ISSN 2047-3338

Evolution of Grid-GIS Systems

Hari Singh[1] and Seema Bawa[2]
[1]N.C. College of Engineering, Panipat, Haryana, India
[2]Thapar University, Patiala, Punjab, India
[1]harirawat@rediffmail.com, [2]seema@thapar.edu

***Abstract*— Grid is the collection of heterogeneous resources, providing good opportunities for the users of virtual organizations forming a Grid for coordinated resource sharing and problem sharing. Much Grid middleware software exists for implementing a Grid framework. Apart from many application areas of Grid Computing, Geographic Information System has been an emerging application area for Grid Technology in recent time. In this paper, the developments in the field of Grid-GIS are discussed after a brief overview of Grid Computing and GIS separately. Some observations are also discussed in the concluding section of the paper.**

***Index Terms*— Grid Computing, GIS, Web-GIS and Grid-GIS**

I. INTRODUCTION

IN earlier days when mainframe computers were used for batch-processing, users submit their jobs and wait for the results, but the processing time of mainframes was fully utilized. With the advancement in technology, processors, memory and other computer components became faster and cheaper, which resulted in PCs easily affordable by many users. Now, users were not waiting for processing their jobs rather the processing power of PCs were not being utilized, which introduced the technology of parallel and distributed processing.

Though conventional parallel computing/processing and distributed computing/processing technologies increased the processing speeds of applications but a recent wide area distributed and parallel computing technology called grid middleware has addressed issues such as security, discovery, storage, execution, information, service integration, resource monitoring, failure detection and recovery, quality-of-services and efficient utilization of the waste going processing power of PCs.

Grid middleware is the system software between applications and operating system. Many Grid middleware systems exist like UNICORE, Alchemi, Globus Toolkit, Legion, Gridbus, Condor, etc. These Grid middleware systems provide a grid-computing infrastructure. Many researchers have compared these Grid middlewares. Gridbus, Globus, Legion, and UNICORE are compared on the basis of architecture, implementation model and several other features [14]. A philosophical and technical comparison between Legion and Globus is done [15]. Performance of Alchemi and Globus is compared [10]. Similarities and differences between Grid computing and such distributed computing systems as P2P,

CORBA, Cluster Com puting, and DCE are explained [2]. With the help of Grid middleware sharing of all kinds of resources, including data resources, computing resources, storage resources, information resources, knowledge resources, etc. has become possible.

This paper describes the evolution of Grid-GIS systems. Section II describes the Grid Computing Technology and the underlying architecture. Section III describes the Geographic Information System (GIS), accessing data from GIS databases, problems of GIS databases such as heterogeneity and interoperability, and spatial data handling language, Geographic Markup Language (GML). Section IV describes the GIS data accessing and processing through Web. Section V describes the integration of Grid Technology and the GIS. Section VI describes the conclusions.

II. GRID COMPUTING

Grid Computing is about coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. The sharing is not only limited to file exchange but rather direct access to software, computers, and other resources. The sharing is highly controlled, with resource providers and consumers defining clearly and carefully as what is shared, who is allowed to share, and the conditions under which sharing occur. A set of individuals or institutions defined by such sharing rules form a virtual organization [16]. Survey of the major international efforts in developing the state-of-the-art Grid Computing Technology is discussed in [8].

Current Distributed computing technologies do not address the concerns and requirements listed above. For example, Internet technologies address communication and information exchange among computers but do not provide integrated approaches to coordinated use of resources at multiple sites for computation. Current technology either does not accommodate the range of resource types or does not provide the flexibility and control on sharing relationships needed to establish virtual organizations.

A virtual organization is an approach to computing and problem solving based on collaboration in computation and data rich environments. It is a network, linking geographically dispersed agencies with partially overlapping objectives. Each such agency is competent in providing one particular service; Grid exploits the powers of such competent resources and provides the best service to the users. Grid has the solution to the limitations of current distributed computing technologies and it complements these technologies, which can use Grid technologies to achieve resource sharing across institutional boundaries.

