

网络环境下矢量数据高效并行可视化方法

郭明强^{1,2} 黄颖² 吴亮^{1,2} 谢忠^{1,2}

1 中国地质大学(武汉)信息工程学院,湖北 武汉,430074

2 教育部 GIS 软件及其应用工程研究中心,湖北 武汉,430074

摘 要:针对网络环境下传统的矢量地图可视化模型中矢量数据可视化效率低,地图服务器集群并行处理能力弱的难题,提出一种基于矢量数据要素空间分布的矢量数据高效并行可视化方法。研究了矢量数据空间分布信息的采集、检索和分析原理,重点阐述了矢量数据实时可视化任务的分解与并行处理流程,最终实现了矢量数据的高效并行可视化,达到了充分利用地图服务器集群中并行计算资源的目的。仿真实验结果证明,矢量数据高效并行可视化方法可以提升网络地图服务集群的并行处理能力,满足大用户量并发访问的需求。

关键词:网络地理信息系统;网络环境;矢量数据;可视化;并发访问

中图法分类号:TP301.6;P208 **文献标志码:**A

网络地图服务是网络地理信息系统中最常见和最基础的功能服务之一^[1]。在对空间数据实时性要求较高的业务系统中,业务办理人员需要的地图图像必须与矢量数据实时保持同步,以确保案件办理的唯一性和准确性^[2-3]。

本文针对网络环境下矢量数据并发访问效率低的问题进行分析,采用了对计算开销较大的矢量数据地图可视化请求任务进行均衡分解的策略,对矢量数据实时可视化任务进行分解,以借助集群环境中多个并行计算资源并行的完成矢量数

据可视化任务,达到矢量数据高效可视化的目的。通过仿真实验与传统的矢量地图可视化模型进行对比,结果表明本文提出的方法能有效地缩短网络环境下大用户量并发访问时的平均响应时间,提升网络地图服务的并行处理能力。

1 矢量数据高效可视化方法

1.1 传统的矢量地图实时可视化模型

传统的矢量地图实时可视化模型中(图 1),

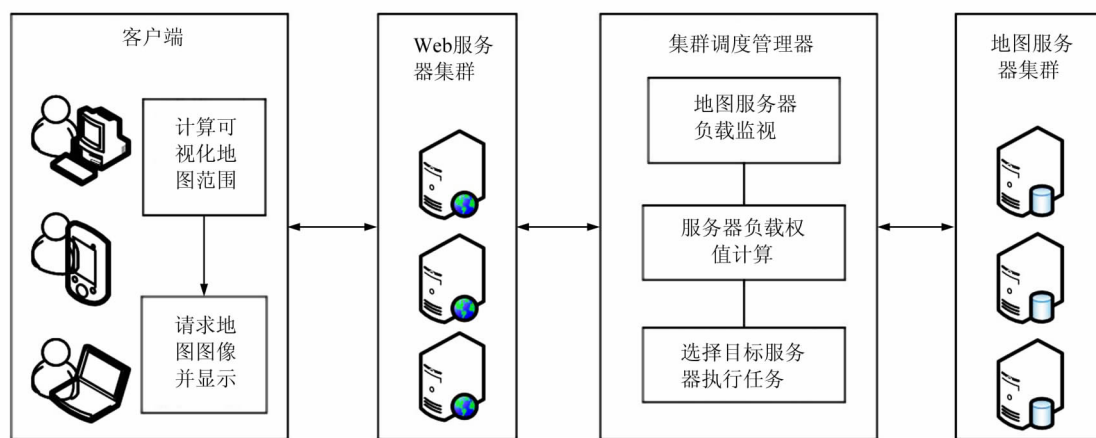


图 1 传统的矢量地图可视化模型

Fig. 1 Traditional Vector Map Visualization Model

收稿日期:2013-07-04

项目来源:国家十二五科技支撑计划资助项目(2011BAH06B04);中央高校基本科研业务费专项资金资助项目(CUGL140833);中国博士后科学基金资助项目(2014M552115)。

