ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.21, No.10, October 2010, pp.2622−2630
doi: 10.3724/SP.J.1001.2010.03647

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于社会网络特征的 P2P 内容定位策略*

黄永生+, 孟祥武, 张玉洁

(北京邮电大学 计算机学院,北京 100876)

Strategy of Content Location of P2P Based on the Social Network

HUANG Yong-Sheng+, MENG Xiang-Wu, ZHANG Yu-Jie

(School of Computer, Beijing University of Posts and Telecommunications, Beijing 100876, China)

+ Corresponding author: E-mail: hyosheng@sohu.com, http://www.cs.bupt.cn

Huang YS, Meng XW, Zhang YJ. Strategy of content location of P2P based on the social network. *Journal of Software*, 2010,21(10):2622−2630. http://www.jos.org.cn/1000-9825/3647.htm

Abstract: Enhancing the efficiency of the file location is important in the study of unstructured P2P network. Flooding and random walks are simple and easily implemented. However, the former will increase the load of P2P network to much and put bounds to the search depth, and the latter's lower network load and deeper search comes at the cost of lower search breadth and more response time. This paper puts forward a strategy of content location and routing of a search request in an unstructured P2P network. By applying the rationale of social network and simulating the ability of the peers of social network, the strategy proposed in this paper, can make better use of the ability of the peers and locate files faster with lower search depth and breadth. Supported by the equivalent hardware environment, the experimental results demonstrate that the time spent on content location can be reduced by more than 50%.

Key words: P2P; content location; social network; forwarding strategy; searching request routing

摘 要: 提高文件的查找定位效率是无结构的 P2P 网络一个重要的研究内容.泛洪法和随机查找法虽然简单和易于实现,但是前者会较大地增加网络负载,而且搜索的深度不能太大;后者虽然可以降低网络负载和适当增加搜索深度,但却以牺牲搜索的广度和增加响应时间为代价.提出一个无结构 P2P 内容分发网络的内容定位和查找请求路由方案.它利用社会网络的基本原理,通过模拟社会网络的特征,发挥节点的能动性,可以在有限的搜索深度和广度内快速查找定位文件.模拟实验结果表明,在相同的硬件环境支持下,P2P 网络文件平均定位时间可以缩短 50%以上.

关键词: P2P;内容定位;社会网络;转发策略;查找请求路由

中图法分类号: TP393 **文献标识码**: A

P2P 系统是为用户(节点)之间直接共享资源(包括文件、存储空间、CPU 等)[1]而设计的,一诞生就以其优越性征服了使用者.德国互联网调研机构 ipoque 称,P2P 已经彻底统治了当今的互联网,其中 50%~90%的总流量都

* Supported by the National Natural Science Foundation of China under Grant No.60872051 (国家自然科学基金); the National Key Technology R&D Program of China under Grant No.2006BAH02A11 (国家科技支撑计划项目); the Program of the Co-Construction with Beijing Municipal Commission of Education of China (北京市教育委员会共建项目专项资助)

Received 2009-01-20; Accepted 2009-04-27

来自 P2P 程序.

无结构的 P2P 网络的内容分发是当前研究的热点,无结构 P2P 内容分发系统的首要任务是在庞大的 P2P 网络中为节点查找文件.泛洪法和随机查找法是 P2P 网络两种典型的文件定位方式,它们分别通过向所有相邻节点发送查找请求(广度搜索)和随机选择相邻节点发送查找请求(深度搜索)来查找定位文件.

随着 P2P 网络规模的扩大,使得节点的相邻节点数目增多,网络节点之间的最大路径增大.在这种情况下,泛洪法的广度搜索和随机查找法的深度搜索的弊端,表现得就越来越严重.泛洪法中大量查找请求和查找结果数据包的传送,严重增加了网络的负载.随机查找法需要查找更大的深度,从而延长了查找时间,不能及时反馈查询结果.本文提出一个基于社会网络特征的 P2P 网络文件定位策略,根据 P2P 网络中节点的社会特征来确定文件查找方式.在这种文件查找方式中,查找请求的发送建立在有效的选择上,可以有效地控制查找的深度和广度.实验结果表明,本方法具有更高的查找效率.

1 相关工作

如何提高文件的查找效率是无结构的 P2P 内容分发网络需要解决的最主要的问题之一.为了满足无结构的 P2P 网络文件查找的需要,研究者提出了许多解决方案,最典型的查找方法是泛洪法和随机查找法.另外,为了克服泛洪法和随机查找法的缺点,也提出了一些介于泛洪法和随机查找法之间的查找定位方式.

