ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.2, February 2006, pp.339−348
DOI: 10.1360/jos170339

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于自组织聚类的结构化 P2P 语义路由改进算法*

刘 业+, 杨 鹏

(计算机网络和信息集成教育部重点实验室(东南大学),江苏 南京 210096)

An Advanced Algorithm to P2P Semantic Routing Based on the Topologically-Aware Clustering in Self-Organizing Mode

LIU Ye+, YANG Peng

(Key Laboratory of Computer Network and Information Integration, Ministry of Education (Southeast University), Nanjing 210096, China)

+ Corresponding author: Phn: +86-25-83791823, E-mail: yliu@seu.edu.cn, http://www.seu.edu.cn

Liu Y, Yang P. An advanced algorithm to P2P semantic routing based on the topologically-aware clustering in self-organizing mode. *Journal of Software*, 2006,17(2):339–348. http://www.jos.org.cn/1000-9825/17/339.htm

Abstract: Structured P2P Networks create a virtual topology on top of the physical topology. The only relation between the two layers is the hashing algorithm, which makes the node's logical ID independent of its physical location. By analyzing the Hash function, some novel logical connections among the destination node, the traditional semantic routing relay node sequence, and the ID of the clustering neighboring nodes are found. In this paper, the SCSRAA (self-organizing clustering semantic routing advanced algorithm) is resented to improve the efficiency of semantic routing. Since the clustering nodes only have local views in self-organizing mode, some rules are proposed for a node to learn other nodes' physical location. The SCSRAA's routing algorithm is described completely. Simulations have verified that the method can improve the semantic routing efficiently.

Key words: P2P; semantic routing algorithm; physical topology; self-organizing; clustering

摘 要: 结构化 P2P 网络是构建于物理网络拓扑之上的一层 Overlay 网络,两层之间的唯一联系是 Hash 散列函数,这种 Hash 关系使得节点的逻辑 ID 号与物理位置之间不存在任何联系.从分析 Hash 散列函数的性质入手,归纳出目的节点、传统(chord)语义路由中继节点序列、聚类邻居节点集三者之间的逻辑关联特性,并将其应用于所提出的基于自组织聚类的语义路由改进算法 SCSRAA(self-organizing clustering semantic routing advanced algorithm)中,从而达到提高语义路由效率的研究目的.针对自组织模式下聚类节点仅存在局部视图的特性,详细讨论了聚类算法及节点获取其他节点物理位置信息的各种规则,给出了 SCSRAA 路由算法详尽的描述及理论分析.仿真实验表明,该算法具有较强的语义路由效率提升能力.

关键词: P2P;语义路由算法;物理拓扑;自组织;聚类
中图法分类号: TP393 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant No.60573133 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.2003CB314801 (国家重点基础研究发展规划(973))

Received 2004-10-08; Accepted 2005-03-11

P2P 网络语义路由效率的研究是推动 P2P 网络进一步发展的关键问题.结构化 P2P 网络的本质是在物理网络之上架构一层 Overlay 网络,在网络拓扑中,节点间是逻辑相邻关系;并通过自组织的方式,依据相应的相邻关系来建立路由表、对等节点的物理位置与逻辑 ID 号 $node_{id}$ 之间的哈希(Hash)散列关系[1],使得两者之间不存在任何联系,即在 P2P 语义路由过程中,报文在节点间传递经常会走一些冤枉路.在此背景下,如何在语义路由的过程中适当考虑加入物理位置的因素,从而提高整个 P2P 网络的语义路由效率,是当前 P2P 网络研究的热点.

Brocade[2]在结构化路由算法的基础上再叠加一层,在 P2P 网络中设置多个超级节点,超级节点维护本管理区域内部所辖的节点列表.语义路由时,报文先从源节点发送到本地超级节点,由超级节点再转发到目的节点.该类结构导致了网络中节点的功能是不对等的,对超级节点的性能提出了更高的要求;同时,需要添加额外的协议处理"节点→超级节点→[(下一跳节点→超级节点)]*→目标节点"序列间报文的转发.依据 P2P 网络中节点间相对物理距离对节点 $node_{id}$ 进行聚类的研究,方案又可细分为集中式聚类和分布式聚类.文献[3]提出了一种集中式聚类算法.该算法使用聚类服务器完成对等节点间的聚类,每个聚类团内有一个代理节点,聚类服务器负责管理所有对等节点的聚类以及聚类团中代理节点的相关状态参数等.集中式聚类算法能有效地对节点按照物理位置进行聚类,但是,聚类服务器本身就是单点瓶颈.文献[4]提出了一种分布式聚类算法.该算法首先在 Internet 上设定一系列参考点,通过 Ping 方式得到对等节点与这些参考节点的距离,根据距离对参考节点序列进行排序,排序结果作为聚类的参数,并依据聚类的结果指导 Overlay 网络的构造,即 $node_{id}$ 值的生成,使得 $node_{id}$ 相邻的节点同时也是物理拓扑相邻的.文献[4]中具体阐述了 CAN 网络的 $node_{id}$ 与物理位置关联的过程.该算法具有自组织聚类的特征.但是,该算法将 $node_{id}$ 的生成与物理位置相关联,不适用于 Chord[5],Pastry[6],Viceroy[7]等利用 Hash 函数生成 $node_{id}$ 的 P2P 网络.

