploads his file to the server. In case if we won't use layer-1 then anybody may get entry to the system access and download any file also if we won't use layer-2 then any hackers or administrators may acquire information from the storage. After applying dual layer mechanism we can achieve following objectives

- To ensure privacy of data in data grids so that right person gets right access to the information.
- To provide dual layer security mechanism i.e. first the login authentication and second cryptography technique with fragmentation of the uploaded data file as per its size.
- To provide a real time secure channel to shield secrecy of information from people with administrative rights.
- To present an easy structural design for systematizing key organization for durable storage of encoded information.
- To present an admission control technique for erasure based coding and privacy of keys depending on ontological privileges.

VI. TASKS INVOLVED IN DATA GRID SECURITY

The implementation can be divided in to four sub task

• Network Module
• Dynamic randomization process
• Secure data share
• Replication data grids

Network Module

Client-server computing or networking is a distributed application architecture that partitions tasks or workloads between service providers (servers) and service requesters, called clients. Often clients and servers operate over a computer network on separate hardware. A server machine is a high-performance host that is running one or more server programs which share its resources with clients. A client also shares any of its resources; Clients therefore initiate communication sessions with servers which await (listen to) incoming requests.

Dynamic Randomization Process

The delivery of a packet with the destination at a node in order to minimize the probability that packets are eavesdropped over a specific link, a randomization process for packet deliveries, in this process, the previous next-hop for the source node s is identified in the first step of the process. Then, the process randomly picks up a neighboring node as the next hop for the current packet transmission. The exclusion for the next hop selection avoids transmitting two consecutive packets in the same link, and the randomized pickup prevents attackers from easily predicting routing paths for the coming transmitted packets.

Secure data share

Secure data partitioning (both secret sharing and erasure coding) and dynamic replication in data grids, in which security and data access performance are critical issues. More specifically, we investigate the problem of optimal allocation of sensitive data objects that are partitioned by using secret sharing scheme or erasure coding scheme and/or replicated. We consider achieving secure, survivable, and high performance data storage in data grids. To facilitate scalability, we model the peer-to-peer data grid as a topology.

Our goal is to replicate the data shares and allocate them to different nodes in the data grid to minimizing cost. We decompose the allocation problem into two sub problems— intra cluster and inter cluster share allocation problems—and deal with them separately and independently.

Replication Data Grids

Data grid is a distributed computing architecture that integrates a large number of data and computing resources into a single virtual data management system. It enables the sharing and coordinated use of data from various resources and provides various services to fit the needs of high performance distributed and data-intensive computing. Dynamic replication to achieve data survivability, security, and access performance in data grids. The replicas of the partitioned data need to be properly allocated to achieve the actual performance gains. We have developed algorithms to allocate correlated data shares in large-scale peer-to-peer data grids Replication is a natural solution for reducing the communication cost (as we have discussed) as well as sharing the access load. In peer-to-peer data grids, replica can be placed on widely distributed nodes to achieve better access performance and load sharing. In cluster-based data grids, caching data on widely distributed nodes is necessary (in addition to replication on cluster nodes) to achieve improved access performance and load sharing. Data partitioning can contribute to reduced storage cost. It has been shown that erasure coding-based schemes can greatly reduce the overall storage cost and effectively share the storage consumption.

VII. SYSTEM DESIGN

System Architecture:

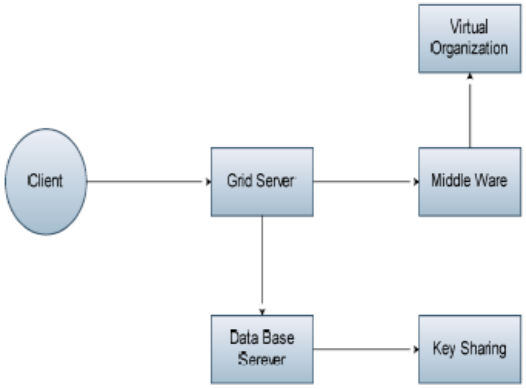

Figure 7. System Architecture

The client can secure the data through the grid server and it will be stored in data grid by data fragmentation and data replication concept, and also can access the data from grid server.

