服务网格中的资源服务选择代理研究*)

聂铁铮 申德荣 于 戈 李 锐 殷 楠 周文生
(东北大学 信息科学与工程学院 沈阳110004)

摘 要 为了解决服务网格系统中出现的服务选择问题,提出了一个资源服务选择代理模型。借鉴OWL-S提出了一种服务需求描述语言(SRDL),使用户能够准确地描述对所需资源服务的要求;基于三种不同服务级别的QoS服务选择算法,用户可按需选择最合适的资源服务,降低作业执行时的出错率;采用候选服务缓存机制增加系统运行效率,同时降低远程服务查询代理的负担。

关键词 网格服务,服务选择,QoS

Study on Resource Service Selection Broker in Service Grid

NIE Tie-Zheng SHEN De-Rong YU Ge LI Rui YIN Nan ZHOU Wen-Sheng
(School of Information Science & Engineering, Northeastern University, Shenyang 110004)

Abstract In this paper, a model for resource services selection broker is presented to solve the problem of service selection in the service grid system. Based on OWL-S, a Service Requirement Description Language, SRDL, is described, with which the users of service gird system can accurately describe the requirement of resource service. The model also create three QoS-based service selection algorithms based on different service requirement level to select the grid services that are best fit the implementation of user's job and reduce the failure rate of user's job. Finally, a candidate service buffer mechanism is utilized in the broker, which increase the system running efficiency, and reduce the burden of long-distance service discovery broker.

Keywords Grid service, Service selection, QoS

1 引言

资源服务选择是网格系统的重要组成部分,直接影响网格提供的服务的质量。资源服务选择的任务是根据用户对作业资源需求描述信息结合服务网格中网格服务的状态信息,为用户作业按需选择合适的资源服务。资源服务选择代理主要目的是为保证用户作业可靠执行,提供优质的资源服务。随着开放网格服务结构(OGSA)[1]概念引入到网格当中,人们提出了网格服务的概念。目前关于网格中资源的选择与调度策略绝大多数都是基于传统的批处理作业,即基于作业的处理时间和运行负载进行预测[2,3],这使得它们无法适应服务网格中对资源需求的不断变化。因为,在服务网格中用户所关心的服务质量(QoS)因素将不仅局限于作业执行时间,作业运行的可靠性、安全性、稳定性、效率以及成本等都将成为重要的服务质量因素,对服务质量的需求等级和质量因素侧重点也会有所不同。对于网格中的资源服务可以使用WSDL1.1[4]和OWL-S[5]来进行服务接口定义和服务信息的描述,但缺少有效的服务需求描述信息方式。

本文针对服务网格中用户作业所需的资源服务选择问题,提出了一种资源服务选择代理模型,根据用户的不同需求对等价的服务过滤和排序,为用户选择符合QoS要求的资源服务,并通过在用户和服务网格间建立服务等级协议(Service Level Agreement, SLA[6])提供可靠的服务质量。对于资源需求的描述问题,通过借鉴OWL-S对相关规范的描述,提出一种用于描述资源服务需求的服务需求描述语言(Service Requirement Description Language, SRDL)。

2 资源服务选择代理体系结构

资源服务选择代理在服务网格系统中是作业执行代理中的一个子模块,如图1所示,它负责在服务查询代理返回的资源服务集合中按需选择满足相应QoS的资源服务。资源服务选择代理主要由资源服务需求描述解析模块、服务质量评估模块和资源请求管理模块组成。各模块的功能说明和资源服务选择代理的运行流程说明如下:

(1)资源服务需求描述解析模块。解析用户作业描述中的服务需求描述。用户的资源服务需求描述中包含两部分信息:服务特征描述信息和服务质量需求描述信息。我们使用服务需求描述语言

*)该课题得到国家863计划CIMS主题(编号:2003AA414210)、国家自然科学基金(编号:60173051)资助。聂铁铮 硕士研究生,主要从事服务网格方面的研究。

