

IndexPeer: 半结构化 P2P 系统资源发现模型 及其 DHT 算法^{*}

刘小虎¹, 蒋从锋^{1,2}, 李 垦¹

(1. 华中科技大学 工程计算与仿真研究所, 武汉 430074; 2. 杭州电子科技大学 计算机学院, 杭州 310018)

摘 要: 提出了一种基于半结构化 P2P 系统的资源发现模型——IndexPeer 模型。该模型采用三层索引结构, 由下至上分别为 virtual organization 层、service peer 层和 indexpeer 层。IndexPeer 充分利用 P2P 系统中的节点性能差异, 具有高效的查找功能, 且对网络的高度动态性有良好的适应性。针对 Index Peer 模型的体系结构, 对该模型的结构化层 DHT 算法进行了研究和设计。基于 small-world 的仿真实验表明, IndexPeer 模型的体系结构及其结构化层 DHT 算法具有较好的鲁棒性和一定的可扩展性。

关键词: 分布式哈希表; 点对点系统; 索引; 资源发现; 搜索

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2008)06-1648-04

IndexPeer: semi-structured resource discovery model and its distributed hash table based search algorithm for P2P systems

LIU Xiao-hu¹, JIANG Cong-feng^{1,2}, LI Ken¹

(1. Engineering Computing & Simulation Institute, Huazhong University of Science & Technology, Wuhan 430074, China; 2. School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: This paper proposed a semi-structured resource discovery model——IndexPeer. IndexPeer was a hybrid hierarchical architecture consisting of three levels: virtual organization, service peer, and index peer from bottom to top. The architecture could bear high heterogeneity, high dynamicity, and high self-adaptivity in P2P system. It proposed a resource discovery model based on DHT for IndexPeer. Simulations were run in a small-world network. The results show that IndexPeer is robust, scalable, and fault tolerant and can greatly improve the resource search efficiency.

Key words: distributed hash table(DHT); P2P(peer-to-peer) system; index; resource discovery; search

P2P 是实现资源共享的一种分布式技术^[1], 是当前研究的热点。P2P 系统中各个节点松散地结合在一起, 节点可以随意加入或退出系统。当前的 P2P 网络一般可以归纳为三种结构: 集中协调式、完全分布式和混合式结构^[2]。集中协调式结构有单点故障的缺陷; 完全分布式 P2P 系统的资源发现查询请求以泛洪(flooding)方式在网络中传播, 会产生过多的通信负载^[3]; 混合式结构引入超级节点(ultra-peer)的概念, 指定一些节点成为与其相邻节点的服务器, 成为超级节点, 各超级节点通过完全分布式结构连接起来。混合式非结构化 P2P 系统对系统拓扑结构的高度动态性有较好的适应性, 但由于搜索查询过程需通过点与点之间的联系来完成, 资源发现及检索效率受查询算法的影响较大。

近年来, 基于分布式哈希表(DHT)的分布式资源组织、管理与发现服务正在应用到 P2P 系统、网格计算、Web 服务等分布式环境中, 成为构建大规模分布式应用的基础结构^[4~7]。DHT 的主要思想是: 首先将每条文件索引表示成一个 (K, V) 对。其中: K 称为关键字(key), 可以是文件名(或文件的其他描述信息)的哈希值; V(value) 是实际存储文件的节点的 IP 地

址(或节点的其他描述信息)。所有的文件索引条目(即所有的 (K, V) 对) 组成一张大的文件索引哈希表, 只要输入目标文件的 K 值, 就可以从这张表中查出所有存储该文件的节点地址。然后再将上面的大文件哈希表分割成很多局部小块, 按照特定的规则把这些小块的局部哈希表分布到系统中的所有参与节点上, 使得每个节点负责维护其中的一块。这样, 节点查询文件时, 只要把查询报文路由到相应的节点即可(该节点维护的哈希表分块中含有要查找的 (K, V) 对)。为了完成上述查询, 节点要按照一定的规则来分割整体的哈希表, 进而也就决定了节点要维护特定的邻居节点, 以便路由能顺利进行。这个规则因具体系统的不同而不同, CAN、Chord、Pastry 和 Tapestry 都有自己的规则, 也就呈现出不同的特性^[8]。基于 DHT 的分布式检索和路由算法因为具有查找可确定性、简单性和分布性等优点, 正成为国际上结构化 P2P 网络研究和应用的热点。

