

基于本体的网格资源匹配算法研究^{*})

卢国明 顾小丰 孙世新 李建平

(电子科技大学计算机学院 成都 610054)

摘要 由于网格动态异构等特点,传统的基于资源属性的精确匹配方法不够灵活,并且扩展性差。近年来,本体论引入到网格计算中,试图在语义层次上采用通用可扩展的信息系统建模工具,使得网格资源分配系统能够高效和精确地检索网格资源信息。通常采用的技术是建立和维护一个集中和一致的网格资源本体。这种集中的本体不适合具有分布式特性的 P2P 网格资源匹配。针对 P2P 网格,本文提出了一种基于分布式网格本体的 P2P 网格资源匹配模型。在该模型中,全局本体由各个节点的独立的本地网格资源本体构成。网格资源匹配操作完全分布式地由节点自主控制。这种方法可扩展性强,更适合 P2P 网格的资源匹配。

关键词 本体,对等网,网格,资源描述框架,离散哈希表

Ontology Based Resource Matching in the P2P Grid

LU Guo-Ming GU Xiao-Feng SUN Shi-Xin LI Jian-Ping

(School of Computer Science and Engineering, UEST, Chengdu 610054)

Abstract Because of the dynamic and heterogeneous characteristic of grid, traditional resource matchmaking which is based on exactly resources attributes matching leads to low flexibility and bad extension. Recent years, the ontology is introduced to the grid computing. Using the information system modeling tool in the semantic level, the grid system highly, effectively and precisely retrieves the grid resources. Usually used method is to establish and maintain consistent and centralized grid resource ontology. However, this kind of centralized ontology is not suitable for the P2P Grid. In view of P2P Grid, this article proposed an ontology-based P2P Grid Resource Matchmaking Model (ORMM). In this model, the global ontology constitutes by the integration of the local ontology located in each P2P node. Grid resources matchmaking is done and controlled completely by the local P2P node without central control mechanism. This model is flexible and extensible, and is suitable for P2P Grid resources matchmaking.

Keywords Ontology, P2P Grid, RDF, DHT

1 引言

网格这一术语于 20 世纪 90 年代中期提出,用来表述一种适用于高端科学和工程的分布式计算体系结构。网格的优势在于它能发现、分配并协商使用网络可以达到的范围内的各种能力(capability)^[1],这些能力使计算机能够提供计算服务、存储服务以及应用服务。

与传统计算机系统的资源管理技术相比,由于网格资源的自治性、分布性、异构性、动态性等特点,网格环境下的资源管理问题是一个非常复杂的问题。当前网格系统中广泛使用的两种资源描述语言 Globus RSL 以及 Condor ClassAds 均采用属性/值对来表示资源的度量或者配置参数,通过比较任务描述和资源描述中相应属性值来求解当前资源与任务的匹配问题。这类技术要求资源提供方与需求方事先达成一致,因此无法提供足够的灵活性,并且当新类型资源加入的时候系统很难扩展。造成这种问题的根本原因是没有在语义层次上采用通用可扩展的信息系统建模工具,使得网格资源分配系统无法高效和精确地检索网格资源信息。本体论(Ontolo-

gy)^[2]作为一种能够在语义和知识层次上描述信息系统概念模型建模工具,自提出以来就引起国内外众多科研工作者的广泛关注。尤其本体论与 Web 结合,导致了语义网^[3]的诞生,在 W3C 的主导下有望解决 Web 信息共享的语义问题。文[4,5]把语义网技术引入到网格资源匹配中来,建立了资源供需双方共享的网格资源本体(Grid Resource Ontology),通过基于该本体的推理获得满足资源需求的资源。与传统方法相比,这类方法具有资源描述灵活、易扩展、推理能力强等优势。当然,它们有一个共同的前提:资源的供求双方共享一个集中的本体。

