

无结构P2P覆盖网络的拓扑优化*

冯国富^{1,2+}, 张金城¹, 姜玉泉¹, 顾庆², 陆桑璐², 陈道蓄²

¹(南京审计学院 信息科学学院,江苏 南京 210029)

²(南京大学 计算机科学与技术系,江苏 南京 210093)

Optimization of Overlay Topology for Unstructured Peer-to-Peer

FENG Guo-Fu^{1,2+}, ZHANG Jin-Cheng¹, JIANG Yu-Quan¹, GU Qing², LU Sang-Lu², CHEN Dao-Xu²

¹(School of Information Science, Nanjing Audit University, Nanjing 210029, China)

²(Department of Computer Science and Technology, Nanjing University, Nanjing 210093, China)

+ Corresponding author: Phn: +86-25-58318603, E-mail: fgfmail@dislab.nju.edu.cn

Feng GF, Zhang JC, Jiang YQ, Gu Q, Lu SL, Chen DX. Optimization of overlay topology for unstructured peer-to-peer. Journal of Software, 2007,18(11):2819-2829. <http://www.jos.org.cn/1000-9825/18/2819.htm>

Abstract: This paper addresses the issue that what is the optimal topology for the purely distributed peer-to-peer (P2P). It is usually assumed that unstructured network will form a Power-Law topology. However, a Power-Law structure is not always the best for all the applications. In this paper, the influence of the overlay topology on P2P search is firstly investigated and the precise relation among the distribution of the degree, the query rate and the success rate is attained. Then an optimal distribution model of degrees in terms of the popularity of data items is proposed to guide the overlay construction. Finally, the experimental results show that the proposed topology is effective to improve the success rate for the Random Walks. The fundamental result is significant for the optimal overlay structure and offers a new understanding of the replica deployment in unstructured P2P.

Key words: unstructured P2P (peer-to-peer); overlay topology; Random Walks

摘要: 研究了全分布无结构 P2P(peer-to-peer)网络拓扑的最优化问题.通常认为,无结构 P2P 网络拓扑属于 Power-Law 结构.然而,Power-Law 并非对所有应用都是最好的选择.首先研究了无结构 P2P 覆盖网络结构对无结构 P2P 搜索的影响,给出了结点度分布、访问频率模式和搜索成功率之间的关系.然后基于数据访问频率分布,给出了结点度的优化分布模型.实验结果表明,该无结构 P2P 拓扑优化结构在提高搜索成功率方面是有效的.该工作对构造合理的覆盖网络拓扑具有重要意义,同时将加深对无结构 P2P 环境下数据部署问题的认识.

关键词: 无结构 P2P;覆盖网络拓扑;Random Walks

中图法分类号: TP393 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant Nos.60573106, 60402027 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z199 (国家高技术研究发展计划(863)); the Colleges Oriented Provincial Natural Science Research Plan of Jiangsu Province of China under Grant Nos.07KJD520052, 06KJD520090 (江苏省高校自然科学研究计划)

Received 2006-09-23; Accepted 2007-02-08

在全分布的 P2P(peer-to-peer)系统中,每个结点维护少量邻居结点信息,构成应用层覆盖网络.查询消息在结点之间转发实现匹配检索.依靠查询消息遍历结点的搜索方式简单,但其搜索效率低,网络开销大.目前,提高无结构全分布体系结构 P2P 系统性能的方法主要有两种:一是通过收集、交换结点信息来提高路由效率;二是通过大量数据复制来提高命中效率.数据复制在搜索过程中对数据副本进行动态管理,使访问频率较高的数据在系统拥有较多的副本,以提高搜索成功率、降低搜索延迟.但数据复制往往会带来网络流量开销.

与数据复制不同,我们假设数据已部署到结点,结点基于存储内容流行程度由此形成不同权重;本文研究了如何基于结点权重构造最优覆盖网络,让热点数据处于更重要的位置,满足更多用户的查询请求,以提高系统搜索性能和服务性能.

本文第 1 节回顾相关工作.第 2 节分析拓扑结构对搜索的影响,并给出覆盖网络连接度分布的优化模式.第 3 节给出模拟实验和性能分析.第 4 节总结全文.

1 相关工作

依靠查询消息遍历结点的搜索方式简单、有效,适应P2P系统的高度动态性,得到了广泛的研究和使用.Gnutella^[1]是最早采用查询消息广播实现网络资源发现的P2P系统.Gnutella能够较好地适应结点的动态加入和退出,但会产生大量冗余消息,占用网络带宽,不具有可扩展性.为此,提出了许多算法来提高消息数量的可控性和查询消息的命中率.

1.1 消息转发模式

无结构P2P通过查询消息转发匹配本地存储来实现搜索定位.在Flooding消息转发模式中,结点将查询消息

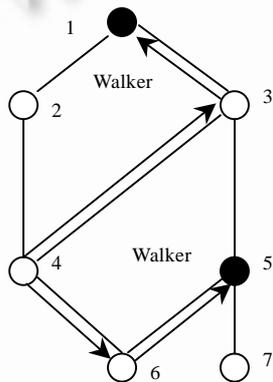


Fig.1 An example of overlay network and the routing algorithm of Random Walks

图 1 覆盖网络示例和 Random Walks 路由算法示例

转发给所有与之相邻的结点,并用一个TTL(time to live)值来控制查询消息转发次数.Modified-BFS(modified breadth-first search)^[2]是对Flooding的“横向”改进,并非每次都转发给所有邻居结点,而是随机选择部分邻居以削弱查询消息扩散.Expanding Ring^[3]是对Flooding的“纵向”改进.查询源首先以较小的TTL进行Flooding搜索,如果搜索失败,则增大TTL继续搜索,直到搜索成功或者TTL达到某一上限值.在Random Walks^[3]中,查询者生成 k 个查询消息,每个消息在结点之间依次转发传递以匹配查询.这样的每一个消息被称为walker,如图 1 所示.同样,每个walker具有一个TTL来限制最多传递的次数.Random Walks路由算法可以通过调整walker的个数在搜索成功率和消息数量之间作出动态折衷,具有更好的灵活性,是最常用的消息转发模式之一.我们采用Random Walks消息模式作为本文的路由算法,研究了该消息模式下结点度分布与结点收到查询消息的能力以及搜索成功率之间的关系.