第一作者:郭明强,博士。主要从事网络空间信息服务关键技术研究。E-mail: gmqandjxs@163.com

通讯作者:吴亮,博士,副教授。E-mail: wuliang133@189.cn

客户端只负责实时计算矢量地图可视化窗口中的地图范围,将请求传递给 Web 服务器集群,然后到达集群调度管理器,集群调度管理负责对地图服务器集群中各服务结点负载状态进行监视,通过计算服务器负载权值选择合适的服务结点处理自客户端的请求任务^[4-5]。这种传统的矢量地图可视化模型存在两个弊端:① 客户端只负责计算矢量地图的可视化范围,其他的计算均由地图服务器完成,使得地图服务器的负荷较重,降低了地图服务器的处理能力;② 集群调度管理器只能按某种负载分配策略将来自客户端的任务分配给合适的地图服务器结点进行处理^[6-7],但不能对单个请求任务进行分解并调度地图服务器集群中的多个服务器同时并行处理,因此,对于大数据量矢量地图实时可视化应用,由于数据量大,数据检索和绘制时间长,传统的矢量地图可视化模型无法满足需求^[8]。

1.2 矢量数据高效可视化原理

为解决大数据量矢量地图实时可视化效率低的难题,本文对可视化任务进行“分而治之”的策略,将可视化任务进行分解并调度多个地图服务器并行完成矢量地图可视化任务。该策略需要解决以下两个关键问题:① 如何对客户端可视化任务根据各个服务器结点的处理能力进行公平的分解,使多个服务结点能够尽量在同一时间内完成任务;② 如何实现客户端与服务器端协作完成可视化任务的分解,避免由服务器端单独完成分解过程造成性能瓶颈。

为了解决以上问题,本文提出了客户端与服务器端协作,对矢量地图可视化任务根据要素实际分布情况进行公平分解的方法。首先,需要在服务器端为大数据量矢量地图构建要素分布信息库,本文按“横向网格,纵向多级”的方式,用四叉树组织形式对矢量地图范围进行分级网格化,存储各级别下每个网格单元的信息及其中分布的要素数据量。一个矢量地图中一般包含多个图层,在某些实际应用中,为了提高可视化效率,一般会为每个图层设置最大和最小显示比率,控制每个图层的显示级别,以减少同时显示的图层数量以减少可视化范围内要素数量;而有些应用中需要在任意显示级别任意地图范围内看到所有要素的分布情况,地图中的所有图层在任意显示比下均可见(例如需要查看任意范围甚至整个城市的城镇地籍矢量地图中所有房屋在全市范围内的分布情况),这种应用场景下,要素分布信息库中每一级网格中存储的总的要素数据量是相同的,本文

提出的分解方法主要是针对这种应用情况,对于第一种应用情况笔者将在下一步继续进行研究。

其次,需要实现的是如何在客户端对矢量地图可视化域进行处理,以从服务器端要素分布信息库中检索到当前可视化域内的要素分布信息网格集合。本文采用对可视化域进行格网化的方法,将客户端矢量地图可视化窗口细分为多个网格单元,在实际应用中根据矢量地图中要素的分布情况设置网格单元数量,如矢量地图中要素的分布较为均匀,可以选择较小的数值;如果要素分布非常不均匀,为了更准确地分析当前可视域内的要素分布信息,以实现可视化任务的公平分解,则需要选择较大的网格单元数量。在应用部署后,可以通过调整该值测试矢量地图可视化任务的平均响应时间,以确定合适的数值,使系统性能最优。

对客户端可视域进行格网化后,计算出客户端可视域网格单元的逻辑范围跨度,将该值和可视化地图范围一起传递给服务器端的集群调度管理器,集群调度管理器根据客户端可视域网格单元逻辑范围跨度检索与之最接近的矢量地图分布信息网格级别,然后根据可视化地图范围计算出客户端可视域内的矢量地图分布信息网格集合,将其与地图服务器负载权值信息一起返回给客户端。客户端获取到以上信息后在本地进行缓存,并立即对网格集合进行扫描分析,根据地图服务器负载权值对当前可视化地图范围进行划分,得到多个可视化任务子集,然后向多个地图服务器并行发送请求,实现矢量地图可视化任务的“分而治之”,缩短大数据量矢量地图图像的处理时间。