P2P 网格是网格发展的一个重要方向,它具有明显的动态和分散(decentralized)特性。很难在分布的资源节点和资源需求节点间共享完全一致的集中本体。事实上,在一个大规模的 P2P 网格系统中,各个网格节点基于自身对网格的需求,仅具有部分的网格资源本体是合理的。考虑如下一种状况:网格节点 A 是运行操作系统 Redhat 的集群,其自身的节点本体认为 Redhat 是 OperatingSystem 类的子类,网格节点 B 的本体认为 Linux 是 OperatingSystem 类的子类,节点 C 本

^{*}基金项目:国家 863 高技术计划项目(2003AA148040)、国家自然科学基金项目(69682011)、国家自然科学基金项目(10471151)。卢国明 博士研究生,主要研究方向:网格计算及语义网相关技术;顾小丰 高级工程师,主要研究方向:并行计算与分布式计算;孙世新 教授、博士生导师,主要研究方向:网格计算、并行分布式计算以及组合优化等;李建平 教授、博士生导师,主要研究方向:信息安全,小波分析,分形理论、神经网络、电子商务与排队论等。

体认为 Redhat 是 Linux 的子类。那么,当某个资源请求“操作系统为 Linux 的集群”的时候,如果没有分布式本体集成技术,基于单个网格节点本体,那么满足资源请求的节点 A 并不会匹配该资源请求。因此,在没有集中本体的 P2P 网格中,为了充分和有效地进行网格资源匹配,必须提出一种分布式网格的资源本体集成技术。文[6]提出一种全新的基于 P2P 网格的分布式本体集成机制,网格本体由分布在各个 P2P 节点上的本体集成,为 P2P 网格进行分布式资源匹配提供了新思路。但是该文献侧重描述网格资源类间的继承关系,缺乏对复杂网格本体的描述能力。因此,本文对该文献采用的方法进行扩充和改进,提出了基于本体的 P2P 网格资源匹配模型 ORMM (Ontology-based Resource Matchmaking Model)。ORMM 完全采用 RDF 语言描述网格资源本体,完全提供了对类以及类属性的支持,并且采用基于 RDF 的查询语言 RDQL 描述资源需求,使得资源查询能够更具体、更加精确。

2 采用技术及论文相关工作

2.1 采用的相关技术

2.1.1 语义网相关技术

语义网^[3]是当前万维网(World Wide Web)的扩展和延伸。语义网的信息具有充分的、完备的语义定义,能够在人与计算机之间建立语义上的理解和合作。在语义网中,本体论具有非常重要的作用。本体^[2]是概念模型的明确的规范说明,是解决语义层次上万维网信息共享和交换的基础。本体的应用需要描述本体并使得它们能够进行信息交换的语言标准。RDF/RDFS 就是这样的建模语言。RDF 定义了一个简单的数据模型,通过性质(Property)和值(Value)来描述资源以及资源与资源之间的相互关系。RDFS 是对 RDF 的一种补充,它提供了定义类和性质,增强了 RDF 对资源的描述能力。RDFS 作为本体建模语言以其简单和较强的表达能力得到广泛的认可。

HP 实验室的 Jena^[8]是一个以 RDF API 为核心的语义网工具包,它支持 RDF 的解析、创建、操作和查询等功能。推理层、查询功能以及网络 API 建立在 RDF API 之上。采用 Jena 内嵌推理器可以非常容易地对 RDF 进行推理。并且 Jena 支持 MySQL、Oracle、PostgreSQL 等作为其后台数据库,可以直接将 RDF 语句以三元组的形式存放在数据库中。对于大型数据库的查询,Jena 以 RDQL 形式提供了一种表达通用查询的机制。RDQL 允许简明地表达复杂的查询,查询引擎执行访问数据模型的繁重工作。

2.1.2 基于 DHT 的结构化 P2P 系统

Chord^[9]是基于 DHT(Distributed Hash Table)的结构化 P2P 系统。每个节点和被存储的资源通过散列(Hash)函数获得一个 ID 标识自己。在由 N 个节点组成的系统中,每个节点只需维护其他 $O(\log N)$ 个节点的信息。Chord 可以在 $O(\log N)$ 内定位到任何资源。