当前,在基于物理地址的提高 P2P 语义路由算法效率的研究中,大多是考虑设置超级节点,或是在聚类后从聚类的节点中选取某个性能、带宽等综合条件最好的节点作为超级节点.这样,一方面原本分散的网络流量又回归至集中,使得超级节点成为单点失效瓶颈.尽管国内外也有人对该类节点失效问题做了一定的研究,但终归使得对等网络中的节点的关系不再对等;另一方面,从聚类的节点中选取超级节点,由于其本身就处于一个完全分散、缺乏集中控制的松耦合环境下,如何选取令牌是一个值得深究的问题.而考虑将地理位置信息集成到 $node_{id}$ 生成值中的方式,一方面使得 $node_{id}$ 在值域空间内的分布不均衡,从而导致了 P2P 应用中资源层资源并不能均衡地负载到各个 $node_{id}$ 节点上,更糟的情况是,致使部分节点出现过载的情况;另一方面 $node_{id}$ 一旦生成,在其生存期内就不能更改,这种方式并不能动态地适应移动节点或者网络物理拓扑本身的动态变化.

本文提出的基于自组织聚类的语义路由改进算法(self-organizing clustering semantic routing advanced algorithm,简称 SCSRAA),从建立相应的 P2P 网络节点物理位置参照系入手,依据节点坐标间的相对距离,来自组织聚类节点的物理邻居节点,并且通过将聚类节点的 Hash 输出值特性与传统基本语义路由过程所遍历的 $node_{id}$ 序列的特性相关联,归纳出目的节点、传统(chord)语义路由中继节点序列、聚类邻居节点集三者之间的逻辑关联特性,对 SCSRAA 路由性能理论分析及推导,均是建立在逻辑关联特性的基础之上的.SCSRAA 算法既不修改原有 P2P 网络的 $node_{id}$ 的值,也不在网络中设置超级节点,在提高路由效率的同时,保证了网络节点的对等性,因此,该解决方案可作为构造实用、可操作的结构化 P2P 语义路由算法中,语义路由效率提升的可选模块.

本文将从理论分析和仿真实验两方面对该算法进行较为详尽的介绍.本文第 1 节简要介绍结构化 P2P 网络语义路由基本过程及其相关概念.第 2 节是 SCSRAA 算法的理论基础、相关定义、算法描述及其理论分析.第 3 节是实验及其结论.最后是总结.

1 结构化 P2P 网络语义路由基本过程

由于本文所做的改进工作建立在 Chord 基本路由算法的基础上,因此,首先需要对语义路由的概念及 Chord 基本语义路由过程作一个简要的介绍.

1.1 语义路由

结构化拓扑的 P2P 网络实现语义路由的方式是:(1) 将具体的对象通过一定的映射算法得到它在系统中的

唯一标志号,通常有两种方式,即随机产生或者根据 SHA-1 或 MD5 哈希函数得到一个 $object_{id}$,$node_{id}$ 为 P2P 网络中节点的唯一标识符;(2) 将 $object_{id}$ 与节点 $node_{id}$ 之间建立 *assign* 映射关系,多采用截取前缀或后缀的方式,$node_{id}$←*assign*($object_{id}$);(3) 将 $node_{id}$ 组织成一个结构化的拓扑(环型 chord、蝴蝶网型 viceroy、二维平面结构 CAN 等),每个节点通过一定的规则与 P2P 网络中其他一小部分节点的 $node_{id}$ 建立连通关系.这样,当查询报文从源节点到达目的节点时,我们称其为完成了语义路由,或者称其为应用层路由.

1.2 Chord基本语义路由过程

Chord 是 MIT 提出来的基于 P2P 网络的信息资源定位、查找模型,节点通过 $node_{id}$ 构成一个弦环拓扑结构.在模型中,节点与其他节点通过 $node_{id}$ 建立一种幂(2^i)相邻关系来建立路由表.例如,系统中节点为{*N*1,*N*8,*N*14,*N*21,*N*32,*N*38,*N*42,*N*48,*N*51,*N*58},*N*8 节点维护的路由表如图 1 所示.查找文件 123.*mp*3 的定位信息的语义路由步骤如下:*OLM*=*Hash*(123.*mp*3)=*K*50,*K*50 在 Chord 模型中放置于节点 *N*51 上,源端 *N*8 发出获取文件 123.*mp*3 的定位信息的报文,根据路由表,通过幂相邻关系逐步逼近的语义路由过程,其路由序列为〈*N*8 $\xrightarrow{N8+32}$ *N*42 $\xrightarrow{N42+8}$ *N*51〉,从源端路由到目的端的逻辑跳数为 2.详细的路由算法可查阅文献[6].由表 1 同时可以看出,当 $node_{id}$ 在值域区间[0,2^n]中不饱和时,在节点维护的路由表中,多个幂相邻关系(finger 函数)将指向同一个 $node_{id}$.因此,语义路由的平均逻辑跳数只与节点规模有关,而与值域范围 2^n 无关,即平均逻辑跳数等于 $\frac{1}{2}\log_2 N$ (其中 *N* 为 P2P 网络中节点规模).

