

在 P2P 网络中构建支持语义搜索的覆盖网络 DOverlay^{*})

杨娟^{1,3} 白云³ 邱玉辉^{2,3}

(四川师范大学计算机科学学院 成都 610068)¹ (西南大学计算机与信息科学学院 重庆 400715)²
(西南大学语义网格实验室 重庆 400715)³

摘要 P2P 网络上构建支持复杂语义请求的覆盖网络是一个新的挑战。支持复杂语义的覆盖网络不仅要满足 P2P 网络的完全分布性、动态性、可扩展性等相关特性,还要尽量提高 P2P 网络的可靠性和低成本维护性。提出的非结构化领域覆盖网络 DOverlay 由于其资源特征是直接由其对应领域本体的概念树转换而来,因此资源可自动形成相关聚类。而基于 DOverlay 的资源搜索策略 D_Search 不仅在减少了网络开销的前提下增大了搜索成功率,而且在 D_Search 中设计的资源评级策略还大大增加了搜索的有效性。

关键词 P2P 网络,本体,语义标签,领域覆盖网络

Constructing DOverlay of P2P Network to Support Semantic Based Searching

YANG Juan^{1,3} BAI Yun³ QIU Yu-hui^{2,3}

(College of Computer Science, Sichuan Normal University, Chengdu 610068, China)¹

(Faculty of Computer and Information Science, Southwest University, Chongqing 400715, China)²

(Semantic Grid Laboratory, Southwest University, Chongqing 400715, China)³

Abstract P2P network is covered with an overlay topology to support the resources' fast locating. Overlays possess the semantic intelligence can manipulate the complicated queries, such as semantic queries. We propose a P2P Overlay based on ontology concept trees which can adapt to the varied characteristics of the P2P networks. The resource clusters can be formed automatically according to their ontology attributes. D_Search policy applied on DOverlay can effectively improve the successful rate of the resource searching while reducing the message cost. The rating mechanism embedded in D_Search can enhance the efficiency of the searching work.

Keywords P2P network, Ontology, Semantic label, DOverlay

1 研究背景

在过去几年中, P2P 模型由于其高度动态性、分散性、容错性以及低维护成本性^[1,2] 成为了一个在 Internet 规模内进行资源(文件、服务、CPU、内存、存储空间、网络带宽)^[3,4] 共享的一个极具吸引力的模型。P2P 网络通过在具有异构操作系统和网络的节点上构建覆盖网络(overlay)来实现上述目的的。由于覆盖网络的灵活可配置性,因此它允许用户在不用更改支撑物理网络的前提下实现资源发现的分布式操作^[7-13]。覆盖网络的有效性直接影响了 P2P 网络中资源发现策略的性能,因为 P2P 网络中节点的高度变化性对资源发现来说是一个极大的困难。

经典的 P2P 网络,如 Gnutella^[14], Napster^[15], KaZaA^[16] 以及 Morpheus^[17] 采用的都是非结构化的覆盖网络。这种网络虽然易于实现和维护,但是由于节点连接的随意性,导致了搜索半径无穷大。并且,为保证搜索成功率,搜索策略只能采用泛洪机制或随机传递机制^[18],使得网络中的消息数量随网络规模的扩大呈对数级别增加^[10]。

然而基于分布式哈希表(DHT)的结构化 P2P 覆盖网络则既可以保证搜索效率,也可以控制消息数量,如 Chord^[3] 和 CAN^[20]。这种采用 DHT 的覆盖网络主要通过资源进行唯一的哈希编码,使得资源定位能快速实现。但是这种结构化的覆盖网络在支持复杂请求(如语义请求)时则遭遇了极大

的困难,因此构建具有语义智能的 P2P 覆盖网络得到了众多研究者的青睐。pSearch^[21] 在 CAN 上面构建了一个分布式的向量空间模型 VSM 和信号语义索引 LSI,通过高维度向量的角度计算来确定文档间的相似度,因此可以实现语义请求中近义词和同义词的搜索。但是,高维度向量的比较则使得降维计算(使用旋转索引)增大了搜索的开销。SSW^[22] 在构建覆盖网络时采用小世界理论^[23] 聚类,通过邻居索引^[24,25] 进行有选择的消息传递,以达到保证搜索效率和控制消息数量的目的。但是与 pSearch 相同,它也面临维度缩减的问题。IMAGINE-P2P^[26] 提出了一个在结构化 P2P 覆盖网络上支持路径请求的模型,它通过在结构化覆盖网络上构建语义层,实现了基于语义的搜索。但是它不能支持同义词或近义词的资源搜索。iVCE^[27] 中构建了一个双层的资源发现支撑平台,但是它无法适应开放的 P2P 网络。pFusion^[19] 是一个非结构化的 P2P 支撑平台,其主要的目的是在消息路由方面提高性能,因此它在语义搜索方面并没有贡献。

