

P2P网络中基于语义和信任的资源发现算法

孙丽丽, 欧阳松

(中南大学信息科学与工程学院, 长沙 410083)

摘要: P2P 对于分布式文件共享具有很好的前景, 但当前的 P2P 系统仍然缺乏有效的信息管理机制。该文在构建超级节点叠加网络时考虑信任和语义的因素, 语义相似的节点尽量分布在同一个域中。在选取超级节点时考虑信任值、节点能力及动态性等因素, 提出一种高效的基于语义和信任机制的 P2P 资源发现算法。

关键词: P2P 网络; 语义; 信任; 超级节点

Resource Discovery Algorithm Based on Semantics and Trust in P2P Network

SUN Li-li, OUYANG Song

(School of Computer Science and Engineering, Central South University, Changsha 410083)

【Abstract】 P2P systems are emergent and promising technique for resource sharing in distributed environment, but they lack information management mechanism. This paper presents a super-peer overlay network, which combines semantics and trusts, and the nodes which have the similar semantics are distributed in the same domain. When selecting the super-nodes, it combines the trusts, capacity and dynamics. An efficient P2P resource discovery algorithm based on semantics and trust is put forward.

【Key words】 P2P network; semantics; trust; super-node

1 概述

随着网络技术的发展, P2P^[1]引起了人们的广泛关注, 它打破了传统的C/S模式, 在网络中, 每个节点的地位都是对等的, 既充当服务器, 为其他节点提供服务, 同时享用其他节点提供的服务, 即具有Client和Server的双重身份。P2P网络具有资源利用率高、扩展性好、动态性强、高分散化、高健壮性等特点, 因此, P2P技术具有广阔的应用前景。

P2P 网络中的一个关键是搜索技术。集中式 P2P(如Napster)具有效率高、开销小的优点, 但是随着网络规模的增大, 目录服务器必然成为服务瓶颈, 会造成单点失败, 同时还存在扩展性问题; 全分布式 P2P(如 Gnutella)不存在单点失效问题, 但是会产生很多的查询流量, 使得网络有太多的额外开销, 并且定位效率低或不适合动态性强的环境; 而混合分布式 P2P(如 Kazaa)选择少数节点作为超级节点, 由超级节点作为局部的集中式服务器来保存部分 P2P 节点的共享资源信息, 可在全分布式和集中式之间作出较好的权衡。

目前, 在 P2P 网络中, 节点的自组织性和匿名性对 P2P 应用获取成功有显著的贡献, 但也造成了一些恶意节点在网络中提供不可靠或者欺诈的服务, 不能保证所有节点都提供诚实的服务, 从而导致了搜索到的资源的不确定性和非法性。因此, 在选取超级节点时, 若能考虑到节点的信任值, 尽量避免恶意节点的欺诈行为, 选择高效、能力强的节点作为超级节点, 就能有效地提高资源搜索的可靠性和安全性, 改善节点的服务质量, 从而进行更可靠的协同计算。

本文采用语义和信任机制构建超级节点叠加网络拓扑结构, 并且在此基础上提出一种基于语义和信任的搜索策略。

2 基于语义和信任的超级节点语义网的构建

在基于语义和信任的超级节点网络中, 网络中的节点根

据语义分布在不同的域组。本文以图 1 为基础, 提出了基于语义和信任的对等网搜索模型, 图中虚线圆表示一个虚拟语义域。

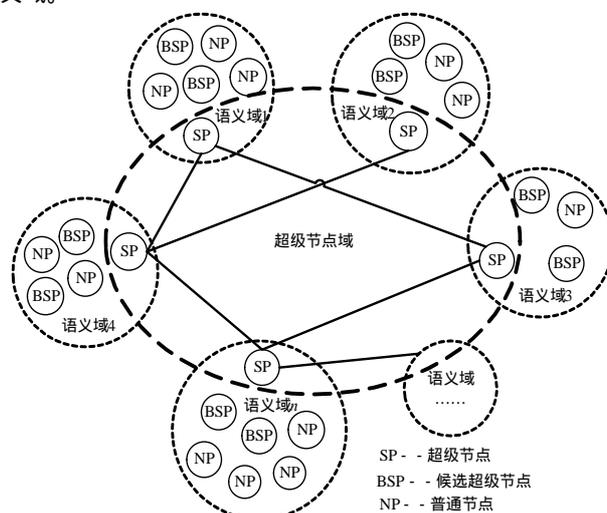


图1 网络结构

(1)域组成员: 每个域组包括 3 种类型的节点: 1)超级节点(Super-Peer, SP), 每个域组有一个超级节点; 2)候选超级节点(Backup SP, BSP), 每个域组设置 2 个候选超级节点; 3)普通节点(Normal Peer, NP), 根据节点能力, 每个域组拥有的普通节点数量可以不同。