2.2 论文相关工作

文[4,5]是比较早将语义网技术引入到网格资源匹配中的例子。文[4]在网格互操作项目 GRIP 中为了解决两个异构的网格平台 UNICORE 和 Globus 的资源互访而建立了一个网格领域本体,通过提取描述网格资源的本体语义,在一定程度上实现了 UNICORE 和 Globus 之间的互操作和资源分配。文[5]采用 RDF 建立了网格资源本体,在此基础上利用

TRIPLE/XSB 进行基于规则的推理,从而实现了资源集合的匹配(gang-matching)。文[4,5]的共同点是建立了资源供需双方共享的网格资源本体(Grid Resource Ontology),通过基于该本体的推理获得满足资源需求的资源。与传统方法相比,这类方法具有资源描述灵活易扩展、推理能力强等优势。当然,它们有一个共同的前提:资源的供求双方共享一个集中的本体。

文[6]提出一种全新的基于 P2P 网格的分布式本体集成机制,网格本体由分布在各个 P2P 节点上的本体集成,每个节点维护通过 DHT 算法映射到该节点的资源类以及该类的实例信息,因此资源的匹配过程转化为对相应类实例集合的集合运算。但是该文献所能够处理的本体仅限于简单的类层次结构,采用的网格本体仅描述了网格资源类间的继承关系,对于网格资源之间的非继承关系(诸如聚合关系)无能为力。并且,该文献通过简单的类的集合运算(并、交、逆)表示资源请求,这种对资源请求的描述不能准确表示资源的需求,不支持网格资源“基数”的描述,在精确描述网格资源的能力方面存在不足。我国有关语义网技术与网格技术融合的研究工作也开展得很早。中科院计算所诸葛海^[7]研究员领导的中国知识网格研究组(<http://kg.ict.ac.cn>),作为织女星网格项目的组成部分,成立于 2001 年,是最早研究知识网格的课题组之一,参与了开创该领域的工作,引起了国外同行的密切关注。他们现在的主要任务是 eScience 的应用研究。

如上所述,语义网与网格技术的融合是当今研究的热点,但是将语义网技术应用到分布式网格资源匹配的研究还不多。我们的目标是在文[6]的基础上,提出并建立一个全面支持网格资源类以及网格资源属性分布式网格本体模型,在此基础上设计支持资源集分配的网格资源匹配系统。本文提出的 ORMM 系统相对于文[6]有如下一些特点:

在本体建模方面采用 RDFS 对网格本体建模,全面支持网格资源类、属性,并扩展支持资源的“基数”描述;将网格资源属性提升到与网格资源类同等重要的程度;在资源查询的表示方法上,采用了类 SQL 语法的 RDQL 语言,能够更具体、更精确地表达资源需求;在网格知识的存储上,采用关系数据库存储大量的网格资源实例信息,缓解了 P2P 节点存储和查询大量实例数据带来的性能瓶颈。

3 ORMM 网格本体建模

ORMM 采用 RDFS 对网格本体进行建模。本体按照类、属性以及语义关系构成,其形式定义如下:

网格本体(GO)是一个三元组: $GO = (C, P, R)$ 。

其中 C 是网格资源类的集合,是描述网格系统中各种资源的 RDF 词汇; P 是网格资源属性的集合,是描述网格资源类之间相互关系的 RDF 词汇; R 是网格类之间的语义关系;这种关系通过词汇 P 进行描述。采用 RDFS,我们可以详细描述网格资源本体类、属性以及通过属性描述本体类之间继承、聚合等各种关系。因此,ORMM 相对于文[6]能够对网格本体进行更为深入和具体的描述。

根据上述的网格本体结构,我们对 ORMM 网格领域本体进行如下建模(如图 1)。在这个网格本体中,我们定义了五类网格资源:Network(网络带宽)、Disk(存储空间)、Memory(内存)、Processor(处理器)以及 OperatingSystem(操作系统)。这五类资源均为 Resource 的子类,因而也继承了 Resource 的三个属性:

- capability:描述该资源的数量;
- authorization:该资源的认证信息;
- period:描述该资源的持续时间;

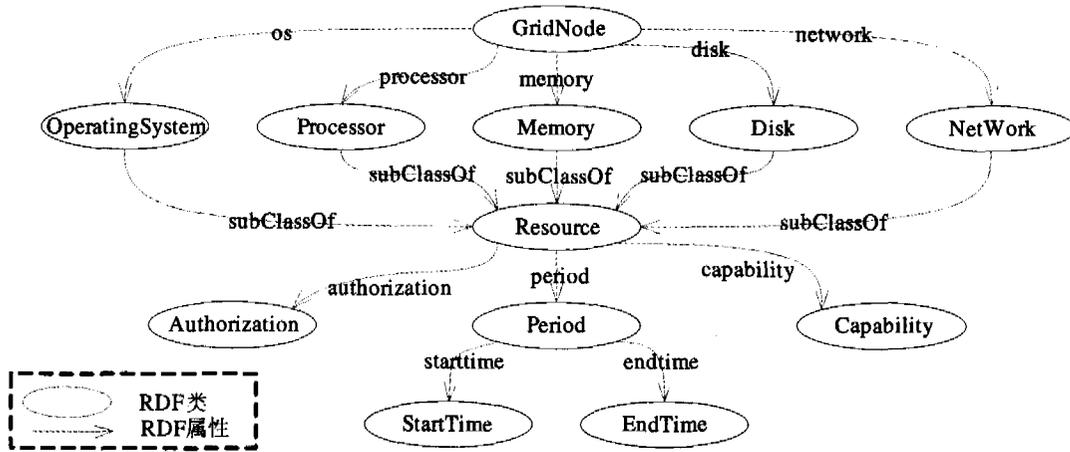


图1 使用 RDFS 对网格领域本体建模

Resource是包括分布在各个P2P节点上的网格资源类的超类。各个P2P节点的网格资源类一定属于图1中Resource某个特定子类。

GridNode(网格节点)对于一个特定的网格节点,通过os、processor、memory、disk和network五个属性与上述五类网格资源相关联。GridNode通过上述五个属性将各自独立的资源类关联起来,是上述五种资源的聚合。它的引入使得ORMM可以按照资源集合为单位进行网格资源匹配。

图1定义了网格资源本体最顶层的信息,定义了ORMM网格本体的框架,因此每个网格节点的本地本体都是该本体的进一步延伸,亦即图1是所有网格节点本体的公共部分,网格节点为图1所示本体的特定添加了特定的子类,从而扩展全局的网格资源本体。

4 ORMM 资源管理模型

ORMM是以RDF/RDFS为网格知识描述语言,以P2P网格Chord系统为基础建立起来的一个分布式网格资源匹配模型。在ORMM模型中,网格知识以RDF三元组形式存放在各个分布式网格节点的关系数据库MySQL中。网格知识以RDF类或RDF属性为关键字在Chord上进行分布式存储和检索。网格知识的推理、查询以及管理其在关系数据库的存储均通过Jena API完成。

如图2所示,整个ORMM由基于DHT协议的网格节点组成。每个网格节点由节点知识库、知识管理器、资源匹配器以及网络接口管理器组成。

• 知识管理器:主要功能是实现网格知识库的存取,并通过网络接口管理器实现全局网格知识的更新与查询。在Jena中,数据库被抽象为数据模型(Model)。Jena支持在同一个数据库管理系统中存储多个数据模型(Model),每个数据模型对应于数据库管理系统中的一张或者多张数据库表(table)。ORMM中,网格知识由本体信息以及资源实例信息组成。因此ORMM的知识管理器相应分为本体管理器和资源管理器两个部分,它们通过Jena API解析并分别负责基于RDF的本体信息和资源实例信息的存取。