本文中构建的基于本体概念的 P2P 覆盖网络 DOverlay 不仅满足 P2P 网络的分散性、高度动态性、可扩展性、容错性等特点,且实现了复杂语义请求的资源搜索(近义词、同义词搜索)。而基于 DOverlay 的资源搜索策略 D_Search 不仅在减少了网络开销的前提下增大了搜索成功率,而且在 D_Search 中设计的资源评级策略还大大增加了搜索的有效性。

本文结构如下:第 2 节给出 DOverlay 的主要实现架构,

^{*}) 本文受“基于语义的资源服务协同和管理研究”(973)项目支持。

第3节给出基于 DOverlay 的有效资源搜索过程,第4节试验部分通过5个度量标准对 DOverlay 的有效性进行验证。最后对本文工作进行总结。

2 可进行资源自组织的覆盖网络 DOverlay

2.1 资源与本体概念树的映射

令每个资源可映射到一个唯一的领域本体,并作为该领域覆盖网络的一个节点成员。由于每个领域本体维护一个属于该领域的资源覆盖网络(overlay),这与 P2P 网络中的传统覆盖网络有所区别,因此我们将它命名为 DOverlay。

令 $A = \{A_i | i \in [1, d]\}$ 为一个本体领域属性值集合,以 Wine Ontology^[5] 为例, $A = \{\text{maker, grape, sugar, flavor, body, color, region, vintage year, descriptor}\}$, 共有9个属性。按本体的一般规律,领域本体的属性种类是有限的。

令 $V(A_i) = \{v_j^i | i \in [1, d], j \in [1, n_i]\}$ 为属性 A_i 的值的集合, v_j^i 是属性 A_i 可能的第 j 个属性值。

令 $M(v_j^i) = \{s_{l,j}^i | i \in [1, d], j \in [1, n_i], l \in [1, h_{i,j}]\}$, $s_{l,j}^i$ 是属性 A_i 的第 j 个属性值 v_j^i 的第 l 个近似词, $M(v_j^i)$ 是属性值 v_j^i 的近似词集合。 $M()$ 函数用于将资源与本体树中的概念建立关系。因此有

$$M_{l \in [1, h_{i,j}]}(s_{l,j}^i) = v_j^i$$

令 $K = \{n_i | i \in [1, k_j]\}$ 为资源 j 中出现的词集, c_j^i 为资源中出现的属性 A_i 的值的个数。 $A(j) = \langle L_1, L_2, \dots, L_d \rangle$ 为资源 j 的属性标记向量,若资源 j 中具有对应属性的值,则该位 $L_i = A_i$, 否则 $L_i = \Delta$ 。具有最大 c_j^i 值的属性定义为 A_j^* , 而最大 c_j^i 值定义为 c_j^* 。

2.2 DOverlay 结构

DOverlay 分为4层结构。每个 DOverlay 维护以属性划分不同的社区,例如 Wine Ontology 中的社区数应为9。每个社区中有多个聚类,每个聚类以一定数量限制的资源节点构成。定义 DOverlay 为一个2元组 $(Nodes, Edges)$, $Nodes$ 是 DOverlay 中资源节点的集合, $Edges$ 是边集合, $Edges = \{e_{ij} | i, j \in Nodes\}$ 。

2.2.1 资源节点上维护的 DOverlay 信息

因为 DOverlay 是以本体属性来划分的社区,且社区中有多个聚类,所以资源节点的标识采用3级标识, Resource ID: CL, ClusterID, LocalID。 CL 表示其所在社区标识, ClusterID 为聚类标识, LocalID 为本地标识。

资源节点 j 上维护的是包括 ResourceID, Neighbor List, MirrorList, CE_j , $\omega(a, b)$, CNum 在内的6部分信息。其中 Neighbor List 是邻居节点列表, MirrorList 是镜像节点列表, CNum 是当前聚类 ClusterID 中的已有节点数。

