Proceedings of the *Open source GIS - GRASS users conference 2002* - Trento, Italy, 11-13 September 2002

# Decentralized GIS Web Services on Grid

Apirak Panatkool*, Sitthichai Laoveerakul*

* Cluster Computing and Grid Technology Research Section, National Electronics and Computer Technology Center, 112 Paholyothin Rd., Klong 1, Klong Luang, Pathumthani 12120, THAILAND, tel. ++66 2 5646900 ext 2277

e-mail {bank,oui}@hpcc.nectec.or.th

**Abstract**

Geographical data is typically generated and stored locally which is then processed by a limited number of specialized computations or services on that site. Nonetheless, to get a global view, one needs to gather information from various locations and process them by using different specific services. Since each location does not contain all the required services, we normally gather data from those locations to be processed on a central server that has all the services. This centralized server paradigm contains some drawbacks such as single point of failure, network congestion, data inconsistency and others. Furthermore, in some real time applications, e.g., traffic control, pollution modeling etc., these problems must be solved.

To alleviate the situation, we propose a decentralized paradigm based on the computational grid model. On this model, data is distributed on the Internet and will be processed by new kind of services called the web services. We adopt the peer-to-peer (P2P) protocol to define the communication between any two adjacent parties. That is each site can assume either client or server role when communicating with the other site. By using this protocol, a web service can be moved to any site to process data thereby resolving the need to have a centralized server.

## 1 Introduction

Geographic information system (GIS) heavily involved with rendering geographic images and accessing pertinent data from database(s) so that is multiple layers of information can be obtained. Different pieces of information will then be overlayed in order to gain new knowledge. GIS problems are considered difficult due to the representation of the information, e.g., images, and the amount of data that one must process in order to understand the problem. Mostly GIS operations are done locally on an offline computer (standalone) connecting to a central database. In order to do the work, we must operate this isolated computer.

Nowaday, personal computers are becoming more powerful as the CPU speed doubles every 18 months as predicted by Moore. Not only does the CPU speed improves, so does the networking speed. Furthermore, the internet allows multiple computers connected to the internet to efficiently talk to each other via TCP/IP protocol. GIS researches also benefit from this improvement as well(The internet GIS concerns with performing GIS operation remotely). We no longer need to operate on the central server. As an example of this centralized paradigm, considering a case in Thailand the Chiangmai province[1] hosts GIS information, this as amount of rainfall precipitation, satellite images of Chaingmai and vicinity etc., this information can be updated and utilized as frequently as you wish. However, if we want to predict the flooding pattern of the nothern provinces, we will need to obtain information from other nothern provinces. As in centralized fashion (see Figure 1), this data must be replicated on the central server which is located in Bangkok.

By so doing, many fundamental data manupilation problems arise. First and foremost, nobody can gaurantee the consistency of replicated data. In addition, we cannot ensure that data from different provinces will be stored in the same format. Therefore, we must download more data than what we actually need in order to make sure that we have more data than what we really need. Secondly, transfering data can be very costly which is going to be very impractical when the GIS analysis is constantly required (note that we cannot reduce the amount of transfered data since we cannot select which data should be transfered). Imagine what will happen when the central server is not operational. This single point of failure is very undesireable in many online systems. One way to alliviate this problem is by not keeping data on the central server where data can be stored on their local servers and the amount of transfer will be minimized. This paradigm is called the decentralized approach.