*N*8+1	*N*14
*N*8+2	*N*14
*N*8+4	*N*14
*N*8+8	*N*21
*N*8+16	*N*32
*N*8+32	*N*42

Fig.1 The Finger table entries for *N*8 in Chord

图 1 Chord 模型中节点 *N*8 的路由表

2 SCSRAA 语义路由模型

2.1 SCSRAA语义路由算法的基本思想

Chord 路由过程是 $node_{id}$ 值逐步逼近的过程.SCSRAA 算法的基本思想是:在 P2P 网络路由转发过程中,用聚类邻居中的节点来替代 Chord 路由算法中的中间转发节点,进而完成 $node_{id}$ 的一次逼近过程.在该过程中,存在一定的概率以 *Cluster_Lan_davg*(见定义 4)的物理路径代价替代 *Wan_Davg*(见定义 3)完成逻辑一跳,从而提高路由效率.

SCSRAA 语义路由模型分为两个相对独立的模块:自组织聚类模块和语义路由算法模块.自组织聚类模块负责按节点的物理位置对节点进行聚类,同时构建及维护本节点的聚类邻居节点路由表.语义路由算法模块通过比较目标节点 $node_{id}$、Chord 路由表中的{$node_{id}$}、聚类邻居路由表中的{$node_{id}$}三者间幂相邻关系来决定当前节点语义路由的下一跳.下面两节分别讨论两个模块的实现思想.

2.2 自组织聚类物理邻居节点模块

定义 **1**(物理拓扑参照系). 为了便于确定对等节点在实际网络中的相对物理分布位置,需要在 Internet 中建立物理拓扑参照系.考虑 v 维,每一维设立一个坐标原点,P2P 网络中的节点通过 Ping 这些坐标原点,从而得到该节点在每一维的坐标值.维数 v 越大,且每一维的坐标原点均匀地散布在 Internet 中,则坐标描述节点的物理位置越精确.

定义 **2**(距离函数). 由于坐标各维特征变量的量纲是一致的——同为网络节点与参考原点的网络传输延时时间,故采用欧氏距离公式作为距离函数,即聚类评判函数.

$$D_v(node_x, node_y)=\left(\sum_{i=1}^{v}\left|X_i-Y_i\right|^2\right)^{1/2},\ node_x\text{坐标为}\langle X_i\rangle, node_y\text{坐标为}\langle Y_i\rangle \quad (1)$$

定义 **3**(***Wan_D_avg***). Chord 路由算法中逻辑一跳的物理距离 *Wan_D*$_{avg}$,为整个 P2P 网络节点两两平均距离.该值与 P2P 网络的规模、网络的延时、网络的物理拓扑等因素有关.

定义 **4**(***Cluster_Lan_davg***). 在节点自组织聚类集合中,节点间的平均距离 *Cluster_Lan_davg*,根据距离函数,在网络拓扑一定的情况下,它与节点自组织聚类过程中的聚类规模有关.

为方便书写,以下记 *Wan_D*$_{avg}$,*Cluster_Lan_davg*,分别为 *D*,*d*.

2.2.1 SCSRAA 聚类算法描述

传统聚类算法的研究需要系统的全局视图,或者全局范围内的反馈机制,并不适用于仅有局部视图的自组织 P2P 网络中的邻居节点的聚类模型.在 SCSRAA 聚类模型中,提出了非对称聚类的概念,即节点收集的邻居节点集中的元素满足非对称特性,即对于节点 $node_a$,$\exists node_b\in\Omega_a$.此时,$node_a$ 不一定存在于节点 $node_b$ 的邻居节点集合 Ω_b 中.

算法 **1**. 聚类算法.

// 初始化邻居节点集 Ω_k,节点最大聚类规模 *Cluster*Max$_k$.

$\theta_k\leftarrow\theta_{default}$; $\Omega_k\leftarrow\{node_k\}$; *Cluster*Max$_k\leftarrow$*Cluster*Max$_{default}$; m_k 为当前聚类规模;

(1) WHILE ($node_k$ is ACTIVE) DO

(2) 按规则 1~规则 3 对应的行为获得 P2P 网络中其他节点物理位置坐标;

(3) if $D_v(node_k,node_j)<=\theta_k$

{

(3.1) $\Omega_k=\Omega_k+\{node_j\}$;

(3.2) 根据规则 3 发布自身及其邻居节点的坐标信息给 $node_j$;

(3.3) 若此时聚类数 m_k 超过 *Cluster*Max$_k$,则移除距离 $node_k$ 最远的点,调整阈值 θ_k,
同时,使得 $\theta_k=D_v(node_k,node_j)_{\max}$,$node_j\in\Omega_k$;

}

(4) $\forall node_j\in\Omega_k$, if ($node_i$ is DEAD) // 定期查询 Ω_k 中元素是否还活着

{

$\Omega_k=\Omega_k-\{node_i\}$;

$\theta_k=\theta_{default}$;

}

(5) ENDDO

规则 **1**(泛洪规则). 自组织管理模式下,定期通过报文扩散方式将节点自身的物理坐标信息向 P2P 网络中其他节点进行发布.

规则 **2**(带内规则). 在报文中嵌入源节点的物理坐标信息,在报文转发过程中,源节点的物理位置坐标值随报文一并转发,路由中继节点获得源节点的物理坐标信息.该过程中完成新加入节点物理坐标信息的局部发布.

规则 **3**(反向规则). 节点 *a* 若欲将节点 *b* 作为邻居集合中的元素,则 Ω_a 中的节点成为节点 *b* 的邻居节点的几率是非常大的.

