Grid-enabled Urban-CA GIS

Qingfeng Guan
guan@geog.ucsb.edu
Department of Geography, University of California, Santa Barbara

Abstract

The development of GIS sciences and technologies motivate the concern of the next generation of GIS, including multi-resources distributed, high-performance computation and data transfer, and collaborative platform of virtual organization for multiple end users.

The development of information technology and GISiences provides chances to improve GIS technology, but introduces challenge to it as well. Conventional distributed/parallel computing technologies do not address all these problems. A recent emerging wide area distributed and parallel computing technology which is called Computational Grid is introduced as a possible solution for the next generation of GIS. Grid computing technology represents a new approach to collaborative computing and problem solving in data intensive and computationally intensive environment, and has the chance to satisfy all the requirements of a distributed, high-performance and collaborative GIS.

A framework of Urban-CA GIS with underlying Grid computing technology is proposed as an example of Grid-enabled GIS. Some particular requirements and challenges of this particular GIS are presented. And some methodologies and Grid computing technologies as solutions of requirements challenges are introduced to enable this distributed, parallel, high-throughput, collaborative GIS application.

As the conclusion, Grid computing has the chance to lead GIS into a new “Grid-enabled GIS” age in terms of computing paradigm, resource sharing pattern, online collaboration.

1 Introduction

Distributed Computing has become one of sub fields of Geographic Information Engineering in the Long-term Research Challenges, 2002 UCGIS Research Agenda. “Digital technology is moving rapidly to distributed computing…GISs have already adapted to several changes in computing architectures… Indeed, we may reach a time when the entire global network is best conceived as a single, integrated ‘computer’." This exactly consists with the basic idea of Computational Grid, which is to provide users a "virtual supercomputer". This "virtual supercomputer" is actually a high-performance distributed parallel computing environment formed by connecting flexibly defined assemblages of computing resources (supercomputers, workstations, individual PCs, etc.) by high performance networks. And these assemblages are supposed to be transparent to users. That means, users use the Computational Grid just like using the electricity grid by plugging an application into the wall, and don't need to be aware of where or what the computing resources are.

GIS users today have access to an unprecedented amount of high-resolution and high-quality environmental data through scanners, remote sensing devices, GPS receivers, etc. Furthermore, these datasets usually belong to different custodians, but could be used by others from different fields with different respective. So GIS can benefit from parallel and distributed computing. But also, the volume and variety of these environmental data, metadata, applications arise issues of interoperability, data and processing standards, which again are general problems in Grid computing. (Kenneth Hawick, et al. 2003)

This paper examines the motivation of applying Grid computing technology on GIS. To date, GIS tends to be more distributed in terms of data repositories, user background and purpose, team work or group collaboration. It also tends to require higher performance to process massive amount of data in relatively shorter time, sometimes even real-time. Grid computing could provide a solution for high-performance distributed GIS. Actually, a few research projects are going on to apply Grid computing on geospatial information processing. (Kenneth Hawick, et al. 2003, Ann Chervenak, et al. 2003, Asvin ananthanarayan, et al. 2003, Giovanni Aloisio, et al. 2003) This paper reviews the research state of Grid computing, and some of applications of Grid computing on geospatial information processing as well. Then it introduces a conceptual framework of "Grid-enabled Urban-CA GIS" as an example of Grid-enable GIS. We believe that Grid computing has the chance to lead GIS into a new "Grid-enabled GIS" age in terms of computing paradigm, resource sharing pattern, online collaboration.

2 Motivation of Grid on GIS

2.1 What's going on with GIS?

Creative motivation comes from development of things. As things go further, new issues show up, new solution are asked, new technologies are required.

After half of a century's development, GIS sciences and technologies have become more and more mature and sophisticated. GISs have been used in more and more fields with different disciplines, and playing more and more important role. As everything else, development of GIS arises issues need to be solved.

As the foundation of an information system, data should be discussed firstly. As just mentioned, GIS users today have access to an unprecedented amount of high-resolution and high-quality environmental data, which includes not only geospatial data, but also more and more non-spatial data such as socio-economic, demographic and commercial data, etc. Recourses of these datasets are getting more diverse. At present, GIS users can get data through scanners, remote sensing devices, GPS receivers, government agencies, social organizations, commercial companies, etc. As the consequence of development of data collecting devices and information networks, the updating of GIS data is getting more frequent. All these result in larger and larger volume of source data involved in GISs. In addition, since increasingly data sources are involved in GISs, data formats used in GISs are getting more various, which means a GIS is required to support multiple data formats. All these issues make it harder to store all datasets in a single site. Usually these datasets belong to different custodians that are dispersed geographically and logically.

