A MODEL OF MULTI-SERVER PARALLEL-PROCESSING IN WEBGIS

Shijun DENG, Lingkui MENG

School of Remote Sensing Information Engineering, Wuhan University, China 430079

Commission Ⅱ, WG Ⅱ/3

KEY WORDS: Multi-server, Parallel-process, AOI, Dynamic Balance Algorithm, WebGIS

ABSTRACT:

As the development of Internet technology, it's become the hot trend to issue all kinds information and to provide information service on the web. This trend makes the geographic information issued on the web and developed greatly. The technique to provide the service of geographic spatial information is called as WebGIS.
WebGIS enables the proceeding of spatial analyzing done on the network and makes all the spatial information data accessed through the Internet. WebGIS enlarges the source of spatial data, enhances the stylish of GIS and promotes the depth and extend of the application of geographic spatial information. It's very significance to the subject of GIS.
As WebGIS applies extensively and provides more functions, more and more people visit the web station that provides the services. The capacity of the server has becoming a bottleneck that prevents the development of WebGIS. To improve the server's capability and response speed, the host machine has been enhanced again and again. Compared with the expensive high-ability host computer or the SMP computer, multi-server has been applied extensive in electron-business and the web information service for the lowest price and high-ability.
This paper proposes a model to solve the problem that single server can't assume heavy burden, which provides the service of WebGIS, and illuminates how to balance the server's burden in this system dynamically. We can serve a great number of clients throughout the world that need the service of WebGIS through the Internet based on the model of multi-server.

1. INTRODUCTION

As the development of Internet and it is becoming the hot field to publish all kinds of information in the current, the technology of WebGIS, which issues the geographic spatial information through the Internet, has developed in the field of geographic information.

In functionally, WebGIS refers to the transplantation of spatial analyze and specialty model to the Internet and the opening to all of the users of the spatial data information. It's significance to develop and construct the GIS for enlarging the source of the spatial data, enhancing the time-ability of geographic information and promoting the application of the geographic information.

As the development of Internet technology and the broadening of it's application field, geographic Information System has been applied extensively in the Internet environment, which gripes more and more eyes on itself. So, it requires the server, which provides the service of GIS, be high-ability to process a large amount of request from the clients throughout the world. To realize this, people have to upgrade and replace their servers to enhance the GIS server's response speed and ability. Compared with the expensive high-ability gigantic computer or SMP workstation, multiple server that work together has been applied extensively in electric business and Internet information publishing because of it's high-ability, extensibility and lower cost.

Currently, It will become one of the hottest fields to be researched on how to take advantage of multi-server wholly and efficiently, to adapt the needs of the WebGIS services and to serve for the country basic construction, such as city planning management, country resource investigating, circumstance protection, traffic construction and management. It will provide the scientific evidence for the decision-making people to make the planning, designing and management decision. In building the basic geographic information system service center we can adopts the model.

2. THE SERVICE MODEL OF MULTI-SERVER/CLIENT IN WEBGIS

The service model of WebGIS has changed from two-tier browser/server to three-tier client/server model. The reference (Yuan, et al., 2000) presents a model of WebGIS based on three-tier client/server model. This paper divided the whole system to three modules in logic about the GIS application platform and the three models are the module of server end, the module of client end and the module of middle function processing.

In the system, there exist a modulating server to receive the client requests unique and compute the servers' burden dynamically and assign the request task to one of the server whose burden is the lightest. The following figure is the model of multi-server/client processing flow-figure:

As is shown in this figure, the main body of this model is divided into three parts, server (including data storing server, transaction processing server), modulating server (also called as middle function part) and the client. The following we'll discuss the function and the purpose of every module detailed.

2.1 Server End

The server is composed of the data server and the transaction-processing server.