关于 $\theta_{default}$, *Cluster*Max$_{default}$ 的设置,分为如下两类情形:

I. 节点处于对 P2P 网络全局统计信息完全未知的自组织模式下:$\theta_{default}\leftarrow$Max*Integer*.聚类算法初期是一个慢收敛的过程,极端的情况是出现形如 $d>D$,随着时间的推移,节点不断获取其他节点位置信息,聚类算法最终使得 $d<<D$,从而提高语义路由效率.*Cluster*Max$_{default}\leftarrow n*(1+\beta(R_c,R_b))$,其中,$n$ 为本节点维护的 Chord 路由表的长度,R_c 为节点的计算资源,R_b 为带宽资源;$\beta(R_c,R_b)\geq 0$.当节点闲置资源比较多时(例如,拥有较多计算资源、带宽资源等),则可以增加聚类规模,为提高整个 P2P 网络语义路由效率多做贡献,语句(4)对聚类节点的维护使得对带

宽资源有一定的占用.因此,节点的带宽资源通过β函数因子也约束着聚类的规模.

II. 当某个面向特定领域应用的 P2P 网络在运行一 wh 阶段以后,存在如下事实:尽管 P2P 网络中节点及其相关的物理拓扑位置的变化曾出现非线性升降趋势,还受到未知随机因素的干扰,但从宏观统计上分析,由于领域应用的网络用户行为存在一定的周期性变化,因而通过相应的建模分析,可以得出 d,D 间的统计函数关系 $g(m,N)$.通过该经验公式指导节点自组织聚类行为,对提高聚类效果是高效的、可行的.这种模式使得节点在聚类初期有较好的收敛值,对于节点生命周期很短的情形比较适用.

2.2.2 节点自组织聚类算法评价

获得 P2P 网络中其他节点物理坐标信息的过程,是在网络性能与网络功能中寻求较优平衡的过程.规则 1 以牺牲网络性能来获得充分的 P2P 网络节点的物理位置坐标信息,节点物理坐标的获取由专门的协议模块周期完成;规则 2 在转发报文中获得所嵌入的源节点的物理位置坐标,其收敛速度较缓慢,另需增加额外的协议来处理报文;规则 3 是一个高效率的反馈过程.对于基于 P2P 网络架构的特定领域应用,需要根据其对网络性能的要求,折衷考虑节点的物理位置坐标信息的发布规则.

语句(3)是 d 不断调优的过程,其收敛速度取决于语句(2)所采纳的规则.语句(4)能够动态地适应节点各种类型的离开、失效行为.算法 1 使得各个节点所聚类邻居节点集的 d 不同,因此使得 SCSRAA 模型中聚类邻居节点集中的元素是非对称的.节点根据语义路由过程中的反馈,动态地修正自身的阈值参数.这种方式适合于节点的自组织管理方式,且易于部署.

P2P 网络节点自组织聚类算法是一个自适应算法,由于联入 P2P 网络的节点的物理拓扑是动态变化的,同时,广域网环境中的网络状况亦是动态变化的,通过 ping 方式得到的节点物理坐标也是动态变化的,通过该聚类算法是能够动态地适应这样的环境变化的.

2.3 SCSRAA语义路由算法模块

在结构化 P2P 网络模型中,节点的 $node_{id}$ 由哈希函数输出得到,函数输入考虑为节点的 IP 地址;节点 $node_{id}$ 值的分布满足哈希函数输出的特征.下面,首先给出 Hash 散列函数[8]的一些与本模型相关的性质.

性质 **1**. 若对于关键字集合中的任一关键字,经哈希函数映射到地址集合中任何一个地址的概率是相等的,即 Hash 函数的输出均匀地散布在值域区间上.对应于 P2P 网络,N 个节点的 $node_{id}$ 的数学期望均匀地散布在 $[0,2^n]$的 N 个连续子区间中.

性质 **2**. 哈希散列函数至少是弱无碰撞性的,即不同关键字的哈希散列值相同的情况很少.Chord 模型中使用的 SHA-1 是美国国家标准与技术研究所于 1995 年公布的哈希散列函数标准.SHA-1 是一个强无碰撞散列函数,对于所输入的任何微小的改变都会引起输出的很大差异.

定理 **1**. 从统计的角度来分析,节点的聚类邻居集合$\Omega_k=\{node_j,j\in[1,m]\}$中,其 $node_j$ 的值均匀分布在值域 $[0,2^n]$上的 m 个顺序子区间中.

证明:节点 $node_j$ 的生成函数为 $Hash(node_j.IPAddr)=SHA\text{-}1(node_j.IPAddr)$,对于聚类集合中的元素,因为其物理位置邻近,使得其 IP 地址也可能是临近的.由性质 2 可知,SHA-1 是一个强无碰撞散列函数,对于输入的任何微小的改变都会引起输出的很大差异,即便是 IP 地址连续的 m 个节点,其 $node_{id}$ 的值仍是均匀分布在$[0,2^n]$的 m 个顺序子空间上.推广到更一般的情况,得证. □

2.3.1 SCSRAA 路由算法描述

下面给出基于自组织聚类的路由算法.为此,我们先对算法中使用的一些主要符号作一简单说明:假设当前节点 $node_k$,运行 SCSRAA 聚类算法后,当前其路由表中维护的邻居节点集合 Ω_k 为$\{node_i,i\in[1,m_k]\}$,其中,$node_k\in\Omega_k$,Chord 路由表中的路由转发下一跳节点集合为$\{node_{ChordRoute},node_{ChordRoute}=finger(node_k+2^j),j\in[1,n]\}$,目标节点为 $node_{Destiny}$.

