第32卷 第24期 计 算 机 工 程 2006年12月
Vol.32 No.24 Computer Engineering December 2006
·网络与通信· 文章编号：1000—3428(2006)24—0112—03 文献标识码：A 中图分类号：TP393

一种基于DHT的动态网格命名解析方案

黄晓涛，叶淮光
(华中科技大学计算机学院，武汉 430074)

摘 要：在分布异构的网格环境下，如何有效地查找网格资源是一个必须解决的问题，网格命名协议是利用全局统一的命名空间来发现网格资源的。该文在网格命名协议中采用了一种查找路由算法DHT的解决方案，可解决网格命名解析服务的可扩展性和灵活性。
关键词：网格命名协议；命名解析；分布式哈希表

A Dynamic Grid Naming Resolution Algorithm Based on DHT

HUANG Xiaotao, YE Huaiguang
(College of Computer, Huazhong University of Science & Technology, Wuhan 430074)

【Abstract】 In grid, which is a distributed and heterogeneous environment, a fundamental problem is to locate required resource. In secure grid naming protocol, it has unique naming space, which can identify grid resource. This article tries to use DHT, which is a common route algorithm in peer-to-peer, as a new naming resolution way for secure grid naming protocol. The result is the features of DHT, including scalability, decentralization, availability, load balance, could be used in grid.
【Key words】 SGNP; Naming resolution; Distributed hashing table(DHT)

网格是一种新兴的基础设施，可以定义为“在动态的、多虚拟组织(VO)上的资源共享和协同问题求解”[1]。与传统的分布式计算相比，网格计算更加强调地理上的分布性，并且往往是针对高性能的计算、海量数据处理以及资源密集型的复杂任务。由于网格系统面对的是分布式异构环境下的问题，通常情况下它的解决方案是在现有各种系统之上形成一个虚拟层，用以屏蔽底层系统之间的差异，给最终用户提供一个透明的全局命名空间，但是随之而来的问题就是如何在这个全局命名空间上唯一地标志一个网格资源。解决这个问题的一个显然的策略是在虚拟层和物理层之间提供一种解析方法，实现在网格中有效的资源定位和查找，所以在网格中应该把命名解析服务作为一项必须具备的基本服务。

当前，网格系统采用一种层次式的全局命名方式解决网格资源的查找问题，但是随着网格应用范围不断的扩大以及网格规模动态变化和不断发展，最终可以扩展到把整个Internet变成一台超大型的虚拟计算机。这无疑是对层次式的命名解析提出了挑战。这种基于静态树形结构的网格规模难以扩展，同时对于网格环境下资源频繁地加入或失效这种动态性难以满足。本文针对网格系统规模动态扩展的问题，尝试把P2P和网格结合起来，将一种P2P系统中的分布式查找路由算法(Distributed Hashing Table，DHT)[3,6]应用到网格系统Legion中，希望能够将DHT的特点，包括规模可扩展性、去中心化、高可用性、负载均衡等，带入到网格系统中来，进而对于Legion以及由Legion项目所提出来的SGNP(Secure Grid Naming Protocol)协议[4,5]进行改进。

1 SGNP协议及命名解析方案

Legion是弗吉利亚大学研制的面向对象的网格项目。相对于其它网格系统，Legion是一个完全面向对象的系统，将网格看成是单一文件系统的单一虚拟操作环境，特别强调安全性，其核心特点是Legion有一个全局命名空间，要为系统中每一个对象提供一个全局唯一的命名，以便在讨论命名解析服务时，就可以在全局意义上来考察命名解析服务，更有针对性，所以选取Legion作为改造的对象，但是这里讨论的思想和方法对于其它网格系统也是适用的。

在Legion中命名空间分成3层：其顶层是用户可读的命名空间，通常是一些字符串，比如在全局空间下网格资源的路径名；第2层是逻辑命名Legion对象标识(Legion Object Identify，LOID)，由一些二进制串组成，用于唯一标志虚拟层上的网格资源，选定以后就不能更改；最下层则是对象地址(Object Addresses，OA)，包括具体通信地址和通信协议，可以随着需要改变的。在Legion中，异构系统之间的互操作性是通过唯一命名来获得的，而每一个Legion对象都被分配了一个唯一的、不可改变的LOID。基本的LOID由一系列可变长度的二进制串组成。

Legion项目在自己经验的基础上，于2002年向网格标准组织GGF提交了一个协议草案Secure Grid Naming Protocol(SGNP)，定义了在网格环境下资源定位的一种方式，并且SGNP协议还引入了一种无需可信第三方的信任机制。SGNP协议的主要内容包括几个方面：