1 相关工作

P2P 系统的资源发现请求是按照一定的扩散策略转发到各个节点进行匹配的。由于 P2P 系统具有高度的动态性、异

收稿日期: 2007-05-30; 修回日期: 2007-08-17 基金项目: 华中科技大学创新基金资助项目(HF04012006271)

作者简介: 刘小虎(1968-), 男, 陕西西安人, 副教授, 博士, 主要研究方向为网格与分布式计算、数值算法、计算机仿真(xhliu@mail.hust.edu.cn); 蒋从锋(1980-), 男, 湖北枣阳人, 博士研究生, 主要研究方向为网格计算、P2P 与分布式计算和计算机仿真; 李垦(1983-), 男, 福建闽清人, 硕士研究生, 主要研究方向为网格计算、P2P 计算及其资源发现。

构性和自治性,资源发现算法的效率和质量成为影响 P2P 系统性能最重要的因素^[9]。当前 P2P 系统中的资源发现与搜索算法一般可分为结构化与非结构化两种。结构化算法包括 Chord、Pastry、CAN 和 Tapestry 等^[10],它们在可扩展性、容错能力和搜索效率等方面存在一定缺陷;非结构化算法的典型代表是 Gnutella^[11]和 Freenet^[12]。其中,Gnutella 将查询请求以广播方式发给所有邻居节点,搜索算法简单、有效,但是通信开销较大;Freenet 通过沿途文件复制和 LRU(least recently used)算法实现相似标志文件聚集。但是,Freenet 大量的沿途复制文件会造成较大的网络通信负载,而且不能保证对稀有文件对象的查询结果^[13]。

文献[13]基于用户节点命中查询的历史信息,提出了一种无结构 P2P 环境下占用带宽较小、易维护的自适应搜索策略 PeerRank。但是,文中没有给出实际 P2P 系统下 PeerRank 的性能表现。考虑到实际 P2P 系统中路由表的更新速度无法跟上网络拓扑的变化速度,文献[7]提出了一种基于 DHT 的加速路由表资源搜索算法 ROAD(routing on active and demand)。ROAD 算法通过构建加速路由表、加快发现服务的速度降低消息转发的延时,并通过幂次序组播算法改善对超级点的依赖性。

2 IndexPeer 模型体系结构

IndexPeer 是一种基于半结构化 P2P 索引技术的资源发现模型。底层以集中式资源发现模型为基础;中间层由服务节点组成非结构化分布式网络;顶层采用结构化 P2P 网络技术,引入 IndexPeer 方法,采用分层的查询机制。IndexPeer 模型将结构化与非结构化 P2P 网络的优点相结合,对于非结构化网络查询代价较高的缺点有较好的改进,并解决了结构化网络节点异质性的问题,能适应 P2P 系统高度的动态性。IndexPeer 资源发现模型分为三层,如图 1 所示。

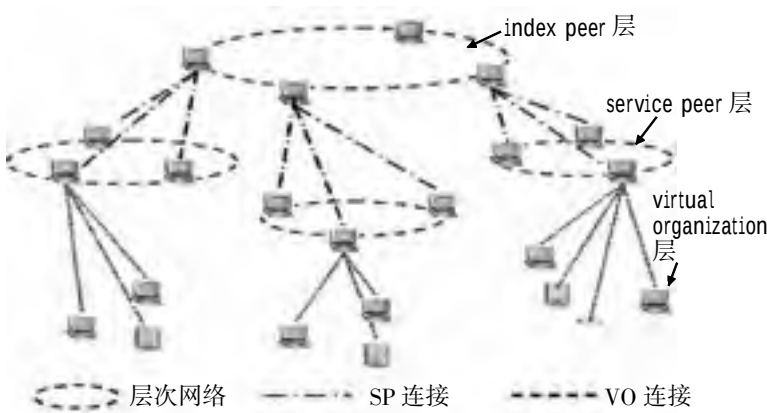


图 1 三层 Index Peer 资源发现模型

底层由提供资源服务的各 virtual organization(VO)节点组成。VO 节点可以采用现行的各种集中式 P2P 结构向外提供服务。各节点需要在上一级的分布式 service peer(SP)中注册信息。VO 节点可以是资源的提供者,同时也可以作为资源的请求者。

