

基于多层覆盖网络结构的资源发现机制^{*})

张忠平 雷炳银 刘欣媛

(燕山大学信息科学与工程学院 秦皇岛 066004)

摘要 资源发现机制是目前网格领域中的一个热点研究问题。对已有的网格资源发现方法进行了分析,指出了这些方法存在的不足,提出了一种基于多层覆盖网络结构的资源发现机制,将具有同一类型资源的网格信息结点组织在一起形成域,把结点数量较多的域通过类型粒度组织成层次形结构。利用类型匹配路由的资源搜索技术将资源发现请求转发到资源所在的域,采用 Topk 技术在相应的域内找出与请求匹配的最佳的 k 个资源,从而改善了资源发现的性能,提高了用户的满意度。

关键词 网格,资源发现,多层覆盖网络,SMR 搜索,TopK 搜索

Grid Resource Discovery Mechanism Based on Multi-layer Overlay Network Structure

ZHANG Zhong-Ping LEI Bing-Yin LIU Xin-Yuan

(College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract Resource Discovery Mechanism is a researched hotspot in grid field at present. The existing resource discovery methods are analyzed, and their insufficiency is pointed out. Focusing on the shortages, a Grid Resource Discovery Mechanism based on Multi-Layer Overlay Network structure is proposed, in which grid information nodes with the same type are organized into regions, and the regions which have lots of nodes are formed hiberarchy by the type granularity. The resource discovery request is transmitted to the region where the resources have been via the Style Matching Routing Source Searching Technique, then the top k resources suit to the request are located by the Topk Source Searching Technique, accordingly the efficiency of resource discovery is meliorated and the degree of user satisfaction is improved.

Keywords Grid, Resource discovery, Multi-layer overlay network, SMR query, TopK query

1 引言

网格中的资源在地理上分布极广、类型异构、数量巨大、动态变化,并且具有多管理域、站点自治等特点。以上特点使得网格环境中的资源发现更加困难,如何对资源进行高效查找和准确定位,是当前网格计算研究的一个热点问题。

本文针对网格环境下的资源发现机制进行了深入的研究,从查询效率和用户对查询到资源的满意度两方面综合考虑,提出了一种基于多层覆盖网络(Multi-Layer Overlay Network, MLON)结构的资源发现机制。该机制将网格内的每个资源映射成 MLON 网络的虚拟结点,且每个资源按类型分别属于不同的域,把结点数量较多的域通过类型粒度组织成层次形结构。MLON 网络能很好地满足网格动态性、异构性、分布性和扩展性的要求。在资源发现时利用类型匹配路由的资源搜索技术和 Topk 技术找出与请求匹配的最佳的 k 个资源。仿真实验结果表明该机制具有较高的资源查询效率和较好的用户满意度。

2 相关工作

目前,在资源发现机制的研究中,国内外的很多组织和研究人员都从不同角度对其进行不同层次的研究。

文[1]中提出了基于虚拟组织(Virtual Organization,

VO)的资源发现机制;文[2]中提出了基于 OGSA (Open Grid Service Architecture) 的双层 VO 的资源发现机制;文[3]提出了基于覆盖网络(Overlay Network, ON)的资源发现机制;文[4]采用路由索引(Routing Indexes, RI)方法,提出了动态网格环境中基于树形覆盖网络(Tree-Structured Overlay Network, TSON)的资源发现机制;它们都采用了基于虚拟结构的分布式的资源发现方式,这种方式能很好地屏蔽网格资源的动态性和异构性。

尽管文[1~4]中提出了一些有效的资源发现方法,但它们仍然存在着一些不足。例如它们在组织虚拟结构的域时都是基于物理区域的划分,在资源查找时都是采用属性匹配的方式,这些特点导致它们在资源搜索过程中网络消耗大,效率低下且用户满意度差。文[5]中虽提出了基于资源类型的非集中式网格资源发现方法,但它采用的是单层虚拟网络,在资源定位时采用贪婪的资源搜索方法,而且也没考虑结点间的通信距离,当某类资源社区很大时,它的搜索效率和用户满意度仍然很差。所以在网格环境下,如何进行高效的查询处理,减小网络消耗,提高用户的满意度,仍是资源发现需要解决的问题。