Gnutella 和 Kazaa 采用的查找方式是泛洪法(flooding),向所有的相邻节点发出查找请求,算法简单且搜索范围广,但网络开销比较大,同时需要限制搜索的深度.文献[2−4]使用随机查找法(random walks),从相邻的节点中随机选择指定个数的节点发出查找请求.这种方式能够减少单个节点查找请求转发的个数,可以增加搜索的深度,但节点的选择是随机的.

为了避免泛洪法的网络负载问题和随机查找法搜索路径选择的随机性问题,LinkNet[5]提出了一种新的可扩展分布式数据结构来支持大规模 P2P 系统中的数据查找.在 LinkNet 中,所有的元素存储在一个有序的双向链表中,并使用虚拟链接来加速查找过程.这一方式可以有效地改善文件查找次数按时间均衡增加的 P2P 网络的查找效率.

PeerRank[6]提出基于概率权值的自适应查询消息转发策略,依据用户节点查询命中的历史信息,赋予节点相应权值作为查询消息路由选择的依据,引导查询快速接近目标资源.文献[7−9]提出使用语义相似性来描述节点,根据语义相似性选择搜索的方向.

Gnutella2[10]提出将节点分为超级节点和叶节点,它由少量超级节点和大量叶节点组成,超级节点之间连接复杂度高,超级节点连接数目较多的叶节点而且拥有和它相连的叶节点存储的文件的索引,每个叶节点只与一个超级节点相连.这样,文件查找的主要任务就都在超级节点中完成,可以有效地减少转发的查找请求个数,降低网络负载.同时,网络的深度也变小,弥补了泛洪法搜索深度不足的问题.

文献[11]提出了一种新的 P2P 网络数据路由模型(self-organizing P2P data routing,简称 SPDR),SPDR 利用具有数据转发能力的 P2P 节点构造虚拟数据路由,使 P2P 节点能够动态地突破网络对通信双方的限制,从而解决当前 P2P 网络中许多节点无法进行数据交换的问题,增强 P2P 网络的通信能力.文献[12−14]提出建立基于兴趣的子网,根据节点拥有的文件的特点,将节点组合成不同的子网,查找先在子网间进行.这两种方式都通过将节点按一定方式组合来实现,它们是以限制节点加入和离开的自由性以及增加网络维护复杂性为代价,降低了文件搜索的网络负载,减小了搜索深度.

2 P2P 网络模型设计

2.1 设计思路

在 P2P 网络的文件查找过程中,如果能够充分发挥节点的能动性,如同在社会中,个人的能动性都被发挥出来,社会就会繁荣一样,P2P 网络的文件查找效率就会有所提高.我们常常将互联网比喻成社会网络,社会网络

(social networking,简称 SN)是指个人之间的关系网络.它的理论基础源于六度分隔理论(six degrees of separation),是由美国著名社会心理学家米尔格伦(Stanley Milgram)于 20 世纪 60 年代最先提出来的.他指出,“你和任何一个陌生人之间所间隔的人不会超过 6 个.也就是说,最多通过 6 个人你就能够找到和认识任何一个陌生人”.社会网络理论在现代科技领域有广泛的应用,它的基本应用就是充分发挥社会网络中各个组成部分的作用[15,16].米尔格伦根据六度分隔理论进行了一系列有趣实验,他给一个朋友写了多封信,然后将这些信寄给他随机选取的人们.实验要人们以一种特殊的方式传递信件,即将信件传送给更接近接收人的人,最后,一些信件到达接收人.奇怪的是,到达的信件平均转手不超过 6 次,这就是著名的小世界实验.

根据六度分隔理论和小世界实验,我们可以推出这样的结论:如果 P2P 网络的节点能够模拟六度分隔理论的查找方式,也可以最多通过 6 个节点查找到任意文件所在的节点,这就直接解决了查找的深度问题.在小世界实验中,每个人传递信件时均发挥了自己的能动性,选择了自己认为更接近接收人的人来传递.这对 P2P 内容分发网络的文件查找有重要的指导作用.本文设计方案的指导思想是,让节点具有一定的智能,无论它与其他节点的连接有多少,选择最有可能到达目标节点的路径转发查询请求.为了保证查找速度更快,因而还要尽量选择向查找速度快并且可信任的方向查找.