(2)域组构建规则: 该模型构建域组的规则和传统的超级

作者简介: 孙丽丽(1981 -), 女, 硕士研究生, 主研方向: 网络与通信; 欧阳松, 教授

收稿日期: 2007-11-30 **E-mail:** friendly@163.com

节点网络不同，每个超级节点所管辖的域是由多个语义相似的节点聚合在一起所组成的一个虚拟语义域，在这个语义域中，选择一个信任值高和能力强并且变动少的节点作为超级节点。所有超级节点还要根据语义和信任值选择语义相似和信任值较高的超级节点相连，构成超级节点域，它是一个语义逻辑上的结构。

采用上述结构，对于每个给定的查询，资源的查找及资源信息的扩散基本上能被限制在相应的超级节点语义域中，当在该语义域中查询不到时，可以根据信任值和语义因素将查询发送到最有可能提供结果的相邻语义域中，从而改善P2P资源发现与组织的性能，降低资源搜索时的网络消息负载，减少资源定位延迟。

2.1 节点的加入

当节点初次加入网络或因系统崩溃而再次加入网络时，可以动态地构建覆盖网络。每个新加入的节点必须至少知道网络中的一个节点，本文将每个节点看作一个语义向量，新加入节点 P_i 的语义向量用 V_i 表示，收到加入请求的节点 P_j 的语义向量用 V_j 表示，定义 f 为语义相似函数。节点的加入过程如下：

(1)新加入的节点 P_i 向网络中已知的一个节点 P_j 发送加入请求(add-message)，add-message中包含 P_i 的语义向量 V_i 。

(2) P_j 将它本身的语义向量 V_j 和新加入节点 P_i 的语义向量 V_i 进行语义相似性比较。如果 $f(V_i, V_j) > t$ (其中， t 为语义相似极限)，那么，节点 P_j 向节点 P_i 发送接受报文(accept-message)，否则，发送拒绝报文(reject-message)。

(3)如果新加入节点 P_i 被 P_j 拒绝，那么 P_i 必须寻找另外一个节点，并且进行同样的语义比较，如果不满足条件则继续寻找，为了避免节点无限制地寻找语义相似的节点，设置一个值 T ，当和 P_i 比较的节点数等于 T ，还没有节点满足条件时，则认为没有任何节点满足条件。

(4)如果没有任何节点满足条件，或者所要加入的语义域的超级节点能力有限，不能容纳更多的节点，那么节点 P_i 单独构建一个新的语义域，新节点作为新构建语义域的超级节点。新的超级节点可以根据语义和信任值等因素与其他适合条件的超级节点建立联系。

2.2 超级节点的选取策略

超级节点选取的优劣直接决定了网络的稳定性和资源搜索的性能，因此，超级节点的选取要考虑信任值、节点能力以及动态性等因素，超级节点可以根据以下公式选取：

$$Peer_{Value} = \lambda_1 Peer_{Trust} + \lambda_2 Peer_{Capacity} + \lambda_3 \frac{1}{Peer_{Dynamic}}$$

其中， $\lambda_1 + \lambda_2 + \lambda_3 = 1$ ， $\lambda_1, \lambda_2, \lambda_3$ 分别表示节点信任值、节点能力、节点动态性3个因素所占的权重。节点的信任值越高，能力越强，越能担当超级节点的任务； $\frac{1}{Peer_{Dynamic}}$ 表示节点动态性越弱，稳定性越好，它们形成的超级节点网络就越稳定。在上述3个因素中，节点能力和动态性的量值可以由节点自身提供，节点信任值的计算在下文中详细叙述。

在语义域中，候选超级节点作为超级节点的备份节点，是同一语义域中 $Peer_{Value}$ 值仅次于超级节点的2个节点，一旦超级节点退出或者出现故障，候选超级节点能够及时替代，避免了信息的丢失和语义域的重建。

对于每个超级节点而言，还要根据自身的资源条件、查询历史等信息尽可能把与自己语义相似和对所管辖语义域的

信任值较高的超级节点保持为邻居节点。

2.3 节点的退出

节点的退出分为正常退出和非正常退出2种情况。当普通节点正常退出时，向它所在语义域的超级节点发送一个请求退出报文，超级节点产生一个删除报文删除此节点在超级节点上的注册信息，和此节点相邻的节点删除此节点的信息；当非正常退出时，超级节点会定期发送一个扫描进程可以获知这个情况，从而及时地进行信息更新。当超级节点正常退出时，发送一个退出报文给所在语义域的候选超级节点，选择 $Peer_{Value}$ 值最大的候选超级节点作为新的超级节点，新超级节点与原超级节点在语义网内的邻居节点以及邻居超级节点建立逻辑连接，并且删除原超级节点，新超级节点向语义域内的所有节点发送广播消息，明确超级节点的身份，其他邻居节点也刷新邻居列表；当非正常退出时，候选超级节点周期性地发送扫描报文证明超级节点的存在，当得知超级节点非正常退出时，候选超级节点能及时接替超级节点的工作，并通知语义网内的其他节点，原超级节点的域内邻居节点和域间邻居超级节点也能通过定期的扫描得知新超级节点的存在，并且及时建立逻辑连接。

