Grid and Data Analyzing and Security

Fatemeh SHOKRI

B.A Software Computer Engineering, Computer Engineering Department, Mazandaran Institute of Technology, Mazandaran, Iran.
fatemeh.shokri@aol.com

This paper examines the importance of secure structures in the process of analyzing and distributing information with aid of Grid-based technologies. The advent of distributed network has provided many practical opportunities for detecting and recording the time of events, and made efforts to identify the events and solve problems of storing information such as being up-to-date and documented. In this regard, the data distribution systems in a network environment should be accurate. As a consequence, a series of continuous and updated data must be at hand. In this case, Grid is the best answer to use data and resource of organizations by common processing.
Keywords*: Grid Computing, Secure Structure, Common Processing*

1 Introduction

As it stands, there are lots of researches have been done regarding using calculation systems based on computer networks. According to their importance and their effects on different aspects of calculation systems, in this regards several researches have been done. In these studies different dimensions of them were studied such as: Evaluation of Job-Scheduling Strategies for Grid Computing (Hamscher et al, 2000), UNICORE: A Grid Computing Environment (Erwin, 2001), Intrinsic vulnerability assessment of the aquifer in the Riana spring catchment by the method SINTACS (Janza and Prestor, 2002), Economic models for resource management and scheduling in Grid computing (Buyyal, 2003), Grid Computing: A Brief Technology Analysis (Smith, 2004), Trusted Grid Computing with Security Binding and Trust Integration (Song, Hwang and Kwok, 2005), Scheduling Algorithms for Grid Computing: State of the Art and Open Problems (Dong and Akl,2006), Introducing Virtual Private Overlay Network services in large scale Grid infrastructures (Palmieri, 2007), Implementation of Computational Grid Services in Enterprise Grid Environments (Richard, Joshi and Eswaran, 2008), Reliability in grid computing Systems (Dabrowski, 2009), Reliable Job Scheduler using RFOH in Grid Computing (Mohammad Khanli,Etminan Far and Ghaffari, 2010), Trust Based Authorization Framework for Grid Services (Singh, 2011), A Time-minimization Dynamic Job Grouping-based Scheduling in Grid Computing (Mishra, Mohanty and Mund, 2012).

Due to the expansion of electronic services that are available to share resources in a virtual operating environment, security and trust are important to the parties participating in this environment. Therefore, for identification and understanding of the environment, we should examine the structure of these environments. By expanding various structures such as integrated data environments, data bases virtual networks, Private networks, wide-area networks, and increasing the range of information sharing in a virtual domain, the need to examine and audit the environment in order to gain trust - (either participant) is recommended. If the users and participants

in virtual environments are demanding an independent unit that has sufficient knowledge of the system and can investigate, review this area and participants should be provided with the strengths and weaknesses of safety control structures.

According to the necessity to audit this environment, the auditor's understanding is the first steps to study and review. Required knowledge is understanding of Mentioned environment. Auditor's understanding of the environment in which helps him to deliver efficient and effective investigation. To properly audit each area of the tool, it must be used in accordance with the environment and must be aware of the work environment. The amount of investigations required for each stage depends on the auditor's previous knowledge of the activities and structure and the system of the environment. If the auditor has sufficient and comprehensive knowledge about Environmental audit, the audit procedures would be a more specific, investigation time reduces and reporting speed increases. In this context, this paper examines the structures of the Grid environment and safety grid types, how to establish a secure connection on the grid, the grid safety standards, safety rules on a shared processing environment.

What is the Grid ?

Grid is said for system that is used for the management and integration of distributed resources and services within Domain Controllers. Grid in virtual organization in comparison with private groups, dependent resources and services has a common goal. To meet the needs of information integration and job control management among virtual organizations has been created.**(**Von Welch et al, 2005)