2.2 设计方案

P2P 内容分发网络中文件的查找和现实世界中人的查找有很大的相似性.现实世界中,人有自己的特点,如年龄、职业、籍贯和工作所在地等等,这些都可以为查找提供有效信息.文件也一样,文件的类型和文件内容所具有的性质,对文件的查找也具有一定的指导作用.比如某文件是视频文件,内容为影片,那么,向存有影片视频文件的节点查找就比较容易找到.现实世界中要很快地找到指定的人,委托的中间人就应该是效率高的和可信任的人.P2P 网络中的搜索也应该这样.为了提高查找的效率,保留有助于查找的路径,删除查询效率低的路径,并添加能够提高查询效率的路径.小世界实验只考虑信件到达收信人的情况而不考虑所花费的时间,P2P 网络的文件搜索时间越短越好,因而在选择搜索方向时也要考虑选择搜索时间短的方向.

设置以下 4 个参数,其中前 3 个参数是作为节点选择转发查询请求路径的依据,第 4 个参数作为节点添加路径的依据:

(1) 文件类型命中效果:将文件分成若干类,统计不同类型的文件向这个邻接节点发出查找请求,查找成功的比率;
(2) 信任度:对从邻接节点发来的信息的信任程度;
(3) 动态响应效率:它反映出从发出请求到得到返回结果的平均时间和反应时间的变化幅度;
(4) 节点文件查找次数:从过往的数据包中获取数据,统计非邻接节点的文件被查询的次数.

2.2.1 文件命中效果

一个节点可以有多个邻接节点,查找文件时不必向每个相邻节点都发送查找请求,只需向最有可能查到文件的一个或几个节点发送查找请求即可.为了区分文件以提高查询的效果,首先根据查找的需要对文件进行分类,我们给出以下定义.

定义 1(文件类型向量). 文件类型向量 $F=\{f_1,f_2,f_3,\ldots,f_n\}^T$,用来表示根据查找需要定义的 n 类文件.

定义 2(文件命中效果矩阵). 文件命中效果矩阵 $M=\{S,U\}$,它是用来描述指定的节点统计它的某一相邻节点对自己文件查找(包括转发的其他节点的文件查找)请求的响应结果.其中,S 是一个向量$\{s_1,s_2,s_3,\ldots,s_n\}^T$,它的每一项是对应于文件类型向量 F 中各分量查找成功的次数.U 也是一个向量$\{u_1,u_2,u_3,\ldots,u_n\}^T$,它的每一项对应于文件类型向量 F 中各分量查找不成功的次数.

定义 3(文件命中效果向量). 文件命中效果向量 $R=\{r_1,r_2,r_3,\ldots,r_n\}^T$,用来描述对文件的查找效果,其中的每一个分量对应于文件类型向量 F 中相应的文件类型.

评价一个节点的某一转发路径的查找效率,在历史查找结果中,查找成功的次数越多,查找失败的次数越少,查询效果就越好.某一转发路径查找成功一次,相应的命中效果分量按设定的权值相应增加;查找失败一次,相应的命中效果分量按设定的权值相应减少.文件命中向量的计算公式为

$$R=M\times W+I,$$

其中,$W=\{\alpha,-\beta\}^T$ 是权重系数向量,$I=\{i_1,i_2,i_3,\ldots,i_n\}^T$ 是初始值向量,$\alpha>0,\beta>0,i_k>0,k\in\{1,2,3,\ldots,n\}$且为常数.

历史的文件命中效果的作用要随着时间的推移而衰减,为了描述衰减函数,首先给出次序时间段的定义:

定义 4(次序时间段). 次序时间段,就是把连续的时间按时间顺序分成可操作时间段,可以表示为$\{t_1,t_2,\ldots,t_i,\ldots,t_n\}$.从 t_1 到 t_n 的次序表示时间的先后顺序,其中,t_1 为系统启动后的第 1 时间段,t_i 为第 i 时间段,t_n 为当前时间段.