[1]Chaingmai is a province located in the nothern part of Thailand

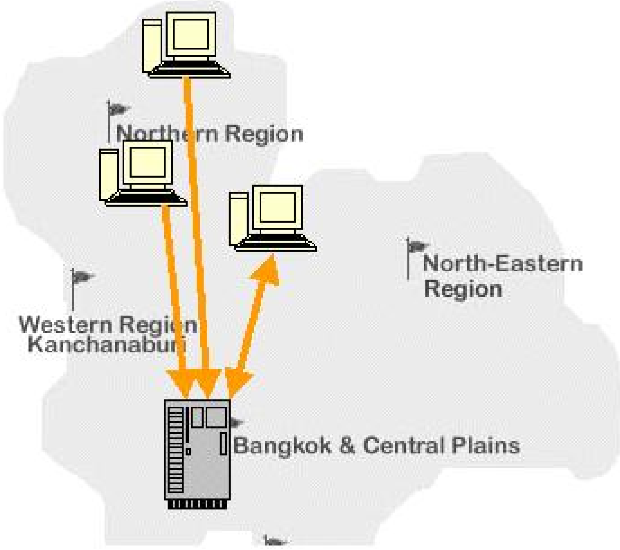


Figure 1: GIS operations on nothern provinces using centralized approach

In this paper, we propose the decentralized concept to be used in GIS. The risk of losing all data from the central server or not being able to operate on the server is going to be reduced. If this concept is adopted, each terminal province will pre-process selected data locally before sending it to subsequent servers to process. Figure 2 demonstrates the decentralized concept where data from nothern provinces are computed and sent to Pitsanulok province to perform some GIS operations. The disadvantage of this approach is that all the terminal server must contain software capable of interacting with each other. We propose the web service technology help us achieve the goal. Web services [2] permit us to create open standard distributed applications since they are built on open standard such as SOAP and XML [4, 6, 7]. These standards offer the clean communication among different programs, called services, situated on a local server. In particular, XML is used to encapsulate data being transfered between systems. Distributed applications created by web services will benefit from passing data in XML format which is a superior way of data representation.

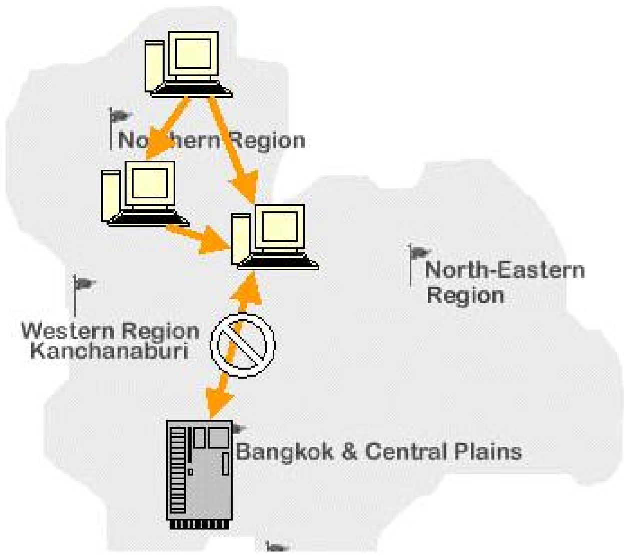


Figure 2: GIS operations on nothern provinces using decentralized approach

The organization of this paper is as follows. Sections 2 and 3 discuss about the two main concepts used in this paper, namely decentralized approach and web services. Section 4 presents the decentralized internet GIS system which is based on the web service technology. Finally, Section 5 draws the conclusion of this work.

## 2 Grid and Decentralized Paradigm

Emergence of the Internet has made it possible to interconnect millions of computer-like devices ranging from powerful supercomputers, computer farms, and all the way to mobile phones. Both computing power and interconnection network speed, according to the current trend, are increasing each year. Imagine how wonderful it will be to be able to draw computing power from these devices in the same way we are using electrical power—not knowing where the (computing) power is generated from. Several attempts have been made to utilize these internet computing resources by means of distributed computing such as SETI@home [10], fightaids@home and others. internet computing Nevertheless, there are many difficulties that need to be resolved before achieving the aformentioned goal.

The prominent challenge to bring up such a system which includes (but not limited to) heterogeneity of networked resources, concurrency of running applications, security, scalability of the system, failure management, and transparency of access for users. Many research efforts propose solutions to tackle these challenges from various view points. The concept of Grid and distributed computing have been around for a long time. There were many attempts to define standard software specifications in order to enable multiple programs distributedly kepted on different systems to be able to work together. These attempts include CORBA, DCOM and DCE where message passing is the norm to coordinate multiple distributed resources [11, 9, 1].

