

# 基于网格技术的资源发现机制的研究与优化

阎文博,张育平,郭朝霞

(南京航空航天大学,江苏南京 210016)

**摘要:**在对 OGSA, P2P 等相关技术和模式分析研究的基础上,提出了一个在资源组织、负载均衡等多方面性能进行优化的网格资源发现模型,并对其体系结构、资源组织方式、资源发现机制作了详细的描述。该机制主要采用集中式与分布式相结合,以对资源服务进行分类为基本思想,结合 P2P 相关技术,具有合理和较高效的资源定位能力及良好的强壮性。

**关键词:**网格;资源发现;P2P;负载均衡

**中图分类号:**TP311;TP393

**文献标识码:**A

**文章编号:**1673-629X(2007)07-0091-03

## Research and Amelioration on Resource Discovery Based on Grid Computing

YAN Wen-bo, ZHANG Yu-ping, GUO Zhao-xia

(Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract:** Resource discovery mechanism has great impact on the efficiency of the resource sharing and collaborating. It is an essential part in grid. On the base of research and analysis on existing technology and models such as OGSA, P2P and so on, the paper presents a model which is advanced in the problems of load balancing, resource organizing. And it also makes a detailed introduction on its architecture, resource organizing mode and discovery mechanism. By using the combination of centralized and distributed way, resource classification and P2P technology, this model has better effect on resource location and strength of the system.

**Key words:** grid; resource discovery; P2P; load balancing

### 0 引言

网格正逐步成为一种新的技术和基础设施,可以充分利用集成的资源,形成一个大规模的计算池,其目的是为了在分布、异构、自治的网络资源环境上构造动态的虚拟组织,并在其内部实现跨自治域的资源共享与资源协作,有效地满足面向互联网的复杂应用对大规模计算能力和海量数据处理的需求。通过网络整合各种多样、异构、动态的资源,实现广域网范围内计算资源、存储资源、信息资源、知识资源等的共享和协作,消除信息孤岛和资源孤岛,向用户提供随处可得的、可靠的、一致的、标准的、廉价的计算能力<sup>[1]</sup>。

网格资源具有分布性、异构性、动态性和自治性的特点。在网格计算环境中,资源的数量是巨大的;并且是动态的、异构的;资源的可用性及性能等会随着实际环境和时间的变化而变化,并且可自由加入和退出。为了能使网格用户方便、高效地使用各种资源,必须

对网格环境下的资源进行有效的管理。如何在动态的环境下、海量的资源中找到满足用户需求的资源是个很重要的问题,这正是资源发现机制要解决的。

随着网格向更大规模的发展,不可靠的资源 and 间歇性的资源参与会占到相当的比例,资源类型也会更加多种多样。而资源发现机制面对的是封装成服务的资源,而不是各种各样原始的物理资源,所以资源发现机制应该不依赖集中、全局的控制,适应资源动态参与的特点,支持间歇性的资源参与方式,并支持基于资源属性的搜索。

### 1 研究背景

#### 1.1 资源发现方式

目前的资源发现机制主要有三种:集中式资源发现,分布式资源发现,以及两者相结合的资源发现机制。

(1)集中式资源发现方式中,只有一个信息节点,共享资源可检索的索引和所有可用系统的状态信息(包括各资源节点可提供的资源类别、位置等)都被聚集在该中心服务器上,每次资源请求都要去中心机询

收稿日期:2006-09-16

作者简介:阎文博(1982-),女,山西人,硕士研究生,研究方向为网格计算;张育平,硕士研究生导师,研究方向为软件工程。

问。这种发现机制的典型代表是 UDDI, UDDI 中心是整个系统的关键,一旦其出现问题,会导致整个系统瘫痪。所以这种机制存在瓶颈且不易扩充。

(2) 分布式资源发现方式没有中央服务器,系统的状态信息也没有被集中在单个的节点上,如 WS Inspection, 一个服务就发布在该服务的交付点。该方式消除了集中式资源发现的通讯瓶颈,但要知道所需服务在哪个节点还需其他机制的支持,使实现复杂且会产生大量的网络通讯。

(3) 集中与分布结合的方式是前两种方式的结合和补充,平衡了集中式搜索的效率和分布式搜索的自治,有利于系统的负载平衡,并且健壮性更好。

## 1.2 开放网格服务体系结构

目前,开放网格服务体系结构(OGSA)已成为网格事实上的标准,通过将 Web 服务的互操作模型引入到网格研究中,确立了以服务作为网格资源新的抽象形式和构造基础,该体系结构定义了一个公开的服务语义(网格服务),规定了如何创建、命名和发现临时网格服务实例;提供了捆绑在网格服务上的协议;并且支持与本地平台设施的集成。根据网络服务描述语言(WSDL)接口以及相关协议,OGSA 定义了建立和组成复杂分布式系统的机制,包括网格服务的生命周期管理、状态管理以及通知等。服务绑定可以支持可靠的请求、身份验证、授权以及委托等。

