基于 Bloom Filter 和概率分发队列的 P2P 网络快速查找算法

程 澜 缑 锦 周 峰
(华侨大学计算机科学与技术学院 厦门 361021)

摘 要 无结构化 P2P 网络资源定位过程中的响应时间、查准率及覆盖率难以同时被优化。提出一种面向有向无环随机网络的基于 Bloom Filter 和概率分发队列的快速查找算法 BFPDQ(Bloom Filter and Probabilistic Distribution Queue),它用 Bloom Filter 表达和传递节点命中资源信息及查找请求信息,计算新查询消息与历史查询消息 Bloom Filter 语义向量相似度,并应用底层网络路径性能信息指导上层转发决策。概率分发队列(Probabilistic Distribution Queue,PDQ)把传统 walkers 表示成为查找消息分发队列,查找请求者协调各分发队列的查找方向和深度,并融合各队列查找过程中得到的定位消息。仿真实验表明,BFPDQ 算法在保持较少冗余信息的同时有效缩短了响应时间。

关键词 P2P 网络,Bloom Filter,概率分发队列,响应时间
中图法分类号 TP393 **文献标识码** A

Quick P2P Search Algorithm Based on Bloom Filter and Probabilistic Distribution Queue

CHENG Lan GOU Jin ZHOU Feng
(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract The strategy of searching resource is a research hotspot in unstructured peer to peer network. It is hard to optimize response time, query hit, and coverage rate for resource location of unstructured P2P network simultaneously. This paper presented a quickly search algorithm called BFPDQ (Bloom filter and probabilistic distribution queue), which is based on probabilistic distribute queue and Bloom filter technology. BFPDQ is mainly used for acyclic random network. Information of resources and requests can be expressed by Bloom filter technology. Meanwhile, performance information of the underlying network's path can be used to guide transmitting strategy for upper layers. PDQ (probabilistic distribute queue) uses distributed queues to substitute traditional walkers to search resources. Requester coordinates direction and depth of those queues and aggregates their resource location messages. Simulation results show that BFPDQ can decrease redundant information, while maintaining a significant reduction in response time.

Keywords Peer to peer network, Bloom filter, Probabilistic distribution queue, Response time

1 引言

目前无结构化 P2P 网络以其简单性、便捷性和易用性等优势在互联网上得到了大量应用。然而,实际应用较多的资源定位策略具有很大的盲目性,在查找过程中容易产生大量冗余消息进而形成网络拥塞。P2P 应用消耗了互联网主干网络中超过 60%的带宽,造成网络负载过重、延迟增大及网络带宽占用率过高等负面影响[1]。

无结构化 P2P 网络节点随机分布、拓扑简单、无强制约束,具有较高的动态适应能力和可扩展能力,但传统资源定位策略存在较大不确定性或冗余性。因此,在全分布自组织 P2P 网络中,如何快速命中目标资源并产生少量查询消息,一直是 P2P 研究的热点与难点。目前,研究人员提出很多关于基于索引的查找算法,其主要思想是收集网络相关信息,建立相应索引信息表,聚合索引表信息,采取概率导向等策略选择最优邻居节点转发查询请求;目的是减少查询消息传播过程中的不确定性[2]。

Bloom Filter 是一种存储复杂度很小的随机数据结构[3],已被广泛应用于无结构化 P2P 网络[4]和物联网[5]。这两类应用的共同点是,每个节点将共享资源用 Bloom Filter(简称 BF)向量表示,将 BF 向量传播至网络,并被网络中全部节点感知。同时,有限逻辑跳步内的节点因保存相应的 BF 向量需消耗大量存储空间。此外,网络感知资源更新消息尤为复杂,且节点间交互动作频繁,这 3 方面因素都消耗大量带宽。针对上述问题,Kumar 等人提出了依指数衰减的 BF 表达和传递节点资源信息,并在此基础上建立弱状态路由机制[6]。这一策略减轻了网络传播负载,但由于引入了指数衰减策略,使得定位效率性能不稳定,甚至将查找消息以一定的概率向错误方向传递。文献[7]对文献[6]中由"噪声"引起的正向错误率导向给出了修正模型。文献[8]提出了 DCBF(Data Copying and Bloom Filter)算法。DCBF 基于有向随机网络,对资源对象进行少量复制,并将副本随机路由给网络中的其它

到稿日期:2011-08-07 返修日期:2011-12-05 本文受国家自然科学基金(61103170),福建省自然科学基金(2010J01335),厦门市科技计划(3502Z20113022)资助。

程 澜(1987—),男,硕士生,主要研究方向为信息管理、人工智能、服务计算,E-mail:chenglanjudy@163.com;缑 锦(1978—),男,博士,副教授,主要研究方向为人工智能、知识工程、数据融合等;周 峰(1987—),男,硕士生,主要研究方向为知识融合、人工智能。