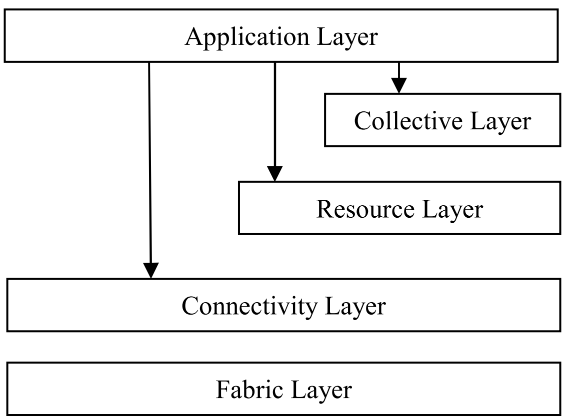

Fig. 1: The Layered Grid Architecture [13]

The architecture of grid computing defines many protocols and APIs between grid nodes to ensure all kinds of operation and application of grid system. The architecture comprises of five layers and each layer has its specific function. From below to upon are basement layer, resource layer, control layer, implementation layer and application layer and is shown in The Layered Grid Architecture, Fig 1 [13]. Components within each layer share common characteristics but can build on capabilities and behaviors provided by any lower layer [16]. Globus is one of the biggest corporation which do their research on grid computing. Globus has constructed Globus Toolkit versions GT2, GT3, GT4 as architecture of grid, which formed the "five-tier sandglass structure [31]. On the basis of five-tier sandglass structure and web services technology, open grid services architecture (OGSA), centered on grid services is presented. Globus Toolkit provides the implementation middleware whose core GT components provide services like security, discovery, storage, execution, information, service integration, resource monitoring, failure detection and recovery, quality-of-services, etc. for Grid Computing [13]. Many Grid resource discovery approaches on the basis of performance factors like scalability, reliability, adaptability and manageability are compared for choosing an appropriate approach to discover a particular resource [27].

Grid-Computing technology for geospatial research and application issues provides significant performance gain at the insignificant cost of network transfer time [20]. Grid technologies relate to other contemporary technologies and Grid concepts and technologies complement and contribute to other technologies [16].

Grids computing provide applications in many disciplines such as Bioinformatics, Financial Modeling, Earthquake, Simulation, and Climate/Weather modeling. Grids offer a way of using the information technology resources optimally inside an organization and can also provide a means for offering information technology as a utility for commercial and noncommercial clients, with those clients paying only for what they use, as with electricity or water. The rapid development of various distributed computing technologies like Internet technology has a profound impact on Geographic Information Systems and these too have been developed rapidly with the progress of computer science, remote sensing, and global positioning system. GIS becomes more and more important in many domains, such as disaster management, military, mapping, and etc. In this paper, application of grid computing on GIS field has been discussed.

III. GEOGRAPHIC INFORMATION SYSTEM (GIS)

A GIS is a computer system capable of assembling, storing, manipulating, analyzing, and displaying geographically referenced information. Hardware, GIS software, and Data are the key components of GIS. Hardware comprises the tools needed to support the many activities of GIS ranging from data collection to data analysis like workstation, digitizer, GPS data logger, Web Server, etc. GIS software is used for creating, editing and analyzing spatial and attributes data like Web GIS. There are two primary types of data that are used in GIS: vector data and raster data. Vector data is spatial data represented as points, lines and polygons. Raster data is cell-based data such as aerial imagery and digital elevation models. Coupled with this data is usually data known as attribute data. Attribute data generally defined as additional information about each spatial feature housed in tabular format. GIS resources (Commercial GISs or Open source GISs [25]) can have many users such as GIS analyst, Web application developer, or simple user.

Under traditional approaches for information access in distributed GIS, data was brought to computation for processing, which unavoidably increases the network burden because in a GIS the spatial data typically has an extremely complex and variable structure and the amount of spatial data is usually large, which involves a great deal of data transfer on the Internet.

The information available in one spatial database may be useful for other GIS application but due to heterogeneity of databases it is difficult for such applications to not only use but also locate useful data. The heterogeneity of spatial information systems has hugely affected the interoperability between different GIS software. Complementary advantages of the different GIS platforms can be fully realized by solving the data sharing and interoperating among heterogeneous platforms. The heterogeneity of existing systems is in term of obscure semantics, diversity of data sets, data modeling concepts, data encoding techniques and storage structures.