Secondly, developments of spatial analysis and GeoComputing should be considered. As research on geographical phenomena and processes goes more sophisticated, new spatial analysis methods and GeoComputing algorithms are developed. Generally, these new methods and algorithms are more complex than previous ones, which means more computation are involved. When these methods and algorithms are implemented with GISs, consequently more CPU time and memory allocation are required. Meanwhile, as a significant trend in recent years, less responding time of GISs is required. Sometimes, even real-time computing or processing are necessary.

Thirdly, as a vital factor of an information system, users of GISs need to be examined. As mentioned, GIS data could be stored in geographically and logically distributed multiple sites. Similarly, GISs and GIS data could be used by users from different fields with different respective, who are located in various geographical positions. And different users could use GISs with different purposes, such as scientific research, governmental decision making support, business management and analysis, etc.

Fourthly, the development of information network technologies makes it possible for GIS users to share resources and collaborate even though they are dispersed geographically and logically. To date, shared resources include not only GIS data, but also methods and models, software components, computing capabilities, services, etc. As for collaboration, a new concept emerged in the last few years, which is called Virtual Organization (VO). A virtual organization is a network linking geographically distributed individuals or agencies who have partially overlapping objectives; based on this network they can pool complementary core competencies (Jagers et al. 1998). Actually, this model has gained currency in business management since middle 1990's (Palmer & Speier, 1997). Since users and data repositories of GISs tend to be geographically dispersed, and collaborations among users with various disciplines are getting more necessary, GIS has become a candidate of application platform of VO.

2.2 What kind of GIS is needed?

According to the reviewing of development of GIS technologies, some characteristics of the next generation of GIS could be presented.

First, the next generation of GIS should be distributed. Not only GIS data could be geographically and logically distributed in terms of repositories, but also GIS functionalities could be distributed geographically and logically. That is to say, software components and Geoprocessing models could be located in different sites, and be requested by geographically distributed users in form of services. So, this GIS should provide multiple custodians/users tools to manage and access massive distributed resources including datasets and services.

Second, the next generation of GIS should be high-performance. It should respond to requests from users, process multi-site data and perform complicated Geoprocessing in relatively short time, even real-time. Since this GIS is built based on distributed computing system, it has to deal with the data transport issue to achieve acceptably high performance. Furthermore, this management of computing resources should be transparent to users, which means that GIS should find out the most efficient way to achieve high performance automatically and intelligently with as less participation from users as possible.

Third, the next generation of GIS should be collaborative. It should facilitate constructing virtual organizations (VO) for geographically dispersed multi-users in order to accomplish common objectives. A GIS should provide protocols and tools to enable the following four characteristic features of a virtual organization (Shao et al, 1998):

- Connectivity, which establishes linkage among participants of the VO.
- Purpose, which creates an overlapping motivation for participants in the VO.
- Technology, which enables connectivity.
- Boundary, which separates participants from those who are not part of the VO.

2.3 What's the challenge?

Challenge always comes with chance. The development of information technology and GISiences provides chance to improve GIS technology, but introduces challenge to it as well.

First, to make GISs to be distributed computing systems, uniform standard, including data/metadata format and service interface, needs to be established to enable interoperability. Since increasing types of data are involved in GISs, it's getting harder and harder to establish a single universal data format for all of them. One solution is to establish a set of standard data formats and protocols for GIS data types varying from geospatial data to non-spatial data, from numerical data to imagery data, from file-based data to database. Another issue is about standard service interface. As mentioned in the last section, distributed GIS could consist of multiple geographically and logically dispersed GIS functions providing services to end-users. These functions or services could come from different developers, and could also be customized by users. Standard interface makes sure that these functions or services can communicate and team up to perform Geoprocessing. Actually, the goal of Open GIS Consortium (www.opengis.org) is to establish standards for GIS data and geospatial data processing service. However this is a much more difficult problem and there has been little progress so far (Hawick et al., 2003).

Second, to make GIS to be parallel computing system, resource management and task allocation need to be concerned in order to achieve "best" load balance among multiple processors. When computing is implemented in parallel paradigm, data or task needs to be decomposed into parts, and loaded to processors efficiently according to performance of processors and bandwidth of networks among processors. The tradeoff between data transport and parallel processing needs to be concern in order to get efficiency.

Third, to make GIS to be a platform of virtual organization, technical concerns and requirements of VOs need to be satisfied (Foster et al., 2001).

- Highly flexible sharing relationships, varying from client/server to peer-to-peer;
- Complex and high levels of control over shared resources, including security issue such as authentication and authorization, implementation of local and global policies;
- Highly variety of sharing resources, including datasets, documents, models, software components, computers, instruments and networks;
- Diverse usage modes, varying from single user to multi-user, from performance intensive to cost intensive.