GIS research has recently adopt the concept of internet GIS in order to be able to share resources and reach out distance users. For years, GIS had relied heavily on centralized paradigm where the complete set of data is stored on single server. Remote computers can access the data via the interconnection networking with the high networking access costs. The grid and data distribution concepts have more improvement to offer based on this technology. By nature, locally produced data are stored and used by residences. This data can also be shared if there are requests for global planning project, e.g., in flood prevention, we must collect information from different water gates in order to globally predict the outcome. To minimize the communication overheads, small data can be replicated while large data can be split and distributed to other servers.

Our propose framework, allows users to select the services from the suitable location and remotely execute them where the data is stored. Number of servers will be utilized to execute corresponding applications. Only meaningful results are expected at the subsequent servers. This integration of multiple services is carried out by web services. With web services, we can create a complex distributed applications from multiple GIS services which provide limited functions to their local data. Using decentralized paradigm on the Grid will naturally support research in GIS.

According to the new framework, there are number of accesses to various servers which are not on the same administrative domain, i.e., each site manages its own security policy. Security issues are very sensitive and should not be neglected. Data will be protected with different security levels so that only authorized users can access certain level of security. Currently, we are investigating the Globus authentication/authorization based on certificate approach. Globus [8] is a grid middleware which help construct the grid. It provides set of toolkits which solve problems on security, resource identification/allocation and some other distributed computing primitives.

With the advent of high speed networking and many initiatives that will link international computing resources via high speed networking links, considerable GIS research tends to adopt the combination of decentralized and grid paradigms. Information sharing can be easily promoted since owners of such information can arrange different level of securities so that only public part will be shared which they cannot do when storing them in one central database.

## 3 Web Services

Web Service (WS) [2] is an emerging technology that will allow multiple internet based applications to interact. The input and output of WS are not for presentation as we see in web document (using browser to view the document content). In other words, the output of a web server is usually in the HTML format which mainly focuses on how document is shown on the viewer while WS concentrates on building distributed applications. The concept of WS is based on service-oriented architecture (SOA) [12] paradigm where a complete application can be constructed from various services which provide different functionalities. The interaction among these services relies heavily on the web communication protocol, namely SOAP (simple object access protocol) which is constructed by a de facto standard XML specification. In fact the three big IT companies, Microsoft, IBM, and SUN are currently adopting many open standards including these two technologies in their products. WS applications can range from simple stock price quoting, weather reporting to the more complex one like supply-chain interaction.

Web Services is so named due to some historical reason but it has so little to do with web applications, web browsers etc. The obvious common ground would be that they both use HTTP (Hypertext Transfer Protocol) to provide the primitive connectivity among different systems. The following example, would help readers understand the underlying differences between these two technologies. Consider the following traveling plan management (TPM) example. For a certain trip, we need to know at least the route to final destination, local weather, car retal company information, availability/rates of room. Using web application approach, one must surf the internet using web browsers to copy the aforementioned information from each web site. You may create a web site which will compile this piece of information and present it to your customers. Figure 3 demonstrates what we have discussed.