节点；对于拥有副本的节点以衰减 BF 向量传递副本信息，使得网络大部分节点感知资源副本及所在节点信息。DCBF 方法虽以较低的延迟命中目标，但也存在不确定性，副本信息和副本更新信息的转发使得网络信息冗余量较大。文献[9]提出了分布式丢弃 BF 技术，即节点对接收到的信息采用分布式丢弃策略，结合概率搜索小组算法，实现多个小组之间协同的并行搜索。该算法能保持低查找开销并取得较好命中效果，但系统中 BF 索引表维护和共享资源对象的更新十分复杂。

历史查找信息对新查找具有一定的指导意义。基于本地索引信息导向新查找的研究获得了广泛关注。利用历史索引信息，采用概率或排序的策略，转发查找请求，其目的是减少查找的盲目性。APS[10]算法考虑了邻居节点返回结果的成败，并依成功比例选择转发邻居节点，增强了定位性能和动态适应能力。但仅考虑邻居节点查找成功率并不能保证查找消息沿命中目标最大可能性的方向传递，所以命中目标的效率较低。BNS[11]算法从历史查找与新查找的语义关系出发，利用贝叶斯网络进行推理，选择推导概率较大的邻居节点转发查找请求，进一步提高了搜索成功率；但由于中文分词本身就十分复杂，获得语义相关词也较难，因此在一定程度上影响了路由查找请求的方向。

Bloom Filter 技术降低了索引信息维护的开销，同时减少了关键字硬匹配的错误率。本文将传统 Bloom Filter 技术的表示范围扩充到有向无环随机网络，使用 Bloom Filter 表达或传递节点命中资源信息、新查找请求信息及历史查找信息，改变了 BNS 算法中通过语义分词构建推理模型的方法。这种策略有效抑制了信息重叠和回流问题，并使得新查找信息与历史查找信息匹配度计算容易计算和处理。本文提出了 BFPDQ(Bloom Filter and Probabilistic Distribution Queue, BFPDQ)算法，其用 BF 技术表示查找消息，运用条件概率公式计算新查询与历史查询消息的相似度，选择相似度较大的邻居节点转发查找请求，并将传统 walkers 表示成查找消息分发队列，查找请求者协调各分发队列的查找方向和深度，并融合各队列查找过程中得到的反馈消息。BFPDQ 实现了各分发队列的协同合作及并行查找，同时考虑了底层网络距离信息对上层路径选择的指导作用。仿真实验表明，本算法能大量减少网络交互信息量，快速导向目标节点，提高查找成功率并对响应结果进行自动排序。

2 定位策略

2.1 设计思路

在 P2P 网络中，任一节点与多个对等节点相连。查找目标文件时不必向每个对等邻居转发查询消息，只需选择与查找请求相关的路径节点转发。为查找目标 g 建立向量描述是实现本文 BFPDQ 算法的基础，其目的是获得与查找目标相关的邻居集合。本文将所有查询请求信息(含本节点保存的历史查找信息 g)用 Bloom Filter(BF)向量定义。

定义 1 查找请求 g 的 Bloom Filter 向量表示为 V，初始状态时，V 是一个包含 m 位的位数组，每一位都置为 0。使用 k 个相互独立的 Hash 函数，分别将查找信息映射到{1,…,m}的范围中。新查找消息的 BF 向量表示为 V_{new}，历史查找请求表示成 V_{old}。

在判断通过历史查找请求消息 V_{old} 是否与新查找消息 V_{new} 相匹配时，不能笼统地回答“是”或“否”，而是应该返回一个表示向量相似度的值。

定义 2 $Match(V_{old},V_{new})$ 表示查找目标 g_{new} 的 BF 向量表示 V_{new} 与历史查找请求 g_{old} 的 BF 向量表示 V_{old} 之间的相关程度。

$$Match(V_{old},V_{new})=\frac{\sum_{i=1}^{m}(V_{old}[i]\times V_{new}[i])}{\sum_{i-1}^{m}V_{new}[i]} \tag{1}$$

相似度值越大，两个查询请求的相关程度就越高，历史查找请求对新查询可起到借鉴作用。计算节点存储历史查找请求信息与新查找请求的相似度，找出较高相似度的历史查询请求序列。

BNS 算法用分词器 WorldNet 获得搜索目标的上位词、同位词和下位词。词汇的语义用同义词集合来表示，并给出同义词集合元素的权值。中文分词处理比较复杂和困难，分词效果欠佳，最终影响转发路径选择。采用 Bloom Filter 向量表示查找信息，不仅节省了存储空间，而且减少了判定相似度的错误率。

2.2 网络模型

在随机有向、无环网络中，节点路由条目沿单个方向传递，路由至目标节点，阻止了信息回流，使主干网络的信息冗余量减少[7]。网络拓扑中每个节点通过有向边与邻居相连，节点代表随机变量，节点相连的有向边代表节点之间的相互关系，概率值表示选择转发的权值。通过条件概率构建选择转发模型，在有限且不确定的条件下进行计算推理。

