

一种有效的非结构化 P2P 网络资源搜索策略

李春秀, 刘方爱

(山东师范大学 信息科学与工程学院, 山东 济南 250014)

摘要:针对非结构化 P2P 网络资源搜索算法中冗余消息数过多、搜索效率低等问题,提出了一种基于邻居-邻居节点的非结构化 P2P 网络资源搜索策略,该策略可以在一定程度上防止路由环路产生,同时综合考虑邻居节点的活跃度和资源命中数,选择下一条邻居节点路径转发查询消息,将查询信息发往目标可能存在的区域,因此减少了发往网络的查询信息包数量,节省了网络带宽。实验结果表明,该算法可以减少大量的冗余查询信息,提高资源搜索的成功率,是一种有效的非结构化 P2P 网络资源搜索策略。

关键词:非结构化 P2P 网络;活跃度;资源命中数;NNRS 算法

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2010)11-0117-05

An Efficient Resources Search Strategy in Unstructured P2P Networks

LI Chun-xiu, LIU Fang-ai

(School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)

Abstract:For the resources search algorithm's too many redundant messages and low efficiency issues in unstructured P2P networks, which proposed a resources search strategy based on neighbor-neighbor node, the strategy can prevent the routing loop to some extent, meanwhile comprehensive consideration of neighbor nodes' activeness and resources hits to choose the next path to neighbor nodes, it sent messages to potential target areas, so the number of queries' packets are decreased and saving the network bandwidth. Experimental results show that, this algorithm can reduce the number of redundant queries and improve resource search success rate, it is an effective search strategy in unstructured P2P networks.

Key words:unstructured P2P networks; activeness; resources hits; NNRS algorithm

0 引言

近年来,随着网络带宽的不断提高和计算机处理能力的不断增强,Peer-to-Peer(简称P2P)作为一种端到端的互联网技术,得到了全面而广泛的发展。在P2P网络中,不再存在中心服务器,所有节点都进行直接的相互连接并处于同等地位,称为对等点。对等点在P2P网络中既是客户机,享用其他节点提供的服务,同时又充当服务器,为其他节点提供服务^[1]。资源搜索问题是P2P网络的核心问题之一。目前,非结构化P2P网络的资源搜索通常采用基于洪泛(Flooding)算法的消息传递机制,但是洪泛算法容易产生大量冗余通信数据包,占用大量网络带宽,导致搜索速度较慢,降低了系统的效率^[2]。

通过对非结构化P2P网络资源搜索机制及现有改进方法的研究,文中提出了一种基于邻居-邻居节点的非结构化P2P网络资源搜索策略(Based on neighbor-neighbor node resources search,简称NNRS),该策略可以在一定程度上解决冗余消息包的问题,避免网络中路由环路的产生,同时对发往邻居节点的查询信息进行一定条件的限制,只向满足条件的节点路径发送查询信息,在一定程度上减少了发往P2P网络的查询信息包数量,节省了网络带宽,提高了资源搜索的速度和效率。

1 常见的非结构化 P2P 网络资源搜索策略

非结构化P2P网络的基本搜索算法是洪泛(Flooding)算法,该算法的特点是路由时进行全网络遍历的盲搜索,网络的拓扑没有固定的结构,各个节点需要维护一个记录邻居节点的路由表,用来传递查询消息。当网络中的某个节点发出查询请求时,首先向它所有的邻居节点传输查询信息,如果某个邻居节点有符合条件的资源,便响应该查询请求,否则,邻居节点

收稿日期:2010-03-08;修回日期:2010-06-19

基金项目:山东自然科学基金(Y2007G11)

作者简介:李春秀(1981-),女,山东济宁人,硕士研究生,研究方向为网格计算、互联网络;刘方爱,博士,教授,博士生导师,研究方向为并行处理、互联网络、网格计算等。

继续向自己的邻居节点转发查询请求,一直重复这个过程,直到搜索过程结束^[3]。为了防止无限制地向外扩张搜索,每个搜索请求都需要带一个非零的 TTL 值,整个搜索过程通过 TTL 值控制消息的生命期。这种搜索算法的节点覆盖率很高,但是产生了过多的冗余消息,占用了大量的网络带宽,不具有可扩展性。为此,研究者们提出了许多改进算法来提高资源搜索过程中消息数量的可控性和查询消息的命中率^[4]。