GIS Server allows one to share one's GIS resources across an enterprise and across the web or can integrate content from one server with content from other GIS servers. Access to the GIS server is embedded inside the web application and typically hidden from the user of the application. GIS resources are the maps, globes, address locators, geodatabases, and tools that can be shared with others by firstly hosting them on the GIS Server system, and then allowing client applications to use and interact with the resources. The main advantages of sharing your GIS resources on a GIS server are the same as sharing any data through any kind of server technology like the data is centrally managed, supports multiple users, and provides clients with the most up-to-date information. In addition to providing access to a particular GIS resource, the GIS server also provides access to the GIS functionality in the form of interaction that the resource contains [18].

Geographic Markup Language (GML), a subset of Extensible Markup Language (XML) is popular for interoperating the spatial data like GML map making, GML data transformations, spatial queries, and geographic analysis [29]. GML stores and transmits entity's spatial and non-spatial information with XML coding form. It unifies various geo-spatial data to one standard data format with properties: data integrality, extensibility, integration ability, and shape changing ability [9].

IV. Web-GIS

Web-GIS is a Geographic Information System (GIS) distributed across a computer network to integrate, disseminate, and communicate geographic information visually on the World Wide Web. It is constructed on Internet and basically can provide data sharing. It serves a user by software. As a Client-Server application, a client is a Web browser and the server-side consists of a Web server that provides a Web-GIS software program. The client requests for GIS data like a map or some geo-processing over the Web to the remote server. The server translates the request into an internal code and invokes the GIS functions by passing on the request to the Web-GIS software. The software returns the result that is reformatted for interpretation by the client browser application itself or with additional functionality from a plug-in or Java applet. The server then returns the result to the client for display, or sends data and analysis tools to the client for use on the client-side [18].

Using mobile agents in distributed GIS architecture solves the problem of limited communication bandwidth and unstable connectivity faced by the GIS application under Internet environment [5]. There are many good reasons of using mobile agents in the system [28].

Integrating distributed WebGIS by mobile agent and GML technology simplifies the structure of clients and reduces the load of network, and handles heterogeneity [22]. SVG using XSLT stylesheet visualizes GML based spatial data in a browser for a Web-Based GIS solution based on Service Oriented Architecture, for addressing the issue of interoperability [19].

Though Web makes accessing GIS data possible on Internet, it still has some problems and weaknesses which are described in section V.

V. Grid-GIS

The grid computing and web services present the solution to the problems and weaknesses of data distribution of GIS through WebGIS [4]. Aggregating spatial resources based on Grid technology, OGC specification, and Web Service Resource Framework (WSRF) for discovering all the grid services encapsulated results in better resource sharing, higher efficiency, and more cooperative than traditional GIS service [26].

Spatial data sharing and seamless integration becomes much easier and more effective with grid computing, because some complex problems of distributed computing, such as system heterogeneity, security mechanisms and resource dynamic management are hidden and wrapped at the Grid computing level. [6].

Grid-GIS is the integration of Grid Computing and GIS. It is the application of grid computing on GIS field and is constructed on the basis of grid computing environment. It serves a user by means of service and provides universal resource sharing, computing resources, data resources, storage resources, information resources and knowledge resources. Grid-GIS is more extensible and scalable with incremental growth of GIS because there is no need to pre-install any special software. Just connect the computer to the Grid and it becomes the part of Spatial-Grid infrastructure. Grid-GIS can provide abundant computational resources, and also a great potential for large-scale spatial data processing, spatial analysis, spatial statistics applications, etc. [6].

The running structure of Grid-GIS is considered the architecture of Grid-GIS and is shown in Fig. 2. It is made up of five layers according to their main functions, and each layer has its specific functions. From bottom to top are basement layer, resource layer, control layer, implementation layer and application layer. The basement layer provides basic structures and protocols of the network and some other protocols of the grid system, which ensures the information communication between grid middleware and grid system. The resource layer contains many resources, which are mainly made up of local resources and remote resources, which have been registered to resource registry center for sharing. The control layer is the core of the whole system, which ensures the whole system to run correctly. The implementation layer is made up of many middleware, which connects with the system by some specific interfaces. And the application layer is the applications of users based on GridGML [21].

Spatial data related resources are gaining popularity for their increased use in many disciplines. These heterogeneous spatial data resources can be exploited by making them resources of Grid Network. Grid GIS can be used for applications on distributed image processing [21].