基于有向网络建立推理模型，是为了判断在某邻居节点是命中目标资源的可能性。以下描述一个简单的有向随机网络推理模型，g 是查找目标，$u_1,u_2,\cdots,u_n$ 是节点上存储的历史查找索引表中的索引项，其中 $u_1,u_2,\cdots,u_i$ 是与 g 具有较高相似度的索引项；节点 n 的邻居节点表示为 $n_1,n_2,\cdots,n_l$。推理模型中，为了方便变量表示，节点和它代表的随机事件采用同一个变量标识。

查找目标 g 同时表示进行查找的随机事件，$g\in\{0,1\}$；$g=1$表示在节点 n 上查找目标资源成功；否则查找失败。u_1，$u_2,\cdots,u_n$ 表示对相应索引项进行查找的随机事件。随机变量 $u_1,u_2,\cdots,u_n\in\{0,1\}$。P2P 网络中节点 n 的邻居节点 n_1，$n_2,\cdots,n_l$ 表示在其上查找的随机事件，$n_1,n_2,\cdots,n_l\in\{0,1\}$。推理模型中节点 $u_1,u_2,\cdots,u_i$ 到节点 g 之间存在有向边，它表示历史查找请求索引项与新查找目标之间的语义匹配程度。模型中节点 $n_1,n_2,\cdots,n_n$ 到节点 $u_1,u_2,\cdots,u_n$ 之间存在有向边，它表示在节点 n 的邻居节点上命中 $u_1,u_2,\cdots,u_n$ 的可能性。推理模型如图 1 所示。

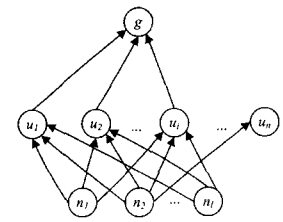

图 1 推理模型

定义 3 u_g 是随机变量 $u_1,u_2,\cdots,u_n$ 中与查找目标 g 具有语义相关的随机变量的集合，即

$$u_g=\{u_{1,g},u_{2,g},\cdots,u_{m,g}\} \tag{2}$$

式中，$u_{i,g}\in\{0,1\}$，$i\in\{1,2,\cdots,m\}$。

利于条件概率公式构建模型的推理过程：

$$\begin{aligned}P(g,n_i)&=\sum_{\forall u_g}P(g,n_i,u_g)\\&=\sum_{\forall u_g}P(g|n_i,u_g)\times P(n_i,u_g)\\&=\sum_{\forall u_g}P(g|n_i,u_g)\times P(u_g|n_i)\times P(n_i)\end{aligned} \tag{3}$$

式中，g 与 $n_i(i\in\{1,2,\cdots,l\})$ 相互独立，即 $P(g|n_i,u_g)=P(g|u_g)$。假定在给出 u_g 时，式(3)可以表示为：

$$\begin{aligned}P(g,n_i)&=\sum_{\forall u_g}P(g|u_g)\times P(u_g|n_i)\times P(n_i)\\&=P(n_i)\times\sum_{\forall u_g}P(g|u_g)\times P(u_g|n_i)\end{aligned} \tag{4}$$

把式(4)中 $P(n_i)$ 移到等式左边可得：

$$\frac{P(g,n_i)}{P(n_i)}=\sum_{\forall u_g}P(g|u_g)\times P(u_g|n_i)=P(g|n_i) \tag{5}$$

式(5)表示在节点 n 的邻居节点 n_i 上发现目标资源信息的可能性的概率表达式。其考虑与新查找存在目标语义相关的索引项在邻居节点上进行查找的影响度。

$P(g|u_g)$ 定义如下：

$$P(g|u_g)=\sum_{k=1}^{m}P(g|u_{k,g}),\text{如果 } u_{k,g}\in u_g,u_{k,g}=1 \tag{6}$$

式(6)计算历史查找请求与新查找请求的语义相似度，即 $P(g=1|u_{k,g}=1)$ 定义为 $Match(V_{k,g},V_g)$，其它情况值为零。

$P(u_{k,g}|n_i)$ 定义如下：

$$P(u_{k,g}|n_i)=\begin{cases}1, & \text{如果 } n_i \text{ 的索引包含 } u_{k,g} \text{ 且搜索成功}\\0, & \text{其他}\end{cases} \tag{7}$$

式中，$P(u_{k,g}=1|n_i=1)$ 表示在节点 n_i 上查找 $u_{k,g}$ 的命中可能性，用以推断在 n_i 上发现 g 的成功率。

2.3 网络距离

定义 4(网络距离) 网络中任意两节点间的往返时延，是评价网络性能的重要指标之一[12]。

此定义表达了节点间的往返时延，但 P2P 网络的灵活性较强，新节点的加入与退出均较为频繁，同时 P2P 网络产生的信息量较大。本文对经典网络距离定义加以补充，使得网络距离的表述更加符合 P2P 网络环境。