算法 **2**. 路由算法.

(1) 若 $node_{Destiny}\in\Omega_k$,则{选择 $node_{Destiny}$ 作为路由下一跳;break;}

(2) 若 $node_{Destiny}\notin\Omega_k$ && $node_{Destiny}\in\{node_{ChordRoute}\}$,则{选择 $node_{Destiny}$ 作为路由下一跳;break;}

(3) 若$\Omega_k-\{node_k\}\neq\varnothing$,则检查邻居集中是否存在$node_{id}$的值落在下一逼近区间$\alpha$中的节点,若存在,则选取一个使得$\|node_{destiny}-node_{id}\|_{\min}$作为路由选择的下一跳;若不存在,则按 Chord 路由算法选择下一跳.

定义 5(下一逼近区间α).

$$\alpha=\begin{cases}[(node_{Destiny}-node_k)/2,node_{Destiny}], & \text{if } (node_{Destiny}>node_k)\\ [(2^n+node_{Destiny}-node_k)/2,node_{Destiny}], & \text{if } (node_{Destiny}<2^n-node_k)\text{ and }(node_{Destiny}<node_k).\\ [(2^n+node_{Destiny}-node_k)/2,2^n]\cup[0,node_{Destiny}], & \text{if } (node_k>node_{Destiny}>2^n-node_k)\end{cases}$$

设 $E(m_k)=m$,即 P2P 网络中节点的聚类邻居集平均规模为 m,$m\in[1,N]$;对于 P2P 网络来说,值域范围 2^n 的值由 Hash 散列函数本身决定,它决定了 Chord 路由表中的表项数目 $Entry$,期望 $E(Entry)=n$,逻辑跳数 $Hops$ 的期望 $E(Hops)(Chord)=\frac{1}{2}\log_2 N$;当聚类规模 $m=N$ 时,即相当于 $E(Entry)=N$,此时,$E(Hops)(N)=1$.$E(Hops)(m)$的函数曲线如图 2 所示.

定理 2. 若 P2P 网络规模 N 充分大,则聚类规模 $m\in[1,n\cdot\log_2 N]$不影响整个 P2P 网络的逻辑跳数期望,$E(Hops)(m)=\frac{1}{2}\log_2 N$.

证明:如图 3 所示,以直线 $E(Hops)(m)=\frac{1}{2}\log_2 N-\frac{\frac{1}{2}\log_2 N-1}{N-n}\cdot(m-n)$ 近似拟合图 3 中的曲线,$N\in[0,2^n)$.

当 $m=n\cdot\log_2 N$ 时,

$$\lim_{N\to\infty}\frac{E(Hops)(m)}{E(Hops)(Chord)}=\lim_{N\to\infty}\frac{\frac{1}{2}\log_2 N-\frac{\frac{1}{2}\log_2 N-1}{N-n}\cdot(n\cdot\log_2 N-n)}{\frac{1}{2}\log_2 N},$$

由 L'Hôpital 法则可得:

$$\lim_{N\to\infty}\frac{E(Hops)(m)}{E(Hops)(Chord)}=1,$$

即 对于$m\in[1,n\cdot\log_2 N]$,当N充分大时,$\exists N_0$,使得$E(Hops)(m)=\frac{1}{2}\log_2 N_0$.得证. □

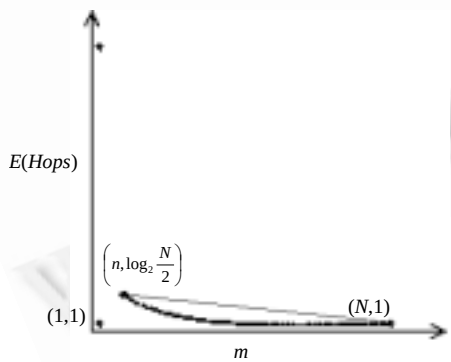

Fig.2 $E(Hops)(m)$~m function diagram

图 2 $E(Hops)(m)$~m 曲线图

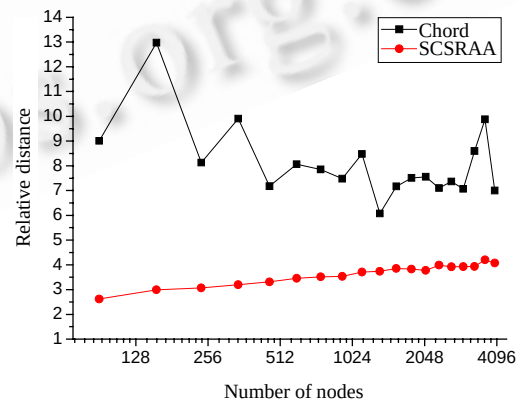

Fig.3 Comparison of routing algorithms between SCSRAA and Chord

图 3 SCSRAA/Chord 语义路由算法比较

2.3.2 路由算法分析

通过将聚类邻居节点的 Hash 函数输出值特性与传统基本语义路由过程所遍历的 $node_{id}$ 序列的特性相关联,归纳出目的节点、传统(chord)语义路由中继节点序列、聚类邻居节点集三者之间的逻辑关联特性,表述如下:

令$\rho=(m_k-1)\cdot 2^{-i}$.根据定理 1 可知,在邻居节点集Ω_k中,有ρ个节点均匀地落在下一跳区间$[0,2^{i-1}]$中;在 Chord 语义路由中继节点序列中,落在下一跳区间$[0,2^{i-1}]$中的节点为 $node_i.finger(node_i+2^{i-1})$.在同为幂次逼近目标节点的路由逻辑一跳过程中,二者的物理路径代价期望分别为 $d,\rho\cdot d+(1-\rho)\cdot D$.下面将计算 Chord,SCSRAA 路由算法所对应的 P2P 网络中任意两节点间报文路由的物理路径期望.

$$m\in[1,n\cdot\log_2 N],$$

$$E(C_i)=E(D_i)=\begin{cases}0, & i=0\\ D, & 0<i<\log_2 N\end{cases},$$

$$E(S_i)=\begin{cases}E(d_i), & \rho\ge 1\\ E(\rho\cdot d_i+(1-\rho)\cdot D_i), & 0\le\rho<1\end{cases}=\begin{cases}0, & i=0\\ d, & 0<i<\log_2(m-1)\\ \rho\cdot d+(1-\rho)\cdot D, & \log_2(m-1)<i<\log_2 N\end{cases},$$

$$E(Chord)=E\left(\sum_{i=0}^{\log_2 N}C_i\right)=E\left(\sum_{i=0}^{\log_2 N}E(D_i)\right)=\left(\int_0^{\log_2 N}t\cdot\frac{1}{\log_2 N}\mathrm{d}t\right)\cdot E(D_i)=\frac{1}{2}\cdot\log_2 N\cdot D \tag{2}$$

$$\begin{aligned}E(SCSRAA)&=E\left(\sum_{i=0}^{\log_2 N}S_i\right)=E\left(\sum_{i=0}^{l\log_2(m_k-1)}E(d_i)\right)+E\left(\sum_{i=\log_2(m_k-1)}^{\log_2 N}E(\rho\cdot d_i+(1-\rho)\cdot D_i)\right)\\&=\left(\int_0^{\log_2(m-1)}t\cdot\frac{1}{\log_2 N}\mathrm{d}t\right)\cdot E(d_i)+\left(\int_{\log_2(m-1)}^{\log_2 N}t\cdot(m-1)\cdot 2^{-t}\cdot\frac{1}{\log_2 N}\mathrm{d}t\right)\cdot E(d_i-D_i)+\\&\quad\left(\int_{\log_2(m-1)}^{\log_2 N}t\cdot\frac{1}{\log_2 N}\mathrm{d}t\right)\cdot E(D_i)\end{aligned} \tag{3}$$

$$\begin{aligned}&=\frac{1}{2}\cdot\frac{(\log_2(m-1))^2}{\log_2 N}\cdot d+\frac{(m-1)}{\log_2 N}\cdot\frac{1}{-\ln(2)}\left[\frac{1}{N}\cdot\left(\log_2 N+\frac{1}{\ln(2)}\right)-\right.\\&\left.\frac{1}{m-1}\cdot\left(\log_2(m-1)+\frac{1}{\ln(2)}\right)\right]\cdot(d-D)+\frac{1}{2}\cdot\frac{(\log_2 N)^2-(\log_2(m-1))^2}{\log_2 N}\cdot D\end{aligned}$$

若 $d=g(m,N)D$,路由提升效率 $\eta=\dfrac{E(Chord)-E(SCSRAA)}{E(Chord)}$ 的解析式为

$$\eta=\frac{1-g(m,N)}{(\log_2 N)^2}\cdot\left\{(\log_2(m-1))^2-\frac{2(m-1)}{\ln(2)}\left[\frac{1}{N}\cdot\left(\log_2 N+\frac{1}{\ln(2)}\right)-\frac{1}{m-1}\cdot\left(\log_2(m-1)+\frac{1}{\ln(2)}\right)\right]\right\} \tag{4}$$

η与值域范围2^n无关,即与 P2P 网络是否饱和无关,仅是聚类平均规模 m 及网络规模 N 的函数,公式(3)对应的 SCSRAA 理论曲线在仿真实验部分给出,如图 4 所示.当 i 较小时,即在 $node_{id}$ 以幂级逼近的路由选择初期,从聚类邻居节点集中选择下一跳节点的概率较高;当 $i\to\log_2 N$ 时,在该路由选择下一跳的算法中,从聚类邻居节点中选择下一跳的概率较小,此时,路由更大的可能是选择原 Chord 路由表项所决定的下一跳节点.同时,我们应该注意到,当$\rho>1$ 时,有ρ个节点是均匀分布在下一跳的区间内的.从ρ个节点中选取一个与 $node_{Destiny}$ 接近的节点,存在常数级的值逼近效果,尽管比起幂级逼近可以忽略,但还是存在的.

3 实验仿真与结论

3.1 实验目的和方法

本实验的目的是测试基于网络物理拓扑聚类的 SCSRAA 算法的语义路由效率.采用在同一路由过程中比较传统 Chord 语义路由算法及 SCSRAA 语义路由算法,以评估其语义路由效果的提升能力.实验中,网络拓扑产生模块采用美国乔治大学的 GT-ITM[9]软件包,原因是 GT-ITM 作为著名的开源代码的网络拓扑产生系统,可以产生多种模型的网络拓扑,例如:Transit-Stub(TS)模型、随机模型等.TS 模型能够较好地模拟 Internet 网络拓扑结构,存在着多个不同网络距离(时延)级别的域,类似于实际网络中存在的多个局域网、城域网的混杂联网模式.