在 OGSA 中,所有的资源都被视为服务<sup>[2]</sup>,所以资源发现最终归结为服务发现。OGSA 对万维网服务框架的服务发现机制进行了扩展,服务注册在服务中介中,服务消费者到中介中查找并获得该服务的 WSDL 文档,然后组织 SOAP 消息调用该远程服务,服务消费者得到所需服务的 WSDL 文档。

## 2 网格与 P2P 技术分析并结合

P2P 最根本的思想在于网络中的节点(peer),既可以获取其它节点的资源或服务,同时又是资源或服务的提供者。P2P 系统中的任何一个 peer 都能够共享系统内的资源,peer 间通过直接交换信息来进行信息和服务的共享,而不需要经过其它的中间实体。它是一种分散的、分布式的资源管理模型。典型系统有:进行文件搜索和共享的 Napster、实现计算资源共享的 SETI@home 和基于 flooding 搜索的 Gridella 等<sup>[3]</sup>。

P2P 与网格都是分布式计算模型,目前均成为计算机领域的研究热点。

(1) 网格融合了分布式系统的许多技术,力图实现对网络上所有能共享资源的访问,整合的资源功能更加强大,而 P2P 是为了使得网络上的所有计算通信实

体间能实现平等互利的通信而构建的,更能够有效地提高互联网边缘普通 PC 的利用率。

(2) P2P 技术为服务共享、分布式计算和信息交流提供了更灵活高效的模式。

(3) 资源发现机制的全局性机制要求是 P2P 系统中比较成熟的,它采用的是主动报告策略,每一个节点周期性地向网络中的其它节点报告它的资源现状,同时发现其邻居节点的相应信息,可以发现空闲节点及一些问题。

(4) 网格动态性的本质需要容错机制保证程序的正确运行,而 P2P 系统具有广泛分布性,分散算法能够避免集中服务所带来的单点失效。

可见,网格和 P2P 技术在广域网中实现异构机器间的合作、动态监视和管理网格资源并适应资源规模的动态变化等方面都有着很多相似之处。两者都具有相同的目标,都是为高性能计算、数据密集型应用等提供远程计算资源的访问能力。而 OGSA 模型采用 Web Services 描述语言(WSDL)来定义网格服务的概念,提供了一个集成 P2P 与网格的机会,将 P2P 模型中的一些技术融合到网格系统中以后,有助于确保网格的可扩展性,并提高资源的定位效率和准确率<sup>[4,5]</sup>。

## 3 资源发现机制的优化设计

基于以上对一些相关问题的分析和研究,为了达到资源发现机制自配置、自主管理、动态发现和容错的要求,以下从多方面考虑和优化,采用集中式与分布式相结合,以对资源服务进行分类为基本思想,结合 P2P 相关技术提出了一种资源组织及发现机制。

### 3.1 体系结构

集中与分布方式结合的发现机制主要思想是:将现有的网格看作由不同的体制管理的小网络的联盟;资源可按节点所分布的地理区域、连通性以及节点之间的相关性等因素进行分簇(cluster)。

基于如上启示,该模型采用与 P2P 相结合的方式,将分布式与层次式结合。把由各服务资源组成的单一网格系统划分为若干个处于对等地位的“小规模网格”系统,每个小规模网格被抽象成 P2P 系统中的一个 peer,相互间为对等互连关系;而每个 peer 内部为一个网格虚拟组织,可以采用不同的网格技术来组织本系统内部的资源。系统模型框架见图 1。

如图 1 所示,每个作为 peer 的资源组织对应 P2P 层的一个 PS(peer service)服务节点,每个 PS 服务和多个 PS 相连接,互为邻居节点,并以 P2P 的方式进行消息交换、交换查询等通讯,用于执行跨虚拟组织的资源发现。而联系节点 CS(contact service)存储有已经注

册 PS 的网格服务句柄 GSH,用于组织 P2P 层的 PS 节点,包括 PS 的加入和离开,如新加入的 PS 向 CS 注册,同时随机获得其他几个 PS 的 GSH,使这些 PS 成为其邻居节点构成覆盖网络。这样就把 peer 内部的集中式和 P2P 层各个 PS 间的分布式方式结合起来,这样既可以满足系统的扩充,又可以减缓分布式机制下产生大量的网络通讯。