定义 5(网络距离) 平均响应时间占响应时间的变化趋势的比重。

上述定义不仅反应了节点间的响应时间，而且增加了对多次响应时间稳定性的考察，由此更能准确地选择最佳转发路径。

网络距离反映了节点 n_i 到节点 n_j 之间的平均响应时间及响应时间的变化趋势。平均响应时间主要代表返回目标结果速度的技术指标；反应时间变化趋势代表节点联系的通信稳定性的技术指标。可通过这两个指标来指导将查找请求转发给响应速度较快和稳定性更好的邻居节点。

文献[13]给出了综合平均响应时间和响应时间变化趋势的计算公式，并很好地兼顾两者之间的关系，建立了一个准确描述变化的计算模型且其处理方式简易，因此本文借鉴了其中的计算模型。以下是通过 N 次历史查询的响应时间经过相应计算处理得到的表达式。

$$T_n^{n_j}=\frac{(n-1)T_{n-1}^{n_j}+\left\lceil\frac{n}{3}\right\rceil t_n^{n_j}}{n-1+\left\lceil\frac{n}{3}\right\rceil} \tag{8}$$

式中，$T_n^{n_j}$ 表示节点 n_j 发送的前 n 次请求的平均响应时间；$t_n^{n_j}$ 表示向节点 n_j 发送的第 n 次请求的响应时间。时间换算计算方法如下所示：

$$t_n^{n_j}=\begin{cases}\dfrac{T}{t}, & t\leqslant T\\0, & t>T\end{cases} \tag{9}$$

式中，T 是设定的最长等待响应时间，t 是实际的响应时间。网络距离需综合考虑平均响应时间和反应时间变化趋势两个指标，以网络距离计算公式如下：

$$E_n^{n_j}=\frac{T_n^{n_j}}{C_n^{n_j}} \tag{10}$$

式中，$C_n^{n_j}$ 描述了节点 n_j 响应时间的变化趋势，兼顾了历史查找请求返回时间变化和最近的响应时间变化。$C_n^{n_j}$ 公式计算如下：

$$C_n^{n_j}=\frac{\sqrt{(n-1)C_{n-1}^{n_j}+\left\lceil\frac{n}{3}\right\rceil(T_n^{n_j}-t_n^{n_j})}}{n-1+\left\lceil\frac{n}{3}\right\rceil} \tag{11}$$

2.4 查找导向路径选择

对于节点 n_i，选择转发查找请求要聚合邻居节点查找目标的可能性和邻居节点的平均网络距离两个指标，以提高查找过程中的命中率，减少因盲目搜索而导致的通信信息冗余。假设向当前邻居节点 n_j 转发查找请求，必须满足以下条件：

$$P(n_j)^s=P(g|n_j)\times E_n^{n_j}\geqslant\alpha \tag{12}$$

式中，α 是设定的阈值。满足式(12)的邻居节点将作为转发路径选择的节点，这不仅可以较快获得返回结果，还可降低网络拥塞的概率。

3 算法描述

3.1 概率分发队列 (probabilistic distribution queue)

为实现协同查找，本文提出一种可扩展的快速定位方法——分发队列。其主要意图是融合查找过程中各中间节点获得的资源信息。因查找请求信息并无统一表达方式，对相同目标资源的查找请求描述方式也存在差异性，故在计算新查找请求与历史索引项的相似度时，不能简单回答“是”或“否”，而应返回由式(12)计算得到的相似度值。在 PDQ 方法中，称资源搜索消息为分发队列 DQ。若一个查找消息在节点 n_i 仅转发给某一个邻居节点 n_j，则称 DQ 从节点 n_i 跳转至节点 n_j；若节点 n_i 将查找消息转发至多个邻居节点，则称 DQ 在节点 n_i 处派出多个分发队列，此时，DQ 将保存在节点 n_i 并以 DQ_i 表示，分派至某节点 n_j 的分发队列以 $DQ_{i,j}$ 表示。如果查找请求发起节点 n_o 最初发出了 k 个查找消息，则认为 DQ_o 初始发出了 k 个分发队列，DQ_o 停留在查找始发节点 n_o。假定 DQ_o 发出的分发队列，经 x 步到达某节点 n，称其为 DQ_n，以下是 DQ_n 查找过程：

1)在中间节点 n，DQ_n 根据式(12)计算当前邻居节点与查找请求的相似度并得到当前局部最大相似度 $\mathrm{Max}_{l,n}$，如果该值大于 DQ_n 全局查找中的已知最大相似度 Max_g，则将此值通知给 DQ_o，DQ_o 通过比较各局部最大相似度，获得聚集后的全局最大相似度 Max_g。