3 网格资源模型

覆盖网络(ON)是通过虚拟或逻辑链路而连接起来的,建

^{*})基金项目:教育部科学技术研究重点项目(205014),河北省教育厅科研计划项目(2006143)。张忠平 博士,副教授,硕士生导师,主要研究领域为网格技术,XML 数据库,数据挖掘;雷炳银 硕士研究生,研究方向为网格技术;刘欣媛 硕士研究生,研究方向为计算网络。

立在一个物理网络上的虚拟网络,负责管理物理网络中的所有资源。多层覆盖网络(MLON)利用树形结构和分层结构来组织覆盖网络,如图1所示。

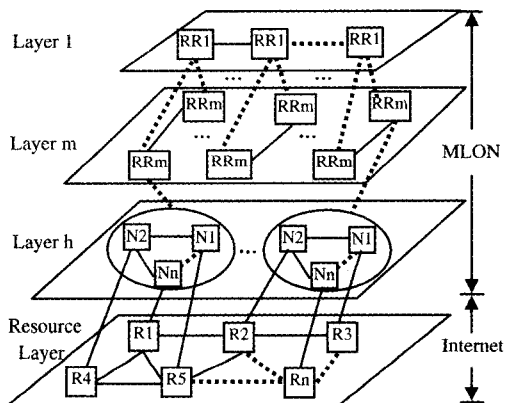


图1 多层覆盖网络结构

定义1(网格资源模型) $G=(N_G, R, A, rf, T)$, 其中, N_G 表示网格中所有结点集合; R 表示网格中所有资源的集合; A 表示资源的属性集合; rf 为资源注册函数; T 表示资源的类型集合, $r_i(a)$ 表示资源 r_i 的类型, 则 $r_i(a) \in t_i$, 即 $R(A) \in T$ 。

假设1 一个资源仅且只属于一种资源类型, 即 $|r_i(a)|=1$ 。

MLON以域和层为基本逻辑管理单位, 通过分类将数量巨大的资源在逻辑上划分成若干区域, 每个区域由资源路由器(Resource Router, RR)来管理, 资源路由器掌握着大量的域内结点服务信息。资源按类型分别属于不同的域, 当一个域中的虚拟结点增大到一定的数量时, 将这个域分裂为两个或多个更小的域, 资源类型的粒度变小; 反之, 当一个域中的结点减少到一定数量时, 将域中的结点合并, 资源类型的粒度变大; 这样就形成了多层覆盖网络(MLON)。

定义2(MLON模型) $G_v=(N_v, F, S, E)$, 其中, N_v 表示MLON所有结点集合, 则 $|N_v| > |N_G|$; F 表示父结点集合; S 表示孩子结点的集合; E 表示邻居结点集合。

假设2 结点 n 的邻居仅且只能是 n 的兄弟结点, 即 $f(n)=f(e(n))$ 。

根据假设1, MLON为每个资源分配一个全局唯一的ID地址, ID与资源的物理地址绑定。如同IP地址一样, ID由域编号(ID_r)和域内结点编号(ID_i)两部分组成, 再将两部分分段, 每段代表某种类型的资源。资源路由器(RR)的ID表示其管理域的域编号, 即 $ID_{RR}=ID_{RR_r}+ID_{RR_i}$, 其中 $ID_{RR_i}=0$ 。

根据定义2, MLON中结点 n_{vi} 有如下特点:

- (1) $\forall n_{vi} \in N_v$, 若 $E_i(n_{vi}) \neq \emptyset$, 则 $\exists E_i \in E$;
- (2) $\forall n_{vi} \in N_v$, 若 $f_i(n_{vi}) \neq \emptyset$, 则 $\exists f_i \in F$;
- (3) $\forall n_{vi} \in N_v$, 若 $\exists ID(f(n_{vi}))=ID_r(n_{vi})+0$ 。