实验过程中,对生成的 TS 拓扑数据文件进行映射,将所有节点一一对应为网络端节点.所生成的 TS 模型中的物理拓扑相关参数,在 N、Transit 域、Stub 域三者之间存在同时缩放的函数关系$\mathcal{H}$,这是符合现实互联网络环境的.

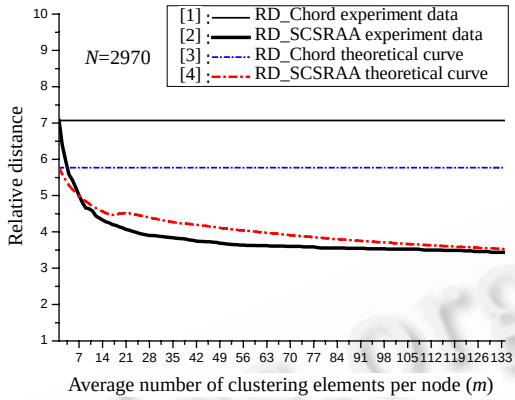

Fig.4 RD Effect of average number of clustering elements per node in SCSRAA

图 4 SCSRAA 中节点的平均聚类规模对 RD 的影响

3.2 实验过程与结果

3.2.1 验证参数

相对距离 RD(relative distance).对于系统中的任意两个节点传递一个报文,设根据 P2P 路由表从源节点路由到目的节点所经过的 IP 层 *hop* 数为 *Dist_p2p*,直接从源节点路由到目的节点所经过的 IP 层 *hops* 数为 *Dist_ip*,则 RD 为二者的比值,有的系统称为 RDP(relative delay penalty).该指标越低,表明该 P2P 系统语义路由的效率越高.

3.2.2 SCSRAA/Chord 语义路由效率比较

3.2.2.1 实验过程

(1) GT-ITM 软件包产生 *n* 个节点的网络拓扑 TS-*n*;*n* 取不同规模;

(2) 随机选取任意 2 个节点作为源节点 $node_{Source}$ 和目的节点 $node_{Destiny}$;

(a) 通过 Dijkstra 最短路由算法计算所有节点间的延时距离 d_{ij};

(b) 计算 Chord 路由算法经过的节点序列;累加给定序列中节点间的 d_{ij},并在此基础上计算 *RD_Chord*;

(c) 计算 SCSRAA 路由算法节点序列;累加给定序列中节点间的 d_{ij},并计算 *RD_SCSRAA*;

(3) 重复步骤(2),完成相关 RD 的统计工作.

在 GT-ITM 产生的拓扑中存在缩放关系$\mathcal{H}$,因此,本实验过程中的聚类算法是基于自组织情形II来验证 SCSRAA 模型的路由效率的.

3.2.2.2 实验结果

对于固定阈值(θ_i=100≈a_0*Lan_davg)所得到的测试结果如图 3 所示,SCSRAA 算法路由效率提升明显.在图 3 中,*RD_SCSRAA* 曲线存在着缓慢增长的趋势,原因分析如下:*Lan_davg* 为 TS 模型所有 Stub 域中节点间的平均物理距离,其值随着 *N* 的增大而增大,由于θ_i 固定,则 a_0 的值缓慢变小,从而导致 *RD_SCSRAA* 的缓慢增长,若是在实验中固定 a_0 不变,则可以消除这一缓慢增长趋势.另外,值得一提的是,在实验过程中,SCSRAA 算法对 Chord 路由过程中最坏情形的提升效果是相当明显的.

3.2.3 SCSRAA 模型中聚类规模、路由效率二者之间的关系

当系统规模 *N* 为 2 970 时,所对应的缩放关系$\mathcal{H}$简化为 $g(m,N)=\frac{d}{D}$,如图 5 所示.图 4 给出了 *RD_SCSRAA*,聚类邻居节点平均规模 *m* 之间的理论曲线(公式(3))及实测数据曲线.实测数据与理论曲线在聚类规模范围 $m\in[1,n\cdot\log_2^N]$ 内的数据较好地吻合,验证了 SCSRAA 模型中对逻辑关联特性叙述的正确性,即验证了定理 1、

公式(3)和公式(4)的正确性.在节点规模为 2 970 的系统中,从聚类规模 m,*RD_SCSRAA* 两个指标综合考虑,聚类规模 $m\in[16,24]$是最优区间,回顾一下实验过程中 2 970 节点拓扑生成时,我们给定的 Stub 中的节点数恰好就是 18,这证实了 SCSRAA 算法与系统的物理拓扑能够紧密关联.当 m=1 时,即聚类集合中除了节点自身以外没有任何其他邻居节点时,SCSRAA 算法回退为传统的 Chord 路由算法.对于其他聚类规模范围,由于节点所维护的邻居节点规模过大而不符合实际情况,本文不予分析.