2.4 What's the approach?

Current GIS technologies and distributed/parallel computing technologies do not address all these concerns and requirements mentioned in the previous section. Present distributed GISs still focus on handling geographically distributed data. Even some of them provide remote spatial analysis or Geoprocessing, the services provided are relatively simple but not extremely intensive on computation. Current parallel GISs are highly centralized. Expensive supercomputers or computer cluster are required, which blocks users who have no access to these high-cost facilities. Most of current distributed computing technologies address communication and information exchange among computers but not the coordinated use of resources at multiple sites for computation. Some of them focus on enabling sharing resource within a single organization, such as CORBA and Enterprise Java.

Fortunately, the recent distributed computing trend indicates a new approach for the next generation of GIS, which is called Computational Grid. The Grid computing technology has the chance to satisfy all the requirements of a distributed, parallel and collaborative GIS.

3 Computational Grid

In the earlier definition, “A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities” (Ian Foster, Carl Kesselman. 1998). At this point, flexibly defined assemblages of computing resources (supercomputers, workstations, individual PCs, etc.) connected by high performance networks (Lan, Wan, Internet, etc.) are used to form a distributed parallel computing environment. These assemblages are supposed to be transparent to users. That means, just like using the electricity grid by plugging an application into the wall, users use the Computational Grid as a virtual supercomputer, and don’t need to be aware of where or what the computing resources are.

Along with continuing research, the concept has been broadened to address social and policy issues. In recent papers, Grid Computing is concerned with "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations." (Ian Foster, Steve Tuecke. 2000) That is to say, Computational Grid is not only an infrastructure for sharing datasets and computing capabilities, but also provides a platform for multi-user to collaborate to solve massive computing problems. Obviously, this wider concept gives Computational Grid more potential to drastically impact the way we think of computation.

Table 1 shows a few of possible usages of Computational Grid. As we can see from this table, Computational Grid could satisfy all of the requirements of the next generation of GIS. GIS could be a candidate of the example of every category in this table.

Category	Examples	Characteristics
Distributed Supercomputing	DIS Stellar dynamics Ab initio chemistry	Very large problems needing lots of CPU, memory, etc
High throughput	Chip design Parameter studies Cryptographic problems	Harness many otherwise idle resources to increase aggregate throughput
On demand	Medical instrumentation Network-enabled solvers Cloud detection	Remote resources integrated with local computation, often for bounded amount of time
Data intensive	Sky survey Physics Data Data assimilation	Synthesis of new information from many or large data sources
Collaborative	Collaborative design Data exploration Education	Support communication or collaborative work between multiple participants

Table 1. I. Foster, C. Kesselman. *Chapter 2 of "The Grid: Blueprint for a New Computing Infrastructure"*, Morgan-Kaufman, 1999

Actually, several research projects on Computational Grid are going on, such as Globus, TeraGrid, Legion, NetSolve, NINF, Nimrod, APST, NEOS, Condor, NWS, etc. Recently, after realizing the importance and promise of Computational Grid, a few industrial companies have been involved in this computation revolution, such as IBM, Sun, Microsoft, and so on. Many commercial and non-commercial products for Grid computing have been created. For instance, the Globus Toolkit, which was developed by Globus project, has become the de facto standard for scientific research and industrial products. So, we already have the standard development tool to build up our own Computational Grid and Grid Computing applications.

4 Grid-enabled Urban-CA GIS

In this section, the framework of an Urban Cellular Automata GIS using underlying Grid computing technology is proposed.

4.1 Cellular Automata Model

Cellular Automata (CA) were introduced as models of self-reproducing systems by John Von Neumann in the 1950s (S. Bandini, etc., 2001). Codd (1968) describes them as elegant mathematical models for a class of processes operating in discrete time and discrete space. From then on, the research on CA has emerged from different perspectives and went into different directions.

A classical CA is a set of identical elements, called cells, each one of which is located in a regular, discrete space. Each cell can be associated with a state from a finite set. The model evolves in discrete time steps, changing the states of all its cells according to a transition rule, homogeneously applied at every step. The new state of a certain cell depends on the previous states of a set of cells, which can include the cell itself, and constitutes its neighborhood.

Many CA models have been developed to be applied in many different fields, including urban modeling, international development, medical modeling, geological modeling, and ecological modeling, and so on (C. Hecker, 1999). In geographical research, CA has been recognized for some time as a powerful tool for GIS modeling and simulation.

4.2 GeoComputing for Urban-CA

Urban cellular automata (Urban-CA) are those cellular automata used to simulate and predict land-use and land-cover change in urban areas. As the consequence of years of study, Urban-CA models have been turned into much more complex ones than those original basic CA concepts. The neighborhood patterns, the transition functions, the linkage to other socio-economic models, etc. have made Urban-CA a huge computing system.

