

# Research of P2P file sharing model and search strategy

ZHAO Guang-li, WU Bei-chun

Information Science and Technology College, Dalian Maritime University, Dalian, China

e-mail: dl\_zhaog1@126.com

**Abstract:** The peers are divided into general peer, aggregation peer and super peer according to its different roles in P2P network, thus a layered virtual network structure is formed logically. On this basis, an improving file sharing model-FSSM is proposed and the mobile agent technology and an idea of the ant colony algorithm in resource search strategy are introduced. Using pathfinding agent and feedback agent method to seek the best resources search path and the optimizing of RPV table at the same time which maintains the aggregation peer can quickly search and locate network resources. The experiment and simulation results indicate that FSSM has distinct improvement and enhancement in the accuracy and efficiency of resource search comparing with Gnutella.

**KeyWords:** Computer software; Application; File sharing; Search strategy; Mobile agent

## P2P 文件共享模型与搜索策略研究

赵广利, 吴蓓春

大连海事大学信息科学技术学院, 大连, 中国, 116026

e-mail: dl\_zhaog1@126.com

**【摘要】**根据 P2P 网络中各对等点角色的不同, 将其划分为普通对等点、集合点和超级对等点, 从而在逻辑上形成了分层的虚拟网络结构, 在此基础上提出了一种改进的文件共享模型-FSSM。在搜索策略上引入移动 Agent 技术和蚁群算法思想。采用探路 Agent 和反馈 Agent 方法寻求最佳资源搜索路径, 同时对集合点所维护的 RPV 表进行了优化, 可以快速的查询与定位。实验仿真结果表明, FSSM 在资源搜索的准确性和搜索效率等方面与 Gnutella 相比具有明显的改善和提高。

**【关键词】**计算机软件; 文件共享; 搜索策略; 移动 Agent

### 1 引言

P2P 技术具有无限的发掘潜力和广阔的应用前景, 因此成为诸多学者研究的热点。文件共享是 P2P 技术兴起的最主要的原因, 也是 P2P 技术应用最广泛的领域之一。目前, 针对 P2P 文件共享系统的研究主要集中在搜索与定位、数据传输、激励机制与安全相关的三个主要问题上。如文献[1]给出了一种激励机制来保证资源下载的成功率; 文献[2]给出了一种 P2P 的文件共享系统信任机制来提高文件共享系统的安全性与服务质量。由于 P2P 本身具有的分布式存储的特点, 使得在资源传送问题上存在的不足相对容易解决, 而难点在于如何快速的对拥有相应资源的对等点进行准确的查询与定位。因此, 如何实现资源的定位是文件共享的一个关键问题<sup>[3]</sup>。如何有效地查找信息, 最新的成果都是基于 DHT 的分布式发现和路由算法<sup>[4]</sup>。但是, 在基于 JXTA 技术构建的网络中, 由于采用 DHT 方法查询, 其查询方式极为复杂且难以实现, 而且查询效

率较低。因此, 为改善现有文件共享系统资源搜索效率低、占用大量带宽以及不支持复杂查询等现状, 本文提出了一种改进的文件共享系统模型-FSSM, 在搜索机制上引入移动 Agent 技术和蚁群算法思想, 避开了 DHT 方法查询的各种弊端, 并对集合点所维护的 RPV 列表进行了优化。最后, 从请求的满意度与反应的满意度两个方面对文中提出的资源搜索算法进行了严格的测试, 证明了该搜索机制的性能、搜索的准确性和搜索效率等方面与 Gnutella 等相比具有明显的改善和提高。

### 2 改进的 P2P 文件共享模型

#### 2.1 网络结构设计

在P2P对等网络中, 将对等点按其在网络中角色的不同分成了三类: 普通对等点、集合点和超级对等点。

1) 普通对等点: 该类对等点提供或下载共享文件。使用邻列表存储与其相邻对等点的信息。当该对等点信任级别比较高、带宽及机器性能等达到一定程度时, 普通对等点可以成为新的集合点。在网络发生变化时(对等点动态的加入与退出), 普通对等点要通知其所连

基金项目: 国家自然科学基金(60672031)资助

Foundation Item: Supported by National Natural Science Foundation of China(60672031)