• 网格资源匹配器:是ORMM的功能核心。通过知识管理器提供的网格知识,资源匹配器通过Jena执行RDQL资源查询语句,实现完全分布式的网格资源匹配操作。

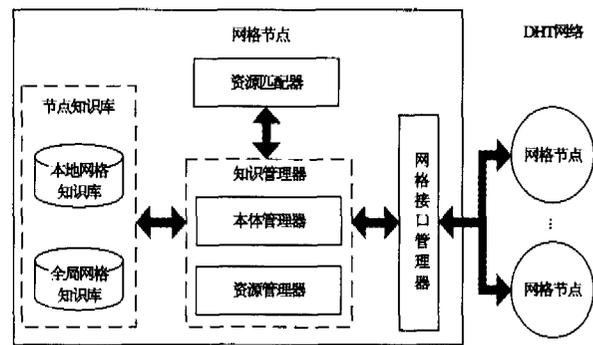


图2 ORMM模型

• 节点知识库:存储当前节点的网格知识。在ORMM中,节点知识库包括存储本地网格知识的本地网格知识库以及存储全局网格知识的全局网格知识库两个部分。相应地,在ORMM中,网格知识库分为四个数据模型:本地本体数据模型LOM(Local Ontology Model)、本地资源实例数据模型LRIM(Local Resource Instance Model)、全局本体数据模型GOM(Global Ontology Model)以及全局资源实例数据模型GRIM(Global Resource Instance Model)。

LOM存储了该节点的本地网格本体信息。这些本体信息来自于该节点自身对网格本体的认知,在节点加入网格以前就已经存在。网格节点加入网格系统时,通过将LOM中的本体信息扩散到整个网格系统上进行分布式存储,从而达到扩展和更新全局本体数据模型GOM的目标(具体算法详见第5节)。存储形式为RDF三元组,通常为如下形式:

- 〈资源类 URI1, RDF 属性, 资源类 URI2〉
- 〈资源类 URI1, rdfs:subClassOf, 资源类 URI2〉
- 〈资源属性 URI1, rdfs:subPropertyOf, 资源属性 URI2〉

LRIM存储当前网格节点对于LOM中相应的网格资源类以及网格资源属性的实例信息,这些信息用于初始化和更新全局网格实例信息。实例信息采用如下RDF三元组形式存储:

- 〈资源 URI1, RDF 属性, 资源 URI2〉
- 〈资源 URI1, RDF 属性, RDF Literal〉

上述资源实例表示方法与文[6]不同,后者采用了为同一资源集合所有资源分配一个标识符的方法。这种方式无法表

示资源集合内部的拓扑结构。例如,查询一个网格节点,该节点包含一个 Linux 集群以及一个图形工作站;Linux 集群的处理器为 Xeon,图形工作站的芯片为 G4。ORMM 采用了更为细致的资源标识方法:每个资源和属性通过 URI 唯一标识。这种唯一标识每一个资源的方式配合 5.2 节的 RDQL 查询,ORMM 可以非常方便地实现上述复杂的网格资源查询。

所有节点上的 GOM 和 GRIS 构成 ORMM 的全局网格知识库。ORMM 以网格资源类名和属性为 DHT 关键字,为网格资源类和属性在网格系统中映射一个网格节点进行存储。由此可见,节点全局网格知识库采用了冗余存储的方式将节点本地网格知识进行了存储。

5 基于本体的资源匹配算法

5.1 基于 DHT 的分布式网格知识库存储

如上节所述,LOM 中存储的网格资源类和属性通过 DHT 算法映射到特定的网格节点进行存储。存储的内容包括网格本体信息和网格资源实例信息,它们分别存储在映射节点的 GOM 和 GRIM 中。

对于被映射的 RDF 资源类而言,GOM 存储陈述该类的直接父类的 RDF 语句;对于被映射的 RDF 属性而言,本体 Model 存储该 RDF 属性的直接父属性。相应地,GRIS 存储映射到该节点的 RDF 类或 RDF 属性的实例 RDF 语句。对于 RDF 类,GRIS 存储该类以及所有子类的资源实例 RDF 语句;对于 RDF 属性,则存储该属性的以及所有子属性实例的 RDF 语句。