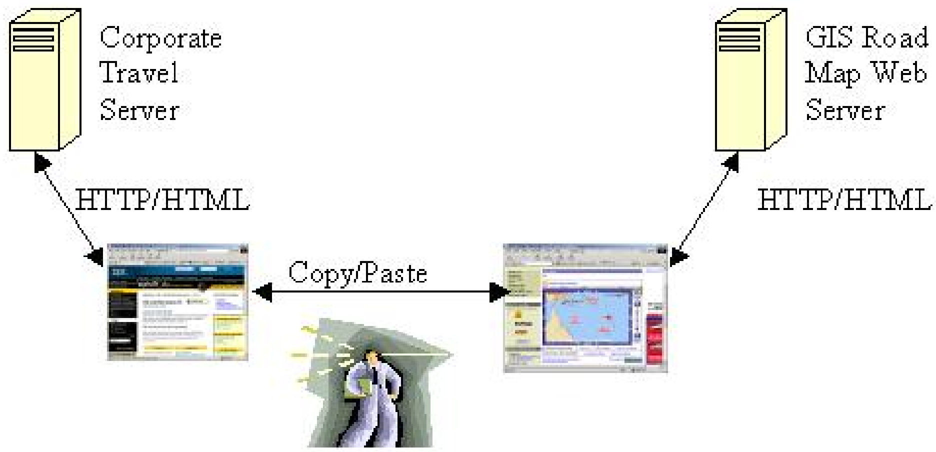


Figure 3: Service integration using web application approach

The information being keyed in to your application is from different sources which can only be provided to you in HTML format. Such information is only good for viewing purposes, e.g., using bold face and larger font with vivid colors to emphasize some important data. Nonetheless, this kind of presentation scheme is suitable for efficient interaction between two applications or also known as services. Data exchange can be done more efficiently with no beautification. Many software developers have tried to simplify the web application output so as to integrate multiple applications. However, as the applications become more complex, they also require more sophisticated input or even producing output that cannot be described by HTML.

Let's consider the previous case study, if we use the Web Service approach, each functionality in Figure 3 is autonomous. They can talk to each other directly without user intervention. In this scenario, data being sent among services are described by using XML format (see Figure 4). The services will be invoked remotely using SOAP, honoring distributed computing sytle.

Currently there are many standalone business services that can be wrapped in web service format. In our case, GIS road map, coporate travel web services from different locations can be integrated, thereby creating applications which can offer more complete services to customers (see Figure 5). In this scenario, there is no single GIS roadmap company that can provide complete information about all local roadmaps. This is very natural in many cases where there is no centralized data center. Web service approach can promote the concept of distributed computing.

## 4 GIS Web Service

There are numerous GIS applications used in Thailand due to the high demands of both government and business sectors to effectively utilize resources. Because of this demand, many software ven-