Grid computing or networks connected to the computer, is a new network model that performs massive computing by using attached processors. Grid uses resource of computers that are connected to networks and can do complex computing with resultant force of these resources. They do this by splitting the operation and consigning the piece to the computer in network to do. (Shahcheraghi and Ahmadinia, 2011a)
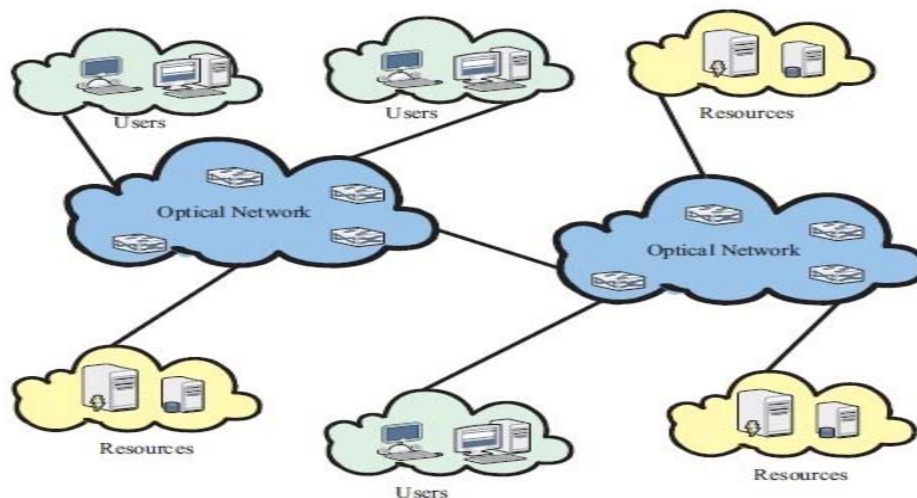

Fig 1. An Example of Grid computing networks

Grid computing networks proposed in the late 1990s as a replacement for conventional supercomputers emerged to address specific problems that require numerical computation and access to large volumes of data that had been distributed.

The main idea was that the networks that are fast enough and using appropriate software, numerous research groups that were

geographically dispersed, could share Computing resources and data management resources in a single system. As result, the system is able to cope with the issues involved with each of these groups could not deal with. (Shahcheraghi and Ahmadinia, 2011c)

Grid computing is a hardware or software structure which provides a range of reliable, stable access to the features and capabilities of off unused or excess of the current needs of the participants in the system. (Foster, 2002)
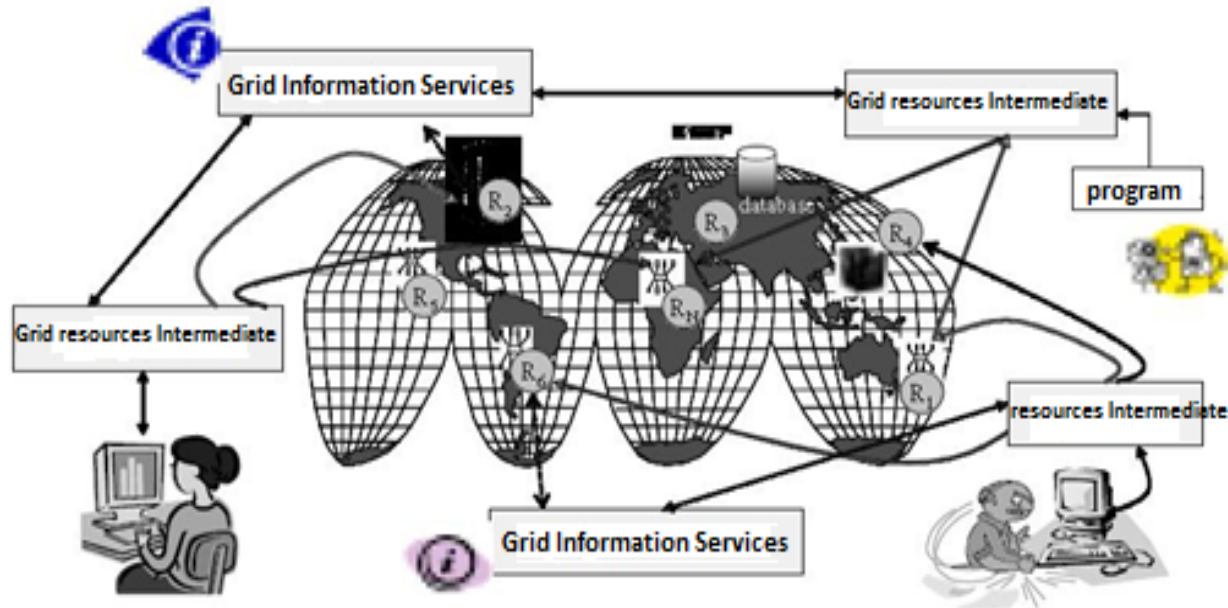

Fig 2. the grid virtualizes geographically disperse resources

In this regard, grid dynamically gathers data Sources from virtual organization (VO) where a service provider that manages and display resource integrally.