接的集合点，以便集合点更新关于该普通对等点的信息。

2) 集合点：首先是一个普通对等点，并动态的维护一个集合列表(RendezvousList)。集合列表中包含其成员的标志和其它信息，从而使得集合点内的对等点可通过集合点搜索到更多的对等点信息。同时集合点会定期向超级对等点发送广告，告知其所负责的那段网络中的变化。这里的集合点在搜索方面相当于 Kazaad 中的搜索对等点。

3) 超级对等点(SuperPeer)：每个对等组内都至少有一个超级对等点。为了避免单点失效和多度冗余带来的复杂性，本文采用二度冗余。一个为主SuperPeer，负责与对等组内的所有集合点相连接并保存所有与其相连对等组内的超级对等点的位置和资源信息；另一个为副SuperPeer，虽然副SuperPeer与对等组内所有的集合点不连接，但它保存了所有主SuperPeer所拥有的资源索引，并由主SuperPeer负责及时更新该资源索引。在正常情况下，一切查询消息都发送给主SuperPeer处理。当主SuperPeer失效时，副SuperPeer根据已拟订的一些规则，主动转换成主SuperPeer继续提供服务。与此同时，再选出一个新的集合点作为副SuperPeer，这样就可避免在单一的SuperPeer结构中，SuperPeer失效后重组的情况，可大大增加P2P网络的可靠性和有效性。其网络的层次结构如图1所示。

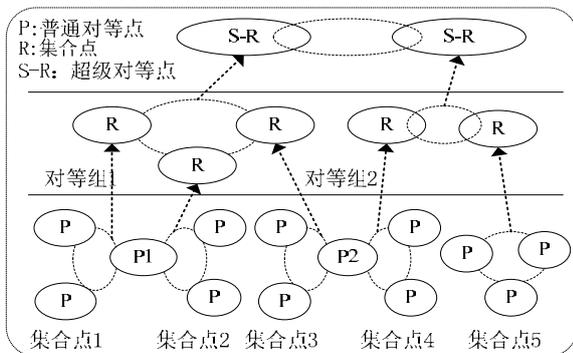


图1 网络层次模型

## 2.2 文件共享模型设计

本模型利用JXTA技术构建了一个跨平台、跨编程语言的P2P网络应用程序的通用平台。该平台为各个对等点间的通信提供了基础，整个文件共享系统可以看作是一个由很多对等点以及连接这些对等点的通路所构成的网络。每一个对等点都可以看作是蚁群算法中的一个蚁巢，Area(在主机上的运行环境)为各类Agent

提供运行环境。每个对等点除了对外提供文档信息服务以外，还可以接收本地用户的文档查询请求，并负责将请求转发到网络上，以及将查询结果返回给本地用户。从而可利用JXTA平台提供的管道通信机制与文件资源所在的对等点进行通信，以实现资源的下载。每个对等点都是由资源管理、移动Agent的调度和拓扑管理三个逻辑模块组成，改进后的文件共享模型总体框架如图2所示。

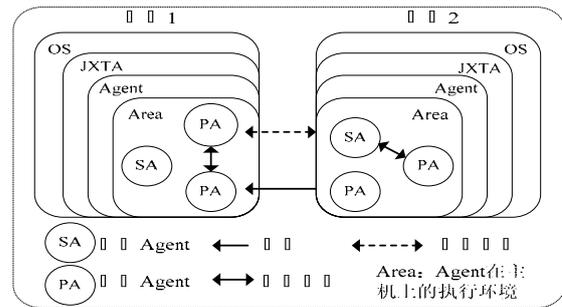


图2 文件共享模型的总体架构

1) 资源管理模块：资源管理是P2P的核心技术之一<sup>[5]</sup>，该模块主要负责本地文档资源的维护，如文档的上传、存储和下载等。重点维护一张XML格式的资源列表，以供其余对等点的查询。该资源列表可以利用JXTA平台提供的通告机制来进行发布。