定义1 定义 $\omega(a, b) = \langle \omega_1, \omega_2, \dots, \omega_d \rangle$ 为资源节点 a 和 b 间连接边的连接权重向量,若资源 a 与资源 b 之间有连接,那么 $\omega(a, b)$ 表示资源 a 与资源 b 的连接分别在 d 个维度上的权重。向量分量

$$\omega_i = \begin{cases} 1 - (c_a^i / c_b^i)^2, & c_a^i \leq c_b^i \\ 0, & c_a^i > c_b^i \end{cases}$$

c_a^i 表示资源 a 在属性 A_i 上产生的值的个数。

$\omega_i > 0$ 表示资源 b 比资源 a 在属性 A_i 这个维度上更有意义。并且,如果消息传递只依靠一个维度的权重,那么搜索可以只局限在本社区中。如果消息传递依靠多个维度权重,那么可以按情况实现快速跨社区搜索。

2.2.2 DOverlay 资源聚类策略

DOverlay 中资源的聚类策略是按照节点的 CE 值进行聚类。如前所述,CE 值的大小决定了资源节点可能的最大度的值。

定义2 定义

$$CE_j = c_j^* / \max_{f \in community(A_j^*)} (\overline{c_f^*})$$

$$\text{MaxLink}(j) = \lceil CE_j \cdot CN \rceil + l_j$$

$\text{MaxLink}(j)$ 表示资源 j 允许的最大连接数。其中 $\overline{c_f^*}$ 为系统在一定时间内根据历史信息所定义的可能出现的平均最大属性值个数。 CN 是约定的一个聚类中允许的最大节点数。

2.3 DOverlay 路由策略

定义3 定义社区 A_i 与 A_j 的边界集 B_{ij} , 有 $B_{ij} = \bigcap_{r \in B_{ij}} M_r = \emptyset$

$$B_{ij} = \{r | r \in community(A_i) \cup community(A_j), c_r^i \neq 0, c_r^j \neq 0, E_r < \beta\}$$

资源 j 在请求信息到达不同社区且属于该社区边界集的节点时,若 j 之前并未与该社区边界集成员建立连接且两个节点都允许建立连接,则建立连接。消息根据需求继续转发。请求 request 消息在资源节点 a 上的转发策略是由 a 与其邻居节点 b 之间的转发概率 $p(a, b)$ 决定的。

定义4 定义 $p(a, b)$ 为资源节点 a 向其邻居节点 b 发送 request 消息的概率,则有

$$p(a, b) = \begin{cases} \lceil c_b^{A_j^*} / c_a^{A_j^*} \rceil \cdot (1 / e^{\alpha c_b - \beta}), & b.CL \neq A_j^*, CE_a \geq \gamma \\ 1, & \text{Otherwise} \end{cases}$$

$$p(a, b) = \begin{cases} 1 & b.CL = A_j^* \\ 0 & \text{Otherwise} \end{cases}$$

其中 γ 为资源节点 a 是否进行全转发的门限值。

定义5 当资源 j 请求到达同社区资源节点 a 时, a 向其邻居节点 b 转发请求消息概率

$$p(a, b) = \begin{cases} \omega_{A_j^*} / \|\omega(a, b)\|, & b.CL = A_j^*, \lceil CE_a \cdot CN \rceil - n_j \neq 0 \\ 0, & \text{Otherwise} \end{cases}$$

其中 $\omega_{A_j^*}$ 表示属性 A_j^* 在向量 $\omega(a, b)$ 中所对应的分量元素取值。

当请求已到达同社区节点 a 时,若 a 的连接未满足,则停止转发消息,否则向 a 的邻居节点中同社区的节点转发,转发的依据是向同社区中聚类系数更大的节点转发,但是为了进行任务均衡,我们用 $\|\omega(a, b)\|$ 来限定它的转发概率。

3 在 DOverlay 中进行基于语义的资源搜索

3.1 用户请求信息

用户请求信息 request = {DomainKey, Parameters} 由两部分构成, DomainKey 标志用户需求信息所在本体域,而 Parameters 部分则存储了用户对需要资源的请求描述。 Parameters 由如下结构组成:

$$\text{Parameters} = (((C_1, OP_1, C_2), OP_2, C_3), OP_3, C_4 \dots) OP_n, C_n$$