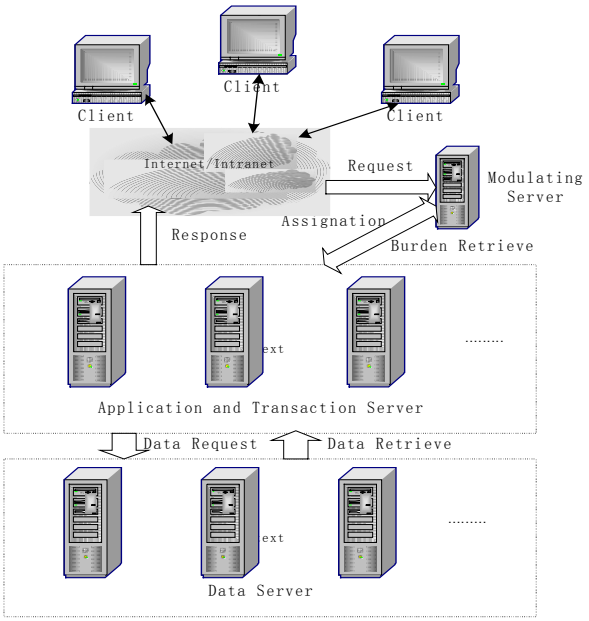

Figure 1. The model of Multi-server/client

2.1.1 Data Server

In the model of multi-server/client of WebGIS service, we must store complicated and large data in distribute serves because of the data is so large in amount and comes from all different sources, such as satellite remote sensing, surveying on the pot, spatial database established from map digitalize.

The data server provides the deposit of storing the spatial database and exchange data with the transaction-processing server. It provides the function of maintenance, appending, deleting and querying the spatial data through the module of data storage and exchanges the data result through high-speed LAN or Internet. The data server processes the data request from the transaction server and then returns the data set that client need to transaction server.

The data can be stored in different physical machine. But the data server can be the same physical machine with the transaction server. It provides the data service for the transaction

2.1.2 Transaction-processing Server

The transaction-processing server processes the web request from client, such as response the data request and provides the GIS spatial data for the client, exchanges the data with the data server, and modifies, appends or get the spatial data through communicating with the module of data server.

It also controls the user's purview and validates the user's statue, and does some complicated mathematic computation and transaction processing such as measuring the spatial distance, spatial transforming, analyzing buffer-area and spatial querying and searching, the computing result will return to the client immediately.

As the enlarging and increasing of web request, transaction server couldn't be single machine in physical and it must be in cluster. It has the unit URL to the web users through the modulating server and provides the service in parallel processing. That will enhances the server's response ability and capacity.

2.2 Modulating Server (Middle Function Module)

The modulating server is used to receive the client requests unique and compute the servers' burden dynamically in this system and assign the request task to one of the server whose burden is the lightest.

This server receives the requests from the client and the burden parameter of every server in the system. This server will use the algorithm of evaluating, which computes the burden of each server, to calculate each server's response probability in real time and then assigns the response task to one of the server which probability is the lightest. This algorithm will be discussed lately in this paper .

2.3 Client

There are three modules in the client: Core of spatial data management, module of GIS integration function, and module of communication service.

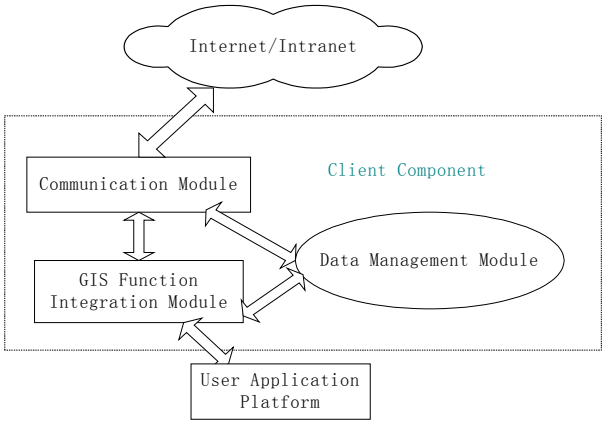

Figure 2. The module of Client

2.3.1 Core of Spatial Data Management

The core of spatial data management is the basic core module of all of the client's modules. This module is in charge of managing local spatial data, responding the request coming from super application program and operating or processing the data correspondingly, for instance, it manages the data obtaining from the server and stores them, modifies them, appends them, deletes them and so on. It can also do spatial analyzing and querying the local spatial data, which is obtained from server, and sending the request that needs data service.