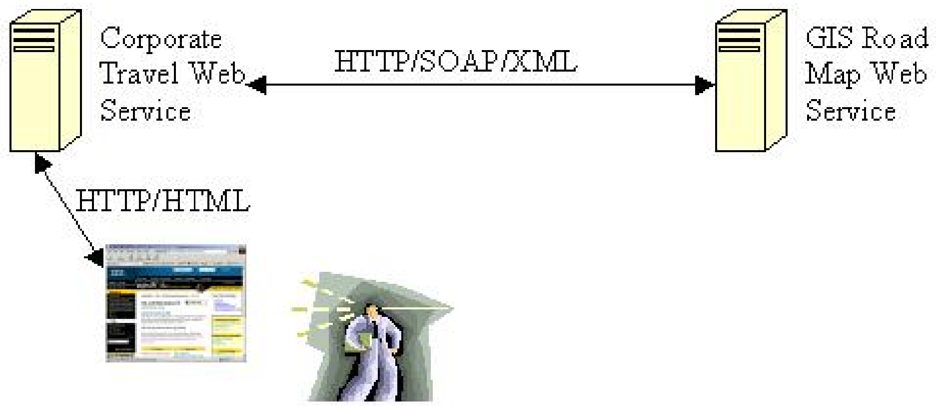


Figure 4: Service integration using web service approach

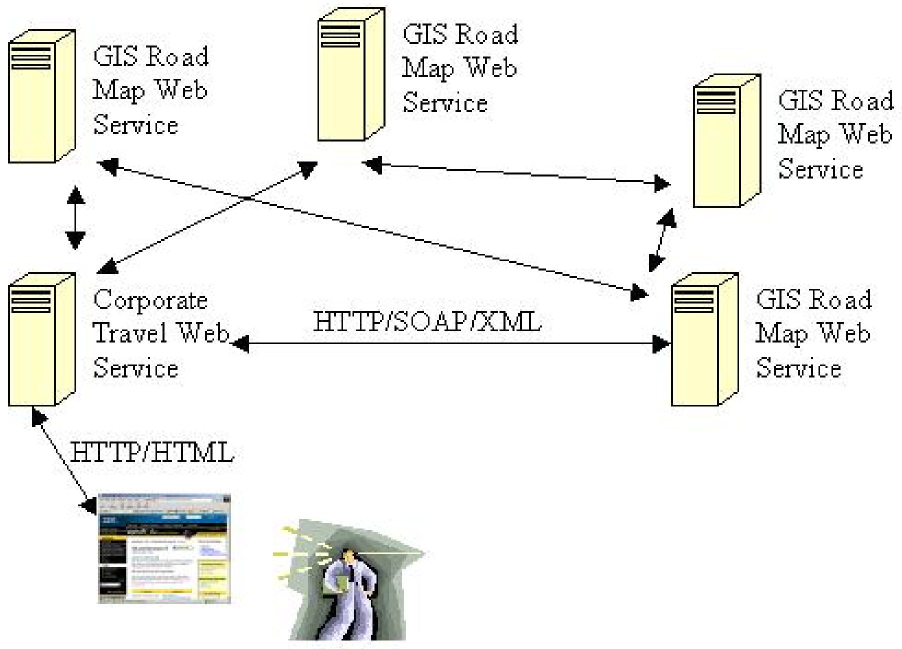


Figure 5: Putting it all together using web services approach

dors are competing to produce software with proprietary GIS data format hoping that it will become standard once they can dominate the market share. This creates many serious problems when conducting a large scale project (national or international level). We propose a GIS application which concentrates on providing GIS services through the Internet. To alleviate the incompatibility of data formats, we adopt the Geography Markup Language [3] (GML standard version 2.0) recommended by OpenGIS Consortium.

Figure 6 shows the interface of the proposed GIS web service. We developed a technology which will convert data from object-relational database[2] (ORDBMS) to GML data. This GML data will be input to web services such as GML graphics processing service and GML maintenance service. In this figure, the GML graphics unit processes GML data and create a 2D image using scalable vector graphic [5] (SVG) technology. This system is created for GIS governmental programs such as SchoolNet where the government would like to provide the internet access to students in remote schools and rural development project for Omkoi[3] district. The left and upper right panes of this figure present GML tree view and features respectively where the lower right pane shows the resulting SVG 2D image of Omkoi district.