首先是它的数据类型，位置独立对象标志(Location-independent Object Identifier，LOID)和物理绑定Binding。SGNP中的LOID是逻辑命名，对应于Legion中的LOID，Binding则对应于Legion中间的OA。在LOID和Binding之间存在一种映射关系。下面是LOID的具体格式：

LOID://<LOIDType>/<DomainResolverID>/<BindingResolverID>/<ObjectID>/<SecurityInfo>

基金项目：国家自然科学基金资助项目(60403027)
作者简介：黄晓涛(1966－)，女，副教授，主研方向：网络集成，网络安全，智能数据及语义Web；叶淮光，硕士生
收稿日期：2006-01-05 **E-mail：**huangxt@mail.hust.edu.cn

SGNP 协议中第 2 个方面的内容规定了网格中的资源，包括网格命名服务和解析层次。其中解析层次就是维护提到的 LOID-Binding 之间的映射关系。由于 LOID 也是采用了层次式的结构，在 SGNP 协议中命名解析时按照层次来解析，每层中都有一些固定的节点来提供解析服务。处于最顶层的解析节点 LOIDResolver 有一个公开的绑定，它的具体物理地址对于加入到这个网格中的所有对象都是显式的。其余层次中的每个命名解析节点上都需要维护一个映射表，其中包含着该节点下一层次中逻辑 ID 与 Binding 之间的对应关系中。另外 SGNP 还提供两种协议：绑定协议和重绑定协议，实际上就是上面所描述的命名解析层次来进行命名解析的过程，以及当 Binding 失效时获得有效 Binding 的过程。

在 SGNP 协议中，命名解析服务是采用一种静态的层次式的命名解析方式。这种方式具有层次清楚、结构简单、便于管理的优点。当网格规模比较大时，出现单点故障后系统的自恢复性比较差，这对于网格系统所要求的可容错性、可扩展性显然有所差距。最重要的是对于网格系统来说，大量存在的是暂时的服务和资源，如何维护解析节点上映射表的有效性和正确性是一项繁杂的工作。相比之下，如能够将这些管理工作交给系统完成，则制约网格规模难以扩展的问题应该可以解决。具体来说，我们希望提供一种动态的命名解析服务，使得网格中的服务和资源既能够被迅速定位，同时对系统而言，这些网格服务和资源能够动态地加入或离去。

2 基于 DHT 的动态命名解析服务

如果把网格资源和网格服务的名字也当作一种资源，那么实际上命名解析的过程也就是查找名字的过程。在传统的分布式系统中，资源查找主要有两种方式：穷举式查找和集中式查找。一般来说，穷举式查找的效率较低，而集中式查找的管理维护成本较高。分布式哈希表 DHT 是新一代的 P2P 系统中所使用的一种新的查找路由算法，DHT 综合了穷举式查找和集中式查找的优点，它通过在限定规模尺度下的节点上分别保存部分信息，并且控制网络中节点的标识，使得网络中每一个节点都能够在一个有效的跳数中到达任意一个节点。由于 DHT 算法既不存在中心服务器，同时又能够有效地查找信息，从而解决了网络规模扩展的问题。我们希望借助于 DHT 算法提供的规模可扩展性、去中心化、高可用性、负载平衡等优点来实现命名解析服务。下面首先以采用 DHT 技术的 P2P 系统 Chord 为例介绍 DHT 算法，然后描述怎样在 SGNP 协议中使用 DHT 来提供命名解析。

2.1 DHT 算法

DHT 即分布式哈希表，它的主要思想就是网络中所有的节点按照特定的拓扑结构组织起来，这样网络中任意两个节点就能够指定的跳数中到达对方，通常情况下是 logN(N 为网络规模)，具体情况和网络拓扑结构有关。

Chord是最先采用DHT算法的一种有结构的P2P系统，它通过对关键字的有控制的存放，实现了在logN数量级的查找性能。在Chord中，节点和关键字都有一个标志符，关键字存放于节点上。在这里，节点充当了一个哈希桶（hashing bucket）的概念。Chord网络的规模有一个确定值 2^m，这个值一旦确定就不能再更改。所有的节点和要存放的内容的键值都是在 2^m范围之内，形成两个重合的标志符环，其中节点的标识和关键字标志由SHA-1 等散列函数产生。存放关键字k的节点n就叫做该关键字k的后继节点，表示为successor(k)。每个节点维护着至多m个表项的路由表，称之为指针表(finger table)，其第i个表项表示为 $s=success(n+2^i-1) \mod 2^m$，这样，对于系统中任何一个节点而言，都可以构造出一棵以该节点为根节点的二叉查找树，这样在二叉树中就可以实现以O(logN)跳数找到任何一个节点。存放关键字时，按照关键字哈希后得到的值存放在相应的节点之上。查找关键字时，按照哈希值查询节点的指针表，得到相应的路由信息后，递归的转发查询请求到后继节点上，最终查找到所需要的信息。而当在节点加入或节点失效时，Chord系统不需要干预即可以 $O(\log^2 N)$ 的复杂度自行恢复到平衡状态。