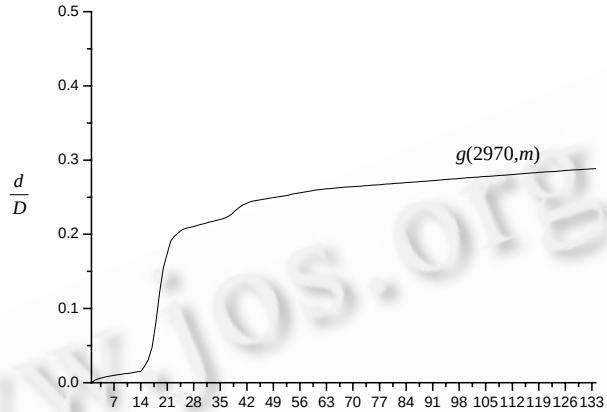

Fig.5 $g(m,N)$~m, N=2970 nodes

图 5 $g(m,N)$~m, N=2970 节点

实验表明,SCSRAA 算法语义路由效率提升效果明显.同时,由于网络拓扑产生后,节点间的延时距离是固定的,因而此仿真并不能验证广域网环境中网络状况动态变化的情形.在真实环境中,通过 Ping 方式得到的节点间的坐标位置是动态变化的.本文提出的自组织聚类算法能够适应这样的环境而动态变化,但本实验过程并不能验证这一特点.

4 结束语

P2P 网络是构建于物理网络拓扑之上的一层 Overlay 网络,两层之间的唯一联系是 Hash 散列函数,节点 $node_{id}$ 值中没有任何关于其物理位置的信息.本文给出了目的节点、语义路由节点序列、聚类邻居集中节点这三者之间的逻辑关联关系,并应用于基于物理位置聚类的语义路由选择算法中.SCSRAA 模型与其他相关研究相比,优点主要体现在:(1) 没有设置超级节点,维护了对等网络节点的对等特性;(2) 无须在 $node_{id}$ 值中嵌入物理位置信息,保证其值均匀地散布在值域空间,使得 P2P 网络上层应用负载均衡,免受过载影响.物理邻居节点的聚类工作及从聚类节点中选择路由下一跳的工作是 SCSRAA 模型中最主要的两项工作.本文所做的工作包括:从理论上分析了提高语义路由效率的可行性;针对自组织邻居节点聚类模块,提出了自组织非对称动态聚类的解决方法,给出了多种切实可行的节点物理位置坐标信息发布规则.实验部分进一步证实了 SCSRAA 算法的有效性.SCSRAA 作为单独的路由性能提升模块,也可以用于例如 Pastry,SkipNet,蝴蝶网 Viceroy 等结构化 P2P 的路由算法中,因此,本文的研究有一定的应用前景.

在今后的工作中,我们将针对节点主动退出行为及容错模型,对 SCSRAA 模型做进一步改善.在实际 P2P 网络应用中,需要建立节点规模,*Wan_Davg*,*Lan_davg* 这三者之间的统计关系模型,用来指导在节点聚类初期阈值的设定工作,以期加快聚类的收敛速度.对于这方面的研究,有待日后结合网络行为学的研究来开展.

致谢 衷心感谢顾冠群教授对本文工作所提出的宝贵意见.

References:

[1] Ratnasamy S, Shenker S, Stoica I. Routing algorithms for DHTs: Some open questions. In: Druschel P, Kaashoek M, Rowstron A, eds. Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems (IPTPS 2002). Berlin: Springer-Verlag, 2002. 174−179.

[2] Zhao BY, Duan Y, Huang L, Joseph AD, Kubiatowicz JD. Brocade: Landmark routing on overlay networks. In: Druschel P, Kaashoek M, Rowstron A, eds. Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems (IPTPS 2002). Berlin: Springer-Verlag, 2002.

[3] Krishnamurthy B, Wang J, Xie YL. Early Measurements of a cluster-based architecture for P2P systems. In: Proc. of the ACM SIGCOMM Internet Measurement Workshop. New York: ACM Press, 2001. 105−109. http://www.imconf.net/imw-2001/proceedings.htm

[4] Ratnasamy S, Handley M, Karp R, Shenker S. Topologically-Aware overlay construction and server selection. In: Proc. of the IEEE INFOCOM Conf. New York: Institute of Electrical and Electronics Engineers, Inc., 2002. 1190−1199. http://www.icir.org/sylvia/

[5] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for Internet applications. In: Proc. of the ACM SIGCOMM 2001 Conf. New York: ACM Press, 2001. 149−160. http://www.acm.org/sigs/sigcomm/sigcomm2001/

[6] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Guerraoui R, ed. Proc. of the 18th IFIP/ACM Int'l Conf. on Distributed Systems Platforms (Middleware 2001). Berlin: Springer-Verlag, 2001. 329−350.

[7] Malkhi D, Naor M, Ratajczak D. Viceroy: A scalable and dynamic emulation of the butterfly. In: Proc. of the 21st annual ACM Symp. on Principles of Distributed Computing. New York: ACM Press, 2002. 183−192. http://www.podc.org/podc2002/

[8] Wang YB, Xue T. Applied Cryptography. Beijing: China Machine Press, 2003. 135−151 (in Chinese).

[9] Zegura EW, Calvert KL, Bhattacharjee S. How to model an internetwork. In: Proc. of the INFOCOM'96. New York: Institute of Electrical and Electronics Engineers, Inc., 1996. 594−602. http://www.cc.gatech.edu/fac/Ellen.Zegura/pubs1.html

附中文参考文献:

[8] 王衍波,薛通.应用密码学.北京:机械工业出版社,2003.135−151.

刘业(1977－),男,江苏建湖人,博士生,主要研究领域为高性能网络,分布式计算.

杨鹏(1975－),男,博士生,主要研究领域为新一代网络体系结构,形式化理论和技术.