The calibration process of Urban-CA models especially involves massive computation. Since higher resolution data sources are available and higher accuracy is required, the demands of higher computing performance are desired. For instance, a particular CA model called SLEUTH has been developed to simulate the urban area land-use change during urban development, and "the model calibration for a medium sized data set and minimal data layers requires about 1200 CPU hours on a typical workstation" (K. C. Clarke. 2003). The traditional single-processor-workstation-based algorithm can no longer satisfy this requirement in practice. Since CA models are theoretically parallel computing models, one solution is to run the Urban-CA model applications on a multi-processor high-performance supercomputer or a cluster of high-performance workstations.

In the latest version of the SLEUTH model, "dynamic memory allocation returned to a flat memory structure optimized for the Cray memory model", and Message Passing Interface (MPI) was deployed to make the model run on a multi-processor supercomputer or workstation cluster. (K. C. Clarke. 2003) Comparative tests showed that the parallel computing model is much more efficient than traditional ones.

However, this model still has a few points could be developed further.

Firstly, the parallel algorithm in SLEUTH is not for the CA model itself. So-called parallel CA model means that, the whole cell space is split into pieces, all these pieces are distributed to several processors, and transition rules are implemented simultaneously on those pieces by those processors. However, the SLEUTH does not do the CA parallelization. It uses task parallelization instead of data parallelization. During the calibration process, the SLEUTH establishes a series of parameter scenarios and distributes them to several processors. Each processor has different parameter set, implements transition rules on the whole cell space. For those datasets of relatively small areas in low resolution, this task parallelization works well. But for dealing with huge datasets of very big areas in high resolution and a very few parameters sets, it will give each processor too much work to do, take a very long time for each processor to finish just one parameter set calibration, and keep other spare processors idle at the same time. For example, if we deal with the 1km*1km data of the whole United States and only 10 parameters sets, by using a 100-processor supercomputer, hence only 10 processors work at one time. And each of these 10 processors works on the whole dataset, it will take thousands of hours to do that. But the other 90 processors just do nothing. So we need a new algorithm combining data parallelization and task parallelization together, to do calibration efficiently on small dataset with complex parameters sets or big dataset with simple parameters sets, or any other situations.

Secondly, this model is basically a centralized control computing. That is to say, even though the computing process (memory allocation, computation tasks) can be distributed geographically, there is only one end user participating in the human-machine interaction to collect the dataset, set up initial scenarios, control computing procedure, interpret and evaluate the result, and so on. The consequence could be that "university researchers make extensive use of computers when studying the impact of change in land use on biodiversity, but city planners selecting routes for new roads or planning new zoning ordinances do not." (Ian Foster, Carl Kesselman. 1998) But in practice, communication and collaboration among people who are familiar with different fields are getting more and more important. As a result, individuals can gather more comprehensive data, employ more appropriate methods, make more reasonable evaluation and achieve more sophisticated conclusion, than they can do currently, even though they may be geographically separated.

Thirdly, to achieve very high performance, this model requires users to have access to supercomputers or a cluster of high performance workstations. Otherwise, it's going to take an incredibly long time to complete the whole computation as mentioned above. For most users, including researchers, city planners and public people, this is not feasible. Users wish to have a low-cost, accessible, steady, reliable and high-speed computing environment/service to implement their ideas. In another word, we need some innovative high-performance computing technology other than cluster-based parallel computing.

In summary, there is a motivation to make Urban-CA GIS to be a distributed, high-performance and collaborative GIS. As mentioned, Computational Grid could satisfy all those requirements of this next generation of GIS.

4.3 Grid-based solution for Urban-CA GIS

As mentioned above, running Urban-CA model in a Computational Grid environment would help users, such as researchers, government officers, and the public people, to overcome local computational constraints, and work together (share data, computing capabilities, knowledge, etc) to approach much better simulating/predicting results in a reasonably short time by using a inexpensive virtual supercomputer.

In the current version of the SLEUTH model, every time user changes a tiny thing in the scenario file, he/she has to rerun the model all over again, and wait a long time to see the difference caused by the tiny change. If SLEUTH can be run on a Computational Grid, users would be able to have meetings online and share data, knowledge and ideas to set up a series of scenarios, run them very quickly (even simultaneously), explore the results, discuss and evaluate the results to approach a sophisticated conclusion (land-use change research, urban planning, etc.).