对消息转发模式的改进仅仅使得查询消息在数量规模上变得可控,并不能提高搜索效率,因为盲目搜索算法中的每个结点满足查询的概率是相同的,搜索成功的能力取决于数据副本的“密度”以及所遍历结点的数量.与以往从路由算法和消息模式角度提高搜索性能的方法不同,本文根据结点权重,赋予结点的不同“网络地位”,让重要结点收到更多的查询,发挥更重要的作用,从而提高系统性能.

1.2 覆盖网络拓扑

结构化 P2P 将结点组织成诸如环、超立方体、蝶形拓扑、小世界等各种规则网络,数据根据内容存放到相应结点,数据内容和存放位置之间具有耦合性.结构化 P2P 具有搜索命中率保证,具有更小的搜索延迟.但针对结点的动态加入和退出,需要对规则网络进行实时维护,维护代价大.

在全分布的P2P系统中,每个结点维护少量邻居结点信息,构成应用层覆盖网络.查询消息在结点之间转发,实现匹配检索.网络拓扑结构对搜索性能具有重大影响.Freenet^[4]通过LRU(least recently used)存储空间管理算法、大量的文件复制和路由表调整使具有相似id的文件聚集,存储相似id文件的结点聚集,用以模拟Small World网络.在文献[5]中,结点随访问行为动态调整结点连接,使具有相近兴趣属性的结点在覆盖网络中聚集以提高搜索效率.无结构P2P搜索中的结点仅随着结点加入和退出作动态演化.Gnutella参考通信延迟随机组织覆盖网络.文献[6]中提到,Gnutella拓扑中结点度呈Power-Law分布.文献[7]旨在建立一种小直径($O(\log N)$)的随机网络以保证较小的平均搜索路径长度.文献[8]充分考虑了节点构造拓扑的理性,即节点只愿意与能为其带来利益的节点建立连接,能够提高网络效率、遏制恶意节点.

目前的系统通常认为无结构 P2P 拓扑为 Power-Law 结构,文献[9]以此提出了基于网络特征的改进措施.文献[10]以降低搜索路径为目标,证明 Random Walks 消息路由下 Power-Law 拓扑对搜索性能并非最优,并提出了平方根规律的结点度分布优化策略.但它做了很强的假设,一方面假设路径长度没有 TTL 限制,另一方面假设每个数据对象在系统中只有一个副本,这是不太现实的.本文分析了 Random Walks 消息模式下无结构 P2P 搜索中的副本分配模式、TTL 设置、覆盖网络结点度分布等对搜索成功率的影响,并提出了在 Uniform 数据分布情况下覆盖网络拓扑的优化结构.

2 覆盖网络拓扑优化

P2P 系统首先将结点组织成应用层覆盖网络,每个结点仅掌握邻居结点信息,查询消息在结点之间转发实现查询消息匹配检索.因此,一个结点对整个系统的“贡献”在于所存储副本被命中且提供服务的能力,体现于该结点能够收到查询的能力以及所存储内容被访问的频率.

2.1 模型假设

假设一个具有 N 个结点和 m 个相异文件对象的无结构 P2P 系统,每个结点存储相异文件集合 ρ . q_i 代表文件 i 的访问频率,也就是在最近一段时间内对文件 i 的访问次数占总访问次数的比例,有 $\sum_{i=1}^m q_i = 1$ 成立.设定文件访问频率遵循 Zipf 分布,并按访问频率大小排序编号,有 $q_1 > q_2 > q_3 > \dots > q_m$.假设数据对象副本数量采用 Uniform 分配,即数据对象在系统中具有相同的副本数量 r .结点 u 的连接度用 d_u 表示,系统所有结点度之和为 D ,平均度为 \bar{d} .平均度稳定性可通过设定结点加入系统时的初始连接度为 \bar{d} 来保证.结点 u 的权重体现在其存储内容与收到的一个查询匹配的能力,用 Q_u 表示,其中, $Q_u = \sum_{i \in \rho_u} q_i$.假设覆盖网络中结点之间的邻居关系构成一个连通无向图 $G=(V,E)$,其中, V 为结点集合, E 为无向边的集合.假设数据文件最初以随机复制方式进行部署,即每个结点所存储数据对象之间不存在必然性联系.采用 Random Walks 查询消息路由算法,并设定 walker 的个数为 w .

2.2 网络拓扑与搜索成功率

如果将一个文件副本放在结点 u ,那么,该副本的“价值”等于查询消息能够遍历到结点 u 的次数以及查询消息与结点内容相匹配的概率.