(SRDL)描述这些信息。

·服务特征描述信息是用户使用自然语言对作业所需服务要实现的业务功能描述。

·服务质量需求描述信息是用户对所需服务的服务质量标准(QoS criteria)的描述,这些服务质量标准必须是可量化的,如可由第三方提供,且是可监测的。

资源服务需求描述解析模块将这两部分信息解析成对象后提交给服务质量评估模块。

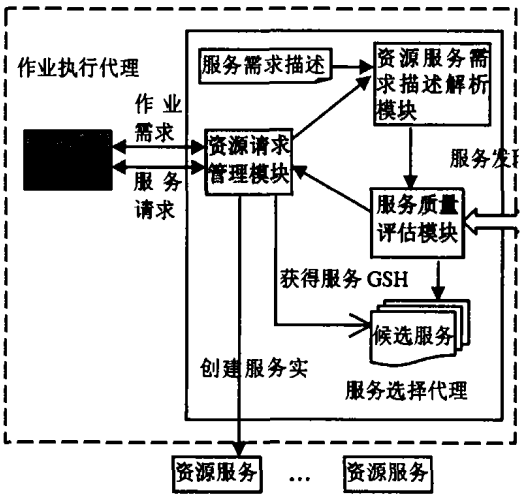

图1 资源选择代理体系结构

(2)服务质量评估模块。负责对服务查询代理返回的资源服务集合过滤,并根据适当的质量评估模型对符合要求的资源服务排序。将资源服务需求描述的解析结果提交给服务查询代理进行基于本体知识查询,并获得所有满足语义描述的资源服务集合。服务选择代理获得的服务信息包括以下几部分:

·服务的 GSH 通过 GSH 调用资源服务的 Factory 接口,为作业创建服务实例。

·服务质量标准的参数值 用于服务质量评估。

·服务接口映射信息 资源服务相对于标准服务的接口映射信息。

服务质量评估模块根据用户所选择的质量评估策略和服务的质量标准参数值对资源服务集合进行排序,其中排序靠前的部分服务作为作业执行的候选服务集合。

(3)资源请求管理模块。负责响应作业调度代理对资源服务的相关请求。当作业调度代理提出需要资源服务实例时,资源请求管理模块在相应的候选服务中按服务质量顺序选择服务,并通过服务 GSH 调用 Factory 方法为作业的执行创建一个服务实例,将服务实例的 GSH 和资源服务的接口映射信息返回给作业调度代理。如果实例创建失败,则选择下一个候选服务,同时把失败服务标记为"无效"。

3 资源服务需求描述

目前,资源服务的提供者可以使用 WSDL1.1 或者 OWL-S 来描述其服务的特征和接口定义,但这两种语言对于描述服务的业务能力和 QoS 等级的能力显得十分有限。而服务网格用户对作业中的资源服务的需求有以下几个特点:

·用户作业是由一组子作业组成,因此一个作业需要多个资源服务来完成;

·用户对同类服务的 QoS 等级和等级的评估方式有不同的要求;

·描述信息多种多样,包括业务功能的语义描述、服务特征约束描述和 QoS 等级描述;

·评价资源服务 QoS 的数据信息,可以来自服务提供者,也可以由第三方监测获得。

```
<srdl:qualityType name=" QosType1" qosAlgorithm ="qos_value_sort">
<srdl:serviceQualityCriteria name="response_time">
    <srdl:criteriaDataProvider>third_part</srdl:criteriaDataProvider>//参数值来源
    <srdl:valueConstraint>//Qos 标准参数约束
    <maxThreshold>10000</maxThreshold>//响应时间的最大值为 10000
    <qualityDirection>min</qualityDirection></srdl:valueConstraint>
    <srdl:criteriaRelationship>参数关系
        <nextCriteria name="bandwidth" relationship="more"/>
        <criteriaWeight>0.01<criteriaWeight></srdl:criteriaRelationship>
    <srdl:unit>ms</srdl:unit>//单位为毫秒
</srdl:serviceQualityCriteria></srdl:qualityType>
```

图2 资源服务需求描述文件 Qos 需求的描述举例

结合以上讨论,针对服务需求的相关特性,我们借鉴 OWL-S 提出了服务需求描述语言(Service Requirement Description Language,SRDL)。用户可以利用 SRDL 描述作业所需服务的业务功能、条件约束(如限定服务提供者的地域为北方)和 QoS 需求。具体说明如下:

(1)服务业务功能描述。主要通过 SRDL 中本体描述(ontologyDescription)标签说明需要服务完

成哪些实际功能，例如“订购机票，股票查询，货物运输”等。

(2)服务条件约束描述。由约束参数名(serviceParameterName)和参数值(sParameter)，也就是通常的“属性”、“值”对的方式给出，这种表述方式可以不受服务类型的限制。

(3)QoS需求描述。可以指定候选服务的排序算法，服务质量标准参数的约束。QoS标准可以是服务的费用、响应时间、吞吐能力等。在质量参数约束中，用户可以指定评估服务的标准参数值的来源、参数值的阈值约束以及同其他服务质量标准的关系。其中阈值约束中可以定义取值范围和质量方向。标准关系中定义关系或自身的权重。有关QoS需求的描述样例如图2所示。

通过对资源服务的QoS需求进行分类，用户可以对作业中同一类业务服务的不同资源服务作不同的质量要求，应用起来更加灵活。其中QoS标准间采用全序关系描述，包括相等(same)、重要于(more)和远重要于(farMore)。标准的权重应用于含有加权计算的策略中。