4.3.1 Particular Requirements of Grid-enable Urban-CA GIS

To build up a Grid-enabled Urban-CA GIS, the following technical and scientific questions need to be addressed:

First one is effective and efficient distributed/parallel algorithms for Urban-CA models, particularly in Computational Grid environment. As just mentioned, the Urban-CA models have been developed into much more complex ones than the original basic CA concepts. For urban geographical research, the cell space could be no longer homogenous. Cells in the research space could have different properties, which makes the situation much more complicated. The transition rules are no longer as simple as the basic CA models. The state of a particular cell is determined not only by the states of its neighborhoods, but also by some other factors, like transportation systems, government policy, and so on. Lots of Urban-CA models have been linked to other socio-economic models. CA models and socio-economic models affect each other nonlinearly. All these issues make Urban-CA more complex and difficult to be parallelized. Data parallelization for Urban-CA models requires much more information to be exchanged among processors than that for original CA models. Information includes cell states, cell properties, new growth center and random walk (K. C. Clarke, 1997). Communication among processors must be effective and efficient enough to approach high-performance computing.

Second one is online collaborative models and methods for designing Urban-CA modelling scenarios, computing controlling, evaluating results, etc. Computational Grid provides users a potential platform to share not only computing capabilities and data, but also ideas, thoughts and concepts. Users of Urban-CA models could include urban geographers, urban planners, government officers, students and general public. They have different backgrounds, different interests, different purposes for using Urban-CA. Especially in urban planning or in constituting area policy, with Urban-CA run in Grid environment, experts from on different fields can share their special data, work with each others to set up space cells, neighborhood patterns, transition rules, simulation scenarios, and analyze the simulation results. As a result the urban planning or area policy will be more considerate and reasonable. Another example is education usage of Urban-CA model. With Computational Grid, instructors and students can work together to set up several different scenarios to simulate the urban land-use change, and they can see the result in a very short time with such an inexpensive high-performance virtual supercomputer.

4.3.2 Technologies and Methodologies used in Grid-enabled Urban-CA GIS

(1) Parallel Algorithm for Urban-CA model

Since the states of neighborhoods may change, it is necessary to exchange that information among processors at each iteration. Thus the parallel processes that are developed to implement CA applications are not independent.

Communication overhead is the key issue for parallel computing. "Communication overhead introduced by the exchange of data among the processors can be minimized if the amount of data exchanged is kept small and if the distance the data must travel is minimized. It is also important to avoid bottlenecks in the communication such as those that occur when all the processors attempt to communicate with one particular processor at the same time." (C. Hecker, etc., 1999)

A typical solution is ghost cells technique. This technique allows processors exchange the cell-states update information as less as which matters for the evolution at the next iteration. A new algorithm will be developed based on ghost cells technique to suit to complexity in Urban-CA model. For example, new growth centers are determined by exchanging ghost cells information. Random walk alone transportation lines requires information exchange among processors, too.

(2) Resource Management

In this Grid-enabled Urban-CA GIS, geographically and logically distributed resources include data replicas, computer, models, software components, etc. "In order to improve the speed of computation, it is also necessary that the computational work be distributed as evenly as possible among the processors". (C. Hecker, etc., 1999) However, in Computational Grid environment, computing capability of processor depends not only on how fast the processor itself is, but also on data transfer bandwidth, local and globe memory allocation, etc. Furthermore, it's possible that duplicated datasets exist at geographically dispersed sites. So, the algorithm of computing distribution is to be dynamic based on each processor's computing capability and performance of networks. The GIS should provide tools to automatically and intelligently find "best" data servers to download data and "fastest" computing workstations to perform data processing and CA calculation.

Globus Toolkit is a product of the Globus project, and provides basic Grid services and protocols that include resource managers, security protocols, information services, communication services, fault tolerance service, and remote data access facilities. Globus Resource Management and Allocation (GRAM) is a subset of Globus Toolkit focusing on resource management on Computational Grid. The most recent version of Globus also provides a Replica Catalog service for discovery and querying of distributed replicated data archives. Specific resource management tools for this Grid-enabled Urban-CA GIS are going to be developed based on Globus Toolkit.

(3) High-performance Geospatial Data Transport

This Grid-enabled Urban-CA GIS requires high-performance data transport among wide area networks to achieve parallel CA computing and multi-users collaboration. The distributed nature of data sources and consumers means that data access and data transfer must be critical components of this Urban-CA GIS. The transfer facilities should be secure, fast, and reliable (A. Chervenak et al., 2003).

The Globus Toolkit provides a set of protocols and APIs that satisfy these requirements, which is called GridFTP. GridFTP functionality includes features that are supported by the FTP standard, extensions that have already been standardized or are under consideration, and some proposed extensions spatially for Computational Grid (A. Chervenak et al., 2003). And GridFTP provides a uniform interface to various storage systems, which means by using this technology GIS could retrieve data from multiple systems even they have incompatible data access protocols.

(4) Metadata Management and Broker

One of main problems for distributed GISs is the heterogeneity of geospatial data models and formats required in different GIS applications and functions. Metadata becomes the key to bridge the heterogeneous environments of distributed GIS databases and services and to provide users with the semantics and syntactic of GIS databases (Plewe. B & Johnson S. 1999).