Some extensions to the Grid-GIS prototypes such as its integration with mobile agent technology and indexing system for spatial data resources improves the mechanism of spatial data discovery and integration [1, 11]. Mobile agent for real time downloading of spatial information in Grid-GIS environment results in good load balance, high processing efficiency and less network communication [23].

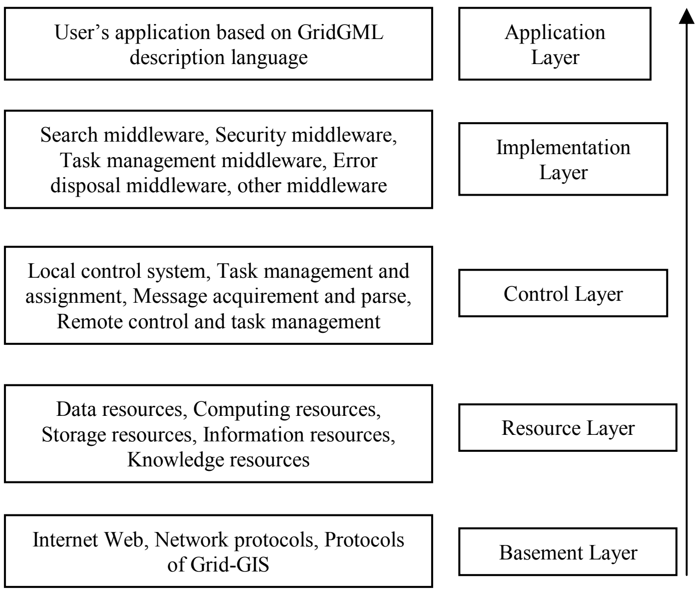

Fig. 2: Architecture of Grid-GIS system [21]

Spatial data index buffer by using R-tree approach provides a method for accessing spatial data effectively in Grid environment [17].

The design of Data Access and Integration (OGSA-DAI) middleware is improved in five aspects: dynamic identification of data source and automatic registration, the speed of data search and data connection through connection buffer memory field between data source and middleware, simplify the document enforcement of middleware through combining client SQL query and the automatic document customization of middleware, access spatial data sources to support direct access, and automatic hand in of updated resource directory from data server to control server [12].

A prototype based on Globus Toolkit 4.0 (for setting the grid environment) and OGSA-DAI service (for spatial data access and integration) is built for proving the architecture and methods for dynamic resource discovery and interoperability [24].

Integrating mobile agent in Grid-GIS solves problems like the lack of interoperability, lack of mechanisms to obtain the balance of system load, and lack of asynchronous computing mechanism of traditional GIS systems, which run in relative low performance status, can not make full use of massive distributed resources, and can not provide flexible services to users [1].

Grid platform helps in solving the GIS interoperability issue when experimented over buffer analysis and polygon merge as two different GIS services to apply on two heterogeneous distributed data sources. The results show that the grid platform not only produces consistent results as compared to using GIS software but it provides a convenient and efficient method. However, the process efficiency is affected for the transfer of processed data among data providers, servers, and clients [3].

The architecture for organizing vector map services built on Globus Toolkit 4.0 and based on Web Service Resource Framework (WSRF) provides a method to amplify the buffering of geospatial information based on WSRF. The virtual organization (VO) consists of five kinds of node: Grid Portal, Certificate Authorization (CA), Domain Manager, Global Resource Catalog Server, Medial Node, and GIS Servers. Grid users submit the global job to Domain Manager through Grid portal. Domain Manager parses the global job to a series of subtasks and transfers these subtasks to corresponding Medial Nodes. Medial Nodes completes their subtasks and merges results in Domain Manager and the final result is transferred to Grid Portal for display. Global Resource Catalog Server manages resource properties to amplify the buffering of geospatial information based on WSRF, which reduces the time of web service invocation on GIS servers, efficiently by Medial Nodes [11].

Three methods namely Message server queue mode, Message combination send mode, and Extended GML method achieves geographical information synchronization in GRID-GIS for the purpose of stability and efficiency of data delivery, especially in unpredictable network environment [7].

A distributed mobile GIS grid model (DMGG) based on mobile agent for real time downloading of spatial information and distributed grid computing in mobile GIS environment has good load balance, high processing efficiency and less network communication [32].

VI. CONCLUSIONS

Grid-GIS Systems have evolved from workstation GIS software to GIS over the web. Earlier Web-GIS Systems made it possible to access and process GIS resources over long distances. Then the integration of Grid computing, an advanced technology with GIS provides advantages over Web-GIS for accessing and processing GIS resources.