2.3.2 The Module of GIS Function Integration

The module of GIS function integration is the basic service module of GIS, which provides special service for every walk of life, such as city planning, city construction and management, traffic management and landlord prospecting, through spatial analyzing and querying, which is referred to spatial transforming, buffer area analyzing, pile-up analyzing and network analyzing.

2.3.3 The Module of Service Communication

This module is the basic communication core of the client, which communicates with the server, sending the data request, which coming from the module of data management to the server and obtaining the data that client needs.

3. THE MODEL OF MULTI-SERVER PARALLEL PROCESSING THE WEB REQUEST

The request that the WebGIS server receives can be divided into three kinds approximately, one is the request of needing data service, one is the request of data operating and the third is the request of data computing. In the mechanism of service system based on three-tiers server/client, the client is not only the machine that receives the result from the server's processing, it also has some data processing ability in use of local computer's processing ability. And so, only when the mission involved large-scale computing and processing, the client sends the data request to the server.

The following we'll talk about the model of multi-server parallel process the web request.

Suppose there are M servers and the N type of request, these N type request can share the same one server, so there will be M× N queues of request. We assume 0≤ i ≤M, 0≤ j ≤N，then we can use q_{ij} to mark the request queue and use M_{ij} to mark the max number of request for each kind of queue. So, that is: $q_{ij} \leqslant M_{ij}$

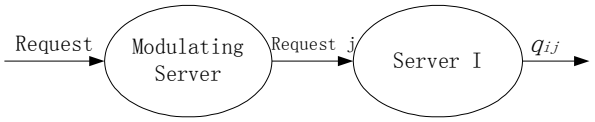

Figure 3. The model of the Response Queue

Assume that the request j's reach velocity is λ_{ij}, it will be sent into the queue of q_{ij}. If the M queues' capacity those responses to the task j get to the upper limit, the procedure will suspend, that is these servers cannot response for the request because of their capacity.

The server's service velocity is V_{ij}, using S_{ij} to denote that the server i responses the request j. Using $T(q_{ij})$ to mark the request's time excursion, the operating velocity u_j, and the probability Pij to express that the request j entering the queue q_{ij}.

When considering the service project that each task will response in the shortest expectation waiting time, that is to say after each kind of request is send into one queue, the server can accomplish the task in the shortest time and return the response result. So the modulating project is the following.

$$q_{ij}: \ P_{ij} = \max(P_{kj}) \quad k = 1,2.....M \tag{1}$$

$$\text{And: } P_{kj} = \begin{cases} SEDR(k) & if\ q_{kj} \le M_{kj} \\ 0 & Otherwise \end{cases} \tag{2}$$

$$SEDR(k) = \begin{cases} \dfrac{T(q_{kj})}{V_k} = \min(\dfrac{T(q_{kj})}{V_k}), k = 0,1.....M \end{cases} \tag{3}$$

In the multi-server's project of assigning the task, there is a modulating server that receive the client's request consolidate. This server compute the Pij value in according to the burden information returned from each transaction-processing server and then assigns each task to the server which Pij value is the largest.

4. DYNAMIC BALANCING THE SERVER'S BURDEN

Dynamic balancing the server's burden is referred that in each time the burden of all of the server will be equality approximately. To balance the server's burden and response for the client's request in real time, there are some methods to realize it.

4.1 Indexing the Data in Hierarchical Scale

When the user enters the system, he will see the whole area of the map and then he will moves to the area interested through zooming in, zooming off and panning. For not all of the users' interesting area is same, the system can't serve each user in the same way. When the system starts, the servers need not to send all of the data of the whole area to the client. To solve this problem, to establish the spatial data database in different scale is the thing the server must do. Although that way makes the capacity of data increase, it can decrease the quantity of data sending to the client.