2) 移动 Agent 管理模块：负责创建和销毁 Agent。在本架构中，主要有两类 Agent，即静态 Agent 和探路 Agent。其中，静态 Agent 主要任务是按用户的要求和搜索内容产生一个或多个带有标识和需要访问节点信息的探路 Agent，为运行到本地的探路 Agent 提供运行环境并与其进行交互。网络中每个对等点上都有一个静态 Agent。而探路 Agent 携带查询信息(移动 Agent 的编号、将要访问节点的 IP 地址/端口对、搜索结果集、组内搜索层次、组间搜索层次、克隆的移动 Agent 数目等)在网络中搜索所需资源，并根据移动到节点的状态来随时更新自身的内容，探路 Agent 是整个搜索机制的关键。事实上，在本架构中，还有一类 Agent，那就是反馈 Agent，主要在探路 Agent 找到符合条件的对等点时，反馈 Agent 会将信息返回到查询起始点。同时，在返回查询起始点的过程中，还会更新原路上各个对等点上的相关信息(RPV 表等)。为了更清楚的表示静态 Agent 和探路 Agent 之间的通信关系，在图 2 中将反馈 Agent 省略掉了。

3) 拓扑管理模块：负责探测邻近节点，维护一张邻列表，且及时更新邻近节点信息，从而决定探路 Agent 移动的下一个节点。由于采用分层的网络结构，图 2

中 JXTA 技术搭建的通用平台更为有效的适应移动 Agent 的生存和运行。因此,该逻辑模块对于文件资源的搜索起着决定性的作用。

### 3 文件共享模型的发现与搜索机制

#### 3.1 发现对等点

如果一个对等点P想通过集合点加入到网络中,必须和一个已在线的某集合点R取得联系,得到该对等组的授权证书后即可加入到网络中。具体的流程如下:

1) 据超级对等点在网络上公布信息,获取想要加入的对等组信息;

2) 生成自己的Peer ID,以IP多播的方式向对等组发送请求;

3) 对等组的集合点R收到请求后对其身份进行验证,通过后对等点P即可加入该对等组中,否则不能加入。

经过以上步骤,对等点P成为集合点R的成员,集合点R成为对等点P的首选集合点。使用JXTA提供的管道机制建立两者之间的管道后,即可进行相互通信。同时,R会找出在逻辑上与P最近的对等点,并将该信息返回给P,从而初始化P的邻列表。加入工作完成以后,此时P还要进行一步判断R是否位于自己的邻列表中,如果不是,就将与R之间的管道删除。

#### 3.2 查询与定位机制

本系统模型是在AntNet算法的基础上,引入移动 Agent技术,以提高文件资源的效率。按照蚁群算法的思想,系统中每个对等点相当于一个蚁巢,而查询请求的转发就要由蚂蚁来实现。有两类网络蚂蚁:前行蚂蚁和后行蚂蚁。在本文提出的模型中,前行蚂蚁就是探路Agent,而后行蚂蚁就是反馈Agent。

##### 3.2.1 邻近查询

根据统计发现,P2P网络中存在着小世界现象,也就是说人们使用和需要查找的数据范围并不会很大,只在自己附近很小的范围内。这里采用有限制的泛洪机制,严格控制TTL(Time to Live,存活时间)。如果TTL定的太小,则不能达到资源点,太大则浪费资源<sup>[6]</sup>。在有限跳数以内,如果还得不到响应的话,就在本地RPV表中找到一个可达的集合点通告,进行下一步的查询。

##### 3.2.2 组内查询

组内查询的重点就是携带有查询信息的移动 Agent如何在组内集合点之间的移动。下面主要给出探路Agent和反馈Agent的具体迁移过程:

1) 当集合点收到搜索请求后,在本地缓存的文件列表中没有找到匹配的文件时,该集合点上的静态Agent就会产生一个携带查询相关信息的探路Agent;

2) 集合点根据动态更新的RPV表,判断组内其它集合点的个数(假设为M)以及网络的性能特点,从而确定组内搜索层次信息L(假设为n层),确定产生的探路Agent数目X,产生数目的公式为: $X + X^2 + \dots + X^L = M$ (其中L, M均为大于0的整数)求出最近似的大于X的最小整数值,集合点克隆出X个探路Agent,每个探路Agent携带部分要继续搜索的节点信息(一个探路Agent携带三个节点的信息)和每个节点与前一跳节点之间的延迟时间信息,在RPV中选择优先级最高的X个节点发送,实现以一定的层次结构并行的进行搜索。