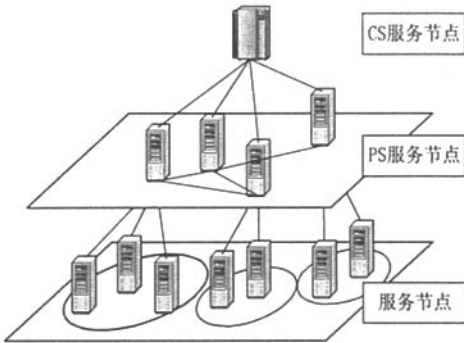


图 1 系统模型框架

### 3.2 资源组织方式

资源的组织方式决定资源发现、资源匹配和资源调度等其他资源管理技术,同时也影响资源的利用率和系统开销。

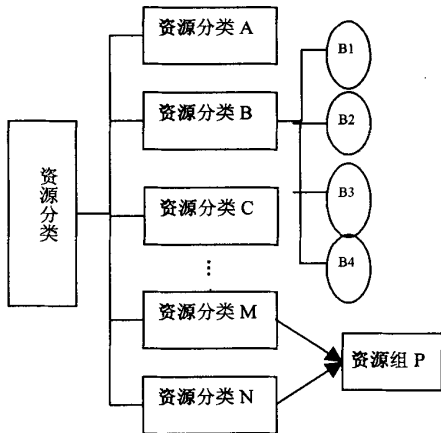


图 2 资源的分类

在混合机制中,资源的分簇可以按节点所分布的地理区域、连通性以及节点之间的某些相关性等各种因素进行。这里采用基于资源类型来组织资源服务的方法。将注册有同类资源的网格节点组织在一起,本质上实现支持基于属性查询的非集中式网格资源发现,从而改善资源发现的性能<sup>[6]</sup>。

首先采用 XML 半结构化语言对资源进行描述,即为每一类资源制定一个统一的元数据标准。资源的元信息主要由资源的属性组成,这样资源实体可以根据元数据标准通过(属性,值)对来进行描述。

根据资源的分类描述,对于网格中的每一类资源,注册该类资源的信息节点形成一个连通的覆盖网络,并将每类资源交由相应资源管理域管理,形成若干的资源管理域。每个功能节点都至少拥有一个同属分类邻节点,这样资源信息就可以尽量在同属分类的节点间扩散,从而改善资源发现以及资源信息更新的性能。

同时,用分布式与层次式结合的资源组织方式,使用分类的层次式资源组目录树记录资源属性信息。在每类资源构成的子树中,把计算能力相当的资源聚集起来保存在同一个资源组中,彼此之间可互为备份资源。一旦运行任务的资源出错,备份资源马上可以接收迁移过来的任务并进行工作。

这样的资源组织方式,既可以加快用户请求的资源定位,又考虑到资源失效情况下的系统提供服务的强壮性。

### 3.3 资源发现过程

一般而言,对于一个特定的网格应用,需要将不同类型的资源组织成虚拟组织(VO)来满足应用的要求,整个资源发现的过程在逻辑上可视为多个单一类型资源发现过程的组合。

结构化的 P2P 机制是基于分布式哈希表进行资源发现的。这里采用的哈希表形式,使用基于哈希表的关键字匹配算法,用基于哈希算法的有目的的路由策略来取代洪泛机制。表结构如表 1、表 2 所示。

表 1 资源注册表

资源名称	Hash 键值	资源地址	资源描述	资源属性
Res M	Hash (Res M)	Loc M	Des M	Att M

表 2 资源哈希表

Hash 值	节点 X
Hash(Res M)	1
Hash(Res N)	0

如表 1、表 2 所示,资源注册表存储了所有直接注册在本节点上的哈希键值及其他资源信息;资源哈希表是二维表,列名是节点列表,行名是整个哈希空间的键值列表,相同键值的资源位于同一行,单元格中的布尔值代表拥有此键值的资源,是否注册或索引在此节点上。

基于这种哈希表结构的资源发现的过程可简单描述为:由收到的用户资源请求计算哈希键值并发送查询指令;按键值从资源哈希表中取出值为“1”的列,找到该资源注册的节点;在资源注册表中查询该节点相关信息返回给自愿请求者。

### 3.4 负载均衡

该资源机制采用集中与分布相结合的发现机制,

(下转第 97 页)

表 3 实验结果 2

	实验结果平均值	实验结果最优值	目前最优值
原算法	30525.04	30355.12	30354.3
改进后的算法	30468.25	30355.12	30354.3

由以上实验数据可以看出原算法和改进后的算法所能搜索到的最优个体是一样的,但是改进后算法得到的最优个体的整体质量更优,因而文中所提出的“动态变异概率”和“基于较优个体的贪婪搜索”对算法整体结果的改进是有效的。如图 4 所示。