2)DQ_n 向局部最大相似度 $\mathrm{Max}_{l,n}$ 对应邻居节点 z 转发查找消息，称其为第一类分发队列(以 $DQ_{n,z}^1$ 表示)。

3)在中间节点 n，DQ_n 向相似度不小于某一阈值的所有

邻居节点转发查找消息，该阈值等于全局最大相似度 Max_g 乘以衰减因子 $\S$ ($0<\S\leqslant 1$)，即向满足 $P(n_j)^s>=\S * Max_g$ 且与第一类分发队列不重复的所有邻居节点转发查找消息，称其为第二类分发队列(以 $DQ^2_{n,z}$ 表示)。

PDQ 过程如图 2 所示。

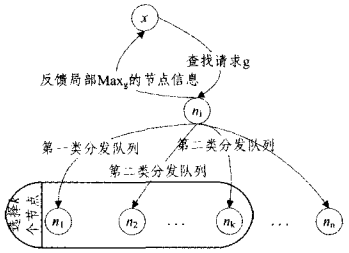

图 2　概率分发队列

在下一跳节点，子分发队列将重复父分发队列的上述操作。查找发起者将各分发队列返回的搜索结果按命中目标文件相似度值的大小排列顺序。

3.2　BFPDQ 算法

P2P 网络中某节点发起查找请求时，对查找信息进行预处理，利用文献[14]提出的 Soundex 算法对查找信息进行编码，使得目标信息表达更确切，并向所有邻居节点转发资源定位信息。邻居节点接收到查找请求后，先在本节点搜索匹配资源，若命中目标资源，则终止查找，并将查询结果返回上一级节点，插入到其索引表中；对预处理后的查找请求信息量化生成 Bloom Filter 向量，同时对索引表中条目进行相同处理生成相应的 BF 向量，还要综合考虑邻居节点的网络距离因素；查找请求 BF 向量与索引项 BF 向量集合逐一进行语义相似度计算，并选择 m 个语义相似度较高的邻居节点作为备选转发对象。计算由式(12)得到的综合导向概率，按 PDQ 方法选取 $k(k\leqslant m)$ 个邻居节点转发查找请求。

持续上述搜索过程，直至初始设定的 TTL 耗尽。节点得到重复查找请求时，丢弃查找请求包以防止多径叠加和冗余消息泛滥。

算法伪代码实现如下：

```
BFPDQ:
main( )
{
  //接受查找请求，并将其转化为 BF 向量
  V=PreOnQuery();
  //在本地搜索
  Result=SearchInLocal(V);
  If (Result){
  reutrn Fresult;//返回查找结果
  }else{
  //按照 BFPDQ 算法搜索
  Fresult=SearchOnBFPDQ(NodeID,V);
  //对返回结果按资源文件相似度大小排序显示结果
  SortBySimilarity(Fresult);
  }
}
PROCEDURE
  SearchOnBFPDQ(Node ID,BF V){
  Result=SearchInLocal(V);
  if(Result){
    return Result;
  }else{
    if(TTL>0){
    //节点 ID 的索引项表示成 BF 向量
    V1=PreQuery();
    //计算索引项条目与查找请求的相似度
    m=ComputeSimilarity(V,V1);
    if(m>a){//a 为阈值
      //计算节点的导向概率值
      f=ComputeProbability(m,NetworkDistance);
      //按 PDQ 算法派出分发队列搜索，
      Neighbor(k)=PDQ(f);
      for(i=0;i<k;i++){
        //各派出的分发队列继续在邻居节点上搜索
        SearchOnBFPDQ(ID,V);
        //插入新完成查询操作索引信息
        InsertRelatedIndex(ID,V,time,NID,Flag);
      }
    }
    TTL--;
    }
  }
}
```

对接收到查找消息的中间节点的处理过程与查找请求发起节点的不同之处在于，无需再次对查找请求进行 BF 向量处理，即整个搜索过程中仅传递存储代价较小的 BF 向量，小数据信息不易导致网络传输拥塞。

3.3　分析

一个优秀的 P2P 系统必须拥有较好的抗干扰性和容错性[15]。新节点初始加入 P2P 网络时，连接至节点度数较高的节点，并建立邻居关系。同时，P2P 具有较强的灵活性，节点增加和退出 P2P 系统均较为频繁。为了减轻系统由此造成的通信负载，节点退出系统时无须通告其邻居节点，由其自行感知判定。在每个节点上维护两张索引表，一张存储其邻居节点的相关信息，其表项有节点 ID、节点状态、与本节点的网络距离等信息；另一张存储其邻居节点历史查找命中结果信息，其包含查找请求信息、查找信息 BF 向量、响应时间、邻居节点 ID 和命中与否等信息。查找路径上的节点在查找结果返回后，更新索引表中的相关索引项。为节约节点的内存空间，须控制索引表大小。索引表管理采用 LRU(least recently used)算法，删除长时间不使用的索引项，以限制其无限的增长。一般情况下，索引表项越多，保存的历史导向信息越多，越有利于提高查找成功率[11]。