在基于 TTL 的搜索算法中,一个结点被遍历的概率与消息模式和网络拓扑相关.在 Flooding 模式中,只有 u 结点有效覆盖区域以内的结点,即与 u 结点距离小于 TTL 的结点,发起的查询才能到达 u 结点;而且从 u 结点有效覆盖区域内结点发起的查询,也必能到达 u 结点.所以在 Flooding 搜索算法中,增加一个副本的“收益”取决于所存放该副本结点的有效覆盖区域内的结点数量.在 Random Walks 模式中, u 结点有效覆盖区域内结点发起的查询不一定遍历到结点 u ,结点 u 被遍历的概率往往与网络结构、该结点所处的位置以及查询在网络上的分布相关.如图 1 所示,当结点 4 收到查询并转发时,使用 Random Walks 模式,其邻居结点收到该查询消息的概率分别为 $1/2$,即结点 4 度数减 1 的倒数(查询消息一般不按原路返回).每个结点发起一次查询,一跳之内到达结点 1 和结点 4 的平均消息个数分别为 $5/6$ 和 $4/3$.

平均情况下的一次查询, u 结点能够收到查询消息的数量 M_u 为

$$M_u = d_u \times \left(\frac{\bar{d}}{D} \times \frac{w}{\bar{d}} + \frac{w \times (TTL - 1) \times (\bar{d} - 1)}{D} \times \frac{1}{\bar{d} - 1} \right) = d_u \times \frac{w \times TTL}{D} \quad (1)$$

其中, \bar{d}/D 代表一般情况下, 查询源与结点 u 一个确定度之间存在一条连接的概率; $w \times (TTL - 1) \times (\bar{d} - 1)/D$ 代表一次查询遍历的结点与结点 u 之间存在一条连接的概率. 其中, w/\bar{d} 意指, 如果结点 u 与查询源之间存在一条连接, 该连接在一次查询中被路由选中的概率; $1/(\bar{d} - 1)$ 代表结点 u 在一次路由中被选中的概率.

式(1)表明, 一般情况下, 一个结点收到查询消息的能力与其连接度之间存在密切联系. 所以, 一个直观的方案, 赋予重要结点以更多的连接度将能提高该结点存储对整个系统的贡献, 提高整个系统的搜索成功率.

但另一方面, 一次查询中对同一结点的访问只有一次遍历是有效的; 太大的结点连接度容易对结点造成重复性访问, 消耗 TTL 而不能增加命中的机会.

在一次查询中, 结点 u 被遍历 t 次的概率为

$$M_{ut} = C_{d_u}^t (w \times TTL / D)^t, \quad 1 \leq t \leq TTL/2 \quad (2)$$

其中, $w \times TTL/D$ 与前面一样, 代表结点 u 一个确定度在一次查询中被连接且被遍历的概率. $C_{d_u}^t$ 代表结点 u 的连接一次被遍历 t 条的所有可能组合.

这样, 结点 u 在一次查询中被有效遍历的概率 EM_u 为

$$EM_u = \sum (M_{ut} / t) = \sum (C_{d_u}^t (w \times TTL / D)^t / t) \quad (3)$$

EM_u 代表 u 结点在一次查询中能够收到的有效消息的平均个数. 当 walker 的个数与 TTL 取值较小, D 取值较大时, 一个结点收到重复消息可以忽略. 本文 t 取值为 1, 并令 $T = TTL \times w$, 则有

$$EM_u = T \times d_u / D \quad (4)$$

2.3 共享单个存储空间时的最佳结点度分布

首先考虑每个结点只共享 1 个存储空间的情况, 结点权重由结点存储数据的流行程度确定, 并假设拥有相同权重的结点具有相同连接度.

按照发起查询分布规律共发起 S 次查询, 则共产生 $T \times S$ 个消息, 平均每个结点发起查询的次数为 S/N . 根据式(4), u 结点在 S 查询中收到的有效查询消息数目为

$$EM_u = T \times S \times d_u / D \quad (5)$$

也就是说, 当系统只有一个数据对象且为结点 u 仅部署一个副本时, 该结点能够命中 EM_u 次查询, 平均能够保证 n 个结点的所有查询, 其中,

$$n = EM_u / (S / N) = T \times d_u / \bar{d} \quad (6)$$

但是, 当在系统中部署第 2 个副本时, 第 2 个结点能够保证查询的新增结点个数小于第 1 个, 因为仅就搜索是否成功而言, 第 2 个副本的部署仅对查询没有保证的结点有效, 对于已被保证的结点来说是无效的. 以此类推, 增加的副本越多, 一个新增副本所能保证查询的新增结点个数就会越少. 用 n_i 代表第 i 个副本部署之后所能保证查询的结点总个数, 用 d_i 代表存储数据 i 副本结点的连接度 (存储相同副本的结点具有相同连接度, 都用 d_i 表示), 将对象的 r 个副本逐一部署, 并令 $\Delta = T \times d_i / \bar{d}$, 则

$$\begin{cases} n_1 = \Delta \\ n_2 = n_1 + \Delta \times (1 - n_1 / N) \\ n_3 = n_2 + \Delta \times (1 - n_2 / N) \\ \dots \\ n_i = n_{i-1} + \Delta \times (1 - n_{i-1} / N) \end{cases} \quad (7)$$

可得 r 个副本部署之后能够覆盖的结点总个数为

$$n_r = \Delta \sum_{k=0}^{r-1} \left(\frac{N - \Delta}{N} \right)^k = N [1 - (1 - T \times d_i / D)^r] \quad (8)$$

因为结点发起查询的概率相同, 所以, 部署数据对象 i 的 r 个副本后带来的搜索成功率 SR' 为查询已被保证

结点与结点总数的比值.

$$SR' = n_r / N = 1 - (1 - T \times d_i / D)^r \quad (9)$$

以上结论是基于系统中只有一个数据对象的假设,所有访问都针对该数据对象;只要收到查询消息,该副本就一定能够命中.当系统中存在多个数据对象时,一个数据对象*i*的副本与查询匹配的概率为 q_i ,所以,数据对象*i*的所有副本对整个系统搜索成功率的贡献为

$$SR_i = q_i (1 - (1 - T \times d_i / D)^r) \quad (10)$$

式(10)表明,数据对象*i*的所有副本对整个系统搜索成功率的贡献是其所在结点连接度的函数.下面结合限制条件 $\sum d_u = D$, 利用拉格朗日乘子法求搜索成功率最大时每个结点的连接度.