Existing researches on Grid-GIS mainly concentrate on the architecture design of Grid GIS, and some conceptual demonstrations. However, few researchers discussed the detailed implementation approach. Not many models exist for improving the integration of Grid and Geographic Information System architecture, some have used the mobile agent technology approach for enhancing the mechanism of spatial data access, retrieval, and processing [1, 23]. The integration of Grid and Mobile Agents over GIS data has not been deeply analyzed for integrated system parameters like dynamic workload estimation for each running agent and dynamic system parameters estimation for each node of the grid in terms of utilization, idle time, memory usage, etc. [1].

Some service discovery frameworks in Grid-GIS systems adopts spatial resource index buffer for decreasing the access time for the spatial resource on GIS server. The failure of the index buffer may increase the searching time for frequently accessed information, which is not desired [1], [11], [17]. In many Grid-GIS systems based on mobile agents, task manager manages the decomposition of a task into sub-tasks and then assemble the results returned by all the sub-tasks. The issue of improving the availability of the task manager becomes an important issue for the consistent operation of the Grid-GIS system [1], [7], [11].

For most of the integrated Grid-GIS models, the need to design a good interface for the interaction of users with the developed system has been suggested for improving the productivity of the user's interaction with the system [4,14,30]. Most of the Grid-GIS systems provide simple GIS services to users; these simple GIS services can be collaborated, and semantically analyzed for providing complex GIS service [24].

In the Geographic Information Grid System (GIGS), a Grid-GIS prototype, the security of mobile agents in the system has been proposed for future research work. Many methods exist for the security of mobile agents but these methods have not been applied over Grid-GIS systems for improving the security of the integrated system [1].

Some of the integrated Grid and Mobile Agent systems have successfully experimented over simple parallel computations such as matrix multiplication of lower orders. The frameworks can be extended to operate for computations having higher complexity and for scalable Grids [8]. Many proposed integrated Grid and GIS systems have successful experimented over simple GIS computations, dealing with vector data. The frameworks can be extended to operate over complex GIS computation, involving raster data and for scalable Grid [3], [9], [12].

REFERENCES

[1] Q. Tong and Z. Shen, "Research of Mobile Agent Technology for GIS-Grid," IEEE Computer Society, Ninth International Conference on Hybrid Intelligent Systems, pp. 190-194, 2009.

[2] Matt Haynos, "Perspective on grid: Grid computing – next-generation distributed computing", http://www. 106.ibm.com/developerworks/library/gr-heritage.

[3] Y. Sun, G. Li, "Interoperability Research of Heterogeneous GIS Based on Spatial Information Grid," IEEE Computer Society, International Conference on Computer Science and Software Engineering, pp. 41-44, 2008.

[4] J. Cheng, W. Li, "Research of the Application of Grid Computing on Geographical Information System," Proceedings of IC-NIDC, IEEE, pp. 1070-1075, 2009.

[5] M. Chen, J. Guan, and X. Huang, "Research on Distributed GIS Based on Mobile Agent," International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, pp.67-73, 2000.

[6] Y. Wang, L. Ge, C. Rizos, and R. Babu, "Spatial Data Sharing on Grid," www. Available online at citeseerx.ist.psu.edu, Accessed on October 2010.

[7] J. Tan, X. Fan, J. Zhu, et al., "Research on Geographical Information Synchronizing in GRID-GIS," International Workshop on Earth Observation and Remote Sensing Applications, IEEE, 2008.

[8] M.Baker et al., "Grids and Grid Technologies for wide-area distributed computing", Software Practie and Experience in press (DOI: 10.1002/spe. 488), 2002.

[9] W. Yaqin, G. Hua, S. Cuiya, and S. Weixing, "Research on Multi-source Heterogeneous Spatial Data Exchange Model Based on Ontology and GML," The Eight International Conference on Electronic Measurement and Instruments, IEEE, Vol. 3, pp. 931–934, 2007.

[10] Md. Ahsan Arefin et. al., " Alchemi Vs Globus: A Performance Comparison," 4th International Conference on Electrical and Computer Engineering, ICECE 2006, 19-21 December 2006, Dhaka, Bangladesh.