基于次序时间段的衰减函数的定义如下:

定义 5(文件命中效果次序时间衰减函数). 对于次序时间段$\{t_1,t_2,\ldots,t_i,\ldots,t_n\}$中的任意时间段 t_i,文件命中效果的衰减倍数为 $d(t_i)$,衰减函数表示如下:

$$d(t_i)=\frac{1}{(n+1-i)^2}.$$

文件的查找成功次数和查找不成功次数都要随着时间的推移而衰减,衰减只对矩阵 M 的各分量产生影响.加入衰减因素后,当前文件命中效果可以表示为

$$R=\sum_{i=1}^{n}d(t_i)(M_i\times W)+I=\sum_{i=1}^{n}\frac{M_i\times W}{(n+1-i)^2}+I,$$

其中,M_i 表示第 t_i 时间段的文件命中效果.文件命中时间距离当前时间越长,对当前命中效果的影响越小.

2.2.2 信任度

在无结构的 P2P 网络中,从其他节点收到的信息的可靠性是不确定的.如果从一个相邻节点收到的信息可靠性不高,就尽量不要向这个节点转发查找请求,甚至删除与这个节点的连接.相反地,如果从一个相邻节点收到的信息可靠性高,在其他条件相同的情况下,就选择向这个节点转发查询请求.

由于网络中节点是完全对等的,没有任何节点是完全可以信任的,因而对一个节点的信任度评价只能从与之交往的历史数据和其他节点对这个节点的信任度评价中抽取.因而,一个节点对指定节点的信任度评价要综合自己对这个指定节点的评价和其他节点对这个指定节点的评价.可以表示为

$$SR_k=\rho R_k+(1-\rho)G_k,$$

其中,SR_k 表示对节点 k 的查找信息的信任度,R_k 是指定节点对节点 k 的信任度,G_k 是选取的其他节点对节点 k 的信任度,ρ用来区分两者的比重.

R_k 是对指定节点对节点 k 提供的信息的历史统计而获得的.对节点 k 每次返回信息的真实性进行验证,如果真实,则增加信任度,否则,加倍减少信任度.验证后的信任值表示如下:

$$T=\begin{cases}1, & \text{当提供的数据真实时}\\ -2, & \text{当提供的数据不真实时}\end{cases}.$$

历史的信任度也应随时间的推移而衰减,次序时间段中,第 t_i 段时间对节点 k 的信任值 R_k^i 可以表示为

$$R_k^i=\frac{\sum_{j=1}^{m}T_j}{m},$$

其中,m 是 t_i 时间段与节点 k 交互的次数,T_j 是第 j 次交互的信任值.

信任度的衰减函数选用上面定义的 $d(t_i)$,结合衰减情况,当前指定节点对节点 k 的信任度可以表示为

$$R_k=\sum_{i=1}^{n}R_k^i d(t_i).$$

其他节点对节点 k 的信任度 G_k 用选取的节点的信任度的平均值来表示:

$$G_k=\frac{1}{l(setG)}\sum_{h\in setG}R_{hk},$$

其中,$setG$ 是选定的节点的集合,$l(setG)$是选定节点的集合中元素的个数,R_{hk} 是节点 h 对节点 k 的信任值.

2.2.3 动态响应效率

动态响应效率反映出节点 i 从向节点 j 发出请求到得到响应的平均时间和响应时间的变化幅度,前者是代表节点响应速度的指标,后者是代表节点响应稳定性的指标.平均响应时间越短,响应时间的变化幅度越小,节点的性能越好.这两项指标的计算是通过历史数据的统计得到的,如果当前向节点 j 共发出 n 次请求,则当前的平均响应时间可以表示如下:

$$T_n^j=\frac{(n-1)T_{n-1}^j+\left\lceil\frac{n}{3}\right\rceil t_n^j}{n-1+\left\lceil\frac{n}{3}\right\rceil},$$

其中:T_n^j 表示向节点 j 发出的前 n 次请求的平均响应时间,即当前平均响应时间;t_n^j 表示向节点 j 发出的第 n 次请求的响应时间.为了计算和表示的方便,这里的响应时间不是真实的响应时间,而是响应时间的换算值,计算方法如下:

$$t_n^j=\begin{cases}\dfrac{T}{t}, & \text{当}t\leqslant T\text{时}\\ 0, & \text{当}t>T\text{时}\end{cases},$$

其中,T 是系统设置的等待响应的最长时间,t 是实际的响应时间.