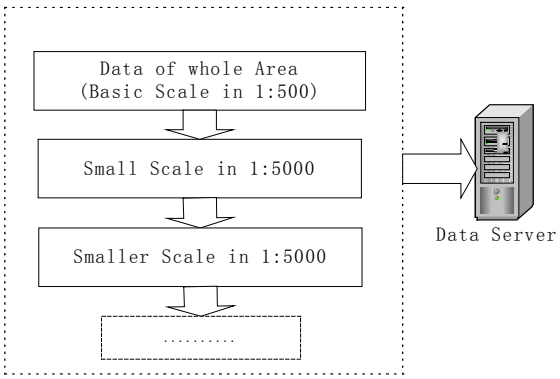

Figure 4. Indexing Data in Hierarchical scale

4.2 AOI (Area of Interested) Data Area Management and the Buffer Area Setting

In order to the client can operate the map data in real time, one common way is to download the current area data in several times to the client. So when the user panning on the map, he can see the map like the map data is in the client machine because the client doesn't need to obtain the data from the server during his operation. But this method has some shortcomings; firstly, in the beginning of the system starts, there is large data to be transfers to the client, the data maybe in several times of which the client needs, and it will cost long time to accomplish this. Secondly, the redundant data needs to be transmitted maybe large, so the server needs more time to response the request from one client and the server will be blocked up.

The spatial data, including geometry data and attribute data, was stored in the data server and every client is a platform that provides the service of all kinds of GIS function. Every user is interested in his own area and function service, so the client work area was named as the Area of Interested (AOI). Each AOI data area can be expressed using the following way:

AOI:（Xmin, Ymin, Xmax, Ymax, Scale）

When one of the users move or pan on the map, his AOI data area will change through his operation. The client then communicates with server in use of the module of service communication and send the message that his AOI data parameter. The server receives the information of client's AOI and then adjusts the correspondence client's AOI data service.

But sometimes, the user will continuously operate (Zoom and pan) on the map according to his need and his AOI area will change continuously. To avoid the frequent requesting data service and sending too many information to the server, we need to set the buffer area of AOI.

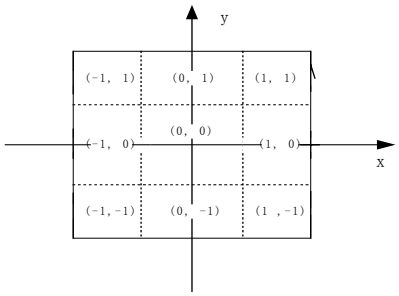

Figure 5. Sketch Map of Descartes coordinate

As in the figure4.2, the shadow part (X1，Y1，X2，Y2) is the current AOI data area, this part is expressed by Descartes coordinate. We establish the current AOI area's buffer area in use Descartes coordinate too. If current area (X1，Y1，X2，Y2) is marked as (0, 0), then (-1, -1) marks the area (X1-（X2-X1）,Y1-（Y2-Y1）, X1, Y1), the same to the former, (-1, 0), (-1, 1), (0, -1), (1, -1), (1, 0) and (1, 1) express the area(X1-（X2-X1）,Y1-（Y2-Y1）, X1, Y1), (X1-（X2-X1）,Y1, X1, Y1+(Y2-Y1)) and as so.

When the client send the data request to the server, the server will return all the data including the current AOI area's data and the buffer area's data in correspondence scale to the client. So when the user operates the map data, he can zoom and pan conveniently and get the result in real time because the data he is operating is in the local computer. Only when the client area is out of the data area's limit or the zoom scale is out of the threshold that system set, the client send the data service request to the server again. Then the server compute the AOI data area and it's buffer area dynamically according to the parameter the client sends and provide the data service correspondingly.