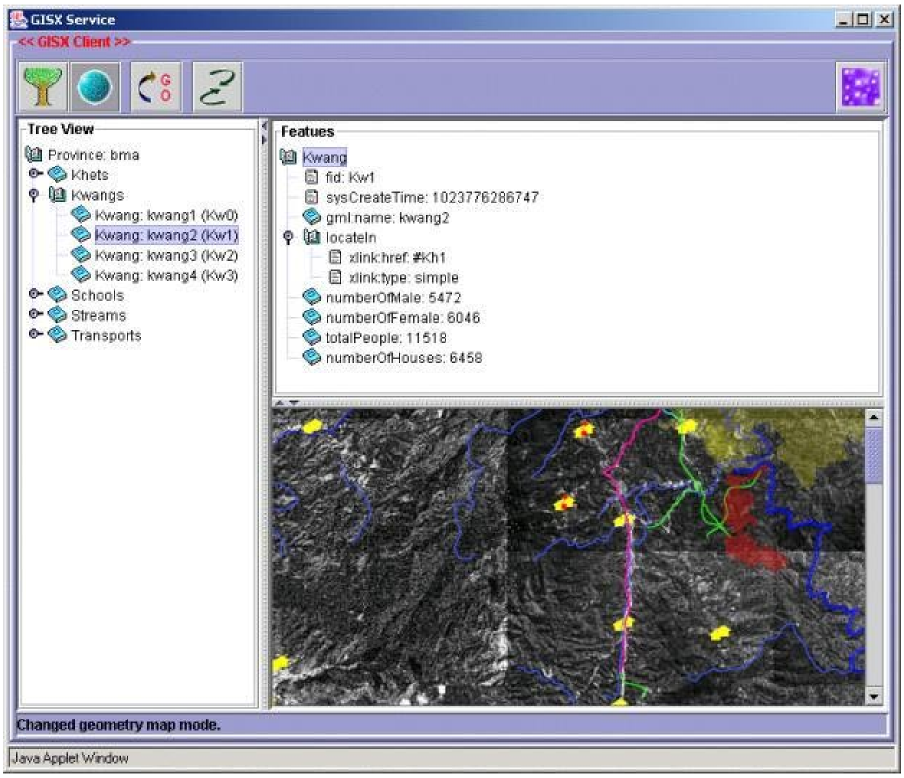


Figure 6: Current development of GIS web service

# 5 Conclusion

GIS in the 21st Century requires high performance of processing power and communication bandwidth. Therefore, in this paper we proposed the decentralized approach to help us distribute jobs and data across the network as well as sharing all resources. By doing so, the collaboration among different organizations can be promoted. Peer to peer network facilitates resource discovery of both geographic data and computing server. In order to create an effective collaboration and communication we proposed to use XML standard format of data excahange across geo-related organizations. The users can apply this technology for analyzing and making decision via the Internet. Web services technology play significant role to create distributed applications. Such technologies allow GIS

[2]We are currently using PostGIS from PostGRES relational database system

[3]Omkoi is a small district of Chiangmai province of Thailand. People who live in this area are suffering from nutritional difficiency.

developers to build independent GIS components which can be run on any platform using any language. The decentralized approach on the grid will increase robustness, reliability, performance and security. The key to success of this system is the collaborative efforts among different sectors of our GIS society.

## Acknowledgment

We would like to thank Cluster Computing and Grid Technology research team at the National Electronics and Computer Technology Center for providing us the ideas to effectively utilize the grid for the GIS web services. Finally, we would like specially thank DrSissades Tongsima for his contribution to this paper.

## References

[1] *Distributed systems concepts and design*. Addison Wesley, 2001.

[2] Jason Bloomberg. Web services: A new paradigm for distributed computing. *http://www.therationaledge.com/content/sep_01/f_webServices_jb.html*, 2001.

[3] OpenGIS Consortium. *Geography Markup Language (GML) version 2.0.* http://opengis.net/gml/01-029/GML2.html, February 2001.

[4] World Wide Web Consortium. *Extensible Markup Language (XML) version 1.0 (Second Edition)*. http://www.w3.org/TR/REC-xml, October 2000.

[5] World Wide Web Consortium. *Scalable Vector Graphics (svg) version 1.0*. http://www.w3.org/TR/SVG, September 2001.

[6] World Wide Web Consortium. *XML Schema Part 1 : Structures. W3C Recommendation*. http://www.w3.org/TR/xmlschema-1/, May 2001.

[7] World Wide Web Consortium. *XML Schema Part 2 : Datatypes. W3C Recommendation*. http://www.w3.org/TR/xmlschema-1/, May 2001.

[8] Ian Foster, Carl Kesselman, and Jeffrey M. Nick adn Steven Tuecke. Grid services for distributed system integration. *http://www.globus.org/research/papers/ieee-cs-2.pdf*, June 2001.

[9] Markus Horstmann and Mary Kirtland. Dcom architecture. *http://msdn.microsoft.com/library/en-us/dndcom/html/msdn_dcomarch.asp*, 1997.

[10] Eric Korpela, Dan Werthimer, David Anderson, Jeff Cobb, and Matt Lebofsky. Seti@home:massively distributed computing for seti. *http://www.computer.org/cise/articles/seti.htm*, 2000.

[11] Object Management Group (OMG). Corba overview. *http://www.infosys.tuwien.ac.at/Research/Corba/OMG/arch2.htm*, 1995.

[12] Michael S. Pallos. Service-oriented architecture: A primer. *http://www.eaijournal.com/PDF/SOAPallos.pdf*, December 2001.