第二层的 SP 以对等的形式构造非结构化 P2P 网络。任一个 SP 在一个或多个 IndexPeer 中注册,并与 IndexPeer 下属的 SP 建立邻居节点关系。SP 在加入和离开时需要在下层 VO 节点中寻找新的替代者。如果没有符合的 VO 节点,便根据各 VO 节点与邻居 SP 的连接响应时间与该邻居 SP 的饱和度,将

VO 节点重新注册到邻居 SP 上,并将结果提交给所属 IndexPeer,以修改其资源对应表(路由表)。SP 相对于 IndexPeer 来说可以自由地加入或离开。SP 将下属 VO 节点的资源信息以关键字 $K_1, K_2, K_3, \dots$ 的形式发往所在的 IndexPeer 注册,并将资源关键字 K 以及 K 所在的资源位置 V 一道存储在 SP 的资源表中。

IP(index peer)层由 IndexPeer 组成结构化 P2P 网络。该网络建立了一个分布式哈希索引表,每个 IP 有特定的 ID 号以及各自下属 SP 的资源对应表。索引表中的 ID 号与节点的网络路由地址有惟一对应的关系。IP 的选取根据其在线的稳定时间与可靠性确定,选择的是高效率、变动少、可靠性高的节点,类似于非结构化 P2P 混合网络中的超级节点。但是与非结构化网络的节点不同的是,IP 的加入与相邻节点建立联系时需要遵循结构化 P2P 网络的构建原则,加入与退出需要对顶层网络进行修复工作。另外 IP 层中,各节点之间的联系通过分布式哈希索引表来检索资源,而不像非结构化网络中通过点与点之间的联系来直接查找。

3 DHT 搜索算法

3.1 需要解决的问题

要将 DHT 算法引入 IndexPeer 资源发现模型中,应该解决以下几个方面的问题: a) DHT 算法在 IP 层的设计应该遵循的原则是在节点失效、遭受攻击和突发性高负载面前都能表现出很好的鲁棒性^[14]; b) 具有良好的可扩展性,能以较低系统开销获得较大的系统规模; c) 可以自我配置,不需要手工干预就可以自动把新加入的节点合并到系统中; d) 提供简单灵活的接口,可以为 SP 层的多个 P2P 系统同时使用。

为解决以上问题,本文提出了一种满足 IndexPeer 资源发现模型的 IP 层改进算法。

3.2 DHT 算法设计

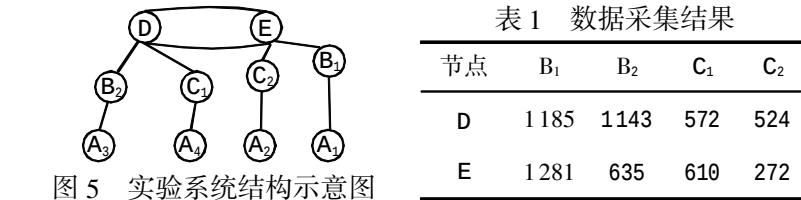
IndexPeer 模型中,关键字标志符 S_{req} 通过哈希关键字本身得到,而 IP 节点的 ID 号则通过散列(hash)节点的 IP 地址得到。所有 IP 节点按照其 ID 号从小到大(取模 2^m 后)沿着顺时针方向排列在一个逻辑的标志圆环上。圆环的映射规则是:关键字标志符为 S_{req} 的 (S_{req}, S_{addr}) 对存储在 IP 节点下的非结构化 SP 网络中,存储方式遵循空间均匀分布原则。该 IP 节点的节点标志等于 S_{req} 或者在环上紧跟在 S_{req} 之后,这个节点被称为 S_{req} 的后继节点,表示为 $successor(S_{req})$ 。因为标志符采用 m 位二进制数表示,并且在 $0 \sim 2^m - 1$ 顺时针排列成一个圆圈, $successor(S_{req})$ 就是从 S_{req} 开始顺时针方向距离 S_{req} 最近的节点。