In addition, it ensures safety of data in these organizations.

Of a project that can use a Grid infrastructure as follows:

1. **SETI@home** is One of the most popular processing cycles network used 3 million computer to obtain information chain.

2. Another known project is"**distributed.net**" Which began in 1997, 1100 successful projects in distributed information can be found in its records.

3. Advanced Computer Facility at NASA, in the heritability of cognitive algorithms, a processor cycle, which

employs the condor, runs nearly 350 workstation of solar and SGL.

4. Up to 27 April 2007, plan to integrate the information in organizations, created based on the mp network and product of processes cycle personal computers is **pcs**, which connected one million the database to each other.

5. The project "**Enabling Grids for E-science**" that is run by the Europe Union, including a connection to science centers in Europe and United States in order to integrate information in scientific projects, use the capacity of useless of the existing scientific resources .Cases which are represented are only a sample of Grid use in the world.

Therefore, the grid uses safety models like GT to create a safe zone. In the following, we examined part of this safety model.

GT2 Security Model for Grid Resource Allocation and Management (GRAM), Monitoring and Understanding (MDS) and is responsible for transferring data from protocol FTP file transfer.

The security model uses the grid security infrastructure (GSI) and the GSI uses three basic elements in its safety mechanism.

The three basic elements should be mutually trusted in formation of a virtual domain

To meet the needs of users in the Grid environment by using Combination of the dynamic style,

Have important roles, as follows:

1. Multiple safety mechanisms for participatory organizations or environments:
   Often it is spoken the important investment in the creation of a mechanism or structure for optimum safety, overshadowed performance some group to be able to create a stronger environment for sharing resources.
2. Creating a dynamic services: Users should create new services (according to sources) dynamically and non-intervention of the controller.

These services should participatory (jointly) and have safe interaction with the other participants. These services need to be done that are not inconsistent with the methods of local control.

3. Creating dynamic and trusted domain: In order to share resources in the virtual domains, should be a multilateral relationship between resources users and participants. The trust must create between members of the participants, It leads to greater trust between participants and biased behavior reduce.
   For the relationships between participants in a virtual organization to be safe,

The relationship of the organization (which forms the foundation of any organization) should be identified and defined. Gt3 is an enhanced version of the GT2 model that allows Programmers and users to work automatically on the Grid. And it has covered the failure and security breaches of previous model.

In line with the standards, two mechanisms and safety policies in main organizations are required:

1. safe technologies based on the use of virtual organizations as a bridge between the participants in the information-sharing environment:
   Research results show that the system uses widely software, In addition to consideration of the mutual benefits, it prevents from side effect and a way-use.
2. Grid security structure:
   Grid security structure has been established to route and support connection of application systems. Definition of part of the extensive services structure and other elements that has contributed to the new competitiveness for achieving safety opportunity, the following is stated. In this regard, safety standards after the information technology like Safety service structure after the integration of GSI and the open grid Services structure OGSA is provided. Using techniques for compiling and expressing as a authorized application way, is expanding.

Safety structure in virtual information environments:

Since the in the virtual information organization, the accumulation and sharing of information would be dealt with. The safety of such a range should be considered as the most important components. Safety Such an environment has secondary components such as authentication,

authorization, and encryption information. In order to better understand the structure of information used in virtual environments,,