[11] Y. Zhao, Y. Fang, B. Chen, M. Yan, and Y. Sun, "A New Grid GIS Prototype for Vector Geospatial Data," Eighth International Conference on Grid and Cooperative Computing, IEEE Computer Society, pp. 367-372, 2009.

[12] X. Liu, X. Huang, and Z. Zhao, "Research on GIS Spatial Database Based on Grid Computing," International conference on Communications and Mobile Computing, IEEE Computer Society, pp. 156-159, 2009.

[13] I.Foster, "Globus Toolkit Version 4: Software for service-oriented systems," J.Comput. Sci. & Technol., Vol. 21, No. 4, pp. 513-520, 2006.

[14] P. Asadzadeh et. al., " Globus Grids and Software Toolkits: A Study of Four Middleware Technologies," http://www.buyya.com/papers/gmchapter.pdf

[15] A.S. Grimshaw et. al., " A philosophical and technical comparison of Legion and Globus," IBM J. RES. & DEV. Vol. 48 No. 2, March 2004.

[16] I.Foster, C. Kesselman and S.Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organiations," Journal of Supercomputer Applications, 15(3), 2001.

[17] Lanzou and Gansu "Spatial Data Resource Index in Grid Environment," Eighth International Conference on Grid and Cooperative Computing, IEEE Computer Society, pp. 116-120, 2009.

[18] T. Horanont, N. K. Tripathi, and V. Raghvan, "A Comparative Assessment of Internet GIS Server Systems," Available online at www.gisdevelopment.net, Accessed on September 2010.

[19] D. Aggarwal "A Robust Service oriented Architecture based solution for GIS Interoperability," http://www.gisdevelopment.net/technology/gis/ma07241.htm

[20] Y. Shi, A. Shortridge, and J. Bartholic, "Grid Computing for Real Time Collaborative Geoprocessing," Symposium on Geospatial Theory, Processing and Applications, Ottawa, 2002.

[21] Z. Shen, J. Luo, et al, " Architecture design of grid GIS and its applications on image processing based on LAN," Information Sciences, Vol. 166, pp. 1-17, 2004.

[22] G. Ji-hong, Z. Shui-geng, B. Fu-ling, and M. Ling-kui, "Building Distributed Web GIS: A Mobile – Agent Based Approach," Wuhan University journal of Natural Sciences Vol. 6, No. 1-2, pp. 474-481, 2001.

[23] T. Gen, "A New Method of Distributed Mobile GIS Grid Model Based on Mobile Agent," International Conference on Computer Application and System Modeling, IEEE, pp. 664-668, 2010.

[24] Q. Yin, Q. Li, X. Li, "Research on Spatial Data Interoperability in Spatial Data Grid Environment," Third International Joint Conference on Computational Science and Optimization, IEEE Computer Society, pp. 351-353, 2010.

[25] R. Lake, "Introduction to GML: Geography Markup Language," Available online at http://www.w3.org/Mobile/posdep/GMLIntroduction.html, Accessed on October 2010.

[26] T.Jing et. al., " Open Source Software Approach for Internet GIS And Its Application," Second International Symposium on Intelligent Information Technology Application, 2008, pp. 264- 267.

[27] A. Sharma and Seema Bawa, " Comparative Analysis of Resource Discovery Approaches in Grid Computing," Journal of Computers, Vol 3 no. 5, May 2008.

[28] D.B. Lange and M.Oshima, "Seven Good Reasons for Mobile Agents," Communications of the ACM, Vol. 42, No. 3, pp. 88-89, 1999.

[29] P. Palathingal, T.Potok, and R. Patton, "Agent Based Approach for Searching, Mining, and Managing Enormous Amounts of Spatial Image Data," Available from http://www.aser.ornl.gov/Publications/FLAIRS05.pdf,

[30] T. Zhang, Q. Qiao, and Q. Wu, " Grid-enabled High Performance GIS Web Portals," International Joint Conferenceon Computational Sciences and Optimization, IEEE Computer Society, pp. 187-191, 2009.

[31] The Globus Project Argone National Laboratory USC Information Sciences Institute, Grid Architecture. Available from: http://www.globus.org.

[32] T. Gen, "A New Method of Distributed Mobile GIS Grid Model Based on Mobile Agent," International Conference on Computer Application and System Modeling (ICCASM), IEEE, pp. 664-668, 2010.