文献[5]提出了迭代深入(Iterative Deepening)方法,它是一种逐步进行的宽度优先搜索,在该算法中进行多次泛洪搜索,每次的搜索深度限制都是递增的,当查询结果满足要求或者已经达到最大深度限制时,整个搜索过程结束。如果搜索能够在较小的深度范围内找到满意的结果,则可以减少网络中传播的消息数量,提高网络带宽的利用率。

文献[6]运用了本地索引(Local Index)方法,节点建立 R 跳距离内邻居节点的数据索引, R 称为索引半径,是一个系统变量。在该方法中,每个节点不但可以处理本节点的查询信息,而且还可以处理该节点周围一跳或几跳的节点信息,因此使得处理查询请求的节点大大减少,总的 CPU 处理时间也相应地减少。

文献[7]中的定向广度优先搜索(Directed BFS)方法不是将搜索请求发送到所有邻居节点,而是根据节点记录的邻居节点的历史信息,选择性地发送查询消息给有良好记录的节点,从而减少了查询消息量,保证了在较短时间内得到所需的搜索结果。

文献[8]提出了随机漫步(Random Walk)算法,该算法在响应时间和网络负担之间做出了一种权衡。节点不是将查询消息发送到所有邻居节点,而是随机选取 k 个邻居节点转发查询消息,中间节点每次再随机选择自己的一个邻居节点转发查询消息,查询请求在路由过程中与请求节点保持联系,查询请求也被称为“Walker”,通过控制 Walker 数量,可减少网络开销。

2 基于邻居-邻居的资源搜索策略(NNRS)

非结构化 P2P 网络的基本搜索算法是洪泛(Flooding)算法,路由时进行全网络遍历的盲搜索,该算法产生了大量冗余查询信息包,极易造成网络流量增大,带来了沉重的网络负担,搜索效率并不高。网络中出现大量冗余消息的原因之一在于存在路由环路,即同一查询信息,多次地发往同一个网络节点,既占用了大量网络带宽,同时又给整个网络系统带来了沉重的负担^[9]。为了解决这一问题,文中提出了一种基于邻居-邻居节点的非结构化 P2P 网络资源搜索策略(NNRS),该策略可以在一定程度上防止路由环路的

产生,减少网络中冗余查询信息的数量,同时按照转移公式,选择更有可能处理查询请求的节点作为下一跳路由,也就是将查询信息发送到最可能具有响应节点的目标区域,进一步减少网络中查询消息的数量和通信开销,节省了网络带宽,提高了资源搜索的速度和查询命中率。

2.1 NNRS 算法简介

在该搜索策略中,非结构化 P2P 网络中的每个节点需要维护 4 个表:本地资源表、热点文件缓存表、邻居节点活跃表(neighbor node active table, NAT),邻居-邻居节点表(Neighbor-neighbor node table, NNT)。本地资源表主要负责本地资源的维护,同时可以提供资源信息的查询。热点文件缓存表主要负责缓存近期经常被访问的热点资源,对于热点资源,可以从该表中直接获取资源的 URL,节省搜索的时间和开销^[10]。邻居节点活跃表(NAT)是一个 $m \times 2$ 的矩阵,如图 1 所示。

	活跃度	资源命中数
邻居节点个数	NAT ₁₀	NAT ₁₁
	⋮	⋮
	NAT _{k0}	NAT _{k1}
	⋮	⋮
	NAT _{m0}	NAT _{m1}

图 1 邻居节点活跃表(NAT)

其中, m 表示所有邻居节点的数目,矩阵中 NAT_{k0} 存储了该邻居节点 k 所具有的活跃度,活跃度越大说明该邻居节点在新近一段时间内比较活跃,它能够接收查询信息的能力较强;NAT_{k1} 存储了邻居节点 k 的资源命中数,资源命中数记录该节点成功命中查询请求的次数,代表了该节点的历史记录信息。邻居节点活跃表(NAT)一方面反映了新近一段时间内邻居节点的情况,代表了节点的新近记录;另一方面又有作为历史记录信息的查询成功次数,这两条信息一起决定查询消息发往的下一个邻居节点,对下一个邻居节点的转发具有指导作用,使得查询信息发往目标可能存在的区域,提高了搜索的效率和命中率。邻居-邻居节点表(NNT)是一个 $n \times 2$ 的矩阵,如图 2 所示。