According to (1), three areas where safety is established in the environment, are described in the following:
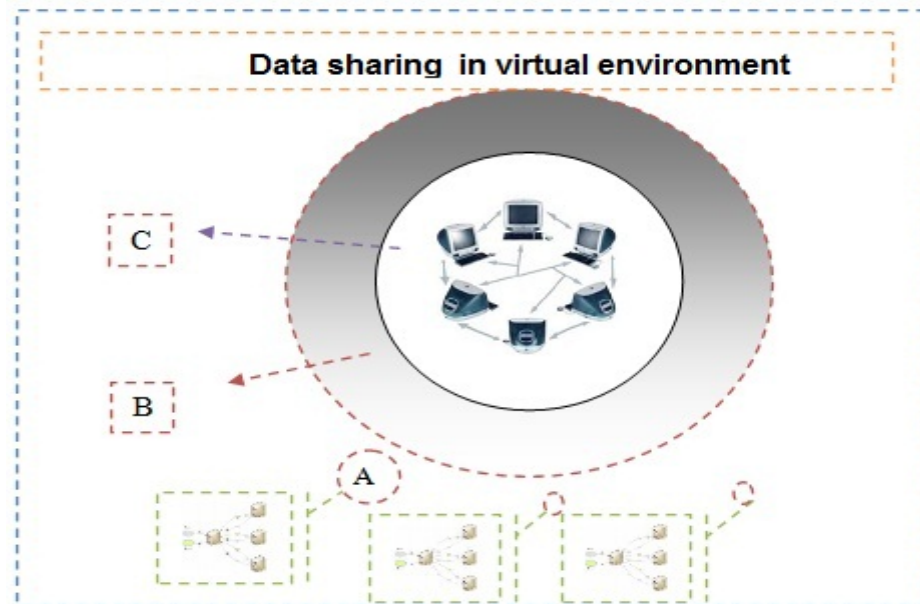

Fig 3. Review of data sharing in virtual environment

1.

ection (A): the set of regulations, restrictions and permits to enter the virtual organization (or range) that Participants (prior to entering corporate area) should sign a security agreement to acquire the licenses. In this area investigate the validity and history participating, and How to provide information and resources will be examined.

2.

ection B)): In this section the control of hardware and software such as Fire walls, crossing servers, and identification ... (To prevent unauthorized entry or exit of non-immunized) are used It should be noted that in addition to controlling the input

and output ofSthe system control of the credibility of the information apply.

3.

ection (C), within the range of participation: Host environment provides, the safety agency that have the applicability of the developed environment, This means that determine illegal information with non-optimal levels of performance, some level of participation, How to access the participatory Sange. However, the safety information structure have specific and exact standard for virtual systems.

These structures are generally focused on two issues.

A)

y default, insists on establishing communication based on the mutual permissions from the secure lines which

are mutual trust. It put forward The role of the safety certificate which determines crucial condition for communication

B)

ritten access agreement to subject of participation: it provides a simple consensus to achieve goals and use safety protocols such as Web Services Description Language, Multilateral relations with web services that communication between units in different work stations accordance with the specific rules of each service and an exchange of messages, Including items discussed in information safety structure. Some of these mechanisms are described below:

1.

eb Services Description Language (WSDL): It is based on business reporting language that is used to describe Web services. Services as the license from network endpoints or ports are defined, and provide a specific structure for the transcripts according to the respective purpose. This language uses a summary of the constantly and relevant messages to Contact in order to be specified Minimum standards for the use of information in the network.

2.

ommunication Lines: Communication Lines Between subscriber's aspect of validity, Safety and reliability must be examined in terms of proper communication.   Communication between the subscribers in this environment (despite the signed agreement) May be Non-secure,

unauthorized and misuse resources in the cooperative organization.

Therefore, the grid safety standards have been raised that the two principal follows.

•

 An open grid service architecture (OGSA) this specification in 2002 is suggested by SETI @ home, the following:

A)

y default, establishing communication based on the mutual permissions which are the secure lines and mutual trust. But not ignored the role of certificate and safety lines that provides condition for secure communication.

W

B)

imple   Object   Access Protocol   (SOAP):
Building a simple environment to achieve the objectives of the agreement and the use of safety protocol such the Web Services Description Language, Multilateral relations Web services which provide relation between units in different workstations tailored to the specific rules of each service and an exchange of messages.

Evaluation and audit Participatory range
Due to the safety issues that have been studied, a reliable control environment, the auditor's opinion will influence the audit test.
System where has High capacity for security and control, Data collection and analysis for the auditor to be more specific; however, investigation of the safety structures such a Participatory environment, the auditor will

needs to follow up its work, In this environment, the auditor is unable to use the traditional methods of auditing, System which is in used, has a structure that is felt, the audit should be quite familiar with the modern technology and Their structures.