动态响应效率 E_n^j 综合考虑平均响应时间和响应时间变化幅度,平均响应时间越短,响应时间的变化幅度越小,节点查找的时间效果就越好,其值只需能够比较节点的查找时间效果,可以用下面的公式来描述:

$$E_n^j=\frac{T_n^j}{C_n^j},$$

其中,C_n^j 体现节点 j 响应时间的变化幅度,节点响应时间的变化幅度既要考虑历史的响应时间变化,又要重视与当前时间最近的时间段的变化幅度,同时还要使动态效率能够体现查找的时间效果.节点 j 响应时间的变化幅度 C_n^j 描述为

$$C_n^j=\frac{1}{n-1+\left\lceil\frac{n}{3}\right\rceil}\sqrt{(n-1)C_{n-1}^j+\left\lceil\frac{n}{3}\right\rceil(T_n^j-t_n^j)^2}.$$

2.2.4 选择发送查找请求或转发查找请求的路径

对于指定的节点 i,选择路径要综合文件命中效果、信任度和动态响应效率 3 个参数,令要查找文件的文件类型为 f,如果指向节点 k 的路径被选中,则满足以下条件:

$$R_f\times SR_k\times E_n^k\geqslant\varepsilon,$$

其中:R_f 是文件命中效果向量中文件类型是 f 的分量;ε 是设定的门槛值,是一个常数.

满足条件的路径被选择作为发送查找请求或转发查找请求的路径,这样不仅选择到更可能找到的文件的路径,而且减少了查找请求的发送路径,降低了网络负载.

同时,定义一个路径使用频度矩阵来描述节点 i 的路径的使用频度.

定义 6(路径使用频度矩阵). 由于节点资源有限,节点连接的路径数必须有一定的限制,设某节点 i 的最大路径数是 m,路径使用频度矩阵可表示为 $ROUTE=\{N,D\}$.其中:N 是向量$\{node1,node2,\ldots,nodem\}$,表示节点 i 连接的节点;D 是向量$\{d1,d2,\ldots,dm\}$,是节点对应的路径的使用次数.

由于节点的路径个数可能小于 m,那么矩阵中的分量没有完全使用.没有被使用的分量置为空,同时,在选择路径时,路径被选用 1 次,路径使用次数即增加 1.

2.2.5 节点文件查找次数

这项指标对自己发出的文件查找请求和转发的文件查找请求得到的回应进行处理,记录查找到文件的节点和文件以及它的查找次数.由于节点的存储空间有限,且这种记录的条数越多,查询效率越低,可以根据节点

的具体情况设置存储的最大条数.这些记录有两个用途:第一,节点自己或收到转发的查找请求,首先在这些记录中查找,如果有相应的文件和它所在的位置(即节点),则直接返回查找到的信息,这样可以缩短查找时间和减少转发的查找信息包的数目;第二,如果某个节点中的文件被查找的频率较高,可以将这个节点加入到自己的路径中,这样,同样可以缩短查找时间,减少转发的查找信息包的数目.

由于存储的记录条数有一定的限制,假设节点 i 设置记录存储的最大条数是 n,当已经存储了 n 条,且 n+1 条也已出现,那么是舍弃第 n+1 条,还是从前 n 条中删除 1 条,将第 n+1 条加入.这是要解决的问题,解决方案是根据时间变化调整记录,根据设置的规则决定记录满 n 条时,若第 n+1 条出现,是舍弃第 n+1 条,还是从前 n 条中删除 1 条然后加入第 n+1 条.下面首先定义记录结构.

定义 7(节点文件查找次数). 节点文件查找次数是一个三元组(fid,nid,fre),其中的 3 项分别描述文件的信息、文件存储的节点和文件被查询的次数,则每个节点都有一个节点查找次数表.如果节点 i 设置节点文件查找次数表记录的最大个数是 l_i,节点 i 存储的这种三元组的个数为 k,则 $0\leqslant k\leqslant l_i$.

节点文件查找次数也随着时间的推移而衰减,衰减函数定义为:

定义 8(节点文件查找次数和路径使用频度的衰减函数). 对于次序时间段$\{t_1,t_2,\ldots,t_i,\ldots,t_n\}$中的任意时间段 t_i,节点文件查找次数和路径使用频度的衰减倍数为 $f(t_i)$,衰减函数表示如下:

$$f(t_i)=\lambda^{n-i},0\leqslant\lambda\leqslant 1.$$

时间每向后转移一个时间段,三元组(fid,nid,fre)中的次数 fre 就都要进行衰减.设(fid,nid,fre_i)为节点 nid 的文件 fid 在 t_i 时间段内的查找次数,则当前时间总的查找次数可以表示为