ORMM 的全局网格知识的分布式存储模型,其实质是利用 Jena 推理引擎派生出网格知识库中未明确表达的其他语句,并将它们以相应的 RDF 类名和 RDF 属性名为索引在网格系统上进行分布式存储。由于网格节点存储了相应 RDF 类和属性的直接父类,因此根据这些信息可以方便地遍历整个本体模型。同时,由于实例 Model 中存储的是所有子类或者子属性的全部实例,ORMM 通过 DHT 算法可以很方便地定位到对应的网格节点,便能够获得该资源的全部实例信息。因此,查找相关资源的过程,即为定位相关 RDF 类和 RDF 属性的网格节点,并检查该节点的 GRIS 中存储的实例语句的过程。

5.2 网格知识库的更新算法

当网格系统有节点加入和退出,以及网格节点资源状态发生变化的时候,全局网格知识库就需要进行更新。当 P2P 网格节点加入网络的时候,需要根据该节点初始的本地网格知识对全局网格知识库进行扩充和调整。这一过程包括接收并存储由 DHT 算法映射到该 P2P 节点的 RDF 类和 RDF 属性,并将该节点自身的本地网格知识在 P2P 网络上进行分布式存储,从而扩充全局网格知识。具体包括如下步骤:

①接收传递到新加入节点存储的全局网格知识。新加入节点接收映射到该节点的 RDF 类和 RDF 属性相关知识。在运行 DHT 协议的 Chord 系统中,新节点的加入会引起部分后继节点(successor)负责存储的全局网格知识传递到新节点进行存储。这些网格知识新加入节点的 GOM 和 GRIS 中。

②将本地网格知识发送到网格系统。新加入网格节点根据保存在本地网格知识库的 LOM 和 LRIS 中的本地网格知识,依次为 LOM 中的每一个类和属性产生一个消息(Message)。该消息包括被描述类(属性)的类名(属性名)、该类

(属性)的直接父类(直接父属性)以及类实例(属性实例)三个部分组成。按照 DHT 算法,将该消息发送到该相应的存储节点。

③当网格节点接收到一个②传来的消息,检查该消息中是否有新知识(包括新的本体信息和新的资源实例信息)。新知识通过知识管理器加入到该节点的全局网格知识库中,并且按照如下方式在 Chord 上进行扩散,以维持网格知识库的一致性。

- 节点接收到新的类实例信息,该节点将会向 GOM 中存储的该类的所有直接父类的存储节点发送消息,要求添加该实例信息。

- 节点接收到新的属性实例信息,该节点将会向 GOM 中存储的该属性的所有直接父属性的存储节点发送消息,要求添加该实例信息。

- 节点接收到 GOM 中某个类的新的直接父类信息,该节点发送该类的所有实例信息到该类的所有直接父类的存储节点;

- 节点接收到 GOM 中某个属性的新的直接父属性信息,该节点发送该属性的所有实例信息到该属性的所有直接父属性的存储节点;

④所有收到步骤 3 发来消息的节点,都必须回复步骤③的发送消息节点,回复内容为该节点 GOM 中存储的相应类或相应属性的父类或父属性信息。这些信息帮助步骤③节点更新 GOM 中的相应类的直接父类或相应属性的直接父属性信息,并删除非直接父类或非直接父属性。例如某个节点存储有属性 P 的父属性有 P1, P2, 当它从存储 P1 属性节点收到关于属性 P1 的回复(P1 的父属性为 P2), 则该节点的网格知识库会删除 P 的父属性 P2, 因为节点现在知道 P2 不是 P 的直接父属性。

步骤③、④会一直进行,直到步骤③节点收到的消息中不再含有新的网格知识为止。

对于 ORMM 而言,全局网格知识库中存储的资源实例反映了该时刻网格系统可用资源。由于网格节点的自治特性,网格节点根据节点自身的资源管理策略或其它原因可能改变相应资源的可用状态甚至发生网格节点离开网络。因此,在这种情况下,全局网格数据库也需要进行更新。