Mode of investigation:
To evaluate the structure, audit examines the safety sectors B, A (figure A),
And study Safety agreement in the safety sector A, consult with The expert system and the company's lawyer, Login licenses as An example of how authentication can be used to test participants.

After investigating first safety layer and ensuring the accuracy of the Agreement, Legal gaps and safety in agreement and accuracy of entrance permits to range participation, Auditor investigates hardware and software structures of participation range and control in terms of quality and quantity. After ensuring hardware and software controls providing the audit's positive comment, third phase is evaluated.

At this stage, the safety mechanisms within the participatory range, addition to applied policies and procedures, need to make sure of the policies and controls. So, the best way to evaluate these controls is using Special audit tests. Because in this tests, auditor introduces himself as a participant to the system and becomes aware of safety rules in the system, and can study and report the strengths and weaknesses in safety range mechanisms within the organization.

Information subscription license services review:
According to figure 3, the auditor has reviewed the CAS, with the use of embedded systems in the database, mode of communication, access Levels to information and authorization on this service, can monitor and recounts error. Considering in the information participation range and resources, a self-adaptive system can be used to increase security. To evaluate the control method, Auditor uses special methods such as artificial intelligence in the evaluation and investigation of the safety structure.

Figure 3 illustrates the process of receiving, storing, indexing and analyzing data. As it can be seen, users can get result of analyzing data in system via their own software and computers and use them for their purposes.
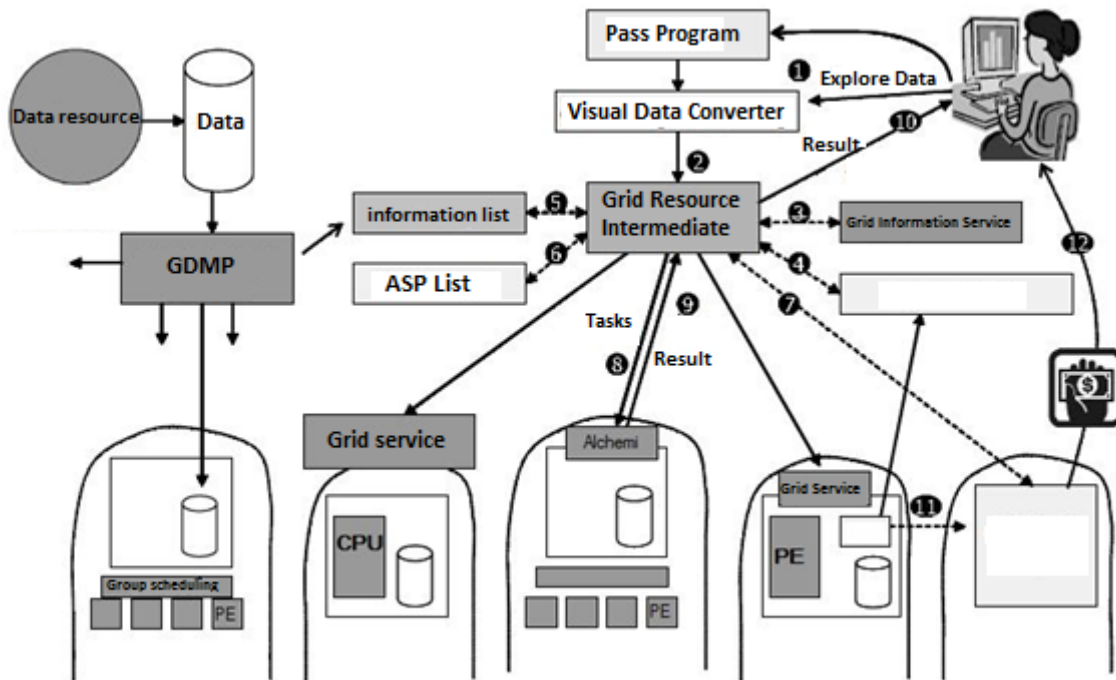
Fig 4. Grid technologies and analyze data distribution

Cooperative computing in grid:

Infrastructure and determining the type of network is data processing methods in computer networks. In other words, the method of processing data in a network is based on computer networks. Network data processing methods are done in three ways: centralized, decentralized, and shared. In Centralized processing, all processing is done at a central computer, and it is suitable for organizations with a centralized structure that requires large computer (in terms of computing and processing power). Decentralized processing is done on personal computers. And instead of using a shared central computer, any user can process information with aid of their personal computer. Difference between centralized and decentralized processing is that the data processing is done in users' computers or terminals. And it creates hardware distribution in organization.(Shahcheraghi and Ahmadinia, 2011b).