MLON中的非叶子结点(即资源路由器)只负责对以它为根的资源子树进行管理, 并为其上层结点提供该子树的信息。在MLON中, 叶子结点与非叶子结点只是逻辑上的区分; 在实际中, 每个资源结点都体现为叶子结点, 非叶子结点实质是物理结点为组建MLON而提供的服务。

4 基于MLON的资源查找算法

定义3(网格资源发现请求处理模型) $K=(G, request$

$(R(A)), l)$, 其中 $request(R(A))$ 表示查找请求, 定义了要查找资源应该满足的各种属性约束; l 表示资源查找过程中的请求转发策略; K 表示在资源发现过程中找到的与查找请求匹配的资源集合。

虚拟网络中各个域内的资源结点间有很大的相识性; 在查找资源时, 用户通常关心的不是某个资源或某类资源, 而是只注意与需求最相关的 k 个结果。本文提出的基于MLON的资源查找算法(Source Searching Algorithm Based on MLON, MLON-SSA)将类型匹配路由的资源搜索技术(Source Searching Technique Based on Style Matching Routing, SMRT)和Topk资源搜索技术^[6](Source Searching Technique Based on Topk Algorithm, TopkT)相结合, 最终找到与查找请求相匹配的资源集合 K 。

4.1 算法描述

MLON中, 非叶子结点管理一个同类资源域, 并且有一个全局唯一的 ID_r 。SMRT是: 当 $request(R_i(A))$ 提交后, 由网络中的资源路由器运用 rf 函数产生一个 ID_{ri} , 用 ID_{ri} 与本域的 ID_r 做与运算, 如果产生的结果等于 ID_r , 说明要查询的资源在该 ID_r 域, 即 $\exists (ID_{ri} \wedge ID_r)=ID_r$, 则 $\exists r_i \in r(ID_r)$; 否则资源路由器采用策略 l 将请求向其父结点传递(无父结点, 向其兄弟结点传递), 直到找到符合请求的资源域。

与请求匹配的资源域内的每个资源可用 m 个属性来表示, 域内的所有资源可以映射到一个 m 维坐标中, 每个资源可用 m 维坐标中的一个点来表示。查找一个资源 $r(a_{m-1}, \dots, a_1, a_0)$ 可以看作是在 m 维坐标中寻找一个满足查询条件的点。因此, $request(R(A))$ 可以转化为在 m 维坐标中寻找满足查询条件且与查询点距离最近的 k 个资源。设 d_i 表示某资源 $r_i(a_{m-1}, \dots, a_1, a_0)$ 与查询点 r 的距离, 则:

$$d_i = \sqrt{\sum_{j=0}^{m-1} (a_j - a_j)^2}$$

在搜索资源时用户往往有对某种属性的偏好, 因此用户在查询时可以根据自己的需要对资源的属性加权。设 $W(w_{m-1}, \dots, w_1, w_0)$ 是资源 r 的属性权值。这时资源 r_i 与查询点 r 的距离 d_i 可表示为:

$$d_i = \sqrt{\sum_{j=0}^{m-1} ((a_j - a_j)^2 \times w_j)}, \text{ 其中 } \sum_{j=0}^{m-1} w_j = 1$$

当网格系统搜索找到逻辑上与请求匹配的最近的资源后, 它还必须能向用户提供物理上通信效率最高的资源。

假设3 网格中结点间的通信带宽和用户请求信息量一定, 则资源结点和用户请求结点间的通信效率与通信距离成反比。

根据假设3, 设 d'_i 表示拥有资源 r_i 的结点与用户请求结点间的距离, 则资源 r_i 与查询点 r 的距离 d_i 可表示为:

$$d_i = d'_i \times \sqrt{\sum_{j=0}^{m-1} ((a_j - a_j)^2 \times w_j)}$$

由上面的公式可以看出, 通信距离 d'_i 和逻辑距离 d_i 成正比, 即通信效率越低的资源在资源发现时越难被选中。

SMRT找到某类资源域后, 将请求传给与 ID_r 匹配的孩子的子结点。当最底层资源路由器RR接收到请求时, 该路由器向其管理的域内的每个结点转发 $request(R(A))$, 叶子结点接收到请求后, 首先检验自己是否满足查询要求, 如果满足, 则计算该资源与查询点的距离 d_i , 然后将 d_i 返回资源路由器; 资源路由器根据 d_i 选择与 $request(R(A))$ 最近的(即最佳的) k 个资源返回给用户。

算法1 基于MLON的资源查找算法

Input: $W, request(R(A)), k$.