其中,第一列为节点 n 的邻居节点,第二列为节点 n 对应的邻居节点的邻居节点。

文中的 NNRS 算法仍然采用 TTL 值来控制网络中查询信息生命期,即为每个查询信息设定一个 TTL_{max} 值,当 TTL 值为零时,该查询消息自动销毁。在 TTL 计算方面,借鉴文献[11]中 TTL 的计算思想,

如果在该节点没有发现目标资源,则该查询消息的 TTL 值减 1,而如果发现了目标资源,则该查询消息的 TTL 值不做任何变动,即 TTL 保持原值,目的是让具有高命中率的查询信息能够访问更多的网络节点,提高资源搜索的命中率。

TTL 值按公式(1)计算:

邻居节点 邻居的邻居节点	
n_1	$n_{x1}, n_{x2} \dots$
\vdots	\vdots
n_k	$n_{x3}, n_{x7} \dots$
\vdots	\vdots
n_n	$n_{x6}, n_{x8} \dots$

图 2 邻居-邻居节点表(NNT)

$$TTL(k) = \begin{cases} TTL(k) - 1, r(k) \in C(i) \\ TTL(k), r(k) \in C(i) \\ 0, otherwise \end{cases} \quad (1)$$

其中, $r(k)$ 表示该节点 k 所拥有资源的集合, $C(i)$ 表示查询信息, 即当节点 k 没有命中查询信息时, $TTL = TTL - 1$, 反之, TTL 的值保持不变, 其他情况下, TTL 的值为 0。

2.2 相关定义

定义 1(冗余消息)。冗余消息为在同一节点被重复处理或在同一条路径上被重复转发的查询消息。

定义 2(活跃节点)。网络中的节点在最近一次活跃期间内发起或转发了到达某个目的节点的查询消息, 则称这个节点为活跃节点。

定义 3(活跃度)。活跃节点所具有的代表活跃程度的一个正整数值称为活跃度, 对于 P2P 网络中的任何一个节点 v , 都有 $\forall v \rightarrow R, R$ 为任意正整数。

定义 4(搜索通信开销)。在 P2P 网络的一次搜索中, 平均每个节点收到的消息个数 f 定义为系统的搜索通信开销^[12]。如公式 2 所示:

$$f = \frac{1}{R} \sum_{i=1}^N m_i \quad (2)$$

其中, R 为查询请求可到达的节点个数; m_i 为节点 i 收到的查询请求个数; N 为系统的规模即网络中的节点个数。

在图 3 中, 若节点 A 为查询信息的发起者, 按照公式(2)可知, Flooding 算法的通信开销为:

$$f_1 = \frac{1}{4} (3 + 3 + 3) = \frac{9}{4}$$

查询消息冗余率:

$$R = \frac{2 + 2 + 2}{3 + 2 + 2 + 2} = \frac{6}{9}$$

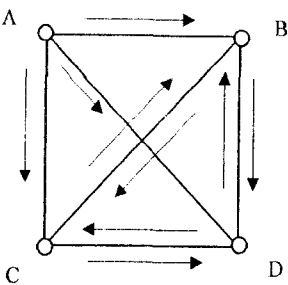


图 3 Flooding 算法查询信息的传输

在 NNRS 算法中, 节点 A、B、C、D 需要维护一个各自的 NNT 表, 如表 1 所示。

表 1 节点 A 的 NNT 表

邻居节点	邻居的邻居节点
B	A C D
C	A B D
D	A B C

表 2 节点 B 的 NNT 表

邻居节点	邻居的邻居节点
A	B C D
C	A B D
D	A B C

表 3 节点 C 的 NNT 表

邻居节点	邻居的邻居节点
A	B C D
B	A C D
D	A B C

表 4 节点 D 的 NNT 表

邻居节点	邻居的邻居节点
A	B C D
B	A C D
C	A B D

节点 A 根据它的 NNT 表, 将查询信息发往它的所有邻居节点(假设节点都满足转移概率), 节点 B、C、D 收到查询信息后, 若都没有目标资源, 则节点 B、C、D 将分别检查各自的 NNT 表, 进行查询信息的转发, 比如节点 B 查看除了节点 A 之外的邻居节点 C 和 D, 是否为节点 A 的邻居, 方法就是在自己的 NNT 表中查看节点 A 是否在邻居节点 C 的邻居节点里面, NNT 表显示邻居节点 C 为 A 的邻居节点, 所以, 节点 B 不向节点 C 转发查询信息包, 同理, 可得节点 B 不向节点 E 转发查询信息包等, 得到图 4。

因此 NNRS 算法的通信开销为: $f_2 = \frac{1}{4} (1 + 1 + 1) = \frac{3}{4}$, 与 Flooding 算法相比, 通信开销减少了 $\Delta f = f_1 - f_2 = \frac{9}{4} - \frac{3}{4} = \frac{6}{4}$, 冗余查询信息减少了 6 个, 查

询消息冗余率 $R = \frac{0}{3} = 0$, 由此可知 NNRS 算法可以避免环路, 减少冗余查询消息, 节省网络带宽, 通信开销较小, 消息冗余率较小。

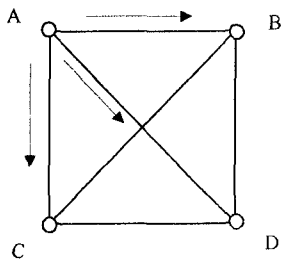


图4 NNRS算法信息的传输

2.3 NNRS算法步骤

在 NNRS 算法中, 每个节点需要维护一个邻居-邻居节点表(NNT), 如图2所示, 该表主要防止网络中路由环路的产生, 减少冗余消息的数量。网络中的节点在转发查询消息时, 首先检查 NNT 表中的待转发邻居节点是否有存在路由环路的现象, 若有, 则不向该节点转发查询信息, 然后计算剩余邻居节点的转移概率, 最后向满足条件的节点转发查询信息, 在一定程度上减少了网络中冗余消息数量, 节省了网络带宽, 提高了搜索的效率和成功率。

NNRS算法的描述步骤如下:

步骤1: 当非结构化 P2P 网络中某个用户通过一个源节点 P_i 进行关键字为 C 的信息查询时, 节点 P_i 首先查找自己的本地资源表和热点文件缓存表, 若找到满足条件的资源, 则将结果返回给查询用户, 并发送消息给它的邻居节点, 更新它们 NAT 表中相应的活跃度和资源命中数, 即各加 1, 否则执行步骤 3。

步骤2: 查看节点的 NNT 表, 是否有存在环路现象的邻居节点, 若有, 则不向该节点转发查询信息, 否则, 转到步骤 3。

步骤3: 节点 P_i 根据转移公式(3)计算每条待转发路径的选取概率, 设定一个概率阈值 $Q^{[13]}$, 且 $Q \in [0, 1)$ 。节点 P_i 根据计算的结果与 Q 值进行比较, 然后选择出满足条件的邻居节点, 向这些邻居节点转发查询信息, 同时 TTL 的值减 1。

$$P_k = \alpha \times \frac{\text{NAT}_{k0}}{\sum_{u \in U} \text{NAT}_{u0}} + \beta \times \frac{\text{TNAT}_{k1}}{\sum_{u \in U} \text{NAT}_{u1}} \quad (3)$$

式中, α, β 为权重, 分别表示活跃度和资源命中数在概率选择中的相对重要性, $\alpha, \beta \in [0, 1)$ 且 $\alpha + \beta = 1$ 。 NAT_{k0} 表示节点 P_i 所维护的邻居节点活跃表(NAT)中节点 k 所对应第一列的值, 为节点 k 的活跃度, NAT_{k1} 表示该邻居节点活跃表(NAT)中节点 k 所对应第二列的值, 为节点 k 的资源命中数, u 为节点 P_i