Globus Metadata Directory Service (MDS) can be used in this Grid-enabled Urban-CA GIS to manage metadata and provide users content-based querying on the distributed geospatial datasets.

According to the above two subsections, an Urban-CA GIS metadata/data sharing platform could be constructed based on the GridFTP and MDS techniques to support remote metadata/data querying and exchanging.

(5) Authentication and Authorization

Security issues in such a wide area distributed GIS is critical, which includes authentication and authorization using community policies as well as allowing local control of resource (A. Chervenak et al., 2003).

Grid Security Infrastructure (GSI), combined with GridFTP protocol, makes sure that sharing and transfer of geospatial data and Geoprocessing (CA computing in this case) are secure in the Computational Grid environment.

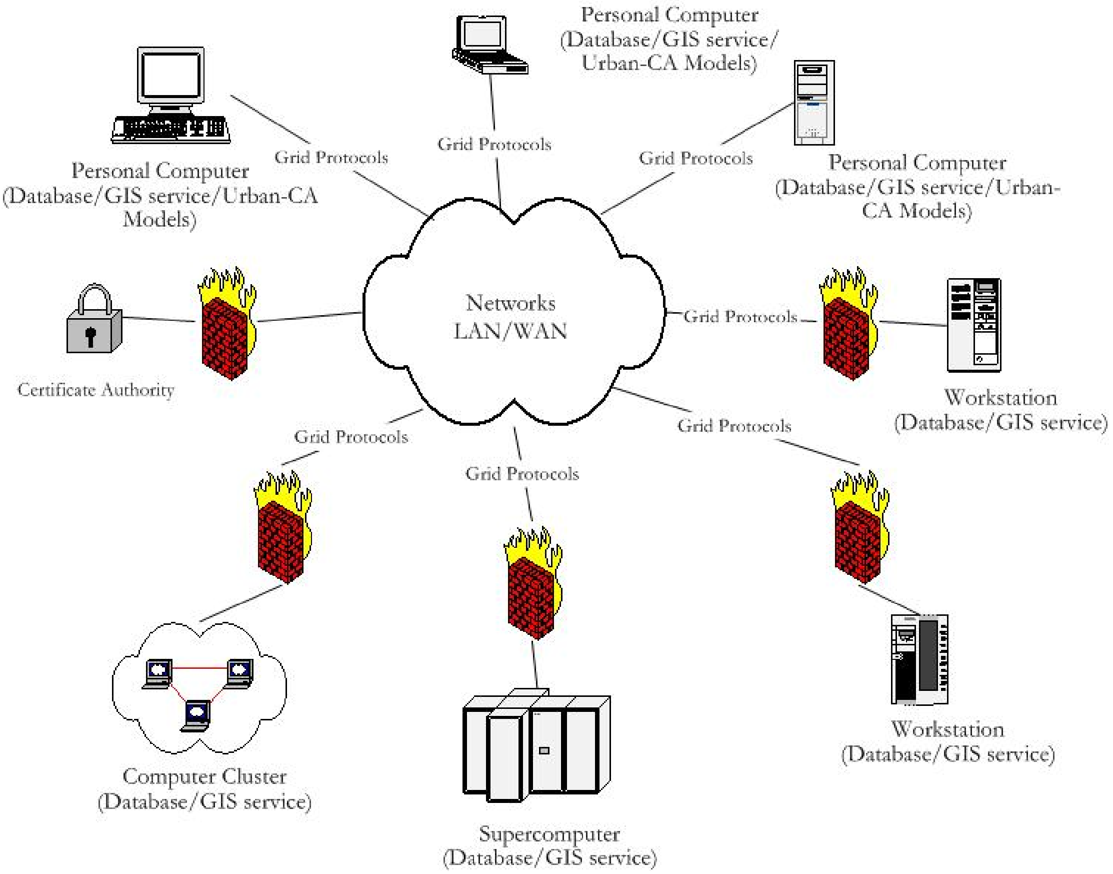

Figure 1 Topology framework for Grid-enabled Urban-CA GIS

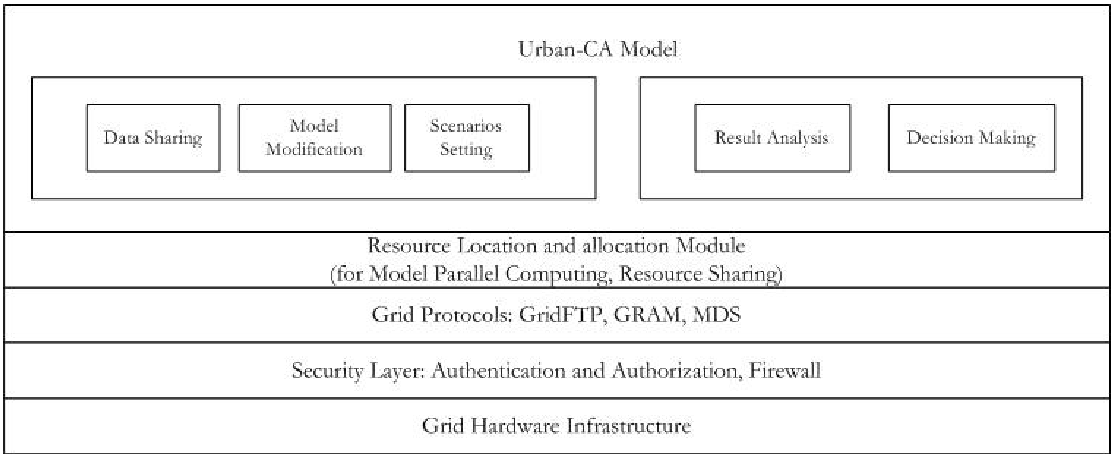

Figure 2 Grid-enabled Urban-CA GIS architecture

5 Conclusion

As the development of GIS sciences and technologies go further, increasingly amount of geospatial and non-spatial data are involved in GISs due to more diverse data sources and development of data collection technologies. GIS data tend to be geographically and logically distributed as well as GIS functions and services do. Spatial analysis and

Geocomputation are getting more complex and computationally intensive. Sharing and collaboration among geographically dispersed users with various disciplines with various purposes are getting more necessary and common. A dynamic collaborative model called virtual organization (VO) is required for GIS application.

Some technical requirements and concerns of the next generation of GIS are proposed in this paper according to the development of GIS technologies. These requirements and concerns include multi-resources distributed, high-performance computation and data transfer, and collaborative platform of virtual organization for multiple end-users.