构造复合函数:

$$F(d_u) = \sum_{u=1}^N (q_u (1 - (1 - T \times d_u / D)^r)) - \lambda \left(\sum_{u=1}^N d_u - D \right), 1 \leq u \leq m \quad (11)$$

其中, d_u 为存储数据对象为*u*的结点的连接度.对每个 d_u 求偏导,并等于0,得

$$\begin{cases} q_u \times \frac{T}{D} \times \left(1 - \frac{T}{D} d_u\right)^{r-1} = \lambda \\ q_v \times \frac{T}{D} \times \left(1 - \frac{T}{D} d_v\right)^{r-1} = \lambda \\ \dots \end{cases} \quad (12)$$

$$\begin{cases} q_u \times \frac{T}{D} \times \left(1 - \frac{T}{D} d_u\right)^{r-1} = q_v \times \frac{T}{D} \times \left(1 - \frac{T}{D} d_v\right)^{r-1} \\ \sum_{u=1}^m r d_u = D \end{cases}, 1 \leq u, v \leq m, d_v > 0 \quad (13)$$

通过对式(13)解方程可得

$$d_v = \left(1 + \frac{T/r - m}{\sum_{u=1}^m \left(\frac{q_v}{q_u} \right)^{\frac{1}{r-1}}} \right) \frac{D}{T}, d_v > 0 \quad (14)$$

2.4 共享多个存储空间时的最佳结点度分布

当每个结点共享多个存储空间时,同一数据的多个副本部署于具有不同结点度和权重的不同结点,每个副本带来的搜索成功率收益受到该数据所在结点以及其他副本所在结点连接度的影响.由于同一结点的多个数据副本之间随机组合,对其中一个副本的最优结点度对其他副本来说,往往并非是最优的.因此,共享多个存储空间时结点度的最优分配并非是对每个数据副本的最优分配的简单叠加,因而也不纯粹是结点权重 Q_i 的函数.鉴于类似的原因,文献[10]提出的基于结点权重统一模式来优化整个覆盖网络度分布的策略被赋予了很强的假设限制,即每个数据对象在系统中只拥有一个副本.

当每个结点共享多个存储空间、采用 Uniform 副本分配模式时,一种次优的启发式结点度分配策略是:将每个结点的每个存储副本作为一个独立的结点,以第 2.3 节的方法计算每个副本连接度的最优分配;然后将每个结点存储副本所分配的最优度相加,即为该结点分配的结点度.结合以上分析结论,在启发式结点度分配中,结点*v*的结点连接度应为

$$d_v = \sum_{i \in \rho_v} \left(1 + \frac{T/r - m}{\sum_{u=1}^m \left(\frac{q_i}{q_u} \right)^{\frac{1}{r-1}}} \right) \frac{D}{T}, d_v > 0 \quad (15)$$

启发式分配方法中的结点度是单共享空间中每个副本最优度分布的一个折衷,因此,其整体性能会受到同一结点数据副本之间访问频率的影响.因为结点度的优化分配就是查询消息在结点优化分配的过程,所以在副

本连接度合并过程中,结节点度相近的副本,在结节点度合并之后可以较好地保持结节点度与总结节点度的比例,维持理论上的查询消息分配量,并保持理论上的成功率最优值.同理,当副本间访问频率差异较大时,合并之后的结节点度会是一个折衷,不能充分发挥热点副本的作用,而对非流行副本来则会造成结节点度的浪费.

已知结节点度优化分布理论模型,文献[10]给出了一种通过结节点局部信息实现覆盖网络最佳度分布的方法.我们在今后的工作中也将做更进一步的研究.

3 实验

3.1 实验环境

实验环境包括 10 000 个结节点、500 个相异数据对象.在同一结节点不存在同一数据对象的两个副本.为了保持覆盖网络连通性,设定每个结节点最小连接度为 1.

随机抽取结节点发起查询,以每 50 000 次查询为一个时间片计算搜索成功率和平均搜索路径长度.采用 Random Walks 消息路由算法.以搜索成功率、平均搜索路径长度为主要测度来衡量系统性能.与 Web 网页访问不同,P2P 访问命中往往会直接下载数据,不会对数据重复性访问,因此,数据访问频率分布采用较平缓的 Zipf 分布.与实验相关的参数及其缺省值见表 1.

Table 1 Parameters and their default values in simulation

表 1 实验环境参数及其缺省值

Parameter	The number of peers	Datstore size c	Routing algorithm	The number of data objects	Query rate	The degree sum D of all peers
Value	10 000	10	Random walks	500	Zipf ($\alpha=0.92$)	80 000

3.2 结节点度、结节点收到消息的能力与搜索

我们首先验证结节点收到消息的能力与结节点度之间的关系.每个结节点贡献一个存储空间,并随机部署副本.结节点度是其权重的函数,分别采用正比例和 Power-Law($\alpha=2.0$)分布函数.TTL 设置为 100.随机选取结节点发起查询,查询消息达到 TTL 才终止,共 1 000 000 次查询,记录每个结节点转发的查询消息数,统计平均每 10 000 次查询中每个结节点收到的消息数量(如图 2 所示).