其中 C_i 表示请求操作数,而 OP_i 表示请求操作符。每一个 C_i 由一个名-值对构成, $C_i = (Aname, Avalue)$, $Aname \in A$, $Avalue \in V(Aname)$ 。 $Aname$ 表示该领域本体中对应的某个属性名,而 $Avalue$ 表示该属性名所对应的某个属性值。 $C_i = (Aname, Avalue)$ 名-值对中 $Avalue$ 值可以为空,此时 $C_i = (Aname, \wedge)$ 表示请求数为包含所有属性值的某个属性。操作数 $OP = \langle or, and \rangle$ 只包含两种操作。分别是 or 和 and 操

作,操作数和操作符满足定义 6 中的 4 条运算规则。

定义 6 操作数 C 和操作符 OP 满足下面 4 条运算规则:

- (1) $C OP C = C$
- (2) $C_1 OP C_2 = C_2 OP C_1$
- (3) $(C_1 \text{ and } C_2) \text{ or } C_1 = C_1$
- (4) $(C_1 \text{ or } C_2) \text{ and } C_1 = C_1$

根据这 4 条运算规则,可以对用户请求进行简化,并过滤掉无用请求信息。

3.2 资源搜索策略——DOOverlay_Search

对用户请求信息 $\text{request} = \{\text{DomainKey}, \text{Parameters}\}$ 中用户的每个搜索条件 $C_i = (Na^i, Va^i)$ 进行并发资源搜索,搜索算法 $\text{Search}(C_i, r, TTL, Re(C_i))$ 描述如下:

$\text{Search}(C_i, r, TTL, Re(C_i))$

1. {
2. $T = TTL - 1;$
3. if $(T \geq 0)$
4. {
5. if $(Va^i \in V_r(A))$
6. { $Re(C_i) = Re(C_i) \cup \{r\};$
7. generate $Level(C_i, r);$
8. }
9. else for each $s \in r.NeighborList \cup r.MirrorList$
10. {
11. send search message from r to s according to definition 3.8;
12. search $(C_i, s, T, Re(C_i));$
13. }
14. else return;
15. }

搜索算法 $\text{Search}(C_i, r, TTL, Re(C_i))$ 表示,在设置的搜索跳数 TTL 限制以内,满足搜索条件 C_i 的资源就要加入到集合 $Re(C_i)$ 中,并给其进行评级,用于最后推荐给用户时综合计算使用。值得注意的是,在算法 $\text{Search}(C_i, r, TTL, Re(C_i))$ 中采用 2.3 节中所描述的消息转发机制。

3.3 DOOverlay_Search 中的资源评级策略

我们从最可能影响到 DOOverlay 中资源使用的 4 个方面(相似度、搜索距离、负载和可信度)进行了分析和建模,形成了定义 7 所描述的综合资源评级策略。

定义 7

$$Level(C_i, r) =$$

$$\frac{w_1 \cdot \text{Sim}(C_i, r) + w_2 \cdot \text{Dis}(s, r) + w_3 \cdot \text{Load}(r) + w_4 \cdot \text{Trus}(r)}{4}$$

其中 $\sum_{i=1}^4 w_i = 1$ 。 $\text{Sim}(C_i, r)$ 表示资源 r 与请求条件 C_i 之间的相似程度,且 $C_i = (Na^i, Va^i)$, Va^i 可映射到集合 $V_r(A)$ 中的词 w_r^i , $(w_r^i, cw_r^i) \in V_r(A)$ 。那么 $cw_r^{Va^i \rightarrow h}$ 表示词 Va^i 及其近义词和相似词在资源 r 中出现的次数,且

$$\text{Sim}(C_i, r) = \frac{cw_r^{Va^i \rightarrow h}}{\sqrt{\sum_{i=1}^{\infty} (cw_r^i)^2}}$$

$\text{Dis}(s, r) = \frac{|T| + 1}{TTL}$ 表示从节点 s 到节点 r 的路径长度。令 (s, T, r) , $T = \{t_i | t_i \in \text{Nodes}\}$ 是从节点 s 到节点 r 的路径。 $\text{Load}(r)$ 是资源 r 的负载量,且定义

$$\text{Load}(r) = \alpha \cdot \left[\tanh\left(\frac{M_r^{(p_r)}}{B(p_r)}\right) \right] + (1 - \alpha) \cdot \left[\tanh\left(\frac{M_0^{(p_r)}}{B(p_r)}\right) \right]^1$$