1.3 矢量数据高效可视化流程

1.3.1 矢量数据要素分布信息采集

要实现本文提出的矢量数据高效可视化,必须对矢量数据可视化任务能够进行空间分布特征分析,而分析的前提是需要提前完成矢量数据要素分布信息的采集。其具体步骤如下。

1) 将矢量数据的各个图层的矢量数据的空间范围进行扩展,生成一个正方形的空间范围,将该范围作为各个矢量图层第 0 级空间分布网格的空间范围。

2) 采用四叉树的组织方式,计算不同级别矢量数据空间分布网格中各个网格单元的空间范围,从原始矢量图层中检索分布于各个网格单元中的空间要素的数量。

3) 将上述过程得到的结果信息存储于矢量数据要素分布信息库中,为矢量数据可视化任务

的高效处理提供数据来源。

该步骤是在需要对矢量数据可视化性能进行优化时对矢量数据进行的预处理,一般可在矢量数据发布前进行,以减少矢量数据实时处理过程中额外的计算开销。

1.3.2 矢量数据可视化任务要素分布信息网格检索

在网络环境下矢量数据可视化请求的处理过程中,要对实时可视化任务进行分解,首先需要对可视域参数进行分析,查找到可视域内的矢量数据要素分布信息网格单元集,具体步骤如下。

1) 将客户端的矢量数据实时可视化请求中的地图范围进行均匀的格网化处理,计算出每个格网单元的空间范围。

2) 将客户端请求的可视域的范围和每个格网单元的空间范围信息发送到服务器端。

3) 服务器在接收到客户端请求的矢量数据可视化范围和客户端可视域格网单元的空间范围后,首先根据可视域格网单元范围信息查找到对应级别的矢量数据要素分布信息网格,然后再根据实时可视化范围从要素分布信息库中检索到当前可视化范围内的矢量数据要素分布信息网格单元集合。

4) 采用文献[4]中的方法计算出当前集群中各个地图服务器的负载权值。

5) 将检索到的矢量数据要素分布信息网格集和各个地图服务器的负载权值反馈给客户端。

1.3.3 矢量数据可视化任务分解与并行处理

要实现矢量数据的高效可视化,不是简单的将矢量数据可视化任务对应的空间范围进行划分,而需要根据矢量图层中空间要素的实际空间分布情况对实时可视域进行公平的分解。具体步骤如下。

1) 客户端根据服务器端返回的矢量数据要素分布信息网格集信息,计算其与客户端实时可视域之间的交集的空间范围。

2) 根据每个要素分布信息网格单元与可视域的交集和要素分布信息网格中的要素数量,计算出各个交集中分布的空间要素数量。

3) 根据服务器端返回的地图服务器集群中各个地图服务器的负载信息,计算出与各个地图服务器负载能力相一致的空间要素总量。

4) 对要素分布信息网格进行逐列扫描,依次为每个地图服务器计算出各自对应的要素分布信息网格集合信息。

5) 计算各个地图服务器对应的要素分布信

息网格集的外包矩形范围,作为各个地图服务器需要处理的矢量数据可视化子任务的子范围。

1.3.4 矢量数据可视化并行处理

客户端将实时可视化任务进行分解后,将计算出的各个地图服务器对应的可视化范围并行地发送给地图服务器集群中的各个计算结点,进行并行处理,各个地图服务器完成各个地图可视化子任务的处理后直接返回给客户端,实现矢量数据可视化任务的并行处理,成倍提高网络环境下矢量数据可视化任务的效率,从而实现矢量数据的高效并行可视化。

