

非结构化 P2P 网络中 IACO 资源搜索算法*

房佩^a, 闫向龙^b, 良梓^a, 吴晓军^{at}

(陕西师范大学 a. 计算机科学学院; b. 物理学与信息技术学院, 西安 710062)

摘要: 针对非结构化 P2P 网络中资源搜索效率低、冗余消息较多的情况, 提出一种基于兴趣因子的蚁群优化算法(IACO)。该算法充分考虑了节点价值对资源搜索的影响, 在传统蚁群算法中引入节点的兴趣因子, 以动态调整节点价值和信息素在计算转发概率时的权重关系, 有效地指导了资源查询路径的生成。仿真结果表明, 该算法可以显著提高资源查询的成功率, 同时在一定程度上减少了网络中的冗余消息量。

关键词: 对等网络; 资源搜索; 节点价值; 蚁群优化算法; 兴趣因子

中图分类号: TP393.08 **文献标志码:** A **文章编号:** 1001-3695(2014)02-0522-04

doi:10.3969/j.issn.1001-3695.2014.02.048

IACO resources search algorithm in unstructured P2P network

FANG Pei^a, YAN Xiang-long^b, LIANG Zi^a, WU Xiao-jun^{at}

(a. School of Computer Science, b. School of Physics & Information Technology, Shaanxi Normal University, Xi'an 710062, China)

Abstract: To solve the problems of low efficiency and more redundant messages in unstructured P2P network resources search, this paper proposed an ant colony optimization based on interest factor(IACO). It gave full consideration to the effects of node value of resources search. It introduced node interest factor into traditional ant colony algorithm to adjust the weight relationship of node value and pheromones dynamically in the forward probability calculation, and effectively guided the generation of resources query path. The simulation results show that the proposed algorithm can improve the success significantly rate and to a certain extent can reduce the amount of redundant messages in the network at the same time.

Key words: P2P network; resources search; node value; ant colony optimization algorithm; interest factor

0 引言

近年来,随着网络资源的不断丰富,P2P网络也应运而生^[1]。P2P网络中没有中央服务器,网络中的每个节点既是客户机,又是服务器,通过节点之间直接交换信息来实现资源共享。相对于传统的C/S模式,P2P网络在容错性、扩展性等方面具有很大的优势,因此被广泛应用于文件共享系统、协同工作、分布式计算等领域。

P2P网络按照拓扑结构可以分为集中式拓扑(如Napster)、全分布式非结构化拓扑网络(如Gnutella)、全分布式结构化拓扑网络(如CAN、Chord)和半分布式拓扑网络(如Kazaa)。其中非结构化拓扑网络在网络拓扑上没有严格的控制,节点加入或离开网络比较随意,不会因为拓扑的调整给整个网络带来很大的负担,所以它是目前运用最为广泛的一种P2P网络结构。在这种拓扑结构中的搜索算法大致可以分为盲目搜索和智能搜索两种。盲目搜索^[2]是指在进行搜索转发时,将消息随机地转发到所有的邻居节点进行资源查找,由于转发的盲目性,会产生大量的查询包,严重地消耗了网络带宽,传统的洪泛算法和在此基础上改进的BFS以及随机漫步算法都属于盲目搜索算法^[3-5];智能算法^[2]是依据节点的已有信息和查询信息,对其邻居节点进行概率分析,有目的地选择转发概

率大的节点进行资源的查询,提高资源查询的效率。文献[6]中利用空间上的局部性即将拥有相似资源的节点聚集成簇,建立不同的兴趣域,实现资源的簇内和簇外搜索;文献[7]中利用节点的信任度,提出一种激励机制,将查询消息转发到信任度高的节点;文献[8]中引入空间向量模型,根据非结构化P2P网络中节点保存的资源信息将节点划分为不同的兴趣域,通过兴趣域进行资源搜索。文献[9]采用分组思想,将P2P网络划分成不同的兴趣组,兴趣组中由中心节点进行统一管理,并且提出了备用中心节点,减少了单点失效对系统造成的危害;文献[10]在P2P网络资源搜索中引入蚁群算法,利用节点信息素和蚁群的正反馈机制智能地选择下一跳路由。

本文在以上研究的基础上,提出一种基于兴趣因子的蚁群资源搜索(ant colony optimization based on interest factor,IACO)算法。通过分析节点价值对资源搜索性能的影响,在蚁群算法中引入节点的兴趣因子,动态调整节点价值和信息素在计算转发概率时的权重关系,选择概率大的邻居节点进行消息转发,在非结构化P2P网络中实现了资源的智能查找。最后通过仿真实验对该算法的性能进行了验证。