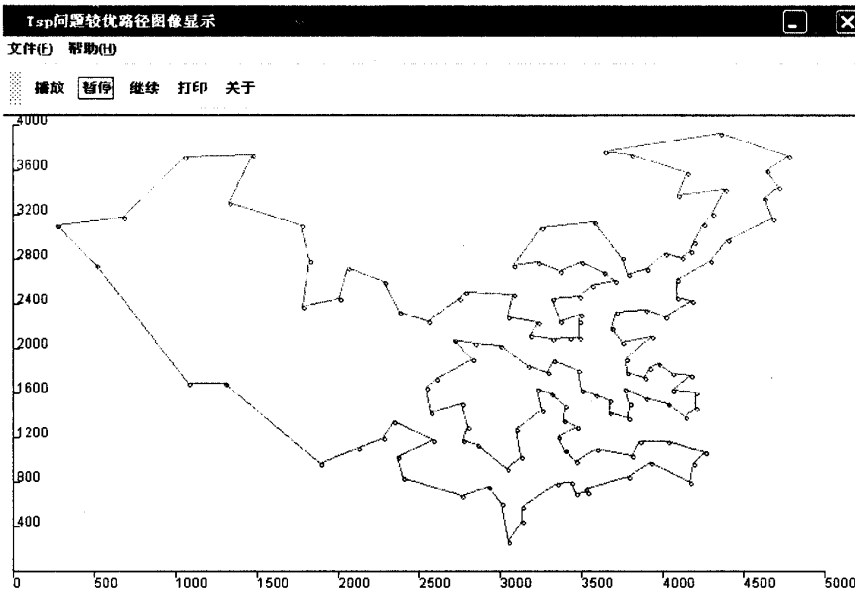


图 4 实验最优路径

### 3 结论

文中对使用逆转算子求解 TSP 的算法进行了改进,提出了就近选择、动态变异概率和基于较优个体的贪婪搜索的方法,并通过求解经典 CHN144 城市实例证明了以上方面对算法的改进都是有效的,使得算法在保证执行效率的同时,提高了最优个体的整体质量。

#### 参考文献:

[1] Michalewicz T G. Inver-over operator for the TSP[C]// Parallel Problem Solving from Nature, PPSN V. Berlin, Germany: Springer-Verlag, 1998:803-812.

[2] 吴斌,史忠植.一种基于蚁群算法的 TSP 问题分段求解算法[J]. 计算机学报, 2001, 24(12): 1228-1233.

[3] 潘正君,康立山,陈毓屏. 演化计算[M]. 北京:清华大学出版社,2004.

[4] 康立山,谢云. 非数值并行算法[M]. 北京:科学出版社,1997.

[5] 宋丹,傅明,朱亨荣,等. 一种改进的遗传算法及其在 TSP 中的实现[J]. 微机发展, 2004, 14(6): 21-23.

(上接第 93 页)

基于分类的思想组织资源,具有较优的负载均衡能力。

首先,采用结构化的 P2P 技术,消除了集中式的索引服务器,实现各 peer 间的资源共享,很大程度上解决了系统可扩展性问题,避免了完全随机搜索的低效,具有很高的路由效率和很强的负载均衡能力。

其次,分布式与层次式结合的资源组织方式,可以在同一个资源组中进行任务转移,如将部分负载转移到轻载节点,实现相似类型资源间的负载均衡。

### 4 总结

文中对网格资源发现及管理的相关技术及已有构架进行了探讨和分析。在此基础上,提出了一个在资源组织方式、资源发现机制、负载均衡等多方面性能进行优化的网格资源发现模型。该模型结构主要采用集中式与分布式相结合的方式、P2P 技术、对资源服务进行分类的思想进行构建,符合资源自配置、自主管理、

动态资源发现和容错等要求,并具有高效的资源定位能力和良好的强壮性。

#### 参考文献:

[1] 董方鹏,龚奕利,李伟,等. 网格环境中资源发现机制的研究[J]. 计算机研究与发展, 2003, 40(12): 1749-1755.

[2] 都志辉,陈渝,刘鹏. 网格计算[M]. 北京:清华大学出版社,2002.

[3] 黄道颖,黄建华,庄雷,等. 基于主动网络的分布式 P2P 网络模型[J]. 软件学报, 2004, 15(7): 1081-1089.

[4] Iamnitchi A. Resource discovery in large resource sharing environments[D]. Chicago: University of Chicago, 2003.

[5] Iamnitchi A, Foster I. On Fully Decentralized Resource Discovery in Grid Environments[C]//In: Lee C A. Grid Computing-Grid 2001. Germany: Springer, 2001.

[6] 朱承,张维明,刘忠,等. 一种基于资源类型的网格资源发现方法[J]. 计算机研究与发展, 2004, 41(12): 2156-2163.