eMicrob: A Grid-Based Spatial Epidemiology Application

Jianping Guo\textsuperscript{1}, Yong Xue\textsuperscript{1,2,*}, Chunxiang Cao\textsuperscript{1}, Wuchun Cao\textsuperscript{3}, Xiaowen Li\textsuperscript{1}, Jianqin Wang\textsuperscript{1}, and Liquan Fang\textsuperscript{3}

\textsuperscript{1}State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences, and Beijing Normal University, Beijing 100101, China
\textsuperscript{2}Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
\textsuperscript{3}Institute of Microbiology and Epidemiology, the Academy of Military Medical Sciences, the Chinese PLA, Beijing 100071, PR China

gjpgis@163.com, y.xue@londonmet.ac.uk

Abstract. The use of Grid technologies allows us to make progress in the prediction accuracy of epidemiological patterns, epidemiological modeling, risk predictions of infectious diseases etc by combining the geo-information and molecular simulation analysis methods. In this paper, we mainly design the eMicrob, in particular, build up the e-Microbe miniGrid deployed in IRSA, CAS and IME, the Chinese PLA. The architecture is as follows: Firstly we review related grid applications that are motivating widespread interest in Grid concepts within the scientific and engineering communities. Secondly we talk about the key methodologies and strategies involved in the construction of eMicrob. In the third section, the system design of the eMicrob, in particular about the architecture of the eMicrob miniGrid is discussed. Finally, we draw some conclusion in the process of the building of eMicrob and make some discussion about the challenges. It has been proven that the methods based on the Grid technologies are revolutionary and high efficient through the experience of the establishment and deployment of the e-Microbe miniGrid.

1 Introduction

Epidemiologists study a diverse range of health conditions as well as the impact that various exposures have on the manifestation of disease. However, we have done some new researches in the epidemiology (science of the public health, http://www.acepideiology2.org/aboutepid/whatis.asp) based on the new emerging technology — Grid computing, integrated with geographical information system (GIS) and remote sensing (RS), which will establish a virtual organization comprised of scientists and engineers from many different fields and discipline. Grid computing, as a new technology, is mainly concerned with “coordinated resource sharing and

\* Corresponding author.
problem solving in dynamic, multi-institutional virtual organizations.” (Foster et al. 2001). A principal objective in the development of Grid computing is to make such resources available to users, regardless of their geographical location or institutional affiliation. While this goal is some way from being achieved, we have developed eMicrob miniGrid in the use of Grid tools, which allows us to make progress in the prediction accuracy of epidemiological patterns, epidemiological modeling, risk predictions of infectious diseases etc by combining the geo-information and molecular simulation.

Now through the eMicrob project we know that patterns and risk predictions can be made through GIS, by which we can study the dynamics of the epidemiological diseases based on above-mentioned environment factors achieved from remote sensing based on the grid platform. The undergoing eMicrob project benefits from grid and spatial technologies in various ways.

In this paper, we firstly introduce the related Grid projects, as good paradigms, which will help us to design the eMicrob system more precisely. The design of the eMicrob miniGrid is outlined and the architecture of the eMicrob miniGrid is discussed in details. At the end of this paper, we draw some conclusions and make discussion about the challenges eMicrob is encountering.

2 Methodology and Strategy

The eMicrob aims at building up a Grid platform (eMicrob miniGrid) in the long run, which will provide secure access to heterogeneous data and expensive resources in different locations, facilitate the study of transmission pattern, spatial distribution patterns, epidemiological trends, even advance the risk prediction precision about the infectious diseases i.e. Severe Acute Respiratory Syndrome (SARS), schistosomiasis (Savioli et al. 1997), avian influenza, malaria and cholera, etc.