更新网格资源实例信息与上述网格节点加入网格系统时的实例信息插入操作一样:首先在 LRIS 中更新本地网格资源实例信息,将更新后的网格资源实例信息发送到相应 RDF 类或者 RDF 属性的存储节点,更改该节点 GRIS 中的相关资源实例,然后按照类层次(属性层次)依次向高层类(属性)存储节点传递更新后的资源实例信息。

当网格节点离开网格系统时,相应的网格知识应当从网格全局知识库中删除。与文[6]中假设一致,节点离开网格系统前,该节点会通知系统,网格系统开始知识迁移和更新。由节点离开网格系统引起的网格知识库更新包括两部分。首先该网格节点负责存储的相应的 RDF 类以及 RDF 属性知识通过 Chord 系统传递到该节点的后继节点进行存储,然后采用更新网格资源实例信息的方法将该节点的网格资源实例信息从全局网格知识库中删除。由于节点实例信息完全存储在退出节点的 LRIS 中,因此根据 LRIS,网格系统依次将分布存储在相应节点全局网格知识库 GRIS 中的资源实例信息删除。具体删除方法与更新资源实例信息采用的方法和步骤完全一致。值得注意的是,ORMM 系统只删除了退出节点的资源实

例信息,由退出的节点补充到全局网格知识库 GOM 中的本体信息并没有删除。因为本体体现的是共享信息模型,所以不应当随着节点的退出而删除。保留这些本体信息有益于网格知识库的完善和扩充,不会带来网格知识库的不一致。

5.3 基于本体的资源匹配算法

网格资源匹配即是查找满足约束条件的网格资源。在基于本体的网格资源匹配中,即为查找网格资源本体中特定资源类的实例。通过 5.2 节的网格全局知识库的更新以后,各节点 GRIS 中存储了相关类/属性的全部实例信息。因此,查询特定资源的实例信息就是通过 DHT 算法定位相应存储节点,然后取得实例信息的过程。对复杂的资源匹配请求,ORMM 资源匹配器将其分解为对单个资源类/资源属性的实例查询进行分别处理,最后在汇总的返回结果(资源实例集合)上执行相应 RDQL 语句,获得满足资源需求的资源实例信息。具体资源匹配过程如下:

①首先由请求资源的网格节点依靠该节点的 LOM 生成资源匹配的 RDQL 查询语句;

②该请求被分解为相关资源类实例和资源属性实例的查询,并发送到相应的网格节点;

③收到实例请求的节点后,查询该节点的全局资源实例数据模型 GRIS,并将符合要求的实例信息返回给查询节点;

④查询节点的资源匹配器将各个节点返回的资源实例信息导入内存数据模型,并采用 Jena 的 RDF 数据查询语言 RQDL 查询最终符合要求的资源 URI。

例如,网格节点 GridNode1 需要查询操作系统为 Linux、处理器为 Intel 且处理器个数大于 24 个的网格资源。相应的 RDQL 查询语句如图 3 所示。

```
SELECT
? GrdNode
WHERE
(? GridNode,<gn:os>,<? OperatingSystem>),
(? GridNode,<gn:processor>,<? Processor>),
(? OperatingSystem,<rdfs:subClassOf>,<gn:Linux>),
(? processor,<rdfs:subClassOf>,<gn:lnetel>),
(? processor,<gn:capability>,<? capability>
AND ? capability>=24
USING
gn FOR <http://www.ccse.uestc.edu.cn/~lugm/schema/>
rdfs FOR <http://www.w3.org/2000/01/rdr-schema#>
```

图 3 资源匹配请求 RDQL 查询

根据 RDQL 语句中涉及到的词汇,包括 os、processor、capability 三个属性,以及 Intel 和 Linux 两个资源类,GridNode1 向相关的 Chord 存储节点发送实例查询信息。