3) 每个接收到探路Agent的集合点,首先搜索自身的文件列表是否满足条件。若有满足搜索要求的情况:就增加探路Agent中的搜索结果集,并向其它正在进行搜索的探路Agent发送更新信息。如果检测结果集已经满足请求节点的要求,就停止向其它集合点迁移,此时生成一个反馈Agent,并将探路Agent中的所有信息复制给它,探路Agent消亡。由反馈Agent携带查询到的信息移到至查询的起点,沿路更新相应对等点的信息;否则集合点上的常驻Agent将根据自己的节点信息更新探路Agent中需要更新的信息,并查看要搜索的节点数目是否大于记录克隆的探路Agent的数目。若大于,就继续克隆出相同数目的探路Agent,并参照本地实时更新的RPV表依次按优先级发送出去;否则,克隆出与要搜索节点的数目相同的探路Agent发送出去;

4) 若没有满足搜索要求的情况,则重复2)、3),直到遍历组内所有的集合点。若仍没有搜索到满足条件的文件,就进入组间查询(具体实现代码略)。

##### 3.2.3 组间查询

在组内查询未能找到满足要求的文件时,起始的查询节点会将探路Agent转发给超级对等点,而后探路Agent将在各个对等组的超级对等点之间进行迁移,然后各个超级对等点再将探路Agent转发给其所在对等组内的各个集合点,进行其所在对等组内的组内查询。

### 4 RPV 表的优化

为了对集合点所维护的RPV表进行优化,选用搜索成功率和节点间的网络延迟作为标准对RPV表中所维护的集合点进行优先级划分。在一个集合点的RPV表中所维护的各个集合点的节点信息中,都包含一个优先级值H。当每次查询结束时,集合点所收到的移动 Agent中会携带有搜索结果和集合点间的网络延迟时

间值。根据这些信息，即可完成优先级的更新，具体方法如下：

假设集合点N在某时刻回收了X个自己转发的同一标识的Agent(分别访问过节点 $N_1$ 、 $N_2$ ...、 $N_X$ )，那么在节点N的RPV表中， $N_1$ 、 $N_2$ 、... $N_X$ 这X个节点的优先级值( $H_{NK}$ ， $K \in X$ )将可能会被更新。如果在节点 $N_K$ 上搜索到匹配的资源，那么 $H_{NK}$ 加1，否则不变。在X个节点中，延迟时间最短的m个节点优先级值加1；延迟时间最长的m个节点优先级减1；剩余节点优先级不变。其中m是由超级对等点根据其所在对等组的网络性能特点动态设置的。这样每个RPV表中各个集合点的优先级的值是动态更新的，集合点在转发移动Agent时，总是会选择最有可能搜索到匹配资源并且延迟最短的节点发送，从而可进一步提高搜索效率。

## 5 文件共享模型测试

### 5.1 测试标准与测试环境

#### 1) 测试过程

从请求结果的满意度出发进行测试，只需在移动Agent中增加一个stepCount变量，用它来记录搜索的步数，与Gnutella查询机制进行比较，测试结果如图3所示。

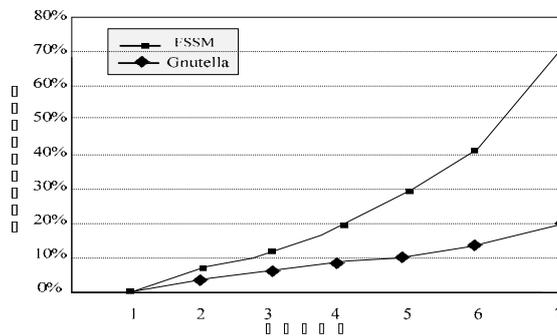


图3 搜索过节点数的百分比与搜索的步数关系图

从时间满意度出发，对各种文件反复进行搜索，搜索成功后记录搜索时间，最后取搜索时间的平均值作为评测的依据，与Gnutella查询机制(泛洪算法)进行比较，测试结果如图4所示。

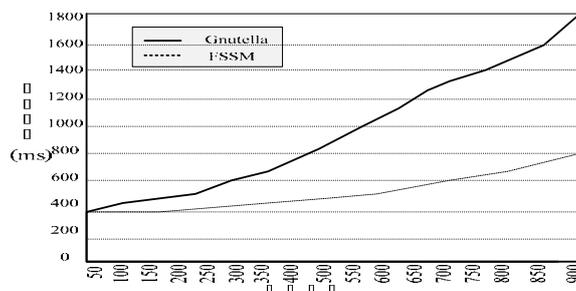


图4 节点数与搜索时间的关系

#### 1) 测试标准

衡量搜索机制好坏的一个重要标准是服务质量。服务质量的好坏是从请求的满意度与反应的满意度两个方面来衡量。请求的满意度是指将前n(用户自定义)个结果返回给用户。如果返回的结果数大于等于n，可以说用户的请求被满足了。这里采用搜索过节点数百分比与算法所用步数的关系来间接反映请求的满意度；从反应的满意度来看，用户所关心的是必须等待多长时间所要求的结果才能到达。