1 传统的蚁群算法

蚁群优化(ant colony optimization,ACO)算法^[11]是在20世

收稿日期: 2013-04-11; **修回日期:** 2013-05-14 **基金项目:** 国家自然科学基金面上项目(11172342); 国家教育部新世纪优秀人才支持计划资助项目(NCET-11-0674); 陕西省自然科学基金资助项目(2012JM8043)

作者简介: 房佩(1987-),男,河南邓州人,硕士,主要研究方向为嵌入式计算机系统(aygyfp@163.com); 闫向龙(1987-),男,陕西神木人,硕士,主要研究方向为智能传感器及传感器网络; 良梓(1987-),女,黑龙江五大连池人,硕士研究生,主要研究方向为智能控制与应用系统; 吴晓军(1970-),男(通信作者),陕西凤翔人,教授,博士,主要研究方向为模式识别与智能系统、复杂系统与复杂性理论。

纪90年代初,由意大利学者Dorigo等人通过研究自然界中蚂蚁的觅食行为提出的一种智能搜索算法。经研究发现,自然界中的蚂蚁在找寻食物时呈现一定的规律性,在这个过程中它们会留下一种类似于信息素的物质,用来指导后续蚂蚁的觅食行为,以后的蚂蚁会感知路径中的信息素浓度来指导自己的下一步搜索行为。所有的蚂蚁都会朝信息素浓度高的路径搜索,这种基于信息素的正反馈机制,使得某一路径上的信息素浓度越来越高,因此它被选择搜索的概率也逐渐变大,搜索的蚂蚁也就越多。蚂蚁的这种集体激励机制可以启发式地促使较短的路径优先被选择。研究发现蚂蚁的这种智能行为比较适合P2P这种动态网络的资源搜索。因此,蚁群算法被越来越多地应用于非结构化P2P网络的资源搜索中。文献[11]在P2P网络资源搜索中引入蚁群算法,按照资源中的关键字将信息素分类,通过不同的信息素智能地选择下一跳路由。文献[12]利用蚁群算法的正反馈机制,充分考虑了邻居节点度和邻居一邻居节点信息,降低了在搜索过程中形成环路的概率,减少了网络中的冗余消息量。

通过研究发现,当前这些算法并没有克服蚁群算法搜索初期资源查询效率低的问题,很少考虑到节点价值以及所要查询的资源与节点的兴趣之间的相似性对资源搜索的影响,而本文的研究则正是基于此展开的。

2 相关描述

通过在蚁群算法中引入兴趣因子来计算节点价值和信息素的权重,提高系统初期节点价值在转发概率中的比重,可以有效地解决传统蚁群算法初期资源搜索效率低的缺点,随着资源查询的进行,兴趣因子可以动态调整节点价值和信息素在转发概率中的权重,更好地形成了有效的查询路径,提高资源的搜索效率。

2.1 节点价值

非结构化P2P网络是一个巨大的动态网络,人们很难知道其中的一个节点相对于整个网络的“全局价值”,所以把一个节点在其邻居节点中的“局部价值”作为该节点的当前价值。非结构化P2P网络中的资源搜索算法都是试探性的搜索,在相同的时间内资源搜索覆盖度越高,则搜索成功率也就越高。本文假定在非结构化P2P网络中的资源是随机分布的,资源搜索覆盖度指的是单位时间内搜索到的资源数量,其中节点的度数和共享资源数是决定资源搜索覆盖度的两个重要因素。因为在P2P网络中,度数高的节点可以确保消息在下一跳转发到相对多的节点,而共享资源多的节点可以保证查询消息在当前查询到相对多的资源,这些都大大地增加了资源查询的命中概率。因此,用节点度数和拥有的资源数来表示该节点的价值,指导生成资源查询路径。

2.2 兴趣因子

在资源搜索的不同阶段,为了动态调整节点价值和信息素在转发概率中的比重关系,提高资源的查询效率,本文在蚁群算法中引入了节点的兴趣因子,用兴趣因子来表示一个节点当前的查询请求与一段时间内历史记录表中记录的查询请求之间的相似度。