收到实例查询请求的节点后,查询该节点的全局资源实例数据模型 GRIM,将符合要求的实例信息返回给查询节点。注意:此时该节点 GRIM 中存储的信息包含相关类及其所有子类的实例信息。比如 Linux 类的存储节点中,关于 Linux 的实例将包括 Linux 类以及全局网格本体知识库中 Linux 的所有子类的如 Redhat、Fedora 等类的实例信息。根据这些实例信息,GridNode1 可以成功地匹配节点自身并不知道的资源类实例(如 Redhat、Fedora 等)。

可以看出,ORMM 的资源匹配有如下特点:

①资源匹配的全分布式。从资源匹配请求的生成到最终匹配结果的获取,无需中央控制节点(专门处理资源匹配请求

的节点)。这消除了由中央控制节点性能局限带来的性能瓶颈。

②资源匹配请求表示灵活。与文[6]采用的类集合运算相比,ORMM 采用类 SQL 的 RDQL 查询语言,简明地表达复杂的查询,可以针对资源类、资源属性以及特定的 RDF 文字(Literal)进行查询,资源需求表达灵活,可扩展性强。并且基于图 1 的本体模型,ORMM 能够查询从单一资源到特定约束条件下的资源集合。

③资源请求本地化。资源请求脚本按照本地网格知识库中的词汇构建。如图 3,GridNode1 的资源请求仅仅使用了节点自身的词汇 Linux,能匹配 Fedora、Redhat 等资源实例。这种匹配方式大大降低了网格节点对全局网格本体的依赖,这种网格节点间的低耦合性大大提高了 P2P 网络的可扩展性。

结论 在网格环境中,其完全的分布式特性和高度的自治性为在 P2P 网络中使用集中一致的 Ontology 带来了很大的挑战。针对 P2P 网络,本文提出了一种基于分布式网格本体的 P2P 网格资源匹配模型 ORMM。在 ORMM 中,全局本体由各个节点的独立的本地网格资源本体构成。网格资源匹配过程完全分布,并且采用 RDQL 查询语言实现了基于资源集的匹配。在网格平衡时(全局知识库更新完毕后),节点仅根据其本地网格本体就能匹配到所有满足需求的资源。ORMM 模型可扩展性强,能够进行复杂的网格资源集合匹配,非常适合 P2P 网络的资源匹配。由于采用 DHT 算法对网格知识按照资源类和属性进行分布式存储,存储根类属性或根属性的 Chord 节点(如存储 Resource 类的节点)可能由于大量的实例信息的查询而成为系统性能瓶颈,这将是下一步工作着重解决的问题。

参考文献

- 1 福斯特等著. 网格计算(第二版). 金海等译. 北京:电子工业出版社,2004
- 2 Gruber T R. A translation approach to portable ontology specifications, [Tech Rep]. Stanford University, Logic-92-1, 1993
- 3 Lee T B, Hendler J, Lassila O. The semantic Web. Scientific American, 2001, 284(5)
- 4 Brooke J, Fellows D, Garwood K, et al. Semantic matching of Grid Resource Descriptions. 2nd European Across-Grids Conference, 2004
- 5 Tangmunarunkit H, Decker S, Kesselman C. Ontology-based Resource Matching in the Grid-The Grid meets the Semantic Web. In: Proc. of Workshop on Semantics in Peer-to-Peer and Grid Computing at the Twelfth International World Wide Web Conference (WWW-03). Budapest, Hungary, 2003
- 6 Heine F, Hovestadt M, Kao O. Towards Ontology-Driven P2P Grid Resource Discovery. In: Proc. of the 5th IEEE/ACM International Workshop on Grid Computing
- 7 Zhuge Z. China's E-Science Knowledge Grid Environment. IEEE Intelligent Systems, 2004, 19(1):13~17
- 8 <http://jena.sourceforge.net>
- 9 Stoica I, Morris R, Liben-Nowell D, et al. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. IEEE Transactions on Networking, 2003, 11