$$fre=\sum_{i=1}^{n} fre_i f(t_i)=\sum_{i=1}^{n} fre_i\lambda^{n-i}.$$

同理,节点向量 $N\{node1,node2,\ldots,nodem\}$对应的使用次数向量 $D_i(d1_i,d2_i,\ldots,dm_i)$表示节点向量 N 在 t_i 时间段内的使用频度,则当前时间总的使用频度可表示为

$$D=\sum_{i=1}^{n} f(t_i)D_i=\sum_{i=1}^{n}\lambda^{n-i}D_i.$$

如果节点 i 某次查找请求或转发的查找请求收到查找成功的回应,文件和节点的描述信息分别是 fid 和 nid,则节点 i 的节点文件查找次数表作如下变化:

1. 将(fid,nid,fre)变为(fid,nid,fre+1),如果存在三元组的前两项分别为 fid 和 nid;
2. 将(fid,nid,1)加入表中,如果不存在前两项分别为 fid 和 nid 的三元组,且记录表中三元组的个数为 k,$0\leqslant k<l_i$;
3. 将记录表三元组中 fre 最小的三元组删除,并加入(fid,nid,1),如果记录表中三元组的个数等于 l_i,且不存在三元组的前两项分别为 fid 和 nid,并且存在 $fre\leqslant 1$ 的三元组;
4. 记录表不作任何改变,如果记录表中三元组的个数等于 l_i,且不存在三元组前两项分别为 fid 和 nid,并且不存在 $fre\leqslant 1$ 的三元组.

定期地以 nid 对三元组进行统计,将 nid 相同的三元组的 fre 相加,如果其和超过设定的门槛值,则增加指向节点 nid 的路径.增加路径的方法是:

- 将节点标识和 fre 分别写入 $ROUTE$,如果矩阵 $ROUTE$ 有空分量;
- 将节点标识和 fre 分别写入 $ROUTE$,如果矩阵 $ROUTE$ 有空分量.

节点查找次数表本身是 P2P 网络中每个节点保存文件在网络中存储位置的小数据库.之所以说它是小数据库,是因为每个节点都存储少量的文件位置信息.这也如同社会网络中的人,他(或她)只掌握社会知识中的小部分.节点本身查找文件或者接到查找请求时,首先从自己的小数据库节点查找次数表中查找,若能找到文件的目标地址,则查找结束;否则,选择路径发出或转发查找请求.

3 模拟实验与分析

根据 P2P 网络运行的需要,仿照 NS2 设计了一个简化版本网络模拟软件,使每一个网络模拟节点具有网络连接、计算和存储能力.并依据以上设计的 P2P 网络模型,用 Java 语言编写了一个模拟软件模拟设计的 P2P 网络模型.以命令行或脚本的形式控制节点的加入、离开、节点间的网络延迟、文件查找和各种信息的查询.模拟实验的硬件环境和系统软件见表 1.

Table 1 Hardware and software of system

表 1 硬件环境和系统软件

Server	HP Proliant DL580 G5 (438087-AA1)
Server type	Intel Xeon MP E7330
Count of processor	4
Size of memory	1GB×4
Operating system	Redhat linux 9.0

模拟实验的过程是用脚本命令来控制的,对于实际的 P2P 系统,任意一个节点进入和离开系统、节点进入后与其他节点的连接情况(网络延迟和网络故障)、某个节点某时是否发出查找请求、发出请求查找哪个文件都应该是随机的.脚本的生成是用程序控制的,控制程序在保证 P2P 系统规模和单位时间查找请求发出数量的同时,也保持上面提到的多种动作的随机性.模拟的 P2P 系统节点的间连接情况、节点加入离开情况和查找发出频率基本设置情况见表 2.

Table 2 Configuration of system simulation

表 2 系统模拟条件

Count of average paths inter-peer	Ratio of count peers enter and exit to peers in system	Count of search request per hour
3	10%	1 500

由于本模型的设计特点是增加节点自身的能动性,每个节点都需要一定的存储和计算能力(假定每个节点需求的存储和计算能力都能被满足),因而模拟实验中每次查找定位需要的时间与模拟软件运行的硬件环境关系密切.我们只在同等条件下检验本模型的对比效果,实验中相关参数的取值见表 3.