经文献[13]研究表明,P2P网络中的节点在资源查询时会呈现一定的特点,即节点感兴趣的内容往往会表现出一定的集

中性,很多节点在一段时间内会对某一个主题的内容表现出一定的兴趣,因此可以通过历史记录表来得到当前的查询请求与过去一段时间内该节点的查询请求之间的相似度,将这个相似度作为节点的“兴趣因子”来决定节点价值和信息素之间的权重关系。如果兴趣因子比较大,则说明当前查询的资源在过去一段时间内出现过的概率比较大,该资源在路径上留下的信息素会比较多,此时信息素在转发概率中所占的比重会比较大;反之,该资源在过去一段时间很有可能没有被请求过,则节点的价值所占的比重会比较大,查询消息会被转发到存在该资源可能性较大的节点上。

在资源搜索过程中,查询请求中可以包含多个关键字,这些关键字往往可以表现出一定的兴趣,因此可以用兴趣域来分别表示当前查询请求和节点历史查询的兴趣类型。兴趣域用关键字出现的频率 w_i 组成的特征向量 d 来表示^[9],其中 $d = (w_1, w_2, \dots, w_n)$,所以查询请求 q 的兴趣域用特征向量表示为 $NE(q) = (w_{q,1}, w_{q,2}, \dots, w_{q,n})$,节点 j 的兴趣域 $NE(j) = (w_{j,1}, w_{j,2}, \dots, w_{j,n})$,其中 n 表示关键字集合。则节点的兴趣因子为

$$\alpha(q, j) = \frac{\sum_{k=1}^n (w_{q,k} \times w_{j,k})}{\sqrt{\sum_{i=1}^n w_{q,i}^2} \times \sqrt{\sum_{i=1}^n w_{j,i}^2}} \quad (1)$$

在系统运行的初期,路径上的信息素浓度均为 τ_{\min} ,此时信息素浓度并不能很好地指导资源的查询,加入兴趣因子,提高节点价值在转发概率中的比重,可以更好地形成有效的查询路径,很好地解决了传统蚁群算法初期资源搜索效率低的缺点,保证了算法自始至终的有效性,提高了系统在整个搜索周期中的效率。

P2P网络是一个动态的网络,网络中的节点和信息实时地发生变化,所以不同时间内节点的兴趣因子也不一样。为了保证节点兴趣因子的实时性,则需要实时更新节点历史记录表中的信息。为节点每一次保存的历史记录增加更新时间戳,当节点的历史记录表存满数据时,则根据记录的更新时间戳将距离当前时间最久的记录置换出表外,以此来保证数据表中记录的实时性和动态性,确保在不同的查询阶段节点兴趣因子的实时性,使得IACO算法的效率和优势得到很好的保证。

3 IACO资源搜索算法

3.1 基本思想

将整个P2P网络中的节点看成是一个一个的蚁巢,每一个查询消息看做是一只蚂蚁,所要查询的资源看做是蚂蚁所要寻找的食物。每一个蚁巢中存放有一定数量的P2P资源、当前信息素浓度、历史查询信息和邻居节点的相关信息。一个节点收到一个查询信息包,相当于一只蚂蚁在蚁巢中寻找食物。首先在本本地查找是否存在该资源,若存在,则发出响应消息;否则向邻居节点转发此查询消息,这时要根据节点的价值、兴趣因子和信息素浓度来计算各个邻居节点的转发概率,从中选取 N 个概率最大的节点来转发查询消息,并且将消息中的TTL值减1,更新搜索禁忌表;当邻居节点收到查询消息时,重复上述过程,直到找到所需资源或者TTL值减为0此次查询结束,然后更新路径中的信息素信息。在系统中还会定期挥发节点中的信息素,避免出现节点中信息素过大的情况。算法的流程如图1所示。

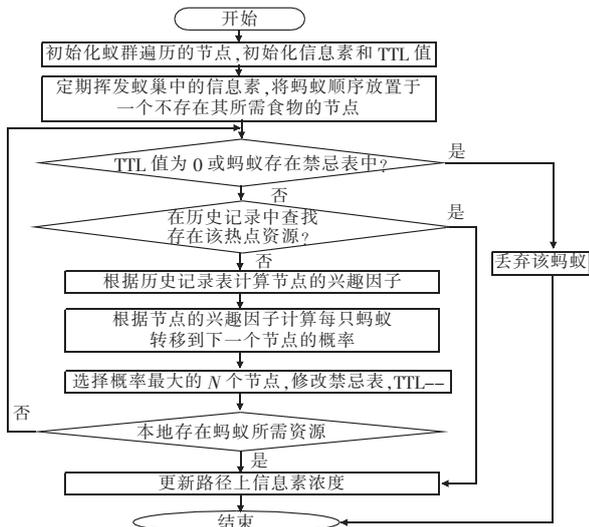


图 1 IACO 资源搜索流程

3.2 算法实现

为了方便 IACO 算法在 P2P 资源搜索时的应用,网络中的每一个节点都需要维护三张数据表:

a) 本地资源表,用来维护本地资源信息,方便其他节点进行资源请求。

b) 历史记录表,记录一段时间内该节点查询过的资源信息,该表中包括的资源信息主要分为查找成功的资源和查找不成功的资源两种。对于查找成功的资源则保存获取该资源的地址;而对于查找不成功的资源则需要保存该资源被请求过的次数,并且为每一条记录保存相应的更新时间戳,当表内空间不足时,按照时间戳将存放时间最长的数据置换出去。该表主要用于流行资源的快捷查询和下一次查询时节点兴趣因子的计算。

c) 邻居节点信息表,它用一个 $n \times 4$ 的二维矩阵来表示,其中 n 表示该节点当前邻居节点的个数,该表主要用来存放邻居节点当前的相关信息,主要包括邻居节点的度数、共享资源数、信息素浓度和历史查询记录,可以通过该表确定节点的价值和它的兴趣因子,用于计算节点的转发概率。

由该算法的主要思想可知,算法利用节点的兴趣因子,从节点价值和信息素两个方面来计算节点转发的概率。当一个节点收到资源查询的消息时,算法具体步骤如下:

a) 判断该消息中的 TTL 值是否为 0 或者在最近一段时间内是否收到有相同标示的查询消息,若有则转到步骤 g), 否则转向步骤 b)。

b) 查找节点 i 的本地资源表中是否存在所要请求的资源信息,若存在则转向步骤 f), 否则转向步骤 c)。

c) 查找节点 i 的历史记录表中所保存的已查找到的资源中是否有该请求资源,若有则转向步骤 f), 否则转向步骤 d)。

d) 判断节点 i 的邻居节点是否在查询消息的搜索禁忌表中,若全部存在则转向步骤 g), 否则转向步骤 e)。

e) 根据式(2)(3)计算不在禁忌表中的每一个邻居节点的转发概率,选择 N 个概率最大的节点作为转发节点,并将这些节点加入到消息的禁忌表中,将 TTL 值减 1, 转向步骤 a)。

$$p_{ij}^v(t) = \frac{d_j + k_j}{\sum_{a \in A} (d_a + k_a)} \quad (2)$$

$$p_{ij} = (1 - \alpha)p_{ij}^v + \alpha \frac{\tau_j}{\sum_{a \in A} \tau_a} \quad (3)$$

其中: p_{ij}^v 表示基于邻居节点价值的转发概率; d_j 和 k_j 分别表示节点 j 的度数和资源数; A 表示节点 i 的邻居节点集合; α 表示节点 j 的兴趣因子; τ_j 表示节点 j 上的信息素浓度。

f) 响应节点按照原路返回请求资源,利用式(4)(5)对路径上的信息素进行更新,信息素更新完成后转向步骤 g)。

$$\tau_{ij}(t) = (1 - \rho)\tau_{ij}(t - 1) + \Delta\tau \quad (4)$$

$$\Delta\tau = \tau \times \mu^n \quad (5)$$

其中: ρ 表示信息素的挥发因子; $\Delta\tau$ 表示信息素的增量; τ 表示初始的信息素增量; μ 为信息素的递减因子; n 表示响应节点距离更新节点的跳数,这样保证了离响应节点最近的节点每一次增加的信息素最多。

g) 此次搜索结束。

根据以上搜索步骤可知,搜索蚂蚁每一次通过选择概率大的节点进行资源查询,使得查询请求被转发到存在目标资源可能性大的区域,这样可以提高资源搜索的成功率。

4 仿真实验

在非结构化 P2P 网络的资源搜索算法中,资源搜索成功率和产生的消息数量是评价搜索算法的两个重要指标。其中搜索成功的次数除以搜索的总次数等于搜索成功率,搜索成功率越高则算法越有效;而资源搜索中产生的消息数量会影响到网络负载和带宽拥塞,在不影响搜索成功率的情况下,减少网络中的消息数量可以改善资源搜索的性能。本文的仿真实验正是基于这两方面对算法进行验证的。