4　仿真实验与分析

本文使用 BRITE[16] 网络拓扑模拟器，获得采用 BA 模型构造的具有幂律和小世界特征的 P2P 网络[17]。通过调整拓扑模型参数构建不同规模的 P2P 网络。为考查 BFPDQ 算法的性能，通过仿真实验将其与随机游走(Random walks)算法[18]、泛洪(Flooding)算法[19]和 BNS 算法[11]在覆盖率、查全率及平均响应时间等性能指标上进行比较。覆盖率定义为某一查找请求可到达的节点个数占节点总数的比例；查准率为查找成功的次数占查找总数的比例；响应时间为查找发起者第一个接收到命中返回结果的时间间隔；平均响应时间为多次查找请求的响应时间的平均值。

由 BRITE 产生的节点总数为 100，边数为 368 的 P2P 网

络拓扑如图 3 所示。

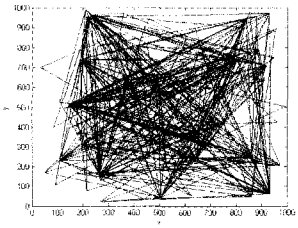

图 3 仿真 P2P 网络

4.1 仿真实验结果

无结构化 P2P 网络中节点分布松散，各节点上的资源分布也不同，因此无结构化 P2P 网络的资源定位策略须具有较强的动态适应性和较高的定位效率。不同系统尺度、相同仿真条件下，比较本文 BFPDQ 算法、Flooding 算法、Random Walks 算法和 BNS 算法在覆盖率和响应时间上的差异，以验证本文算法的可行性。在不同系统规模中，初始设定 TTL 值为 6，对网络中的节点可以设置状态位，标识节点加入或退出网络。在 BNS 算法和 Random Walks 算法中，Walks 数目 k 设置为 6[20]。

由图 4 可知，BNS 算法和 Random walks 算法的系统覆盖率较低，绝大部分情况下不足节点数的 20%；而 BFPDQ 算法系统覆盖率为 40%左右，保证了查找请求尽可能多地遍历与目标资源相关的节点。Flooding 算法的系统覆盖率较高，同时遍历绝大部分节点也造成大量通信开销；其产生的巨大信息量，使得整个网络拥塞严重，降低了响应时延。除覆盖率外，本文还考察了各算法的查准率。由于单一统计查准率的差异较大，图 5 是节点总数为 2000 时多次查询的查准率，图中每个点的值是连续发出 100 个查找请求的平均查准率。由图 5 可见，连续多次发出查找请求后，BFPDQ 能稳步提高查准率，较 BNS 算法平均高 10%。与 BFPDQ 相比，Flooding 算法查准率仅为 10%左右，即查询消息的利用率较低，绝大部分为无用查询消息。从结果来看，Random Walks 算法查准率较低且搜索成功率也不稳定，这主要是因为该算法采用的是随机路由转发模式。综上，BFPDQ 的查询消息实用率较高，减少了盲目查找的冗余信息，具有较高的查准率。

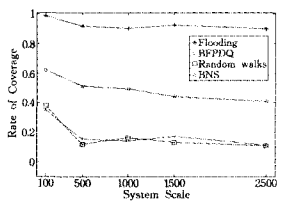

图 4 系统规模与系统覆盖度的关系

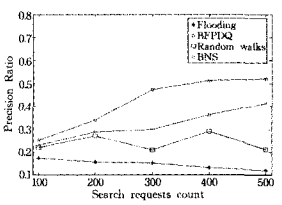

图 5 系统规模与查准率的关系

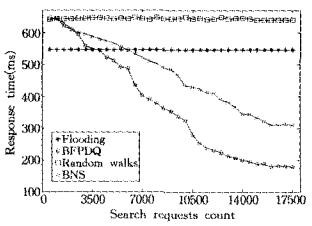

图 6 2000 个节点的仿真情况

BFPDQ 通过每个节点对系统历史查找信息的处理达到降低查找的深度和广度，以提高命中效率，降低响应时间。与其它定位策略相比，每个节点的计算处理时间相对较长。图 6 是 P2P 系统在运行的过程中，节点总数在 2000 个时的响应时间效果对比图。单次命中的响应时间随机性较大，图中每个点的值是连续发出 500 个查找请求的响应时间的均值。

由上述结果可知，在 P2P 网络中，若能充分发挥每个节点的能动性，则查找命中率将随查找次数的增加而大幅度提高，并逐渐趋于稳定。P2P 网络中“热点”资源与查找需求都具有一定的“兴趣”特征，服从 Zipf 分布[21]。实验结果也表明，BFPDQ 虽在 P2P 系统开始运行的较短时间内平均响应时间稍长，但系统运行稳定后，查询命中率快速提升。同时，对于“兴趣”资源，查询时间可缩短 60%以上。