4 基于QoS的服务排序算法

基于QoS的资源服务选择算法应用于服务质量评估模块中，对符合要求的候选服务进行排序。用户可以根据对资源服务需求程度选择不同的算法为其作业挑选最适合的资源服务。

本系统提供了三种可供选择的算法：(1)简单的基于多QoS标准间重要度关系的排序；(2)简单的基于QoS标准参数值范围约束；(3)基于QoS标准参数值的排序。这些算法的优点是适用范围广、易于用户描述需求、运算开销小，但是要求QoS标准的性能随取值单调变化。将用户指定的每一个QoS标准(QoS Criteria)看作是服务质量空间的一个维。一个有N个QoS标准的服务其服务质量空间就包含N个维，$W=\{a_1,a_2,\cdots,a_n\}$，这里a_i代表服务的一个QoS标准取值，一个候选的资源服务在N维服务质量空间中用一个点表示$P_i(a_1^i,a_2^i,\cdots,a_n^i)$，$a_j^i$表示服务$s_i$在QoS标准j上的值，m个资源服务在服务质量空间x的一个维上的投影集合表示为$QoS^x=\{s_1^x,s_2^x,\cdots,s_m^x\}$，$s_i^x$表示服务表示服务$s_i$在QoS标准x上的值，同$a_x^i$。

(1)简单的基于多QoS标准间重要度关系的排序

该算法同文[7]中对资源最优化配置所使用的算法类似。用户首先需要在服务需求描述中对QoS标准定义一个重要度的全序关系，如假设一个服务有x，y，z三个QoS标准，用户对其定义了重要度关系：x>y≫z，表示x比y重要，y远远比z重要，同时用户还可以对QoS标准作阈值限制。算法的具体步骤如下：①如果用户定义阈值限制，首先根据阈值限制过滤掉不满足要求的资源服务。②从重要等级最高的QoS标准开始比较，在此项QoS标准上性能优秀的服务会被排列在前面，如对于三个服务有$s_1^x>s_2^x>s_3^x$则认为s_1优于另两个服务。对于在同一个QoS标准上性能相近的服务集合，将比较它们在下一等级QoS标准上的值。

关于性能相近的定义：最优服务在此向量上的值$s_k^x=Max(\sum s_i^x)$，(这里Max代表性能最好)满足$s_j^x\in(Max(\sum s_i^x)*(1-\alpha)$，则认为$Max(\sum s_i^x))$的服务$s_j$同$s_k$在这项上性能相近。其中$\alpha$的取值由QoS标准间的关系决定，“远重要于”相对于“重要于”的α取值要小。

(2)简单的QoS标准参数值范围约束

该种方法需要根据用户制定的各QoS标准阈值限制，对查询后得到的资源服务过滤，最后保留符合阈值限制的服务。即只要资源服务s_i有$S_{min}^x<s_i^x<S_{max}^x$(任取$x\in\{a_1,a_2,\cdots,a_n\}$)，就认为资源服务符合要求。

(3)基于QoS标准参数值的排序

此种算法是根据资源服务在QoS空间中的位置，结合QoS标准的权重和QoS标准间的关系对服务进行比较的方法。用户按需为每一个维设定一个阈值(threshold)，首先过滤掉不满足阈值约束的服务。然后计算资源服务对应的点$P_i(a_1^i,a_2^i,\cdots,a_n^i)$到空间各维阈值焦点$P_{threshold}(a_1^t,a_2^t,\cdots,a_n^t)$的加权距离，维度x的权值表示为$w_x$。使用加权的原因有两种：(1)维度的重要度不同；(2)维度的计量单位不同，需要统一到同一数量级上。服务对应点P_i的质量可以用它到$P_{threshold}$的距离D_i表示，公式如下：