Table 3 Values of related parameters

表 3 相关参数取值

α	β	ρ	λ
6	1	0.7	0.8

P2P 系统的查找定位行为对存储空间和计算能力的需求分布在不同节点上,本文提出的查找策略是通过每个节点对系统资源位置的掌握和处理降低查找的深度和广度,以提高查找的效率.与传统的查找策略相比,对单个节点存储能力和计算能力要求更高,单个节点的相对处理时间较长.在模拟实验中,大量模拟节点的行为在模拟环境中发生,对模拟环境所处的服务器的内存和计算能力要求较高.为了排除硬件对模拟实验结果的影响,要根据模拟实验的硬件环境合理选择节点的规模.图 1~图 3 是系统在运行的过程中,节点的个数大约稳定在 50 个、100 个和 200 个时的效果.单次查找定位的时间随机性较大,图中每个点的值是连续发出 400 个查找请求的查找定位时间的平均值.图中的横线是没有使用本文设计架构的一般的无结构 P2P 网络查找定位时间模拟结果.

从上面的模拟结果可以看出,在 P2P 网络中,若能发挥节点的能动性,则文件的查找定位效率会随着它们文件查找次数的增加而大幅度地提高并逐渐趋于稳定.同时,节点的个数越多,系统稳定时需要的查询时间相对也长一些.模拟结果表明,本文设计的模型虽然在系统开始运行的较短时间内查找平均时间稍长,但系统运行一段时间及稳定后,查询定位效率大幅度提高.在系统稳定后,查询时间可缩短 50%以上.

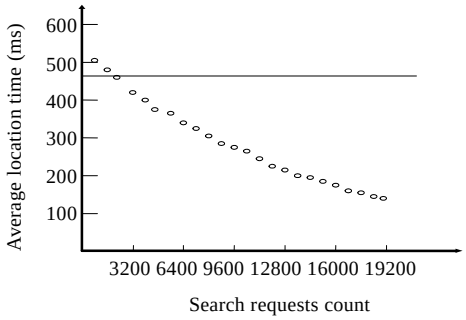

Fig.1 Simulation for 50 peers

图 1 50 节点模拟情况

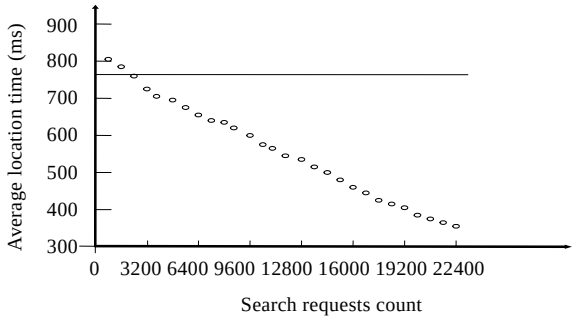

Fig.2 Simulation for 100 peers

图 2 100 节点模拟情况

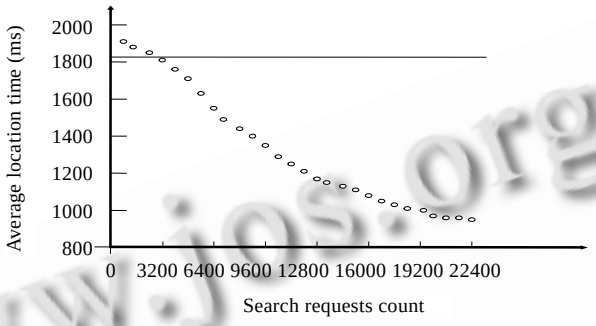

Fig.3 Simulation for 200 peers

图 3 200 节点模拟情况

4 结束语

本文根据社会网络的特征,设计了一个新的 P2P 网络内容查找和定位方式.在这一定位方式中,要发挥节点在内容查找和定位过程中的能动性.P2P 网络中的节点在发出和接收查找定位请求和查找定位回复中获取有用的信息,用来更好地为查找定位请求服务.模拟结果也表明,这一新方案可以提高 P2P 网络中内容查找定位的效率.

References:

[1] Androutsellis T, Stephanos S, Diomidis S. A survey of peer-to-peer content distribution technologies. ACM Computing Surveys, 2004,36(4):335−371. [doi: 10.1145/1041680.1041681]

[2] Kalogeraki V, Gunopulos D, Zeinalipour-Yazti D. A local search mechanism for peer-to-peer networks. In: Charles N, David G, Konstantinos K, Sajda Q, Han van D, Len S, eds. 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: Gupta M, ed. Proc. of the 16th ACM Int'l Conf. on Supercomputing (ICS 2002). New York: ACM Press, 2002. 254−261.