其中 p_r 表示资源 r 所在主机 peer, $R(p_r)$ 表示 r 所在主机 peer 上的所有资源集合, $M_r^{(p_r)}$ 表示经过主机 p_r 的所有需转发消息的数量, $M_0^{(p_r)}$ 表示需主机 p_r 进行处理的消息的数量, $B(p_r)$ 表示主机 p_r 的带宽(单位时间内可传递的消息数)。 $\text{Trus}(r) = \frac{\sum_i U_s(r)}{\sum_i \text{Re } co(r)}$ 表示资源 r 在当前 P2P 网络中的可信度, $U_s(r)$ 表示资源 r 被使用的次数, $\text{Re } co(r)$ 表示资源 r 被推荐的次数。

4 模拟试验和性能仿真

4.1 评价指标

我们通过以下几个常用来评价搜索效率的指标来衡量,并与 SSW 策略^[22]进行对比。

搜索路径(hops):在找到目标资源时所经历的平均跳数。

搜索耗费(message numbers):找到目的资源时所产生的平均消息量。在普通泛洪或流言消息转发机制中,它们的搜索路径值也许很小,但是搜索耗费却相对很高。

查找失败率(fail ratio):在 TTL 限制内未找到目标节点产生连接从而成为孤立节点的节点数量。

4.2 搭建模拟环境

我们构建了一个关于 Wine 本体的 DOOverlay,分别由 5 台主机 peer 进行维护。为了测试搜索的有效性,我们将网络上真实存在的资源文档抽象成语义标签,并按比例 1:4 构建随机生成的语义标签。试验中所创建的 DOOverlay 所涉及到的相关参数描述如表 1 所示。

表 1 参数设置

Parameter	Description	default
N	The total resource number in DOOverlay	1000~5000
CN	Size of a cluster	50
\bar{c}_f	Average value of c_f^i of a period in a community in which resource f locate	30
γ	The shreshold of deciding weather need to flood	0.2~0.25
β	The shreshold of Border nodes set	0.2

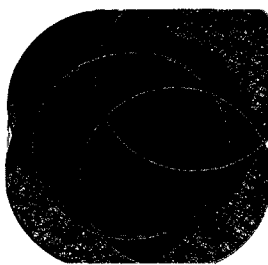


图 1 5000 个节点组成的 DOOverlay 仿真图

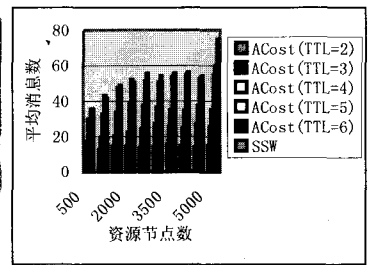


图 2 不同 TTL 时 DOOverlay 的平均消息耗费($\gamma=0.2$)

图 1 是按照表 1 所生成的 DOOverlay 仿真图。图中的圆圈也是代表不同的社区。其中 4933 个资源节点插入成功,67 个失败。共产生 15934 条边,平均度为 3.23。

¹ $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

4.3 DOverlay 的搜索路径与消息耗费

如前所述,搜索路径与消息耗费都是评价搜索策略优劣的指标。但是,通常很难使得两个指标都能令人满意,一个指标的提 高总是要以另一个指标的降低作为代价。正如使用泛洪和流言机制,虽然可以使得搜索路径控制在一个很好的范围内(2~4 跳),但是它们所付出的代价是大量的消息耗费。因此,如何在搜索路径和消息耗费中取得一个很好的平衡,是构建 DOverlay 以及 DOverlay_Search 所探讨的根本目的。下面我们就 DOverlay 在不同 TTL 值和 DOverlay 中不同资源节点数的情况下所产生的最大消息耗费和平均消息耗费情况进行分析和说明。

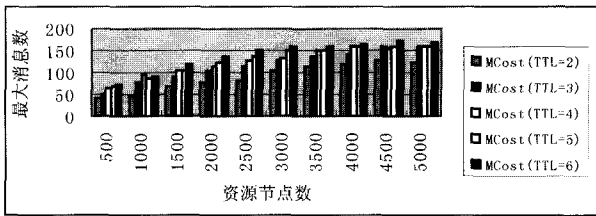


图 3 搜索路径与最大消息耗费($\gamma=0.2$)

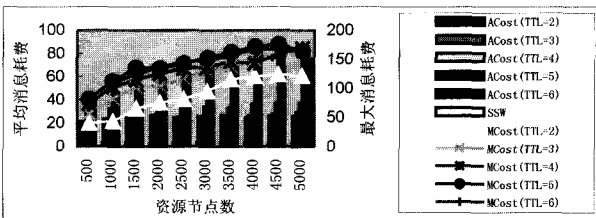


图 4 不同 TTL 时 DOverlay 的消息耗费($\gamma=0.25$)