#### 2) 测试环境

台式机30台，主机配置：CPU:P4 2.4G 内存：512M。30台机器分为两个对等组，每个对等组内一个超级对等点，四个集合点，十个边缘节点，在每个节点上(服务点除外)都共享100个文件作为共享资源。

### 5.2 测试过程与结果

## 2) 测试结果分析

从请求的满意度来看,由图 3 测试结果可知,通过 7 步搜索,FSSM 搜索资源可达 70%左右,而 Gnutella 搜索资源只达到 20%左右;从反应的满意度来看,由图 4 测试结果可知,搜索到同样的资源数,FSSM 搜索所花费的时间明显比 Gnutella 低。本系统采用的搜索机制尽管是无保障的查询机制(在尽可能短的时间内查找到最多的资源信息),但是由于引入了移动 Agent 技术与蚁群算法,并对各种对等点所维护的 RPV 表进行了优化,在搜索的准确性和搜索效率方面具有明显的提高。

## 6 结论

本文在改进的P2P文件共享系统模型基础上,引入移动Agent技术与蚁群算法的思想,提出了一种新的资源搜索策略:采用探路Agent和反馈Agent技术寻找最佳资源搜索路径;采用主/副SuperPeer二度冗余与动态生成与切换技术,避免了单点失效和多度冗余带来的复杂性。文中详细阐述了如何使用移动Agent携带查询信息在对等组内和对等组间进行资源搜索,同时也给出了在移动Agent的迁移过程中如何与所到节点进行实时交互,以实现迁移路径的动态选择。文中提出的资源搜索与定位机制符合小世界(Small world)理论。经试验仿真测试结果表明,本文给出的文件共享模型与搜索算法可对网络资源快速的查询与定位。改善了现

有文件共享系统中在资源搜索上存在的大量冗余消息,占用大量的网络带宽以及不支持复杂查询的现状。

## References (参考文献)

- [1] Zhuang lei, Chang Yucun, Dong Xiguang. Incentive mechanism in peer-to-peer file sharing system[J]. Application Research of Computers. 2009, 26(1), P265-268(Ch)庄雷, 常玉存, 董西广. 一种P2P文件共享系统中的激励机制[J]. 计算机应用研究. 2009, 26(1), P265-268
- [2] Xu Shukui, Wang Hui, Sun Xiao, Jiang Zhihong. Study On Trust Mechanism Of P2P-Based File-Sharing System[J]. Computer Applications and Software. 2008, 25(6), P276-282(Ch)徐树奎, 王晖, 孙晓, 姜志宏. 基于P2P的文件共享系统信任机制研究[J]. 计算机应用与软件. 2008, 25(6), P276-282
- [3] Cai Yong, Dai Jianyong. An Improved Searching Algorithm for Peer-to-Peer Network[J]. Microcomputer Information. 2009, 25(3), P240-242(Ch)蔡勇, 戴建勇. 一种P2P资源搜索的改进方法[J]. 微计算机信息. 2009: 25(3), P240-242
- [4] Luo Wenjie. Peer to Peer summarize (the second edition)[EB/OL]. <http://www.acejoy.com/space/html/72/n-72.html>罗杰文. Peer to Peer 综述(第二版本)[EB/OL]. <http://www.acejoy.com/space/html/72/n-72.html>.
- [5] W u Xiangning, Wang Yuan. Peer-to-Peer Document Sharing System Based On Ant Colony Algorithm[J]. Computer Engineering and Applications. 2007, 43(20), P145-148(Ch)吴湘宁, 汪渊. 基于蚁群算法的P2P文件共享系统[J]. 计算机工程与应用. 2007, 43(20), P145-148
- [6] Zhu Shoude, Pengjian. Improvement on the Serch Strategy of Network resources During P2P File Transferring[J]. Microcomputer application. 2008, 24(9), P48-50(Ch)朱守德, 彭健. P2P文件传输中网络资源搜索策略改进[J]. 微型电脑应用. 2008, 24(9), P48-50