的任意一个邻居节点, U 为节点 P_i 的所有邻居节点。

步骤4: 当查询信息到达一个新的节点 P_j 时, 先对该节点的本地资源表和热点文件缓存表进行搜索, 若有目标资源则产生回应信息, 同时 TTL 的值保持不变, 并发送更新消息给它的邻居节点, 更新它的活跃度和资源命中数, 即进行加 1 操作, 然后转到步骤 2; 否则 TTL 的值减 1, 若 $\text{TTL} < \text{TTL}_{\max}$ 则转到步骤 2, 否则转到步骤 5。

步骤5: 整个查询搜索算法结束。

3 实验结果

P2P 网络的性能评价指标主要有: 检索通信开销、冗余消息数、响应时间、搜索成功率、带宽利用效率等。文中根据 2 个评价指标: 消息冗余率和查询成功率, 对比 NNRS 算法与 Flooding 算法。

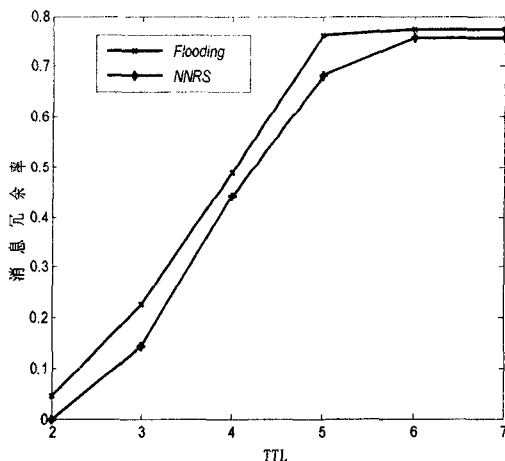


图5 消息冗余率的比较图

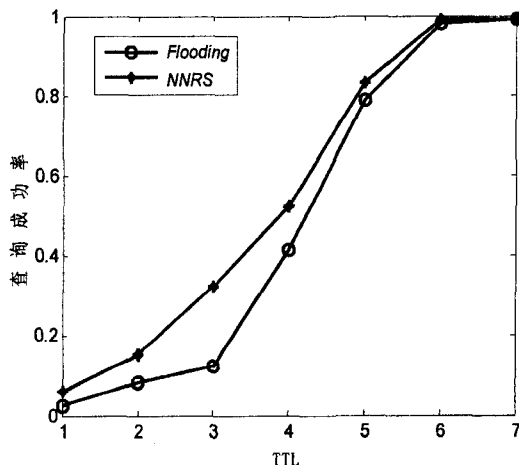


图6 查询成功率的比较

实验采用 Java 语言编程实现, 实验结果如图 5 和图 6 所示。在图 5 中, 与 Flooding 算法相比, NNRS 算法依据 NNT 表, 进行转发查询消息的时候, 在一定程度上防止了网络环路的产生, 减少了冗余信息数量, 因

此消息冗余率较低,能够在一定程度上节约网络带宽。在图 6 中,NNRS 算法依据 NAT 表选择下一个要转发的邻居节点,对查询消息的转发起到了指导性作用,将查询消息尽可能的发往目标资源所在的区域,提高了搜索成功率,可以看出,在相同的实验环境下,NNRS 算法的搜索成功率要比 Flooding 算法高一些。

由实验结果可知,与 Flooding 算法相比,NNRS 算法在搜索成功率和减少冗余消息数量上都有所改进,是一种有效的非结构化 P2P 网络资源搜索算法。

4 结束语

针对非结构化 P2P 网络基本搜索算法洪泛(Flooding)算法,产生大量冗余查询消息的问题,提出了一种基于邻居-邻居节点的非结构化 P2P 网络资源搜索策略(NNRS),该策略可以在一定程度上防止路由环路产生,同时结合转移公式,选择更有可能处理查询请求的节点作为下一跳路由,减少了发往网络中冗余查询消息的数量,提高了资源搜索的成功率,通过实验验证了该策略的优越性。

参考文献:

- [1] 王新生,李学,贾冬艳.基于蚁群算法的非结构化 P2P 资源搜索机制[J].计算机工程,2009,35(7):189-194.
- [2] 程希.物理位置与兴趣相结合的 P2P 搜索技术研究[D].南京:南京理工大学,2007.
- [3] 李运娣,冯勇.分布非结构化 P2P 网络资源定位研究[J].计算机工程与应用,2007,43(6):156-158.
- [4] 冯国富,张金城,姜玉泉,等.无结构 P2P 覆盖网络的拓扑

(上接第 116 页)

参考文献:

- [1] Comon P. Independent component analysis, a new concept[J]. Signal Processing, 1994, 36(3): 287-314.
- [2] 郭武,朱长仁,王润生.一种改进的 FastICA 算法及其应用[J].计算机应用,2008,28(4):960-962.
- [3] Cardoso J F. Blind signal separation: statistic principles[J]. Proceeding of the IEEE, 1998, 86(10): 2005-2009.
- [4] 杨福生,洪波.独立分量分析的原理与应用[M].北京:清华大学出版社,2006.
- [5] Zarzoso V, Comon P. Comparative speed analysis of FastICA [C]//In Proc. ICA-2007, 7th International Conference on Independent Component Analysis and Signal Separation. London, UK: [s. n.], 2007: 293-300.
- [6] Hyvarinen A, Oja E. A fast fixed-point algorithm for independent component analysis[J]. Neural Computation, 1997, 9

优化[J].软件学报,2007,18(11):2819-2829.

- [5] Yang B, Garcia - Molina H. Efficient search in peer-to-peer networks [C]//Proceedings of the 22nd IEEE International Conference on Distributed Computing Systems (ICDCS). Vienna, Austria: [s. n.], 2002.
- [6] Yang B, Garcia - Molina H. Improving search in peer-to-peer networks [C]//In: Sivilotti PAG. Proc. of the Int'l Conf. on Distributed Computing Systems. [s. l.]: IEEE Computer Society, 2002: 5-14.
- [7] Kalogeraki V, Gunopulos D, Yazti D Z. A local search mechanism for peer-to-peer networks [C]//In: Proc. of the 11th International Conference on Information and Knowledge Management. New York: ACM Press, 2002: 300-307.
- [8] Lü Qin, Cao Pei, Cohen E, et al. Search and Replication in Unstructured Peer-to-Peer Networks [C]//Proc. of the 16th International Conference on Supercomputing. New York, USA: [s. n.], 2002: 84-95.
- [9] 张欣璐,刘广钟.无结构对等网络资源搜索算法[J].上海海事大学学报,2008,29(2):78-81.
- [10] 谢鲲,张大方,谢高岗,等.基于轨迹标签的无结构 P2P 副本一致性维护算法[J].软件学报,2007,18(1):105-116.
- [11] Fang Guangwei, Zheng Xiao. Improving Query Mechanisms for Unstructured Peer-to-Peer Networks [C]//Bond P. ChinacomBiz 2008, CCIS 26. [s. l.]: [s. n.], 2009: 60-67.
- [12] 邢长明,刘方爱.基于 P2P 的网格资源发现机制研究[J].计算机技术与发展,2008,18(8):21-23.
- [13] Saroiu S, Gummadi K P, Dunn R J, et al. An Analysis of Internet Content Delivery Systems [C]//Proc. of the 5th Symposium on Operating Systems Design and Implementation. Boston, USA: [s. n.], 2002: 315-327.
- [7] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis[J]. IEEE Transactions on Neural Networks, 1999, 10(3): 626-634.
- [8] Hyvarinen A. A family of fixed-point algorithms for independent component analysis [C]//In Proc. IEEE Neural Networks for Signal Processing (NNSP) Workshop. Amelia Island, FL: [s. n.], 1997: 388-397.
- [9] 曾生根,朱宁波,包晔,等.一种改进的快速独立分量分析算法及其在图像分离中的应用[J].中国图像图形学报,2003,8(A):1159-1165.
- [10] 王小敏,曾生根,夏德深.独立分量分析算法的改进算法[J].电子器件,2008,31(5):1681-1684.
- [11] 张守成,李宏伟,刘永凯.一单元 ICA-R 快速算法[J].计算机工程与应用,2009,45(2):158-161.
- [12] 马建仓,牛奕龙,陈海洋.盲信号处理[M].北京:国防工业出版社,2006.