2 实验设计与结果

2.1 实验参数

本文使用位于高速局域网内的 IBM 刀片中心部署服务器端环境,网络地图服务器集群、集群调度管理器、Web 服务器之间使用千兆局域网络连接,采用城镇地籍矢量数据对本文算法进行实验,并与传统的矢量地图可视化任务调度模型进行了对比。各图层的要素几何类型和要素数量如表 1,所有图层要素数据量总计为 7 465 146。

表 1 各图层要素数量
Tab. 1 Number of Each Layers

序号	图层名	要素几何类型	要素数量
1	街坊	区	1 597
2	宗地	区	431 237
3	房屋	区	882 896
4	综合线	线	118 590
5	界址线	线	3 148 458
6	综合点	点	3 591
7	控制点	点	40 436
8	界址点	点	2 838 341

2.2 实验结果分析

1) 不同类型图层分解的有效性。为了验证本文提出的方法对点、线、区图层都是有效的,使用表 1 中所示的房屋、界址线、界址点等 3 个图层对本文提出的方法进行测试,为避免客户端可视域格网数的影响,将实时可视化范围划分为固定的 128 个单元格。测试结果数据如表 2 所示。

从表 2 中所示结果数据可以看出:本文提出的方法针对点、线、区图层均能进行有效均衡的分解,特别是对于界址点图层,分解后的平均偏差能达到 0.06%。

2) 网络环境下矢量地图服务性能测试。本文提出的方法的目标是提升网络环境下矢量数据可视化的效率,为了验证本文方法的有效

表 2 点、线、区图层测试结果

Tab. 2 Test Result of Point, Line and Polygon Layers

图层	要素数量	分解结果序号	分解后实际要素数量	偏差 /%	平均偏差 /%
房屋	882 896	1	219 237	0.67	1.26
		2	217 650	1.39	
		3	216 428	1.95	
		4	218 496	1.01	
界址线	3 148 458	1	788 615	0.19	0.62
		2	781 624	0.70	
		3	780 359	0.86	
		4	781 393	0.73	
界址点	2 838 341	1	709 626	0	0.06
		2	710 015	0.06	
		3	708 744	0.12	
		4	709 956	0.06	

性,使用不同的并发访问用户数对传统的矢量地图可视化模型和采用本文提出的方法进行改进后的模型进行对比测试,测试结果如表 3 所示。可以看出,改进后的模型与传统模型的平均响应时间的差值随着并发用户数的增大而增大,说明并发访问量越大,改进后的模型在相同时间内的处理能力提高越明显。说明本文提出的方法能有效的提升网络环境下矢量数据的并发访问请求处理能力。

表 3 不同并发用户数测试结果

Tab. 3 Test Result of Different Number of Users

并发数	改进模型平均响应时间/s	传统模型平均响应时间/s	差值/s
1	0.625	0.84	0.215
5	3	4.5	1.5
10	6.5	9	2.5
15	10	14	4.0
20	12.5	18.5	6.0
25	16	23	7.0
30	22.5	32.5	10.0
35	29	39.5	10.5
40	32.5	45	12.5
45	36	50.5	14.5
50	40	58.5	18.5

3 结 语

本文针对传统矢量地图可视化模型对于矢量数据实时可视化效率低,并行处理能力弱的问题。对矢量数据实时可视化任务的特点进行分析,针对矢量数据特有的空间分布特性,通过预先采集矢量数据的空间要素分布信息,构建要素分布信息库,通过客户端与服务端端的协作,实现矢量数据实时可视化任务的分解,调度集群中的多个服

务器并行处理计算开销较大的矢量数据实时可视化任务,以提高网络地图服务器集群的并行处理能力。使用不同的虚拟用户数对采用本文提出的方法的网络地图服务模型进行了仿真测试,试验结果证明,采用本文提出的矢量数据高效可视化方法进行改进后的网络地图服务模型具有更好的抗高负载能力,可以提升网络地图服务器集群对矢量数据并发访问的处理能力,满足大用户密集访问的应用需求。