图 4—图 6 给出了不同评价指标的对比图，BFPDQ 虽系统覆盖率表现一般，但在消息查准率和平均响应延时上有较好表现，在产生较低消息量和覆盖率的同时，能快速提高响应速度和查准率。

结束语 本文提出一种新的 P2P 网络快速定位算法。该算法充分发挥了各节点的能动性，综合考虑了历史查找索引信息和邻居节点间的网络距离信息，并以此为新查找请求提供了较可靠的转发路径信息；运用分发队列方法选择转发路径，实现了多分发队列的协同、并行搜索，还对响应结果按各匹配相似度大小排列返回结果。

参 考 文 献

[1] ipoqu[OL]. http://www.alexa.com/siteinfo/ipoqu.com

[2] Kalogeraki V, Gunopulos D, Zeinalipour-Yazti D. A local search mechanism for peer-to-peer networks[C]// Proc. of the 11th ACM Conf on Information and Knowledge Management. New York: ACM Press, 2002: 300-307

[3] Bloom B H. Space/time trade-offs in hash coding with allowable errors[J]. Communications of the ACM, 1970, 13(7): 422-426

[4] Rhea S C, Kubiatowicz J. Probabilistic location and routing[C]// Proc. of the IEEE INFOCOM 2002. Washington: IEEE Computer Society, 2002: 1248-1257

[5] Hebden P, Pearce A R. Data-centric routing using bloom filters in wireless sensor networks[C]// Proc. of the 4th Int'l Conf on Intelligent Sensing and Information Processing. Washington: IEEE Compu-ter Society, 2006: 72-77

[6] Kumar A, Xu J, Zegura E W. Efficient and scalable query routing for unstructured peer-to-peer networks[C]// Proc. of the IEEE INFOCOM 2005. Washington: IEEE Computer Society, 2005: 1162-1173

[7] 郭得科. 基于 Kautz 图和 Bloom 滤波的对等网络研究[D]. 长沙：国防科技大学，2008

[8] 朱桂明，郭得科，金士尧. 基于副本复制和 Bloom Filter 的 P2P 概率路由算法[J]. 软件学报，2011，22(4)：773-781

[9] 张一鸣，卢锡城，郑倩冰，等. 一种面向大规模 P2P 系统的快速搜索算法[J]. 软件学报，2008，19(6)：1473-1480

[10] Tsoumakos D, Roussopoulos N. Adaptive probabilistic search for peer-to-peer networks[C]// Proc. of the Third International Conference on Peer-to-Peer Computing. Sweden: IEEE Computer Society, 2003: 102-109

[11] 钱宁，吴国新，赵生慧. 基于贝叶斯网络的无结构化 P2P 资源搜索方法[J]. 计算机研究与发展，2009，46(6)：889-897

（下转第 94 页）

改善有限精度造成的短周期效应提供了解决的可能性。

预处理后的超混沌序列是非常理想的伪随机序列，密钥集合中不存在大量的弱密钥(密钥与输出之间存在超出一个好密码所应具有的相关性)和等效密钥(由一个密钥与输出能推导出另一个密钥与输出)。量化处理采用一种通过均分区间一次迭代生成多比特二进制序列的方法，提高了生成序列的速度和随机性。预处理和量化过程是一种不可逆变换，破译者无法根据截取的密文去重构产生序列密码的混沌系统动力学模型、初始状态等，从而基于相空间重构的攻击、基于回归映射的攻击和基于混沌同步的分析方法都将不起作用，这一加密算法有很高的抗破译能力。

本文中的超混沌加密属于流加密，对分组加密的攻击方法是无效的。同时，对选择明文/密文攻击方法，由于混沌的单向性和混沌信号的迭代处理，异或操作后密钥流的推断几乎不可能。该加密算法没有 S-box 空间，临时变量也比较少，而且通过循环产生密钥流，循环过程中需要寄存的变量有限，运行时占用的空间很少，另外，加密和解密过程是可以重用的，这样所占用的空间就大大缩小。

采用 3DES 加密算法，解决了 DES 算法密钥太短的问题，同时也克服了 DES 中存在的弱密钥和半弱密钥的缺陷[14]；另外，将超混沌序列加密后的密文作为 3DES 加密级的输入，使得 3DES 算法的输入和输出间不存在唯一的明文-密文对[6]，从而提高了保密性。在信道中传输的密文是经过双重加密的，可以对抗文献[15]中提出的对混沌系统的识别和对初始值确定的破译方法，从而发挥了两者各自的优势，提高了加密的复杂度。