Output: 最佳的 k 个资源.

Begin

(1) 向 MLON 发送 $request(R(A))$;

(2) $r_i(a) = request(R(A))$;

(3) $ID_n = rf(r_i(a))$;

(4) while($(ID_n \wedge ID_r) \neq ID_r$)

(5) 资源路由器 RR 采用策略 l 向父结点或兄弟结点传送 ID_n ; /* 搜索匹配资源的域 */

(6) while ($s(s(ID_r)) \notin \emptyset$)

(7) 向孩子结点传送 ID_n ;

/* 找到匹配资源的父结点 */

(8) RR_j 采用策略 l' 向孩子结点转发请求;

(9) 叶子结点 r_j 计算与 r_i 的距离 d_j , 并返回 d_j ;

(10) Topk = SelectMinD(d_0, d_1, \dots, d_n);

End

4.2 算法时间复杂度分析

假设网络的资源类型数为 n_t , 即 $|T| = n_t$, 与请求匹配的最低层资源域的结点个数为 n_v , MLON 的高度为 h , 资源返回数为 k .

由算法 1 可知, 资源发现的过程可分为两个阶段: (1) 用户请求查找资源域阶段, 最坏的情况下用户请求从叶子结点传送到 MLON 的顶层, 然后遍历顶层的资源路由器, 最后请求又传送到 MLON 的 $h-1$ 层资源路由器, 请求与结点比较的次数为 $(2h+n_t)$. 由于顶层的资源路由器所管理的资源数目都有一个阈值, 因此深度 h 为一常数. (2) 请求在匹配资源域传播阶段, 当域内的所有的资源均满足请求, 则比较的次数为 $(n_v+k(n_v-k))$, 所以算法的时间复杂度为 $O(n_t+n_v)$.

5 性能分析

在仿真环境下, 分别采用本文提出的资源发现机制及文 [5] 中的 RC 方法, 利用相同的转发策略进行资源查找, 对两种机制下的实验结果进行比较. 与文 [5] 类似, 在仿真过程中, 忽略一次资源发现过程中网络拓扑的动态变化, 并假设所有网络中的资源信息有效.

实验环境: $|R| = 50000, |N_G| = 10000, |T| = 100$, 假设所有结点拥有资源数量均匀分布, 且 $\text{Max}(N_{v_i}(R(A))) = 5000, \text{Min}(N_{v_i}(R(A))) = 50$.

根据上述的实验环境, MLON 中的结点 n_{v_i} 需满足下面两个条件:

(1) $\exists f(n_{v_i}) \in \emptyset$, 则 $50 \leq |N_{v_i}| \leq 5000, de(n_{v_i}) = 4$

(2) $\forall S(n_{v_i}) \in \emptyset$, 则 $50 \leq |N(f(n_{v_i}))| \leq 500, de(n_{v_i}) = 4$

其中, $de(n_i)$ 表示 MLON 中结点 n_i 在某层与其兄弟结点的度.

为了使两种资源发现机制之间具有可比性, 我们引入资源密度^[5]的概念:

定义 4 资源密度(Resource Density) 是满足查找条件资源数量占全体资源总数的比例, 即 $d = \frac{|N_R|}{|N|}$, 其中, d 表示 $request(R(A))$ 对应的资源 r 的密度, 单位为 $1/1000$; N_R 表示与 $request(R(A))$ 匹配的资源信息结点集合; N 表示全体资源信息结点集合.