2.4 信任值的计算

信任是指节点依据历史交往记录形成的对另一节点未来行为的能力、诚实、可靠等方面的主观期望^[2]，是2个节点之间一对一的主观关系。信任机制^[3]是指用户通过自己的过去经历或他人的推荐来选择符合自己要求的交互端的一种机制。采用信任机制能激励用户提供高质量的可靠服务，节省时间和通信开销，促进网络的良性发展。

P2P结构和网格结构存在某些相似之处，因此，本文采用网格中基于信任域的信任模型，该模型将信任关系分为域内信任关系和域间信任关系，设置不同的策略来处理这2种不同的信任关系。

(1)域内信任关系基于同域节点间的交易。任一节点的信任数据库维护2张表：直接信任表和推荐表。在某个节点需要计算另一个节点的信任值时，若2个节点有直接信任关系则直接采用这个信任值；若无则查推荐表，在可以提供推荐的节点中找到一个与对方节点有过直接信任关系的节点，使用该节点的推荐信任值。本文中，节点信任值的计算方法为

$$Trust_a(b) = \begin{cases} dir_a(b) & a \text{对} b \text{有直接信任} \\ rec_x(i_1) \times rec_x(i_2) \times \dots \times rec_x(i_n) \times dir_a(b) & a \text{对} b \text{有推荐信任} \\ 0 & a \text{对} b \text{无信任关系} \end{cases}$$

其中， $Trust_a(b)$ 表示 a 对 b 的信任值， $dir_a(b)$ 表示 a 对 b 的直接信任值； $rec_x(y)$ 是 x 对 y 的推荐因子，表示 x 对 y 提供的推荐信息的信任程度，而 y 是 x 的推荐者。

(2)域间信任关系基于域间节点的交易。信任评价的对象以域为单位。域间信任值是它们的直接信任值和其他域的推荐信任值的综合。每个语义域的超级节点维护2张表：1)域间信任关系表，存储与该域有过直接交易的域的整体信任值；2)包含域内所有节点的权值表，权值表给域内每个节点赋一个权值，用来表示不同节点对该域信任值的影响因子。

若语义域 M 和语义域 N 之间存在一定的信任关系，域内节点的行为可能增加或减少相应的信任值：

$$Trust_M(N) = \begin{cases} dir_M(N) & M \text{对} N \text{有直接信任关系} \\ a \times dir_M(N) + b \times \frac{\sum_{i=1}^n dir_{X_i}(N)}{n} & M \text{对} N \text{没有直接信任关系} \end{cases}$$

其中， X 表示除 M, N 外的任意一个其他域； n 表示除 M, N

外的其他域的数目, a , b 分别表示域间直接信任关系和间接信任关系的权值。

3 基于语义和信任的资源发现算法

资源搜索策略主要包括每个语义域中的消息转发策略和语义域间的消息转发策略。在每个语义域中, 节点的信任值反映了节点间的历史交互信息。因此, 在每个语义域中, 首先采用基于Directed-BFS^[4]的搜索策略, 在选择邻居节点转发查询消息时选择信任值较大的一组节点转发, 采用这种启发式策略能够提高查询效率, 避免恶意节点的欺诈行为, 如果没有节点响应, 可以在所属语义域中重新进行洪泛查询; 在语义域间转发查询消息时, 选择语义相近和信任值高的超级节点转发。搜索算法流程如图 2 所示。

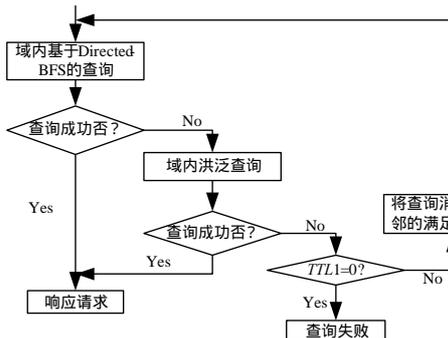


图 2 搜索算法流程

具体搜索过程如下:

(1) 请求节点发出查询消息 $Search(P, query, Trust_{min}, TTL, TTL1)$, 其中, P 是资源搜索节点; $query$ 是一个资源探测报文; $Trust_{min}$ 表示请求节点需要的最低信任值; TTL 为域内查询的跳数限制, $TTL1$ 用来限制消息转发的语义域的数目, 避免消息在不同语义域中无限制地传播。节点 P 首先检查自己的文件数据库, 如果该节点有满足条件的资源, 则查询成功, 停止查询; 否则, 节点查询信任数据库, 选择 $Peer_{value} > Trust_{min}$ 的一组邻居节点转发查询消息, 如果有多个节点满足查询条件, 则选择信任值较高的节点响应请求, 算法结束, 停止查询, 并且更新节点的信任数据库。