In Participatory range of grid systems, processing could be decentralized or distributed, and participants are able to process information jointly (In this environment, a participant can process with use of information from other participants.

This process has the following features:
  ❖ cooperation and resource sharing in processing
  ❖ a continuous network for exchanging sources information between Participants
  ❖ the creation of shared files for cooperative processing
  ❖ a secure environment for processing, without a processing intermediary or handler

Group management Grid system:

The management team make in an important contribution in creating a resource sharing database, Information transfer facilities and systems, developing Local parallel methods of data transmission and combining resources.

This group causes centralized control system that has a complete knowledge of the system states, User demand and exact and comprehensive control on components of participants in the system. This group is fully aware of the protocol goals and safety agreements to support the system. Determining location and time of the participant's access to information in the range is duties of this group. The considerable quality of these services is established in the participatory range.

Conclusions

Grid can be considered as a new experience in virtual research that requires extensive data management. In this article, we review a part of this process. But since Grid is new structures and concepts; for this reason more research is needed in this regard to evaluate more features of this structure. Due to Grid, managers are facing the new opportunities. Therefore, it is important that we continuously assess these systems.

In this regard, I put forward some researches which have done based on Grid technology.

Table 1: A summary of the most important researches on Grid from 2006 up to 2012

Date	Title	Authors	Result
2012	A Time-minimization Dynamic Job Grouping-based Scheduling in Grid Computing	Manoj Kumar Mishra, PrithvirajMohanty, G. B. Mund	This paper proposes "A Time-Minimization Dynamic Grouping-Based Job Scheduling in Grid Computing" with the objective of minimizing overhead time and computation time, thus reducing overall processing time of jobs.
2011	Trust Based Authorization Framework for Grid Services	Sarbjeet Singh	This paper describes different facets associated with trust issues among different entities in a grid environment and proposes a trust model to establish and manage trust relationships
2010	Scientific Data Sharing Using Clustered-Based Data Sharing in Grid Environment	1Rohaya Latip, 2Hamidah Ibrahim and 3Feras Ahmad Al-Hanandeh	In this study, we introduced a new protocol, named Clustered-based Data Sharing (CDS) for data sharing in a large dynamic network such as grid computing by using Clustered-based techniques to improve the accessibility.
2009	Reliability in grid computing systems	Christopher Dabrowski	This study surveys work on grid reliability that has been done in recent years.
2008	A Multi-Agent Architecture		This paper proposed a multi-agent architecture that addressed resource management and application execution with

	for QoS Support in Grid Environment	Ali Rezaee, MasoudRahmani, SaeedParsa, SaharAdabi	support for Quality of Services (QoS) in grid environment.
2007	Introducing Virtual Private Overlay Network services in large scale Grid infrastructures	Francesco Palmieri	In this paper, we propose a novel network resource abstraction for delivering dynamic on-demand Virtual Private Overlay connection services, into large-scale Grid environments.
2006	Implementation of Load Balancing Algorithm in a Grid Computing	1Abdallah Boukerram and 2Samira AitKaciAzzou	This paper describes the complete Implementation of an algorithm of load balancing in an environment of grid computing. The implementation of the algorithm is realized on a cluster of processors in a logic of portability on grids.

References

[1]AlexX.Liu
//www.cse.msu.edu/~alexliu2132 , Engineer Building Department of Computer Science and Engineering ,Michigan State University

[2] Christopher Dabrowsk, 2009, Concurrency and Computation: Practice & Experience - A Special Issue from the Open Grid Forum, Volume 21, Issue 8,Pages 927-959, doi:10.1002/cpe.v21:8

[3] Dietmar Erwin, 2001, UNICORE - A Grid Computing Environment, Uniformes Interface für Computing Ressourcen(Final report – in German) http://www.unicore.org.