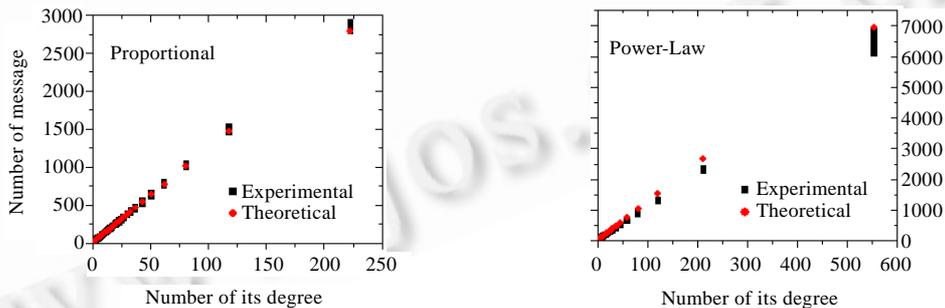


Fig.2 Relation between the number of messages that a peer receives and its degree

图 2 一个结节点收到的查询消息数量与其结节点度之间的关系

如图 3 所示为平均 10 000 次查询中,结节点收到的查询消息数量与其连接度之间的关系,理论结果通过式(1)给出.可以看出,结节点收到的消息数量与其连接度成正比,实验结果与理论结果较好地吻合,具有一致性.

如图 4 所示为平均 10 000 次查询中,结节点收到的有效查询消息数量与其连接度之间的关系,理论结果通过式(3)给出,其中 $t < 4$.可以看出,实验结果与理论结果能够较好的吻合,具有一致性.而且当结节点连接度分布较为均匀时,结节点收到的有效查询消息数量与其连接度接近正比.因此,第 2.2 节关于“有效消息正比于结节点连接度”的假设在一定程度上是合理的,尤其对于连接度小于 200 的绝大多数结节点.

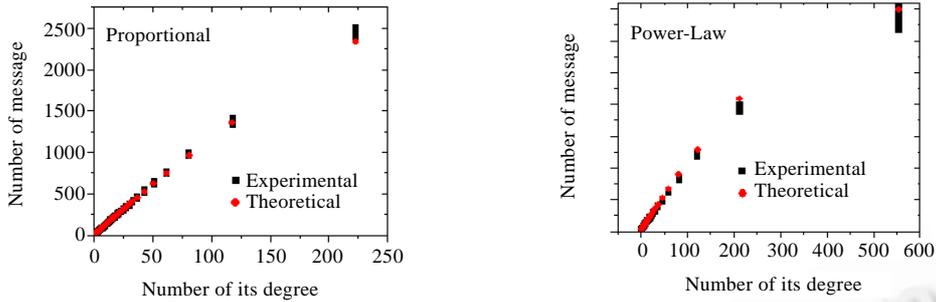


Fig.3 The number of effective messages a peer received vs. the number of its degree
图 3 结点收到的有效消息数量与其连接度之间的关系

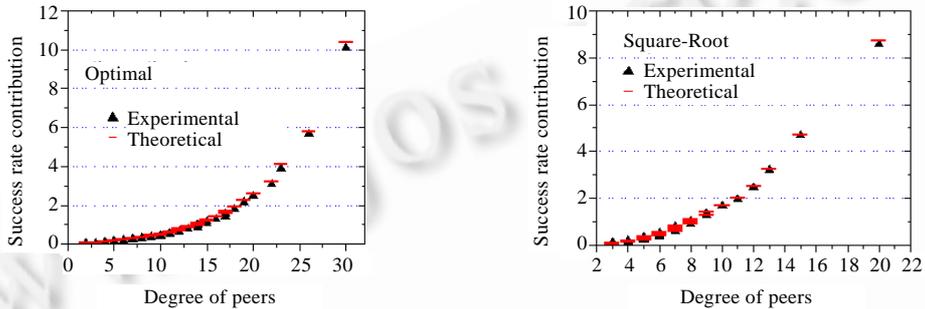


Fig.4 Relation between the success rate, access frequency and the degree
图 4 搜索成功率与数据访问频率、结点度之间的关系

式(10)所表述的搜索成功率与结点连接度之间的关系是结点度可以进行优化以及优化结论正确与否的依据,图 4 是两者之间关系的验证.每个结点共享一个存储空间,网络拓扑分别采用基于式(14)的优化网络(图 4 的左图)和根据文献[10]给出的基于结点权重的 Square-Root 网络(图 4 的右图).结点权重由所存储副本的访问频率给定.共有 200 个文件对象,结点总数为 40 000,TTL 为 25,walker 的个数为 4,理论值由式(10)给出.图 4 说明,在两种拓扑下,实验结果和理论结果都能很好地吻合.它说明结点度、收发消息能力和搜索成功率之间存在着密切的关系,通过调整结点度调整查询消息在结点分布是可行的.式(10)说明,一个数据副本随着分配结点度的增多,单位结点度增量带来的搜索成功率增量逐渐减少,因此,依据数据副本的流行程度,通过调整结点度分配优化整体的搜索成功率是可行的.

3.3 每个结点共享单个存储空间时的结点连接度分布

以下是对每个结点共享单个存储空间时,不同拓扑下搜索性能的实验验证.其中,TTL 取值为 25,walker 的个数为 4.

图 5 显示出各种覆盖网络拓扑下搜索成功率随结点总度数的变化情况.Square-Root 根据文献[10]“结点度正比于结点权重平方根”生成.Constant 拓扑中的所有结点的连接度都相同.Power-Law 根据偏好依附算法生成.Constant 和 Power-Law 拓扑中的结点度独立于结点权重.每条边关联的两个结点都是随机生成的.