Challenges of building such a GIS are presented, such as standard data formats and GIS service interface to facilitate distributed GIS, load balance issue for parallel computing, flexible sharing relationship, multi-resource management, various usage modes and security issues in virtual organization.

Current distributed/parallel computing technologies do not address all these problems. A recent emerging wide area distributed and parallel computing technology called Computational Grid is introduced as a possible solution for the next generation of GIS. Basically, the Grid computing concept is intended to enable coordinate resources sharing and problem solving in dynamic, multi-organizational virtual organizations (Yi Shi et al., 2002) by linking computing resources with high-performance networks. Grid computing technology represents a new approach to collaborative computing and problem solving in data intensive and computationally intensive environment (Yi Shi et al., 2002), and has the chance to satisfy all the requirements of a distributed, high-performance and collaborative GIS.

A framework of Urban-CA GIS with underlying Grid computing technology is proposed as an example of Grid-enabled GIS. Some particular requirements and challenges of this particular GIS are presented. And some methodologies and Grid computing technologies as solutions of requirements challenges are introduced to enable this distributed, parallel, high-throughput, collaborative GIS application.

As the conclusion, Grid computing has the chance to lead GIS into a new "Grid-enabled GIS" age in terms of computing paradigm, resource sharing pattern, online collaboration.

References

1. Ann Chervenak, et al. High-performance remote access to climate simulation data: a challenge problem for data grid technologies. Parallel Computing. 2003. 29. 1335 - 1356
2. Anthony, G. O. Y., Xia, L. Sustainable land development model for rapid growth areas using GIS. Int. J. Geographical Information Science, 1998, 12(2): 169-189
3. Asvin Ananthanarayan, et al. Data webs for earth science data. Parallel Computing. 2003. 29. 1363 – 1397
4. Batty, M., Longley, P. A.. Fractal Cities. London: Academic Press, 1994
5. Batty, M., Xie, Y.. Urban Analysis in a GIS Environment: Population Density Modeling Using ARC/INFO. A. S. Fotheringham and P. A. Rogerson (editors)