4.3 The Algorithm of Multi-server Balancing the Burden Dynamically

During the server responding to the client's request, the server's burden maybe are not equal to each other for every kind service's serving time and the estimated time are not equal. Now let's use the service time to weigh the server's burden, then Ti denotes the burden of the server I and $Q_{ij(k)}$ denotes it's task queue that will be responded. So the total burden is:

$$T_i = \sum_{j=0}^{N} \sum_{k=0}^{b_{ij}} Q_{ij(k)} \tag{4}$$

In the server system, the most burden server is server Mi, which Mi satisfy with:

$$M_i = \{ M \mid M \in \max(T_1, T_2 \cdots \cdots T_M) \} \tag{5}$$

And in the server system, the lest burden server is server Mk, correspondingly Mk satisfies with:

$$M_i = \{ M \mid M \in \max(T_1, T_2 \cdots \cdots T_M) \} \tag{6}$$

The difference between the most burden server and the lest burden server is:

The burdens Ti will be computed out by the modulating server in real time and then adjust all of the servers' burden. The following formula is the presentation of how the modulating server equipoise the burden of the server.

$$(\Delta T \geq \delta) \vee (T_{mi} \geq \delta \wedge T_{mk} = 0) \tag{7}$$

In this formula, δ is the threshold that giving by the system.

Using this way, every server's burden in this system will be in equal response time approximately.

5. DISCUSSION

The model of multiple servers can be divided into modulating server and web application server in logic. The modulating server's main function is to accept the requests from client and

assign them to the web application. The web application server then processes the request and provides the service in GIS. In this model, modulating server is unique, but application web server is multitude.

Based on multi-server service, this paper suggests some methods to parallel-process the request of the client to construct a base geographic information service center. And this paper analyzes the species of the web request and explains three kinds of methods to realize the dynamic balance of the server burden correspondingly.

This paper considers that the web request has three types. One is the request of data service, which is to get the data of a new area from the data server; the other is the request of area operation, which needs operate on the area, such as area zooming in and zooming out, area spanning; the third is the request of data computing, which needs the server compute out the result of analyze such as area overlapping, spatial best route analyzing.

Based on previous analyses, this paper suggests three methods to balance the burden of the server to make the server provide better service for clients.

One is to index the spatial data in various scales. The data of an area stored in spatial database in various scales maybe need more space, while the client request the data of the area, the server can't send all of the data of this area, one of the reason is that the client don't need them all, the other reason is that the capacity of the net can't do this. So, indexing the spatial data in various scales can make the server spend a little time to get the data the client needs and send them through the net.

The other is to set the area of interested (which is called as AOI) of client. The client's current vision area is in some scale, this area has eight-adjacent domains. AOI method looks the client's current vision area and it's adjacent area as an AOI, which is meaning that this area is the client's active area, so the server can send all of the data to the client in corresponding scale.

The third is to balance the burden of the server dynamic in the task of computing. When the client sends the request of computing to the server, the task will be assigned to one of the servers. But in the proceeding in serving the client, there maybe exist the tasks that spend more time than estimated, so some of the server will be in idle, some will be in busy. To solve this problem, this paper suggests an algorithm to balance all the servers' burden.

The model of this paper expounds adapts to construct the base center geographic information that possesses of large spatial data and serves large quantity clients.

REFERENCES

CHANG Yan-qing, 2000. Discussion about the Organization of Larger Amount of GIS Data. *Information of Remote Sensing & Application Technology*,No. 2.

HAN Haiyang, Huang Jianming, Gong Jianya, Yuan Xiangru, 1999. Distributed Processing and Management of Multi-data Source,Hypermedia Geo Information. *Journal of Wuhan Technical University of Surveying and Mapping*, Vol.24, No.3.

JIANG Xiao-hong, Pan Zhi-geng, Yang Meng-Zhou, Shi Jiao-ying, 2000. Divisional Data Management Mechanism for DVE Systems in Multiple Server/Client Model. *Chinese Journal of Computers*, Vol 23, No. 8.

LIU Zhao-ying, FANG Bin- xing, HU Ming-zeng, ZHANG Yi.An 2001. Effective Dynamic Load Balancing Method. *Journal of Software*, Vol. 12, No.4.

SHAN Zhiguang, DatQiong-bai, Lin Chuang, Yang yang, 2001. Integrated Schemes of Web Request Dispatching and Selecting and Their Performance Analysis. *Journal of Software*, 12 (3).

YUAN Xiangyu, Cheng Nengcheng, Gong Jianya, 2000. A Distributed Hyper Map Model for Internet. *Geo_spatial Information Science*, 13(4), pp. 9-15.