图 5 说明本文给出的拓扑结构优化策略是有效的.在结点度独立于结点权重的 constant 和 Power-Law 拓扑中,在每次遍历结点数量相同的情况下,搜索成功

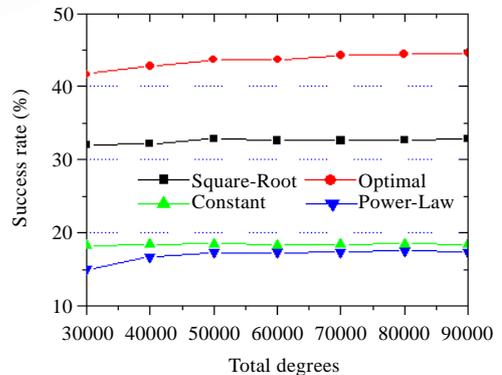


Fig.5 The success rate under the different topologies
图 5 各种拓扑下的搜索成功率比较

与否的关键因素在于数据副本在结点中的“密度”;由于每种拓扑中的数据分配和副本密度都相同,所以后两者的搜索成功率基本相当.而且在总结点度较小的情况下,Power-Law 分布更容易造成对大度数结点的重复访问,所以总结点度在 30 000 时,Power-Law 的搜索成功率偏低.由于 Square-Root 考虑了结点权重,因此搜索成功率有所提高.Optimal 曲线在单存储空间时的搜索成功率在理论上是最优的,高于所有其他拓扑的搜索成功率.

图 5 同时说明,搜索成功率仅与结点度在结点的分配比例相关,而与绝对连接度数值关系不大,因为根据式(6),一个结点收到查询消息的能力取决于连接度与平均结点度的比值.

图 6 是各种拓扑下搜索成功率随 TTL 和 walker 个数变化的曲线.由于每次遍历的结点数即 $T=TTL \times walker$ 维持不变,所以,TTL 和 walker 变化对各种拓扑下搜索成功率的影响都不大.当 walker 的个数等于 16 时,搜索成功率有较大的下降,这是因为这时的 walker 的个数远大于平均结点度,使得 walker 之间产生不同程度的路径重叠冲突造成的.

图 7 反映了系统中数据对象数量对搜索成功率的影响.在总存储空间不变的情况下,随着数据对象数量的不断增长,每个数据对象在系统中的密度逐渐减小,每种拓扑下的搜索成功率都有不同程度的降低,而 Square-Root 的搜索成功率下降比 optimal 曲线更为明显.这是因为在本文给出的优化策略中,结点度会综合副本数量、数据对象数量等多种因素,对网络拓扑进行最优化.而且如式(10)所示,最优结点度分配的确与数据对象数量和数据副本数量存在一定关系.而 Square-Root 仅仅根据结点权重一个因素,不能适应副本数量和数据对象数量的变化.这同时也说明,Square-Root 的拓扑优化能力是有条件、有限度的.

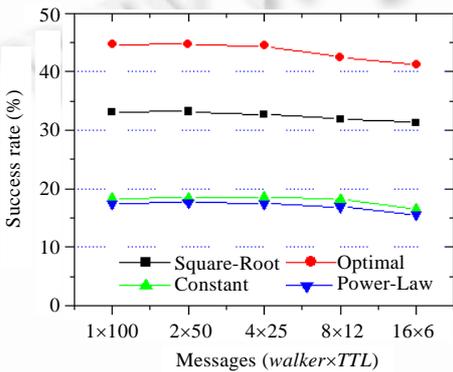


Fig.6 The success rate vs. TTL and the number of walkers

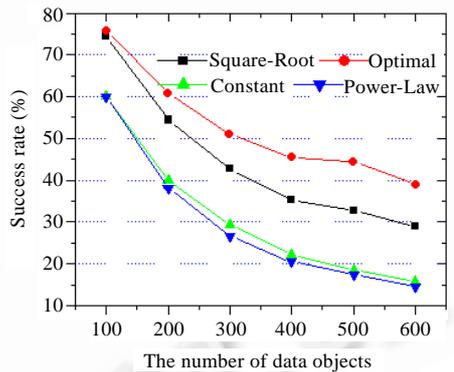


Fig.7 The success rate vs. the number of objects

图 6 搜索成功率与 TTL 和 walker 个数之间的关系 图 7 搜索成功率与数据对象数量之间的关系

图 8 给出了各种拓扑下平均路径长度和系统数据对象数量之间的关系.TTL 设定为 100,walker 的个数设置为 1.可以看出,在系统存储空间确定的情况下,数据对象数量增加,每个数据对象的副本数目会相应减少,数据副本“密度”降低,命中查询的概率降低,所耗费的消息数量增多.图 8 显示,constant 曲线和 Power-Law 曲线几乎重合,这是因为对于独立于结点权重的覆盖网络,除非结点度的巨大差异造成对大度数结点的重复性无效访问,其搜索性能完全取决于访问频率分布和数据副本在系统中的密度.所以,在相同副本“密度”情况下,constant 拓扑和 Power-Law 拓扑具有近乎相同的搜索成功率和平均搜索路径长度.而本文提出的优化网络拓扑具有最好的性能,这是因为拓扑优化策略使搜索成功率达到最高,使搜索失败而达到 TTL 的 walker 数最少,减小消息量消耗,降低搜索路径长度.同时,结点度优化策略根据数据访问频率分配给结点以最合适的结点度,会尽量满足热点数据的性能,提高命中率,降低搜索延迟,使得整体平均搜索路径变短.而 Square-Root 拓扑下的平均搜索路径最长,这说明,在每个数据具有多个副本时的 Square-Root 结点度分布并非最优.图 9 也表明,在基于结点权重的 Power-Law 结点度分布中,Square-Root 的性能要低于正比网络(结点度正比于结点权重)的性能.Square-Root 结点度分布更趋向于均匀分布,不能充分发挥热点数据的作用.而当指数过大时会造成热点数据的连接度浪费,并

损害非热点的性能.如图 9 所示,平均搜索路径长度呈先下降后上升的趋势.