Spatial Analysis and GIS, Taylor and Francis London, 1994: 189—219

6. Batty, M.. Urban Modeling in Computer-Graphic and Geographical Information Systems Environment. Environment and Planning B, 1992, 19, 663-688
7. Batty,M., Xie, Y.. From Cells to Cities. Environment and Planning B, 1994, 21: s31-s48
8. C. Hecker, D. Roytenberg, J. –R. Sack, Z. Wang. System development for parallel cellular automata and its applications. Future Generation Computer System. 1999, 16: 235 – 247
9. C1arke, K.C., Riggan, P. Brass, J.A. A cellular Automaton Model of Wildfire Propagation and Extinction. Photogrammetric Engineering and Remote Snesing, 1995, 60, 1355—1367
10. Clarke, K. C., Gaydos, L. J., Hoppen, S., Methods and Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Barbara, 1996: 21-26
11. Clarke, K. C., Gaydos, L. Loose-coupling a cellular automaton model and GIS: long-term urban grow prediction for San Francisco and Washington/Baltimore. Int. J. Geographical Information Science, 1998, 12(7): 699-714
12. Clarke, K. C., Hoppen, S., Gaydos, L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning B, 1997, 24: 247-261
13. Couclelis, H. Cellular Worlds: a Framework for Modeling Micro-Macro Dynamics. Environment and P1anning A, 1985, 17: 585—596
14. Couclelis, H. From Cellular Automata to Urban Models: New Principles for Model Development and Implementation, Environment and P1aaning B, 1997, 24: 165—174
15. Couclelis, H. Macrostructure and Microbehavior in a Metropo1itan Area. Environment and P1anning B, 1989, 16: 14l — 154
16. E. F. Codd. Cellular Automata. Academic Press. New York, 1968
17. Eneelen, G. White, R. Inge Uljee, Panl Drazan. Using Cellular Automata for Integrated Modeling of Socio-environmental Systems. Environmental Monitoring and Assessment, 1995, 34: 203-214
18. Foster I, Kesselman C, Tuecke S. The anatomy of the Grid: enabling scalable virtual organization. Intl. J. High Performance Computing Applications. 2001. 15(3): 200-222
19. Giovanni Aloisio, Massimo Cafaro. A dynamic earth observation system. Parallel Computing. 2003. 29. 1357 – 1362
20. Ian Foster, Carl Kesselman, Steven Tuecks. The Anatomy of the Grid. Supercomputer Application. 2001.
21. Ian Foster, Carl Kesselman. Computational Grids.
22. Ian Foster, Carl Kesselman. Globus: A Metacomputing Infrastructure Toolkit. http://www.globus.org/
23. Ian Foster. WHAT IS THE GRID? A THREE POINT CHECKLIST. DAILY NEWS AND INFORMATION FOR THE GLOBAL GRID COMMUNITY. 2002,

1, 6
24. Karl Czajkowski, etc. From Open Grid Services Infrastructure to WS-Resource Framework: Refectoring & Evolution. 2004, http://www.globus.org/
25. Kenneth A. Hawick, et al. Distributed frameworks and parallel algorithms for processinsg large-scale geographic data. Parallel Computing. 2003. 29. 1297 – 1333
26. Palmer JW, Speier C. A typology of virtual organizations: an empirical study. Proceedings of the Association for Information Systems 1997 Americas conference. Auguest 15 – 17, 1997. Indianapolis, USA. http://hsb.baylor.edu/ramsower/ais.ac.97/papers/palm_spe.htm
27. Plewe, B. and Johnson, S. Automated Metadata Interpretation to Assist in the Use of Unfamiliar GIS Data Sources. In: Goodchild M., Egenhofer, M., Fegeas, R., and Kottman, C. Interoperating Geographic Information Systems. Kluwer Academic Publishers, Boston. 1999. 203-214
28. S. Bandini, G. Mauri, R. Serra. Cellular automata: From a theoretical parallel computational model to its application to complex systems. Parallel Computing. 2001, 27: 539 - 553
29. Shao YP, Liao SY, Wang HQ. A model of virtual organizations. Journal of Information Sciences. 1998. 24(5): 305 – 312
30. UCGIS. 2002 Research Agenda - Distributed Computing. http://www.ucgis.org/priorities/research/research_white/1998%20Papers/distrib.html
31. White, R. W. and Engelen, G. Cellular Automata Modelling of Fractal Urban Landuse Patterns: Forecasting Change for Planning Applications Budapest. Eighth European Colloquium on theoretical and Quantitative Geography, September, 1993: 12-16
32. White, R. W. Cellular Automata and Fractal Urban Form: a Cellular Modeling Approach to the Evolution of Urban Land Use Patterns. Environment and Planning A, 1993, 25: 1175-1189
33. White, R. W. Cellular Automata and Fractal Urban Form: a Cellular Modeling Approach to the Evolution of Urban Land Use Patterns. Stockholm: Seventh European Colloquium on Theoretical and Quantitative Geography, 1991
34. White, R. W., Engelen, G. Cellular Synamics and GIS: Modeling Spatial Complexity. Geographical Systems, 1994, 1: 237-253
35. White, R. W., Engelen, G. High-resolution integrated modeling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban System, 2000, 24: 383-400
36. Wu F. SimLand: a Prototype to Simulate Land Conversion Through the Integrated GIS and CA with AHP-Drived Transition Rules, Int. J. Geographical Information Science, 1998, 12 (1): 63-82
37. Xie, Y. A Generlized Model for Cellular Urban Dynamics, Geographical Analysis, 1997, 28: 350-373
38. Xie, Y. Analytical Models and Algorithms for Cellular Urban Dynamics. Ph. D. Thesis, University of New York at Buffalo, 1994