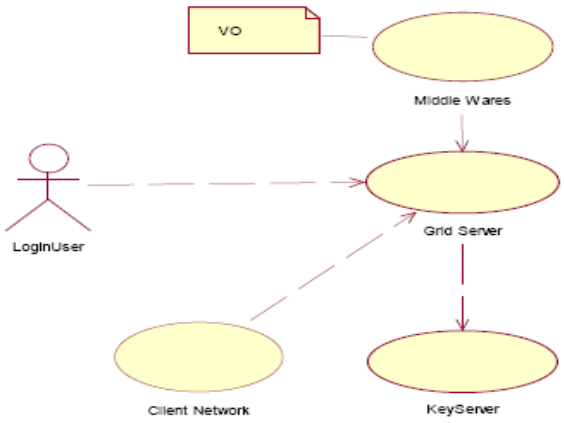

Figure 8. Usecase diagram

Data flow diagram:

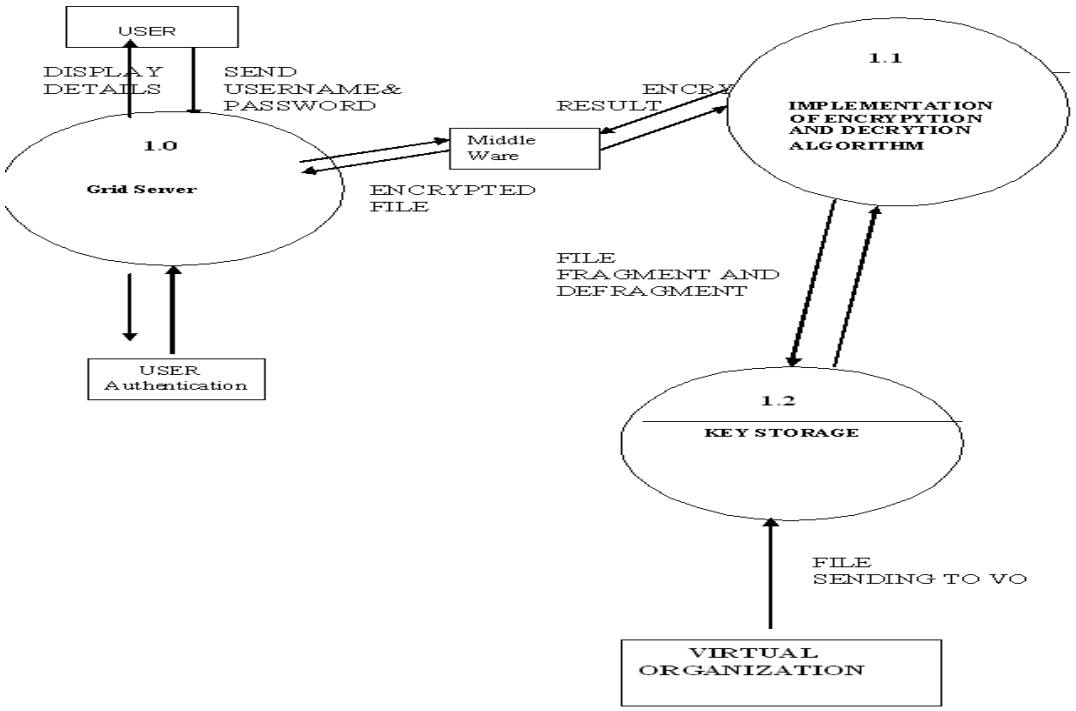

Figure 9. Data Flow Diagram

VIII. CONCLUSION

Grid computing is a modern concept that not just speeds up computing and cut costs, but causes a paradigm shift in computing. However, several challenges still weigh down the technology. Resolving security problems with grid computing is one such major challenge. It requires an adequate understanding of both the security issues in grid computing implementation as well as the solutions presently available to address these. This paper addresses the security needs of both resource consumer and resource provider. This analysis on the one hand will help to increase the reliability in grid computing and on the other hand will lead to develop new applications based on grid computing.