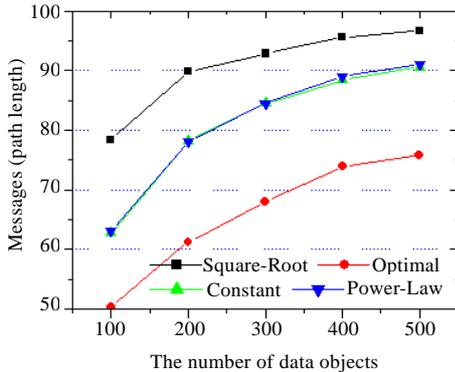


Fig.8 Average path length vs. the number of data objects

图 8 平均搜索路径长度与数据对象数量之间的关系

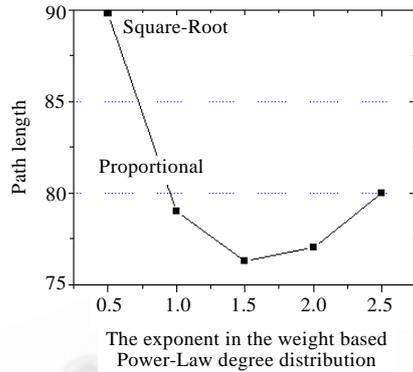


Fig.9 The path length vs. the exponent under the node weight based Power-Law degree distribution

图 9 在基于结点权重的 Power-Law 度分布中,搜索路径长度与其指数之间的关系

3.4 每个结点共享多个存储空间时的结点连接度分布

下面给出每个结点共享多个存储空间时,各种拓扑下的搜索性能比较.系统共有 500 个数据对象,每个结点共享 10 个存储空间,TTL 设置为 7,walker 的个数设置为 3.

图 10 是每个结点共享 10 个存储空间时各种拓扑下的搜索成功率比较.与共享单个存储空间时一样,系统中总的结点连接度数量对系统搜索性能影响不大.基于结点权重的 Square-Root 结点度分布比独立于结点权重的 Power-Law 拓扑具有更高的搜索成功率.本文给出的拓扑构造策略仍然保持了最高的搜索成功率.如公式(14)所示,在本文给出的优化拓扑中,不仅数据对象的访问频率是影响网络拓扑的重要因素,系统中的数据对象数量、每个数据的副本数量、系统存储空间数量等也构成了优化网络拓扑的重要考量因素.

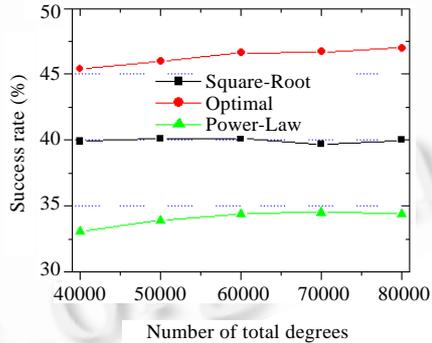


Fig.10 The success rate under different topology

图 10 各种拓扑下的搜索成功率比较

在从每个结点共享单个空间的最优结点度分布过渡到共享多个空间的过程中,每个结点存储内容访问频率之间的差别大小是影响其优化效果的重要因素.本文在进行结点度分配时,首先将每个副本作为一个单独的结点并分配给最合适的结点度.如果需要每个副本在存储空间合并后保持原来的贡献,则每个副本在合并前后需要保持所分配结点度与平均结点度的比例.当一个结点的几个副本流行程度差异较大时,所分配的理想结点度差异会较大,合并后就会偏离与平均结点度的既定比例,不能充分发挥每个数据副本的功能;当一个结点的几个副本流行程度差异较小时,各个副本具有相似的理想结点度,在合并之后容易保持原来的与平均结点度的比例.

在图 11 中,Closer 中每个结点的 10 个副本的访问频率在顺序上是连续的;Randomly 中每个结点数据副本的访问频率是随机的;在 Farther 中,每个结点数据副本之间的访问频率的差异尽可能地大,访问频率之间顺次间隔 50 个次序.正如上面所分析的,图 11 显示,每个结点上数据副本的访问频率之间差异越小,其搜索成功率越高,优化效果也就越明显.

图 12 显示了每个结点共享空间个数对搜索成功率的影响,其中,Rand 是指每个结点所共享空间个数符合泊

松分布,平均空间为 10.可以看出,本文给出的覆盖网络拓扑优化策略始终保持了较高的搜索成功率.Optimized 在 Rand 实验中的搜索成功率略高于每个空间全部为 10 时的搜索成功率,可能的原因在于,当存储空间实行泊松分布后,每个结点存储数据访问频率间平均差异减小的缘故.Square-Root 拓扑下的搜索成功率也提高了 4%,可能的原因在于结点之间权重差异分布进一步明显,使更多的数据副本集中匹配大连接度带来的大量查询消息,进一步发挥了高权重结点的作用.本文提出的覆盖网络拓扑优化策略可以适应结点共享存储空间数量的变化,具有一定的适应性.