为了验证 IACO 算法的有效性,本文采用 PeerSim^[14] 仿真工具进行实验,PeerSim 是一种支持结构化和非结构化 P2P 网络模拟的软件,具有很好的模拟效果。为了增加模拟的真实性,P2P 网络拓扑选择最常见的 Gnutella 网络,网络中的节点和资源分布都服从 Zipf^[15] 分布。在 Gnutella 网络中随机生成网络节点,每个节点的度数不大于 7,随机进行 100 次周期的资源查询,在查询过程中随机生成若干节点,采集相关数据进行算法的性能比较。实验相关参数如表 1 所示。

表 1 实验相关参数

实验参数	取值	参数说明
ρ	0.1	信息素的挥发系数
μ	0.8	信息素的递减因子
τ	1.0	更新节点时信息素增量
τ_{min}	1.0	系统初始的信息素浓度
N	4	选择转发的邻居节点个数

目前在传统蚁群算法的基础上也有许多改进的算法,但大多数的改进算法和传统的蚁群算法存在一样的缺点,在系统运行的初期,由于路径上的信息素指导作用不明显使得资源的搜索效率较低,IACO 算法则利用节点的兴趣因子,增加节点价值在系统运行初期的权重,提高算法初期的资源搜索成功率;随着资源查询周期的变化,动态调整节点价值和信息素在转发概率中的权重关系,通过节点查询的兴趣性倾向,可以提高资源查询的效率。本文从搜索成功率和消息总量两个方面,将 IACO 算法与传统的 ACO 算法和洪泛算法进行比较。通过 100 次周期的运行,采集了大量的数据,如表 2 所示。

表 2 中将每 10 个周期内数据的平均数作为当前算法下资源搜索的成功率和消息总量。从表中的数据可以看出,在任一

周期内,IACO算法的搜索成功率都大于Flood和ACO算法,其中在系统运行最初的10个周期内,IACO算法的搜索成功率则远远大于其他两种算法,并且在系统运行的整个阶段,IACO算法产生的消息总量均少于其他两种算法产生的消息总量。

表2 三种算法性能对比结果

周期数	搜索成功率/%			消息总量		
	IACO	ACO	Flood	IACO	ACO	Flood
1~10	35.6	15.7	14	10370	13952	14075
11~20	48.9	35.5	34.3	54956	58701	71886
21~30	59	51.9	46.3	91761	95620	126685
31~40	64.8	60.4	55	138654	150086	183994
41~50	69	65.6	60	168071	191580	274641
51~60	74.5	69.4	65.1	243204	262021	326224
61~70	77.6	70	67.3	328179	360018	437491
71~80	79.5	74	70.3	387334	425023	542647
81~90	81.5	75	71.6	453230	471759	609200
91~100	80.4	74.5	71	489973	562498	689240

图2给出三种算法资源搜索成功率的对比,图2(a)为资源搜索成功率随搜索周期变化的情况,图2(b)为资源搜索成功率随TTL值变化的情况。从图中可以看出,在搜索周期相同或者TTL值相等的情况下,IACO算法在搜索成功率方面有明显的优势,这是因为在不同的阶段节点的兴趣因子对消息的转发有不同的指导作用,使得消息被尽可能转发到存在请求资源的区域。

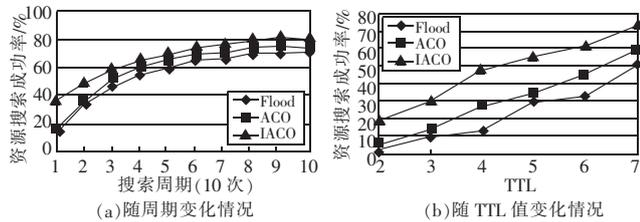


图2 资源搜索成功率对比

洪泛算法由于自始至终都采用盲目搜索,所以在资源搜索的初期搜索成功率十分低,随着搜索周期的增加,资源搜索的成功率逐步增加,到第80次后资源的搜索成功率稳定维持在70%左右;传统的蚁群算法在资源搜索初期,各个节点的信息素还没有建立起来,尚不能指导节点进行有效的搜索,所以此时的搜索成功率与洪泛算法的成功率相仿,但随着搜索的进行,节点的信息素也逐步地建立起来,为资源的搜索进行有效的指导,逐步提高资源搜索的成功率,并维持在75%左右;IACO算法引入了节点的兴趣因子,在搜索初期充分考虑了节点价值对资源搜索的影响,大大提高了资源搜索的成功率随着搜索周期的增加,搜索的成功率维持在80%左右。通过改变查询消息的TTL值也可以看出,在TTL值相同的情况下,IACO算法的搜索成功率也明显高于Flood算法和ACO算法。