REFERENCES

[1] A. Chervenak, E. Deelman, I. Foster, L. Guy, W. Hoschek, C. Kesselman, P. Kunszt, M. Ripeanu, B. Schwartzkopf, H. Stockinger, and B. Tierney, "Giggle: A Framework for Constructing Scalable Replica Location Services," Proc. ACM/IEEE Conf. Supercomputing (SC), 2002.

[2] A.S.Syed Navaz, C.Prabhadevi, V.Sangeetha, Data Grid Concepts for Data Security in Distributed Computing, International Journal of Computer Applications (0975 – 8887) Volume 61– No.13, January 2013, 6-11.

[3] Baker, Mark, Rajkumar Buyya, and Domenico Laforenza. "Grids and Grid technologies for wide-area distributed computing." Software: Practice and Experience 32, no. 15 (2002): 1437-1466.

[4] Foster, Ian, Carl Kesselman, and Steven Tuecke. "The anatomy of the grid: Enabling scalable virtual organizations." International journal of high performance computing applications 15, no. 3 (2001): 200-222.

[5] Global Information Grid, Wikipedia.

[6] H. Stockinger, "Distributed Database Management Systems and the Data Grids," Proc. 18th IEEE Symp. Mass Storage Systems, 2001.

[7] I. Foster and A. Lamnitche, "On Death, Taxes, and Convergence of Peer-to-Peer and Grid Computing," Proc. Second Int'l Workshop Peer-to-Peer Systems (IPTPS), 2003.

[8] K. Ranganathan and I. Foster, "Identifying Dynamic Replication Strategies for a High Performance Data Grid," Proc. Second Int'l Workshop Grid Computing, 2001.

[9] Karonis, Nicholas T., Brian Toonen, and Ian Foster. "MPICH-G2: a Grid-enabled implementation of the Message Passing Interface." Journal of Parallel and Distributed Computing 63, no. 5 (2003): 551-563.

[10] Krauter, Klaus, Rajkumar Buyya, and Muthucumaru Maheswaran. "A taxonomy and survey of grid resource management systems for distributed computing." Software: Practice and Experience 32, no. 2 (2002): 135-164.

[11] Krauter, Klaus, Rajkumar Buyya, and Muthucumaru Maheswaran. "A taxonomy and survey of grid resource management systems for distributed computing." Software: Practice and Experience 32, no. 2 (2002): 135-164.

[12] L. Xiao, I. Yen, Y. Zhang, and F. Bastani, "Evaluating Dependable Distributed Storage Systems," Proc. Int'l Conf. Parallel and Distributed Processing Techniques and Applications (PDPTA), 2007.

[13] L. Xiao, I. Yen, Y. Zhang, and F. Bastani, "Evaluating Dependable Distributed Storage Systems," Proc. Int'l Conf. Parallel and Distributed Processing Techniques and Applications (PDPTA), 2007.

[14] M. Baker, R. Buyya, and D. Laforenza, "Grids and Grid Technology for Wide-Area Distributed Computing," Software- Practice and Experience, 2002.

[15] M. Tu, "A Data Management Framework for Secure and Dependable Data Grid," PhD dissertation, Univ. of Texas at Dallas, http://www.utdallas.edu/~tumh2000/ref/Thesis-Tu.pdf, July 2006.

[16] Mohd Samsu Sajat, Suhaidi Hassan, Adi Affandi Ahmad, Ali Yusny Daud, Amran Ahmad, Implementing a Secure Academic Grid System – A Malaysian Case, Proceedings of the 10th Australian Information Security Management Conference, Novotel Langley Hotel, Perth, Western Australia, 3rd-5th December, 2012.

[17] Montagnat, Johan, Ákos Frohner, Daniel Jouvenot, Christophe Pera, Peter Kunszt, Birger Koblitz, Nuno Santos et al. "A secure grid medical data manager interfaced to the glite middleware." Journal of Grid Computing 6, no. 1 (2008): 45-59.

[18] www.globus.org, 2008.