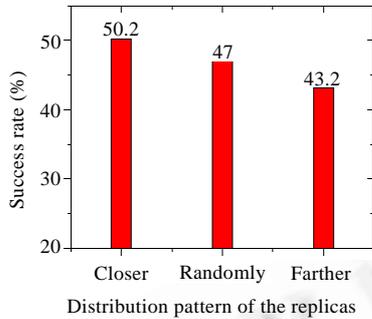


Fig.11 The impact of the diversity in query rate of objects in a peer to the success rate

图 11 同一结点数据访问频率差异性对搜索成功率的影响

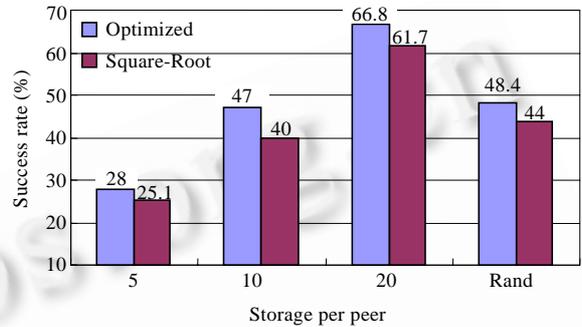


Fig.12 The impact of the datastore size of peers' to the success rate

图 12 结点存储空间大小对搜索成功率的影响

本文通过调整结点连接度分布,让高权重结点收到更多的查询消息,以命中更多查询发挥更大作用.以上实验也表明,本文提出的 P2P 覆盖网络拓扑优化方法在提高搜索成功率、降低搜索路径长度方面是成功的,而且对 TTL、walker 的个数、数据对象数量和共享空间数量的变化具有适应性.但与之相应的一个问题是结点负载均衡.由于一个结点可以根据工作负荷主动决定是接受还是拒绝另一个结点的连接请求来抑制负载增加,也可以通过数据复制实现负载动态转移,达到既发挥热点数据作用、又不致结点负载过度的目标,因此,本文不再对负载均衡做进一步的讨论.

4 结论

本文研究发现,无结构 P2P 中结点连接度与收到的查询消息数量、命中查询次数具有密切关系.因此对流行程度不均衡的数据对象副本来说,结点度成为一种可分配的资源.本文试图解决无结构 P2P 搜索中的一个问题:在结点总度数确定、结点内容确定的前提下,如何优化结点度分配、构造合适的覆盖网络拓扑结构以提高搜索成功率?本文首先研究了结点收到查询消息能力与结点度之间的关系,并将搜索成功率问题转化为网络拓扑的覆盖问题.在明确结点内容、结点连接度和搜索成功率之间关系的基础上,提出了一种基于结点内容访问权重的覆盖网络优化策略.实验表明,本文给出的优化拓扑可以提高系统搜索成功率,而且高于通常使用的 Power-Law 随机网络拓扑.同时,本文的结论也可以作为构造其他大规模分布式应用中应用层覆盖网络的指导,具有一定的理论和现实意义.

References:

- [1] Gnutelliums LLC. Gnutella protocol specification version 0.4. 2006. http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf
- [2] Kalogeraki V, Gunopulos D, Zeinalipour-Yazti D. A local search mechanism for peer-to-peer networks. In: Proc. of the 11th Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2002. 300-307.

- [3] Lü Q, Cao P, Cohen E, Li K, Shenker S. Search and replication in unstructured peer-to-peer networks. In: Proc. of the Int'l Conf. on Measurements and Modeling of Computer Systems. New York: ACM Press, 2002. 84–95.
- [4] Clarke I, Sandberg O, Wiley B, Hong TW. Freenet: A distributed anonymous information storage and retrieval system. In: Federrath H, ed. Proc. of the Workshop on Design Issues in Anonymity and Unobservability. Berlin: Springer-Verlag, 2001. 46–66.
- [5] Condie T, Kamvar S, Garcia-Molina H. Adaptive peer-to-peer topologies. In: Lambrix P, Duma C, eds. Proc. of the 4th Int'l Conf. on Peer-to-Peer Computing. New York: IEEE Press, 2004. 53–62.
- [6] Ren Y, Sha CF, Qian WN, Zhou AY, Ooi BC, Tan KL. Explore the “small world phenomena” in pure P2P information sharing systems. In: Proc. of 3rd Int'l Symp. on Cluster Computing and the Grid (CCGrid). IEEE Computer Society, 2003. 232–239.
- [7] Pandurangan G, Raghavan P, Upfal E. Building low-diameter P2P networks. IEEE Journal on Selected Areas in Communications, 2003,21:995–1002.
- [8] Tian HR, Zou SH, Wang WD, Cheng SD. Incentive compatible adaptive topologies for P2P networks. Journal of Software, 2006, 17(4):845–853 (in Chinese with English Abstract). <http://www.jos.org.cn/1000-9825/17/845.htm>
- [9] Adamic L, Huberman B, Lukose R, Puniyani A. Search in power law networks. Physical Reviews E, 2001,64:46135–46143.
- [10] Cooper BF. An optimal overlay topology for routing peer-to-peer searches. In: Proc. of the 6th Int'l Middleware Conf., Middleware 2005. Berlin, Heidelberg: Springer-Verlag, 2005. 82–101.

附中文参考文献:

- [8] 田慧蓉,邹仕洪,王文东,程时端.激励一致的自适应 P2P 拓扑构造.软件学报,2006,17(4):845–853. <http://www.jos.org.cn/1000-9825/17/845.htm>



冯国富(1977—),男,山东沂水人,博士,讲师,主要研究领域为分布式计算,协议测试.



顾庆(1972—),男,博士,副教授,CCF 高级会员,主要研究领域为分布式计算,软件质量,软件测试.



张金城(1962—),男,博士生,教授,主要研究领域为 CIMS,计算机审计.



陆桑璐(1970—),女,博士,教授,CCF 高级会员,主要研究领域为分布式计算,高性能计算.

姜玉泉(1961—),男,教授,主要研究领域为计算机审计,分布式计算.



陈道蓄(1947—),男,教授,博士生导师,CCF 高级会员,主要研究领域为分布式计算,软件工程.