从图 2 可以看出,当 DOverlay 给定不同的 TTL 限制时,其消耗的平均消息数是随 TTL 限制增加而增加的,这与实际情况相符合,搜索跳数越少,其能够产生的消息量也越少,当然其能够成功找到目标资源的可能性也越小。这里我们并没有将搜索成功率纳入到讨论的范围,我们会在后面专门讨论查找失败率与消息耗费之间的关系。从图中我们可以发现,当 TTL=6 时,其平均耗费的消息数与 TTL=5 时相比较已经没有太大差别了。其实,在试验中发现,只要 TTL ≥ 5 后,其耗费的平均消息数已经没有明显的变化了。这里为了表示清楚,我们略去了 TTL ≥ 7 以后的数据。我们将最后的消息数与文献[22]中的 SSW 策略结果进行了对比,可见在消息耗费上,DOverlay 是有显著优势的。

图 3 是不同 TTL 限制时 DOverlay 所产生的最大耗费。之所以要把最大消息耗费和平均消息耗费的数据分开讨论,目的在于产生不同的结论,如前所述,讨论 DOverlay 的平均消息耗费之目的在于比较 DOverlay 与 SSW 在消息耗费上的整体性能,而讨论 DOverlay 的最大消息耗费的目的是在于从 DOverlay 不同 TTL 限制所消耗的最大消息数上可以看出 DOverlay 搜索的有效 TTL 步数。从图 2 已经发现,当 TTL ≥ 5 后,其平均消息耗费数就已经没有显著增加了。而图 3 是多次试验的平均最大消息耗费,从图中可以发现从 TTL ≥ 4 以后,其最大消息耗费就没有明显变化了。图 3 更加证明了 DOverlay 的搜索有效跳数为[4,7]之间。

图 4 是我们扩大了消息转发的阈值 γ ,也就是说放松了消息转发的条件。很明显,无论是平均消息耗费还是最大消息耗费,与图 3 相比都有明显的增加。并且与 SSW 相比,性能也有所下降,也就是说 DOverlay 的消息耗费性能在 $\gamma > 0.2$

后会有所下降。换言之,也就是当 $\gamma=0.2$ 时,DOverlay 在消息耗费上会取得很好的性能效果。当然仅仅把消息耗费控制在理想的范围内是不够的,搜索跳数也是评价 DOverlay 的另一个指标。从图 5 同样可以看出当 TTL 在[4,7]区间内,消息耗费不会有明显的变化,也就是说 TTL 始终可以被控制在[4,7]范围内,并取得很好的消息耗费性能。当然,评价 DOverlay 的指标除了消息耗费和 TTL 以外,更重要的是获得很好的搜索成功率。下面,我们通过分析 TTL 与查找失败率之间的关系来评价 DOverlay 的性能。

4.4 查找失败率

我们通过分析不同网络规模下不同的 TTL 跳数内 TRM 的搜索性能,并与 SSW 策略相比较。结果如图 5 和图 6 所示。

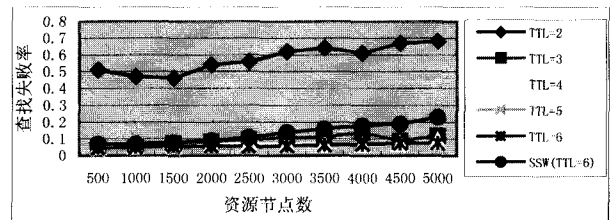


图 5 TTL 与查找失败率($\gamma=0.2$)

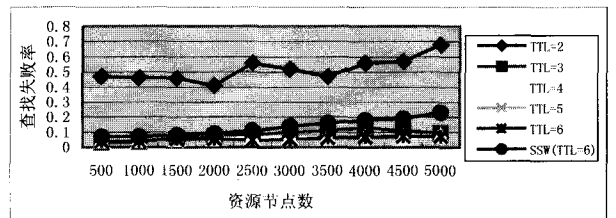


图 6 TTL 与查找失败率($\gamma=0.25$)