2.2 使用 DHT 提供动态命名解析服务

从 P2P 的技术路线而言，强调有着更大规模程度上的参与者，以及对于所有的计算实体之间都是平等的，从本质上来说就有一个去中心化的趋势，系统的动态性要比较好。DHT 技术正是为了符合上述条件，创造出来的一种查找路由技术，针对网格规模扩展和动态性的问题，希望基于 DHT 提出一种动态命名解析方案，来改进 SGNP 协议中原有的层次式目录的命名解析方案。具体说来，在 SGNP 协议中，把 LOID 当作键值，相应的 Binding 当作键值所对应的内容，把按照 SGNP 协议构建的网格系统中的节点当作存储<Loid，Binding>值对的节点，则可以采用 DHT 来实现分散式存储方式，并且可以做到按键查询，从而实现命名解析服务。在我们的实现当中，DHT 的算法采取类似于 Chord 中的 DHT 实现方式，即所有系统中的节点构成一个 Chord 环，关键字构成另一个重合的 Chord 环。

下面考虑这样的场景，在网格中存在着大量的网格对象($>2^{30}$)和网格节点($>2^{20}$)，对于Legion这种采用全局统一视角的网格系统，这是合理的。理想的情况是网格系统中的节点都参与进来，所以假定默认情况下，所有加入网格的节点都提供命名解析服务。这个过程应该在节点初始化时完成。如果加入网格的节点不愿意提供命名解析服务，那么它应该有一个向提供命名解析服务Chord环的注册机制，这里暂不考虑这种情况。由于不同于层次式的命名空间，DHT算法也是一种哈希表，因此它要求命名空间是一维的，需要为系统中节点和网格对象提供一个唯一的ID，按照DHT算法的要求，这个ID的生成是对LOID进行哈希后得到，具体的哈希算法采取SHA-1 算法。其中SHA-1 算法可以提供比较好的负载均衡性。同时为了提高系统的冗余性，所有节点的命名和Binding都应该存放在不止一个节点上，给定一个阈值k（比如k=3）由LOID散列出来的值放在与其Node-ID值最相近的k个后继节点上。当发生节点失效时，不会产生命名不能解析的情况。同时为了防止死链接，每隔一定时间(比如 12h)就要在系统中重新发布一下的命名绑定。

在命名解析时，首先把要解析的 LOID 值经过 SHA-1 算法变换后得到在 Chord 环中的 ID，然后根据这个 ID 去查找存放<Loid，Binding>值对的 Node-ID，再在这个节点的本地表中去查找 LOID 对应的 Binding，从而完成命名解析过程。这个命名解析过程不同于层次式的命名解析，层次式命名解析的树形层次是固定的；而基于 DHT 的命名解析方式虽然也是一种树形的层次式解析，但是它可以从任意一个节点动态生成一个二叉树。为了显示区别，把这种命名解析方式称之为动态命名解析。同时，这种动态性也体现在对于系统而言，节点加入退出时，整个网格系统可以自适应这个过程，只要存储关键字的节点没有失效，那么一定可以保证这个关键字在 O(logN)就可以被找到。动态命名解析流程如图 1 所示。

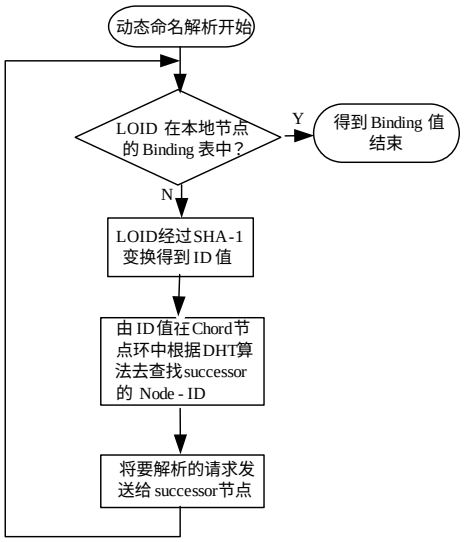

图 1 动态命名解析流程

3 结束语

P2P 网络和网格都是大规模的分布式系统中发展很迅速的技术，二者有着不同的背景和技术路线。但是二者有着近似的目标，并且也出现了逐渐融合的趋势。在本文中提出来使用一种在 P2P 网络中常用的技术 DHT，来解决一个网格系统 Legion/SGNP 中所需要的一种常见的服务——命名解析服务，试图将 P2P 网络中一些很好的特性带到网格中来，比如，P2P 的自组织性，负载均衡，规模可以扩展得非常大，最重要的是，P2P 网络中所提供的自组织性还可以有效地减少管理配置工作，减少人工干预，这不仅可以使网格更具有灵活性，同时这也更符合网络的精神。但是，另一方面，也应该看到网格和 P2P 的技术背景还是很不一样的。网格是针对特定用户的，从规模上来讲，大多数只是中等规模的网络，只是它的物理跨度可能很大。同时，虽然基本上所有的网格系统中都要提供类似于命名解析的服务，但是具体各个网格系统的解决方案是不一样的，有些并不提供全局统一的命名空间，从而本文中所提出的解决方案也就失去了意义。所以本文中所提出来的方案还是有其局限性的。