SP 层网络中提供一个随机的资源分配算法,如依据 SP 号的泊松分布,将该资源的地址 S_{addr} 存储在资源表 $res(S_{req}) = S_{addr}$ 上。当 SP 节点正常退出时,它将资源表上的资源信息转移到邻居 SP 节点中,以保证资源始终处于 IP 节点下。这种规则可以在小范围内准确定位资源的位置,SP 节点的加入和退出对资源的区域定位没有影响。

图 2 给出了一个 $m=6$ 的圆环,环中分布了 10 个 IP 节点,存储了 5 个关键字,节点标志前加上 ID 而关键字前加上 S_{req} 以

64 MB; VO 节点要求主频 ≥ 100 MHz, 内存 ≥ 32 MB。

仿真过程如下: 节点通过运行在每个仿真机上的节点状态监测进程获取节点运行参数, 以 10 s 为一个周期从本节点收集。采样结果通过一定的映射规则计算得出, D、E 节点赋值为 1, 可充当 IP 节点; B₁、B₂、C₁ 和 C₂ 节点赋值为 2, 可作为 SP 节点; A₁、A₂、A₃ 和 A₄ 赋值为 3, 为 VO 节点。通过比较, 采集 SP 节点到 IP 节点的连接情况, 依次将 SP 节点加入 IP 节点, 得到 10 次 CPU、存储量和带宽综合平均值, 如表 1 所示。



资源要求的实现过程如下: 假如由 A₁ 提交新资源 K₁, 通过 DHT 相关监听函数取得函数关键字, 此处选用 SHA-1 哈希算法。计算结果将 K₁ 分配给 E。SP 层采用随机算法将 K₁ 分配给 C₂ 存储。当 A₂ 在系统中提交资源查询请求时, 在 C₂ 的哈希表中得知 K₁ 的存在, 将结果返回给 DHT 相关监听函数, 查找得出 K₁ 的 E 节点位置, 并提出资源请求。E 节点将 C₂ 地址返回, C₂ 提供 A₁ 地址连接。至此, 完成资源查找。

5 结束语

P2P 系统中资源发现机制的可用性和有效性取决于其是否适应 P2P 系统的高度动态性和查询的高效性。本文提出的 IndexPeer 模型采用三层半结构化索引网络, 充分考虑了 P2P 系统中节点性能差异, 按节点的处理能力分配职能, 克服了结构化 P2P 网络的缺点。顶层的结构较稳定, 修改代价较小, 下层节点采用非结构化网络, 可自由加入和退出, 能很好地适应 P2P 系统的高度动态性。

参考文献:

[1] ARAMESWARAN M, SUSARLA A, WHINSTON A B. P2P networking: an information sharing alternative[J]. Computer, 2001, 34(7): 31-38.

[2] ANDROUTSELLIS-TEOTOKIS S, SPINELLIS D. A survey of peer-to-peer content distribution technologies [J]. ACM Computing Surveys, 2004, 36(4): 335-371.

[18] LINDSEY S, RAGHAVENDRA C S. PEGASIS: power-efficient gathering in sensor information systems[C] // Proc of IEEE Aerospace Conference. Montana: IEEE Aerospace and Electronic Systems Society, 2002: 1125-1130.

[19] MANJESHWAR A, AGARWAL D P. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks[C] // Proc of the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing. 2001: 2009-2015.

[20] MANJESHWAR A, AGRAWAL D P. APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks[C] //Proc of International, Parallel and Distributed Processing Symposium. Florida: IEEE Computer Society, 2002: 195-202.