We take advance of the following methodologies and strategies to design and construct the eMicrob. First, we build up a science team comprising of 3 kinds of technician, i.e. developer of code, epidemiologist and specialist from spatial domains. Second, we build our different database based on the SRB (http://www.npaci.edu/DICE/SRB/) technology. Moreover, we deploy the miniGrid on the platform – HIT-SIP developed by us .now we will expound on it. HIT-SIP (Wang et. al., 2004) Grid platform in Institute of Remote Sensing Applications, Chinese Academy of Science is an advanced High-Throughput Computing Grid testbed using Condor developed by Department of Computer Sciences, University of Wisconsin-Madison. Heterogeneous computing nodes include two sets of Linux computers and WIN 2000 professional computers and one set of WIN XP computer provide stable computing power. The Grid pool uses java universally to screen heterogeneous characters. In general, users can use the heterogeneous Grid and share its strong computing power to process remote sensing images with middlewares as if in one supercomputer. Across the HIT-SIP platform, we can process and analyze the remote sensed images, for example, supervised classification and unsupervised classification and retrieval of various environment factors, which will be used in such discipline as biology, public health, and other spatially related disciplines.
3 Architecture of the eMicrob MiniGrid

To build the e-Microbe miniGrid, we deployed the miniGrid between IME Grid center (8 nodes) and IRSA Grid center (10 nodes). Both centers are required to use such universal grid platform as Globus, Condor or Condor-G as a starting point in order to develop the web data portals and computing portals, moreover, the epidemiologist and RS and GIS scientists can use them easily.

The framework consists of the following components:

- eMicrob User Interface and Web Portal Service, including data portal
- All kinds of databases and Model Pools
- Data Cache
- Application Modules

---

Fig. 1. The architecture of the eMicrob miniGrid
Based on the above-mentioned database components and computer tools, the architecture of the eMicrob miniGrid is given in Figure 1.

4 Conclusions and Results

The eMicrob miniGrid has provided a new platform of epidemiology investigation on which we collect remotely sensed data and other related data. Then integrate them into epidemiology applications researches, based on which analyses of microbe and/or virus combined with the real-time experimental data, are made and visualization are resulted in.

The science outcome based on the eMicrob miniGrid not only include the outcomes of spatial domain, such as we can disclose potential links between remotely sensed factors and specialized diseases (Beck et al. 2000), which has been proven prevalent in the research of epidemiological diseases, but also include ones of epidemiological disease such as disease controlling and prediction, and rules of the microorganism variances, the character of hepatitis B virus or the coronavirus causing SARS, taking it for example, can be analyzed on such eMicrob miniGrid platform.

We think that the scientist, both from spatial information and epidemiological fields, will greatly promote their researches in the epidemiological disease and in the end will provide more scientific information to the government officials to do correct decision-making. Meanwhile, through the experience of the design of the eMicrob and the implementation of the miniGrid, we recognize that the method based on the eMicrob miniGrid has made the study of the epidemiological diseases change deeply and will have a great potential impact for it can make the epidemiology and microbe discipline to make great stride in the near future.

Acknowledgements

This publication is an output from the research projects “CAS Hundred Talents Program” and “Innovation Program” funded by IRSA, CAS. The authors wish to acknowledge related departments and individuals.

References

Foster I., Kesselman C., Tuecke S., 2001, “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”. International Journal of Super-computer Applications, 15(3): 200-222.

Beck L. R., Bradley M, Lobitz Byron L. Wood, 2000, Remote Sensing and Human Health: New Sensors and New Opportunities. Emerging Infectious Diseases, Vol. 6, No. 3, May–June 2000:217-226.

Savioli, L., Renganathan, E., Montresor, A., Davis, A., Behbehani, K., 1997, Control of schistosomiasis – a global picture. Parasitology Today, 11, 444-448.

Jianqin Wang, Xiaosong Sun, Yong Xue, Yanguang Wang, Ying Luo, Guoyin Cai, Shaobo Zhong, and Jiakui Tang, 2004, Preliminary Study on Unsupervised Classification of Remotely Sensed Images on the Grid. Lecture Notes in Computer Science, Vol. 3039, pp.995-1002.