图 5 是 $\gamma=0.2$ 时的 TTL 与查找失败率之间的关系。从图中可以看到,只有当 TTL 值相当小(=2)时,其失败率才很高;当 TTL ≥ 3 时,其查找失败率是很低的。同样,与 SSW 策略的查找失败率相比,不难看出,我们的查找性能是优于 SSW 的。而图 6 则是扩大了其消息转发门限的搜索性能,与 $\gamma=0.2$ 时相比,搜索性能并没有显著提高,因此可以看出 $\gamma=0.2$ 是 TRM 的最优性能配置,既能获得很好的资源资源搜索效果,又能将消息耗费和 TTL 控制在一个理想的范围内。

结束语 本文提出了一个构建于完全分布式 P2P 的覆盖网络 DOverlay。DOverlay 满足 P2P 网络的分散性、高度动态性、可扩展性、容错性等特点,并且实现了复杂语义请求的资源搜索(近义词、同义词搜索)。在实现资源自组织的 DOverlay 上使用导向性路由策略可以在保证成功率的前提下将消息数控制在一个理想的范围内。由于基于 DOverlay 的搜索策略 DOverlay_Search 引入了综合评级机制,使得 DOverlay 的资源推荐更加有效和准确,并通过模拟试验和仿真对 DOverlay 的性能进行了证明和评价。但由于资源与本体概念之间的映射还介于半自动和手工操作之间,因此对资源进行自动本体特征抽取是我们正在进行的工作。

参考文献

[1] Theotokis S A, Spinellis D. A Survey of Peer-to-Peer Content Distribution Technologies. ACM Computing Surveys, 2004, 36 (4):335-371