从图3可以看出,随着查询周期的增加,网络中的消息总量持续增加。

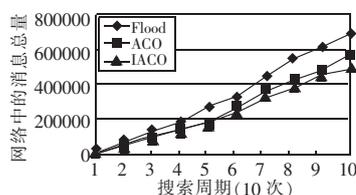


图3 网络中的消息总量随查询周期的变化

洪泛算法转发查询消息时采用盲目转发,所以它的消息增量最大;传统的蚁群算法对消息的转发有一定的指导性,减少了不必要的消息转发,所以它的消息增量远远低于洪泛算法;

而IACO算法在传统蚁群算法的基础上,充分考虑了节点价值和节点兴趣因子对资源搜索的影响,更好地指导了查询消息的转发,所以它的消息增量最低。因此IACO算法大大降低了网络中的消息总量,从而很好地减少了网络中的冗余消息数量。

5 结束语

本文针对P2P网络中的资源搜索提出了一种基于兴趣因子的蚁群优化算法。通过引入节点的兴趣因子,动态调整节点价值和信息素在转发概率中的比重,从而降低了节点转发时的盲目性,提高了资源搜索的效率。仿真实验结果表明,IACO算法可以大大提高资源的搜索成功率,并减少了网络中的消息冗余量。由于增加了计算节点转发概率的复杂性,在一定程度上会影响节点的响应时间,因此,如何降低节点的响应时间和均衡节点的负载会成为下一步工作的重点。

参考文献:

- [1] VINCENZA C, MICHELE M, GIUSEPPE M, *et al.* Social behaviors applied to P2P systems: an efficient algorithm for resources organization [C] // Proc of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. 2006: 65-72.
- [2] 钱宁, 吴国新. 无结构化P2P网络资源搜索机制研究综述[J]. 计算机科学, 2010, 37(4): 7-10.
- [3] JIANG Hong-bo, JIN Shu-dong. Exploiting dynamic querying like flooding techniques in unstructured peer-to-peer networks [C] // Proc of the 13th IEEE International Conference on Network Protocol. 2005: 122-131.
- [4] KALOGERAKI V, GUNOPULOS D, ZEINALIPOUR D. A local search mechanism for peer-to-peer networks [C] // Proc of the 11th International Conference on Information and Knowledge Management. New York: ACM Press, 2002: 300-307.
- [5] LV Qin, CAO Pei, COHEN E, *et al.* Search and replication in unstructured peer-to-peer networks [C] // Proc of the 16th International Conference on Super Computing. 2002: 84-95.
- [6] 刘业, 杨鹏. 基于自组织聚类的结构化P2P语义路由改进算法[J]. 软件学报, 2006, 17(2): 339-348.
- [7] 关志涛, 曹元大, 李剑. 混合对等网络中基于激励的可靠性搜索算法[J]. 北京邮电大学学报, 2008, 31(2): 42-45.
- [8] 赵新慧. 基于兴趣域的非结构化P2P模型[J]. 计算机与数字工程, 2011, 39(9): 75-78.
- [9] 侯秀红, 蔡道家. 基于兴趣分组的P2P网络拓扑模型[J]. 计算机工程与应用, 2009, 45(6): 124-127.
- [10] 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.
- [11] DORIGO M, MANIEZZO V, COLORNI A. Ant system: optimization by a colony of cooperating agents [J]. IEEE Trans on Systems, Man, and Cybernetics-part B: Cybernetics, 1996, 26(1): 29-41.
- [12] 李春秀, 刘方爱. 基于蚁群算法的非结构化P2P网络资源搜索策略[J]. 计算机工程与应用, 2012, 48(4): 97-99.
- [13] 赵新慧, 冯锡炜, 石元博. 基于兴趣域中心节点的P2P网络拓扑[J]. 科学技术与工程, 2011, 21(11): 5228-5231.
- [14] IRUM K, SYED F, YOUSAF B. PeerSim: an efficient and scalable tested for heterogeneous cluster-based P2P network protocols [C] // Proc of the 13th International Conference on Modelling and Simulation. 2011: 420-425.
- [15] RAO Wei-xiong, CHEN Lei, FU A W-C, *et al.* Optimal proactive caching in peer-to-peer network: analysis and application [C] // Proc of the 6th ACM Conference on Information and Knowledge Management. 2007: 663-672.