[4] Xu Y, Ma XJ, Wang C. Selective walk searching algorithm for Gnutella network. In: Diane W, ed. Proc. of the 4th Annual IEEE Consumer Communications and Networking Conf. New York: IEEE Communications Society, 2007. 746−750.

[5] Zhang KL, Wang S. LinkNet: A new approach for searching in a large peer-to-peer system. Chinese Journal of Computers, 2006, 29(4):611−617 (in Chinese with English abstract).

[6] Feng GF, Mao YC, Lu SL, Chen DX. PeerRank: A strategy for resource discovery in unstructured P2P systems. Journal of Software, 2006,17(5):1098−1106 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/1098.htm [doi: 10.1360/jos171098]

[7] Halevy AY, Ives ZG, Suciu D, Tatarinov I. Schema mediation in peer data management systems. In: Dayal U, Ramamritham K, Vijayaraman TM, eds. Proc. of the Int'l Conf. on Data Engineering. Bangalore: IEEE Computer Society, 2003. 505−516.

[8] Kementsietsidis A, Arenas M, Miller RJ. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In: Halevy AY, Ives ZG, Doan AH, eds. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press, 2003. 325−336.

[9] Qiu ZH, Xiao MZ, Dai YF. A user behavior based semantic search approach under P2P environment. Journal of Software, 2007, 18(9):2216−2225 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/18/2216.htm [doi: 10.1360/jos182216]

[10] Gnutella2. http://en.wikipedia.org/wiki/Gnutella2

[11] Liu Y, Shen B, Zhang HK. A novel data routing scheme based on self-organizing for P2P Networks. High Technology Letters, 2004,14(12):16−20 (in Chinese with English abstract).

[12] Xue GT, He XJ, Jia ZQ, You Y, Li ML. Research of using interest-subnet grouping algorithm and improvement on content location of gnutella. Journal of Shanghai Jiaotong University, 2004,38(12):2108−2111 (in Chinese with English abstract).

[13] Kunwadee S, Brace M, Zhang H. Efficient content location using interest-based locality in peer-to-peer systems. In: Ibrahim M, ed. Proc. of the 22nd Annual Joint Conf. of the IEEE Computer and Communications Societies. Bangalore: IEEE Computer Society, 2003. 2166−2176.

[14] Xi T, Zhang DL, Zhe Y. Efficient content location based on interest-cluster in peer-to-peer system. In: Min W, ed. Proc. of the IEEE Int'l Conf. on e-Business Engineering. Bangalore: IEEE Computer Society, 2005. 324−331.

[15] Paul C, James D, Victor G. Social network model of construction. Journal of Construction Engineering and Management, 2008, 134(10):804−812. [doi: 10.1061/(ASCE)0733-9364(2008)134:10(804)]

[16] Coyle CL, Heather V. Social networking: Communication revolution or evolution. Bell Labs Technical Journal, 2008,13(2):13−18. [doi: 10.1002/bltj.20298]

附中文参考文献:

[5] 张坤龙,王珊.LinkNet:一种用于大规模 P2P 系统查找的新方法.计算机学报,2006,29(4):611−617.

[6] 冯国富,毛莺池,陆桑璐,陈道蓄.PeerRank:一种无结构 P2P 资源发现策略.软件学报,2006,17(5):1098−1106. http://www.jos.org.cn/1000-9825/17/1098.htm [doi: 10.1360/jos171098]

[9] 邱志欢,肖明忠,代亚非.一种 P2P 环境下基于用户行为的语义检索方案.软件学报,2007,18(9):2216−2225. http://www.jos.org.cn/1000-9825/18/2216.htm [doi: 10.1360/jos182216]

[11] 刘云,沈波,张宏科.一种新的 P2P 网络自组织数据路由模型.高技术通讯,2004,14(12):16−20.

[12] 薛广涛,贺小箭,贾兆庆,尤晋元,李明禄.使用兴趣子网划分算法对 Gnutella 中资源定位机制的改进.上海交通大学学报,2004,38(1):2108−2111.

黄永生(1979－),男,山东曹县人,博士生,助教,主要研究领域为 P2P 内容分发网络,个性化推荐服务.

张玉洁(1969－),女,讲师,主要研究领域为网络服务,数字媒体.

孟祥武(1966－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络服务,语义 Web.