一步的研究重点。

参考文献

- [1] Pansiot J J, Grad D. On Routes and Multicast Trees in the Internet. *ACM SIGCOMM CCR*, 1998, 28(1): 41-50
- [2] Cheswick B, Burch H, Branigan S. Mapping and Visualizing the Internet // *Proc. 2000 USENIX Annual Technical Conf. San Diego, California, June 2000*: 1-12
- [3] Internet Mapping Project. Available at; URL: <http://research.lumeta.com/ches/db/>
- [4] Siamwalla R, Sharma R, Keshav S. Discovering Internet Topology. Technical Report [EB/OL]. CS Dept., Cornell University, July 1998. Available at; URL: <http://www.cs.cornell.edu/skeshav/papers/discovery.pdf>
- [5] Govindan R, Tangmunarunkit H. Heuristics for Internet Map Discovery // *Proc. IEEE INFOCOM 2000. Tel-Aviv, Israel, Mar. 2000*: 1371-1380
- [6] Chang H, Jamin S, Willinger W. Inferring AS-level Internet Topology From Router-Level Path Traces // *Proc. SPIE ITCOM 2001-Scalability and traffic control in IP networks. Denver, Colorado, Washington: SPIE Press, Aug 2001*: 196-207
- [7] Bejerano Y, Rastogi R. Robust Monitoring of Link Delays and Faults in IP Networks // *Proc. IEEE INFOCOM 2003. San Francisco, California, 30 Mar. -3 Apr. 2003*: 134-144
- [8] Cooperative Association for Internet Data Analysis. <http://www.caida.org/>
- [9] Donnet B, Raouf P, Friedman T, et al. Efficient Algorithms for Large-Scale Topology Discovery // Eager DL, Williamson CL, Borst SC, Lui JCS, eds. *Proc. of the ACM SIGMETRICS 2005. New York: ACM Press, 2005*: 327-338
- [10] Donnet B, Raouf P, Friedman T, et al. Deployment of an Algorithm for Large-Scale Topology Discovery. *IEEE Journal on Selected Areas in Communications (JSAC)*, Special Issue on Internet, Special Issue on Internet Sampling, 2006, 24(12): 2210-2220
- [11] Donnet B, Huffaker B, Friedman T, et al. Implementation and Deployment of a Distributed Network Topology Discovery Algorithm. Tech. Rep. arXiv cs.NI/060362, Mar. 2006: 1-12
- [12] Waddington D G, Chang F, Ramesh V, et al. Topology Discovery for Public IPv6 Networks. *ACM SIGCOMM Computer Communications Review*, 2003, 33(3): 59-68
- [13] 6Bone Registry. <http://www.viagenie.qc.ca/en/ipv6/registry/>
- [14] 杨柳, 李振宇, 张大方, 等. 冗余最小化的 IPv6 拓扑发现方法. *计算机研究与发展*, 2007, 44(6): 939-946
- [15] Liu Zhenshan, Wang Qingxian, Luo Junyong. A new distributed topology discovery technology for IPv6 networks // *Proc. IEEE SoutheastCon, March 2007*: 627-632
- [16] Barford P, Bestavros A, Byers J, et al. On the Marginal Utility of Network Topology Measurement // *Proc. 1st ACM SIGCOMM Workshop on Internet Measurement (IMW 2001). San Francisco, California, Nov. 2001*: 5-17
- (上接第 68 页)
- [2] Balakrishnan H, Kaashoek M F, et al. Looking Up Data in P2P Systems. *Comm. ACM*, 2003, 46(2): 43-48
- [3] Stoica I, Morris R, et al. Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications // *Proc. 2001 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communicating. 2001*: 149-160
- [4] Rowstron A, Druschel P. Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems // *Proc. IFIP/ACM Int'l Conf. Distributed Systems Platforms. 2001*: 329-350
- [5] <http://www.schemaweb.info/schema/SchemaDetails.aspx?id=62>
- [6] Singh M P, Huhn M N, et al. Service-oriented computing: semantic, processes, agents. Ltd, 2005
- [7] Cuenca-Acuna F M, Nguyen T D. Text-based Content Search and Retrieval in ad hoc P2P Communities // *Int. Workshop on Peer-to-Peer Computing. 2002*: 220-234
- [8] Zeinalipour-Yazti D, Kalogeraki V, et al. Exploiting Locality for Scalable Information Retrieval in Peer-to-Peer Systems. *Elsevier Information Systems Journal*, 2004, 30(4): 277-298
- [9] Gupta A, Sahin O, et al. Meghdoot: Content-based Publish/Subscribe over P2P Networks // *Proc. Fifth ACM/IFIP/USENIX Int'l Conf. Middleware. 2004*: 254-273
- [10] Zhuge H, Li X. Peer-to-Peer in Metric Space and Semantic Space. *IEEE Tran. Knowledge and Data engineering*, 2007, 19(6): 759-771
- [11] Yang B, Garcia-Molina H. Efficient Search in Peer-to-Peer Networks // *Proc. 22nd Int'l Conf. Distributed Computing Systems. 2002*: 5-14
- [12] Tang C, Xu Z C, et al. Peer-to-Peer Information Retrieval Using Self-organizing Semantic Overlay networks // *Proc. of the ACM SIGCOMM 2003. 2003*: 175-186
- [13] Kalogeraki V, Gunopulos D, et al. A Local Search Mechanism for Peer to Peer Networks // *Proc. of the 11th Int. Conf. on Information and Knowledge Management. 2002*: 300-307
- [14] <http://www.gnutella.com>
- [15] <http://www.napster.com/>
- [16] <http://www.kazaa.com/>
- [17] <http://www.morpheus.com>
- [18] Lv Q, Cao P. Search and replication in unstructured peer-to-peer networks // *Proceedings of ACM International Conference on Supercomputing. 2002*: 84-95
- [19] Zeinalipour-Yazti D, Kalogeraki V, et al. pFusion: A P2P Architecture for Internet-scale Content-based Search and Retrieval. *IEEE Tran. Parallel and Distributed Systems*, 2007, 18(6): 804-817
- [20] Ratnasamy S, Handley M, et al. A scalable Content-Addressable Network // *ACM SIGCOMM'01. 2001*: 161-172
- [21] Tang C, Xu Z, et al. pSearch: Information Retrieval in Structured Overlays. *Computer Communication Review*, 2003, 33(1): 89-94
- [22] Li M, Lee W C, et al. Semantic Small World: An Overlay Network for Peer-to-Peer Search // *Proc. of the 12th IEEE Int'l Conf. on Network Protocols. 2004*: 228-238
- [23] Watts D J. 小小世界. 中国人民大学出版社, 2006
- [24] Yang B, Garcia-Molina H. Improving search in peer-to peer networks // *Proceedings of International Conference on Distributed Computing Systems. 2002*: 5-14
- [25] Li M, Lee W C, et al. Neighborhood signatures for searching P2P networks // *Proceedings of International Database Engineering and Application Symposium. 2003*: 149-158
- [26] Zhuge H, Sun X P, et al. A scalable P2P Platform for the Knowledge Grid. *IEEE Tran. Knowledge and Data Engineering*, 2005, 17(12): 1721-1736
- [27] 刘志忠, 王怀民, 等. 一种双层 P2P 结构的语义服务发现模型. *软件学报*, 2007, 18(8): 1922-1932