参 考 文 献

- [1] Wang Hao, Yu Zhanwu, Li Rui, et al. Cost Based Load Balancing for Network Geographic Information Service[J]. *Acta Geodaetica et Cartographica Sinica*, 2009, 38(3): 242-249(王浩,喻占武,李锐,等. 基于开销代价的网络地理信息服务负载均衡算法研究[J]. *测绘学报*, 2009, 38(3): 242-249)
- [2] Mustafa N, Krishnan S, Varadhan G. Dynamic simplification and Visualization of Large Maps[J]. *International Journal of Geographical Information Science*, 2006, 20(3): 273-302
- [3] Yang Bisheng, Purves R, Weibela R. Efficient Transmission of Vector Data over the Internet[J]. *International Journal of Geographical Information Science*, 2007, 21(2): 215-237
- [4] Huang Ying, Xie Zhong, Wu Liang, et al. A Web-GIS Model Based on Cluster Scheduling Load-balancing Algorithm[J]. *Earth Science—Journal of China University of Geosciences*, 2010, 35(3): 407-414(黄颖,谢忠,吴亮,等. 基于聚类调度负载均衡的 WebGIS 模型[J]. *地球科学(中国地质大学学报)*, 2010, 35(3): 407-414)
- [5] Zhang Zhongju, Fan Weiguo. Web Server Load Balancing: A Queueing Analysis[J]. *European Journal of Operational Research*, 2008, 186(2): 681-693
- [6] Li Zhongmin, Yu Zhanwu, Zhu Li. A Spatial-Data-Content-Based Dynamic Load Balancing Method[J]. *Geomatics and Information Science of Wuhan University*, 2009, 34(5): 622-624(李忠民,喻占武,朱莉. 基于空间数据内容的动态负载均衡方法[J]. *武汉大学学报·信息科学版*, 2009, 34(5): 622-624)
- [7] Zhu Li, Shen Weiming, Li Rui, et al. Genetic Algorithm-Based Dynamic Load Balancing for Server Cluster in Network GIS[J]. *Geomatics and Information Science of Wuhan University*, 2011, 36(6): 722-725(朱莉,沈未名,李锐,等. 利用遗传算法的网络 GIS 集群服务器动态负载均衡算法[J]. *武汉大学学报·信息科学版*, 2011, 36(6): 722-725)

- [8] Lehto L, Sarjakoski L T. Real-time Generalization of XML-encoded Spatial Data for the Web and Mobile Devices [J]. *International Journal of Geographical Information Science*, 2005, 19(8/9): 957-973

An Efficient Method for Parallel Visualization of Vector Maps Under the Network Environment

GUO Mingqiang^{1,2} HUANG Ying² WU Liang^{1,2} XIE Zhong^{1,2}

1 Faculty of Information & Engineering, China University of Geosciences(Wuhan), Wuhan 430074, China

2 GIS Software and Application Project Research Center of the Educational Department, Wuhan 430074, China

Abstract: In order to solve the efficiency problem in vector map visualization with the traditional vector map visualization model, and to improve the capability of parallel processing for vector maps, a new approach for parallel visualization of vector maps is proposed, based on vector data spatial distribution information. This research focuses on the principles related to the collection, retrieval and analysis of vector data spatial distribution information. A workflow for decomposition and parallel processing are elaborated in this paper. Eventually, when all of the computing resources in a server cluster are sufficiently utilized, the parallel visualization of vector data can be realized. Simulation results demonstrate that the new approach proposed in this paper can improve the capability of a map server cluster, to meet the needs of massive concurrent users.

Key words: WebGIS; internet environment; vector data; visualization; concurrently access

First author: GUO Mingqiang, PhD, specializes in CyberGIS. E-mail: gmqandjxs@163.com

Corresponding author: WU Liang, PhD, associate professor. E-mail: wuliang133@189.cn

Foundation support: The National Key Technology Support Program of China, No. 2011BAH06B04; the Fundamental Research Funds for the Central Universities; the China University of Geosciences (Wuhan), No. CUGL140833; the China Postdoctoral Science Foundation, No. 2014M552115.