[4] Fangpeng Dong and Selim G. Akl, 2006, Scheduling Algorithms for Grid Computing:State of the Art and Open Problems, Technical Report No. 2006-504, doi=10.1.1.69.3660

[5] Foster, Ian, What is the Grid? A Three Point Checklist, Argonne National Laboratory & University of Chicago, July 20, 2002

[6] Francesco Palmieri, 2007, Introducing Virtual Private Overlay Network Services in Large Scale Grid Infrastructures, Journal of Computers, Vol 2, No 2 (2007), 61-72, Apr 2007, doi:10.4304/jcp.2.2.61-72

[7] Leyli Mohammad Khanli, Maryam Etminan Far, Ali Ghaffari, 2010, Reliable Job Scheduler Using RFOH in Grid Computing, Journal of Emerging Trends in Computing and Information Sciences, Vol. 1, No. 1, pp43-47.

[8] Manoj Kumar Mishra, PrithvirajMohanty and G B Mund. Article: A Time-Minimization Dynamic Job Grouping-based Scheduling in Grid Computing, 2012,International Journal of Computer Applications 40(16):16-25.doi: 10.5120/5064-7419.

[9] MitjaJANzA&Joerg PRESTOR, 2002, Intrinsic vulnerability assessment of the aquifer in the Ri`ana spring chatcmentby the method SINTACS, GEOLOGIJA 45/2, 401–

406, Ljubljana, doi:10.5474/geologija.2002.039

[10] R.J.A. Richard, Ajay A. Joshi and C. Eswaran, 2008, Implementation of Computational Grid Services in Enterprise Grid Environments, American Journal of Applied Sciences, Volume 5, Issue 11,Pages 1442-1447, DOI: 10.3844/ajassp.2008.1442.1447

[11] RajkumarBuyya, David Abramson, Jonathan Giddy and Heinz Stockinger, 2003, Economic models for resource management and scheduling in Grid computing, Concurrency and Computation: Practice and ExperienceSpecial Issue: Grid Computing Environments, Volume 14, Issue 13-15, pages 1507–1542, DOI: 10.1002/cpe.690

[12] Roger Smith, 2004, Grid Computing: A Brief Technology Analysis, CTOnet.org.

[13] Sarbjeet Singh, 2011, Trust Based Authorization Framework for Grid Services, Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, No. 3, pp136-144.

[14] Shahcheraghi, Azinsadat and Ahmadinia, Hamed, Managing Integration Data in Grid (February 9, 2011). Rayaneh Magazine, Vol. 21, pp. 68-72, 2011. Available at SSRN: http://ssrn.com/abstract=2088074

[15] Shahcheraghi, Azinsadat and Ahmadinia, Hamed, Survey Sharing Information in Dynamic Information Environments - Part 1 (September 23, 2011). Ecommerce and Computer Magazine, Part 1, Vol. 9, No. 59, pp. 45-49, 2011. Available at SSRN: http://ssrn.com/abstract=1968066

[16] Shahcheraghi, Azinsadat and Ahmadinia, Hamed, Survey Sharing Information in Dynamic Information Environments - Part 2 (September 23, 2012). Ecommerce and Computer Magazine, Part 2, Vol. 9, No. 60, pp. 45-49, 2011. Available at SSRN: http://ssrn.com/abstract=2079970

[17] Shanshan Song, Kai Hwang , Yu-kwong Kwok, 2005, Trusted Grid Computing with Security Binding and Trust Integration, Journal of Grid Computing, DOI: 10.1007/s10723-005-5465-x

[18] Volker Hamscher, UweSchwiegelshohn, AchimStreit, RaminYahyapour, 2000, Evaluation of Job-Scheduling Strategies for Grid Computing, Grid Computing — GRID 2000, Lecture Notes in Computer Science Volume 1971, 2000, pp 191-202

[19] Von Welch, Frank Siebenlist, Ian Foster, John Bresnahan, Karl Czajkowski, JarekGawor,Carl Kesselman, Sam Meder, Laura Pearlman , Steven Tuecke, 2005,Security for Grid Services, Universities of Chicago, Southern California & Argonne National Laboratory.