(2) 当 $TTL=0$ 时, 如果没有任何节点响应查询消息, 那么, 节点 P 重新设置 TTL 值为 $TTL2$, 在所属语义域中进行洪泛查询, 如果有多个节点满足条件, 选择信任值较高的节点响应。当 $TTL2=0$ 时, 没有任何节点响应, 则转到(3)。

(3) 查询节点将查询消息转发到所属语义域的超级节点, 超级节点将查询消息转发到信任值最高且没有查询过的邻居超级节点所属的语义域, 并且 $TTL1-1$, 在该语义域中继续(1)和(2)的查询, 如果查询成功, 则停止, 返回查询结果, 并且更新节点的信任数据库; 否则, 循环(3)的操作, 直到 $TTL1=0$ 。

(4) 当 $TTL1=0$ 时, 没有任何节点响应查询消息, 则可认为查询失败, 并且向发出请求的节点发送 fail-message 消息, 同时更新节点之间的信任值以及域内节点的权值。

4 仿真及结果分析

本模拟实验是在一台 PC 上完成的, 主要与协议 Gnutella 进行比较。实验过程中产生 4 个语义类别的文件。每个节点管理 100 个文件, 分别属于 2 个语义类别, 比例为 0.8:0.2。仿真实验主要从 2 个方面比较分析了 Gnutella 环境下的洪泛算法和本文构造的资源搜索算法。为了简化实验过程, TTL ,

$TTL1, TTL2$ 采用相同的值。

查全率: 由图 3 可知, 由于对节点进行了有效的组织, 并且选取了合适的超级节点, 因此在相同 TTL 的情况下, 本文算法的查全率均高于 Gnutella。

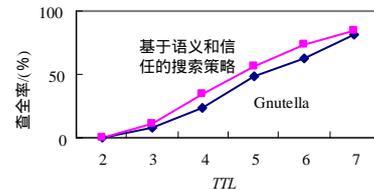


图 3 查全率

带宽利用率: 在搜索过程中, 如果减少转发消息的节点数, 但又不影响搜索结果, 则会减少网络带宽的占用, 提高带宽利用率。所以, 用访问节点数量的多少来衡量网络的搜索效率。通过多次提交查询请求, 统计访问节点的数量, 得到实验结果如图 4 所示。图 4 中显示在返回的文档数量相同的情况下, 本文算法的节点访问量远远低于 Gnutella, 随着返回文档数的增加, 这种变化越来越明显。因此, 基于语义和信任的搜索策略能够更高效地利用带宽, 系统的可伸缩性也更好。

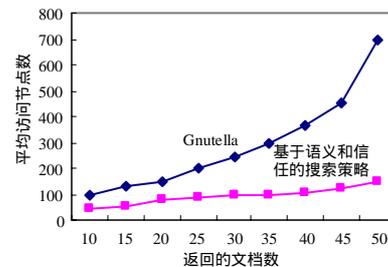


图 4 平均访问的节点数

5 结束语

本文构造的基于语义和信任机制的超级节点网络能够避免采用提供虚假信息的欺骗性节点。这种设计的最大优点在于选取的超级节点的能力强、稳定性好, 因此, 能提供更好的服务, 在一定程度上保证网络的良好性。

由于相似语义的节点属于同一个语义域, 查询时先在所属语义域中进行查询, 避免了在整个网络中无目的的查找, 在查询时考虑到节点的信任值, 一次成功或失败的查询能够更新节点的信任值, 从而对以后的查询提供更好的指导, 能够提高查询速度、查全率, 有效减少网络负载, 提高宽带利用率。

参考文献

- [1] Milojicic D S, Kalogeraki V, Lukose R, et al. Peer-to-Peer Computing[R]. HP Laboratories, Technical Report: HPL-2002-57, 2002.
- [2] Wang Yao, Vassileva J. Trust and Reputation Model in Peer-to-Peer Networks[C]//Proceedings of the 3rd IEEE International Conference on Peer-to-Peer Computing. [S. l.]: IEEE Press, 2003.
- [3] Resnick P, Kuwabara K, Zeckhauser R, et al. Reputation systems[J]. Communications of the ACM, 2000, 43(12): 45-48.
- [4] Beverly Y, Garcia-Molina H. Improving Search in Peer-to-Peer Networks[C]//Proceedings of the 22nd International Conference on Distributed Computing System. Vienna, Austria: IEEE Press, 2002.