结束语 提出一种基于超混沌加密技术和 3DES 加密技术相结合的级联加密技术方案。对超混沌实值序列进行了预处理和量化处理，预处理改善了超混沌系统在有限精度实现时的短周期现象，得到了自相关特性良好的输出序列；量化处理提高了序列密码产生的速度和随机性。基于超混沌系统设计的加密算法具有很大的密钥空间、较好的安全性和较强的抗破译能力。该方案利用了超混沌加密技术和 3DES 两者各自的优势，比任意一种加密技术单独使用时的保密性能都好。将这种级联加密技术应用于 Outlook 2007，实现了电子邮件内容的加解密。实验结果表明，这种保密技术可以在网络信息传输中很好地完成加密和解密过程，安全性能好，应用方便简单。

参 考 文 献

[1] Zhou Hong, Ling Xie-ting. Problems with the chaotic inverse system encryption approach[J]. IEEE Trans on Circuits and Systems I: Fundamental Theory and Applications, 1997, 44(3): 268-271

[2] Kanso A, Smaoui N. Logistic chaotic maps for binary numbers generations[J]. Chaos Solitons & Fractals, 2009, 40(5): 2557-2568

[3] Wang Xing-yuan, Yang Lei, Liu Rong, et al. A chaotic image encryption algorithm based on perceptron model[J]. Nonlinear Dynamics, 2010, 62(3): 615-621

[4] Cruz-Hernandez C, Lopez-Gutierrez R M, Aguilar-Bustos A Y, et al. Communicating encrypted information based on synchronized hyperchaotic maps[J]. International Journal of Nonlinear Sciences and Numerical Simulation, 2010, 11(5): 337-349

[5] 赵耿，方锦清. 现代信息安全与混沌保密通信应用研究的进展[J]. 物理学进展，2003，23(2)：212-255

[6] 丘水生，陈艳峰，吴敏，等. 混沌保密通信的若干问题及混沌加密新方案[J]. 华南理工大学学报：自然科学版，2002，30(11)：75-80

[7] 刘明华，冯久超. 一个新的超混沌系统[J]. 物理学报，2009，58(7)：4457-4462

[8] 张池平，施云慧. 计算方法[M]. 北京：科学出版社，2002：371-384

[9] Wang Ying, Han Chun-yan, Liu Yuan-yi. A parallel encryption algorithm for color images based on Lorenz chaotic sequences [C]//Proceeding of the 6th World Congress on Intelligent Control and Automation. Dalian, China, 2006: 9744-9747

[10] 姚洪兴，李萌，杜贤利. 一种基于超混沌的二值序列生成方法[J]. 科学技术与工程，2008，8(13)：3508-3512

[11] 刘金梅. 多个混沌系统构造密码算法的理论及应用研究[D]. 广州：华南理工大学，2009

[12] Shannon C E. Communication theory of secrecy systems[J]. Bell Systems Technical Journal, 1949, 28: 656-715

[13] 潘勃，冯金富，陶茜，等. 基于超混沌映射和加法模运算的图像保密通信方案[J]. 计算机科学，2009，36(8)：273-275

[14] 宋震，等. 密码学[M]. 北京：中国水利水电出版社，2002

[15] Sobhy M I, Shehata A-E R. Methods of attacking chaotic encryption and countermeasures[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. 2001, 2: 1001-1004

(上接第 61 页)

[12] Shavitt Y, Tankel T. Big-bang simulation for embedding network distances in Euclidean space[J]. IEEE/ACM Transactions on Networking (TON), 2004, 12(6): 993-1006

[13] 黄永生，孟祥武，张玉洁. 基于社会网络特征的 P2P 内容定位策略[J]. 软件学报，2010，21(10)：2622-2630

[14] Zaharia M A, Chandel A, Saroiu S. Finding content in file-sharing networks when you can't even spell[OL]. http://research.microsoft.com/workshops/PTPS2007/paper/ZahariaChandelSaroiuKeshav.pdf, 2007

[15] Gaeta R, Sereno M. Generalized Probabilistic Flooding in Unstructured Peer-to-Peer Networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2011, 22(12): 1-8

[16] BRITE[OL]. http://www.cs.bu.edu/brite

[17] Jovanovic M A. Modeling large-scale peer-to-peer networks and a case study of Gnutella[M]. University of Cincinnati, USA: 2001

[18] Lv Q, Cao P. Search and replication in unstructured peer-to-peer networks[C]//Proc. of the 16th ACM Int'l Conf on Supercomputing(ICS 2002). New York: ACM Press, 2002: 254-261

[19] Gnutella[OL]. http://www.limewire.com/developer/gnutella_protocol_0.4.pdf

[20] Jiang S. LightFlood: Minimizing redundant messages and maximizing scope of Peer-to-Peer search[J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(5): 601-614

[21] 孙新，刘玉树，刘琼昕. 具有位置感知和语义特征的 P2P 网络模型[J]. 电子学报，2010，38(11)：2606-2610