$$D_i^2=\sum_{j=1}^{n}((a_j^i-a_j^t)w_j)^2$$

或

$$D_i=\sqrt{\sum_{j=1}^{n}((a_j^i-a_j^t)w_j)^2}$$

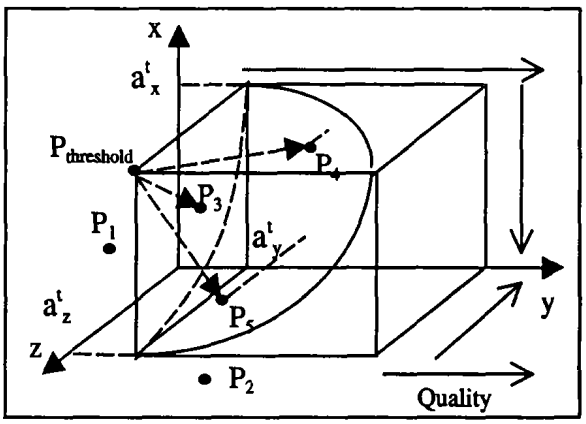

图3 基于Qos标准参数值的排序

其中 a_i^j 是服务 i 在 j 维上的值，a_t^j 是 j 维上的阈值，D_i 的值越大说明服务的质量越优秀。下面是应用三维质量空间对服务质量评价的一个例子。假设需求的资源服务有3个 QoS 标准{x，y，z}，有5个由查询获得的服务，如图3所示。首先过滤掉不满足阈值要求的服务 P_1、P_2，通过 P_3、P_4、P_5 到阈值点 $P_{threshold}(a_x^t, a_y^t, a_z^t)$ 的距离可以看出，P_4，P_5 的性能要优于 P_3。在维度 x 上的重要度大于其他两个维度，则根据 P_4、P_5 在维度 x 上的投影，得出 P_5 的性能高于 P_4。因此认为 P_5 最适合用户作业执行，并按 P_5、P_4、P_3 顺序排列。

另外，系统根据用户对服务的需求划分为相应的等级，分别对应不同的 SLA 描述与规则。表1为不同的服务等级对应的 SLA 的描述及相关规则。

表1 不同服务等级的 SLA 的描述及相关规则

资源服务需求等级和评估算法	SLA 描述及相关规则	
范围约束级	Sort Algorithm:qos_value_range	Selection Algorithm:best fit
（基于多 Qos 标准间重要度的关系）	Candidate Service List Length:10	Service Availability:85%
关系约束级	Sort Algorithm:qos_relation_sort	Selection Algorthm:random
（简单的 QoS 标准参数值范围约束）	Candidate Service List Length:15	Service Availability:90%
高级约束级	Sort Algorithm:qos_value_sort	Selection Algorithm:best fit
（基于 QoS 标准参数值的排序）	Candidate Service List Length:5	Service Availability:95%

资源服务选择代理根据用户的资源服务需求可确定采用的服务选择算法和相应的 SLA 描述，实现资源的自动选择。其中从服务需求到 SLA 描述的映射对于用户是完全透明的。

结论 在服务网格的作业执行代理中加入资源服务选择代理，虽然会占用少量系统资源，但减轻了服务网格系统的负担。从整体的系统开销看，资源服务选择代理在服务网格处理用户作业时所带来的正面效应是十分显著的：

(1)用户可以通过门户工具使用 SRDL 合理地描述作业执行中所需要的资源服务。

(2)通过三种候选服务排序算法来适应不同等级用户对资源服务的不同 QoS 需求，解决网格中最常见“什么样的服务更适合用户作业”的问题，保证运行作业的资源服务质量。

(3)降低了用户作业执行期间的出错率。由服务选择代理创建服务实例后再提交给作业的调度代理，可以提早发现资源服务的错误，同时使用其他候选服务来继续执行用户作业。

(4)降低资源服务提供者运行负担。只有当作业调度代理提出资源需求后才创建服务实例，因此服务提供者可以提高其服务运行效率，获得更多的盈利。

(5)减少与服务查询代理间远程通讯并降低服务查询代理的查询负担。一次性查询后将可用性和可靠性高的服务作为候选服务，可以避免频繁地连接服务查询代理进行查询。

在我们构建的服务网格系统中，通过使用 SRDL 帮助用户描述作业执行对资源服务的需求，并使用资源服务选择代理为用户选择最适合作业执行的资源服务。资源服务选择代理中采用基于 QoS 的服务选择算法使用户作业执行时可以获得稳定的、可靠的、可用的服务。资源服务选择代理的应用使作业执行效率提高，出错率下降。但是资源服务选择代理在资源服务的选择算法和作业需求的描述上依然存在不足，这些都将在未来的研究中得到完善。

参考文献

1 Foster I,Keselman C. The Physiology of the grid:An open grid services architecture for distributed systems integration. http://www.globus.org/research/papers/ogsa.pdf,2002

2 Liu Chuang,Yang Lingyun,Foster I,Angulo D. Design and Evaluation of a Resource Selection Framework for Grid Applications. HPDC11,July 2002

3 Frey J,Tannenbaum T,Foster I,Livny Miron,Tuecke S. Condor-G:A computation management agent for multi-institutional grids. Cluster Computing,2002,5:237～246

4 W3C. Web Service Definition Language (WSDL). http://www.w3.org/TR/wsdl,2001

5 Martin D,Burstein M,Denker C,Hobbs J,Kagal L,et al. OWL-s1.0 Release 2004 http://www.daml.org/services/owl-s/1.0/,2004

6 Ludwig H,Keller A,Dan A,King R. A service level agreement language for dynamic electronic services. Electronic Commerce Rsearch,2003,3:43～59

7 Al-Ali R,Hafid A,Rana O F,Walker D W. On QoS Adaptation in Service-Oriented Grids. In:1st Intl. Workshop on Middleware for Grid Computing,2003