[21] XU Ya, HEIDEMANN J, ESTRIN D. Geography informed energy conservation for Ad hoc routing[C] //Proc of the 7th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2001: 70-84.

[22] YU Yan, ESTRIN D, GOVINDAN R. Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor net-

[3] OO B T, HUEBSCH R, STOICA I, *et al.* The case for a hybrid P2P search infrastructure[C] //Proc of the 3rd International Workshop on Peer-to-Peer Systems. Heidelberg: Springer-Verlag, 2004: 141-150.

[4] 李运娣, 冯勇. 基于 DHT 的 P2P 搜索定位技术研究[J]. 计算机应用研究, 2006, 23(10): 226-228.

[5] GUPTA I, BIRMAN K, LINGA P, *et al.* Kelips: building an efficient and stable P2P DHT through increased memory and background overhead [C] //Proc of the 2nd International Workshop on Peer-to-Peer Systems. Berlin: Springer-Verlag, 2003: 160-169.

[6] LI Jin-yang, STRIBLING J, MORRIS R, *et al.* A performance vs cost framework for evaluating DHT design tradeoffs under churn [C] // Proc of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Piscataway: IEEE Press, 2005: 225-235.

[7] 杨峰, 李凤霞, 余宏亮, 等. 一种基于分布式哈希表的混合对等发现算法[J]. 软件学报, 2007, 18(3): 714-721.

[8] RATNASAMY S, SHENKER S, STOICA I. Routing algorithms for DHTs: some open questions[C] //Proc of the 1st International Workshop on Peer-to-Peer Systems. Heidelberg: Springer-Verlag, 2002: 45-52.

[9] 施晓秋. 非集中式 P2P 系统中资源搜索与现存问题分析[J]. 计算机工程, 2007, 33(5): 91-93.

[10] BALAKRISHNAN H, KAASHOEK M F, KARGER D, *et al.* Looking up data in P2P systems[J]. Communications of the ACM, 2003, 46(2): 43-48.

[11] ZHU Ying-wu, HU Yi-ming. Enhancing search performance on Gnutella-like P2P [J]. IEEE Trans on Parallel and Distributed Systems, 2006, 17(12): 1482-1495.

[12] CLARKE I, SANDBERG O, WILEY B, *et al.* Freenet: a distributed anonymous information storage and retrieval system[C] //Proc of the ICSI Workshop on Design Issues in Anonymity and Unobservability. Berkeley: International Computer Science Institute, 2000: 311-320.

[13] 冯国富, 毛莺池, 陆桑璐, 等. PeerRank: 一种无结构 P2P 资源发现策略[J]. 软件学报, 2006, 17(5): 1098-1106.

[14] GODFREY B, LAKSHMINARAYANAN K, SURANA S, *et al.* Load balancing in dynamic structured P2P systems[C] //Proc of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies. Piscataway: IEEE Press, 2004: 2253-2262.

[15] ZHANG H, GOEL A, GOVINDAN R. Using the small-world model to improve Freenet performance[C] //Proc of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies. New York: IEEE Press, 2002: 40-49.

works, UCLA/CSD-TR-01-0023[R]. [S. l.]: UCLA Computer Science Department, 2001: 1-11.

[23] CHEN Ben-jie, JAMIESON K, BALAKRISHNAN H, *et al.* SPAN: an energy efficient coordination algorithm for topology maintenance in Ad hoc wireless networks[C] //Proc of the 7th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2001: 85-96.

[24] SUBRAMANIAN L, KATZ R H. An architecture for building self configurable systems[C] //Proc of the 1st ACM International Symposium on Mobile Ad hoc Networking and Computing. Piscataway: IEEE Press, 2000: 63-73.

[25] RODOPLU V, MENG T H. Minimum energy mobile wireless networks [J]. IEEE Journal Selected Areas in Communications, 1999, 17(8): 1333-1344.

[26] YOUNIS M, YOUSSEF M, ARISHA K. Energy aware routing in cluster-based sensor networks[C] //Proc of the 10th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems. Washington DC: IEEE Computer Society, 2002: 129-136.