图 2(a) 所示为两种机制在一次资源查找时涉及结点数

的比较. 从图中可以看出, 当与查询请求类型相同的资源数量较少时, 涉及到的结点数量基本相同; 但与查询请求类型相同的资源数量较多时, MLON 机制涉及到的结点数远远少于 RC 机制, 尤其当资源密度较小时, MLON 机制具有更加明显的优势.

图 2(b) 所示为两种机制在一次资源发现过程中查找跳数上的比较. 可以看出当 $N_i(R(A))$ 较小时, 两种机制的跳数基本相同; 但当 $N_i(R(A))$ 较大时, 随着资源密度的增加, 本文提出的 MLON 发现机制的查找跳数明显小于 RC 机制的查找跳数, 即 MLON 发现机制具有较高的效率.

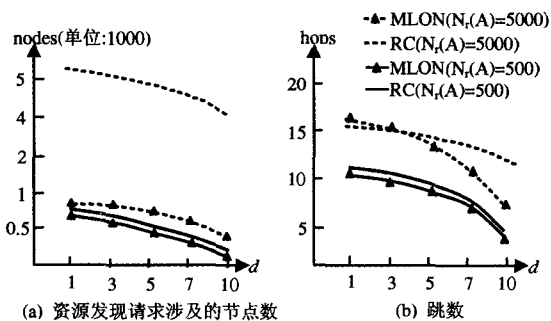


图 2 MLON 与 RC 的比较

在资源定位时, RC 采用的是贪婪的资源搜索方法, 它往往给用户返回属性偏差过大的结果, 即 $RC: |a_{return} - a_{request}| \rightarrow \infty$, 这个结果对用户来说往往并非最佳的, 而且资源结点和请求结点间的通信效率也得不到保证, 本文采用的 Topk 搜索技术, 它能根据用户的需要的一次返回 k 个最合适的资源, 即 $Topk: |a_{return} - a_{request}| \rightarrow 0$, 用户可以根据自己的需要在结果中挑选最佳的且通信效率最高的资源. 所以本文提出的资源发现机制在用户满意度方面要优于 RC 方法.

由以上仿真结果可以看出, 使用该文提出的资源发现机制, 可在数量较多某类资源中, 以较少的结点访问量和跳数, 取得较好的查询效率, 同时还可以查找到最合适的 k 个资源.

总结 本文基于物理资源的类型划分建立了一个层次结构的虚拟网络 MLON, MLON 网络能很好地满足网络动态性、分布性和扩展性的要求并能屏蔽资源间的异构性; 提出了一个基于 MLON 结构的资源发现机制, 该机制首先利用类型匹配路由的资源搜索技术将资源发现请求转发到资源所在的域, 然后采用 Topk 技术在相应的域内找出与请求匹配的最佳的 k 个资源, 仿真结果表明 MLON-SSA 能够以较高的效率搜索到与请求匹配的且用户满意度较好的 k 个资源.

参考文献

- 1 Iamnitchi A, Foster I T. A Peer-to-Peer Approach to Resource-Location in Grid Environments. Grid Resource Management, Kluwer, 2003
- 2 Talia D, Trunfio P. Peer-to-Peer Protocols and Grid Services for Resource Discovery on Grids. Grid Computing, The New Frontier of High Performance Computing, Advances in Parallel Computing, Elsevier Science, 2005(14)
- 3 Mastroianni C, Talia D, Verta O. A Super-Peer Model for Building Resource Discovery Services in Grids; Design and Simulation Analysis. In: Proc. European Grid Conference (EGC2005), LNCS, 2005(3470):132~143
- 4 Marzolla M, Mordacchini M, Orlando S. Resource Discovery in a Dynamic Grid Environment. In: Proc. DEXA Workshop 2005, 2005, 356~360
- 5 朱承. 基于资源类型的非集中式网格资源发现方法研究: [博士学位论文]. 长沙: 国防科学技术大学, 2004
- 6 水超, 贾焰, 邹鹏, 等. 一种 P2P 环境下高效 Topk 资源搜索技术. 计算机研究与发展, 2006, 43(Suppl): 108~116