网格和 P2P 具有不同的技术优点，但是目前在 P2P 系统和网格系统之间还缺乏交流，需要在具体的网格环境下去实现并验证本文中所提出来的这种方法。同时，DHT 技术在实现上也有多种选择，选择一个更适合于网格环境下的实现也是很重要的。今后，这两方面的研究与改善将是我们研究工作的重点。

参考文献

1 Foster I, Kesselman C, Tsudik G. The Anatomy of the Grid: Enabling Scalable Virtual Organizations[J]. International Journal of Super Computer Applications, 2001, 15 (3): 200-222.

2 Czajkowski K, Fitzgerald S, Foster I, et al. Grid Information Services for Distributed Resource Sharing[C]. Proceedings of the 10th IEEE International Symposium on High-performance Distributed Computing, 2001.

3 Stoica I, Morris R, Karger D, et al. Chord A Scalable Peer-to-peer Lookup Service for Internet Applications[C]. Proceedings of ACM SIG2 COMM´01, San Diego, California, USA, 2001.

4 Grimshaw A S, Natrajan A, Humphrey M A, et al. From Legion to Avaki: the Persistence of Vision[M]. Chichester: John Wiley & Sons, 2003: 266-298.

5 Apgar J, Grimshaw A S, Harris S, et al. Secure Grid Naming Protocol: Draft Specification for Review and Comment[EB/OL]. http://sourceforge.net/projects/sgnp.

6 董 芳, 费新元, 肖 敏. 对等网络 Chord 分布式查找服务的研究[J]. 计算机应用, 2003, 23(11): 25-28.

(上接第 100 页)

4 结论

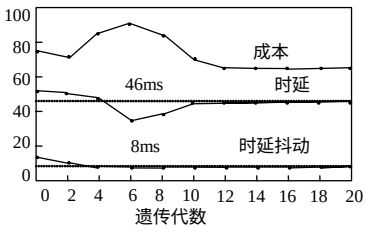

图 9 D＝46，J＝8，L＝0.001，B＝70 时组播树成本、时延、时延抖动随遗传代数变化曲线

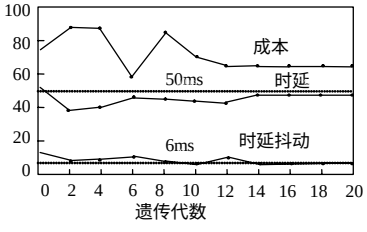

图 10 D＝50，J＝6，L＝0.001，B＝70 时组播树成本、时延、时延抖动随遗传代数变化曲线

针对 QoS 组播路由问题，本文给出了带宽、时延、时延抖动、丢包率以及成本最小 QoS 组播路由模型，并利用改进的遗传算法进行求解。该算法主要有以下特点：(1)采用预处理机制，有效地减少了算法编码空间和搜索空间，显著地提高了算法的搜索效率；(2)特殊的树结构编码，使染色体长度达到固定，简化了编码操作，简化了复杂的编码和解码过程；(3)通过对交叉操作和变异操作的改进，使算法迅速跳出局部最优解，向全局最优的方向发展，同时加快了算法收敛的速度(见图 9、图 10)。

参考文献

1 王征应, 石冰心. QoS 组播路由的启发式遗传算法[J]. 电子学报, 2001, 29(2): 253-256.

2 Xiang F, Junahou L, Jieyi W, et al. Qos Routing Based on Genetic Algorithm[J]. Computer Communication, 1999, 22(9): 1394-1399.

3 Ravikumar C P, Bajpai R. Source-based Delay-bounded Mult Isting in Multimedia Networks[J]. Computer Communication, 1998, 21(2): 126-132.

4 石 坚, 邹 玲, 董天临等. 遗传算法在组播路由选择中的应用[J]. 电子学报, 2000, 28(5): 88-89.

5 刘 莹, 刘三阳. 基于遗传策略的实时多点传送路由算法[J]. 西安电子科技大学学报, 2000, 27(2): 215-218.

6 杨 云, 徐永红, 李千目等. 一种QoS路由多目标遗传算法[J]. 通信学报, 2004, 25(1): 43-51.