基于内容相似度的对等网络信用模型研究*)

陈颖熙[1] 李贤有[2] 郭 陟[1] 顾 明[1]

(清华大学软件学院 北京 100084)[1] (中国北车集团大同电力机车有限责任公司 山西大同 037038)[2]

摘 要 由于缺乏有效的信用管理机制,对等网络节点间存在着大量的欺诈行为,从而严重影响了整个网络的服务质量。本文在已有模型的基础上提出了一种对等网络信用模型,它把节点共享内容的相似度作为信用评估的重要因素,节点可以根据交易经验自适应地调整与邻居节点的连接,从而提高下载服务的质量。模拟实验结果表明,该模型能够有效处理节点的恶意行为,并且具有良好的扩展性和较低的运行开销。

关键词 对等网络,信用,相似度,自适应

Research on Peer-to-Peer Trust Model Based on Content Similarity

CHEN Ying-Xi[1] LI Xian-You[2] GUO Zhi[1] GU Ming[1]

(School of Software, Tsinghua University, Beijing 100084)[1]

(China North Locomotive (Group) Corp. Datong Electric Locomotive Co., Ltd. Datong Shanxi 037038)[2]

Abstract Due to lack of effective trust management mechanism, there are a lot of deceptive behaviors in P2P networks, which seriously decrease the quality of the network services. In order to improve the quality of the download services, this paper proposes a novel P2P trust model which introduces content similarity to evaluate the peer's trust, and adaptively update the connections with its neighbors based on transaction experience. Experimental results demonstrate that this model can deal with the malicious attacks and provide good performance in terms of communication overhead and scalability.

Keywords Peer-to-peer network, Trust, Similarity, Self-adaptation

1 引言

对等网络(Peer-to-Peer,P2P)模式克服了传统 C/S 模式的不足,形成了资源边缘化、网络可扩展、负载均衡等特点。网络中节点地位平等,且扮演双重角色,既是服务请求者也是服务提供者。因此,对等节点可以最大限度地查找并下载网络上的丰富资源。这也正是如今文件共享系统(如 BitTorrent、Maze、eMule、KaZaa[1] 等)受到人们极大青睐的重要原因。然而,也正是由于开放、匿名的运行策略,P2P 网络维护开销大、管理困难。自私节点的存在使 P2P 网络中的欺诈行为随处可见:恶意节点提供虚假文件甚至是网络病毒,free-riding 节点只下载资源但不共享资源。据 KaZaa 统计显示其网络中有超过 50% 的最热门的音乐文件被恶意篡改并在网络上共享[2]。之所以存在着这些问题,归根结底是因为传统 P2P 网络缺乏有效的管理,节点间没有信任机制的约束,节点并不会因为自私行为影响到自己的既得利益。因此,在 P2P 网络中采取有效的信用管理方法是解决问题的关键所在。通过引入信用模型可以保护良好行为节点的根本利益,同时抵制恶意节点,这有利于规范 P2P 网络节点行为,鼓励节点向其它用户提供优质可靠的服务。本文在已有模型的基础上提出了基于内容相似度的对等网络信用模型 CSBTrust(Content Similarity Based Trust)。该模型利用节点间的内容相似度来评价节点提供服务的能力,并根据节点间的交易历史自适应动态地调整网络连接,提高节点搜索下载服务的效率。模拟实验表明 CSBTrust 模型在拥有恶意节点的 P2P 网络中能够大大降低下载到虚假文件的概率,从而提高了下载服务的质量。

2 相关工作

在 P2P 网络中,目前主要分为两大类信用模型:

(1)基于本地信誉度的信用模型[3,4]。这类模型采用节点历史交易的反馈信息来评价节点信誉。为了计算本地信用,PeerTrust[3] 模型需要设置复杂的参数和一个通用的信用矩阵。虽然文中提出了分布式实现模式和信用数据管理模式,但是这种模式在结构化 P2P 网络中开销非常大。

(2)基于全局信誉度的信用模型[5,6]。这类模型中最典型的是 S. D. Kamvar 提出的 EigenTrust[6]。为获取节点的全局信誉度,该算法通过邻居节点(包括邻居的邻居)满意度的不断迭代,最终形成一个全局信誉度视图。然而 EigenTrust 没有对迭代的收敛性做出确定性保证;它引入的 pre-trusted 节点如何选择难以决定,在实际执行过程中可操作性差;没有对造成交易失败的节点在信任度上做出惩罚[5]。另外,随着网络规模的不断扩大,EigenTrust 进行的全局迭代会使整个网络的开销巨大。请求目标节点本地信用的消息将导致洪泛(flooding),甚至影响迭代效果。因此,EigenTrust 比较适合于小规模网络。

3 CSBTrust 模型

3.1 模型说明

在 P2P 环境下,节点共享的内容存在某种相似性,可以

*)基金项目:国家"973"计划基金项目(2004CB719406)资助。陈颖熙 硕士研究生,研究方向为 P2P 网络;郭 陟 博士后,研究方向为计算机系统安全与软件体系结构;顾 明 教授,研究方向为操作系统、分布式应用系统支撑平台和电子商务等。

认为相似节点对某类服务的评价标准是类似的，也就是节点更应该相信与它具有相似评价标准的节点提供的推荐。这正如现实生活中，我们更愿意和与自己有相同兴趣爱好的人交流并倾听他所提出的观点和建议。因此，CSBTrust 模型在节点信用推荐中引入了内容相似度因素。另外，还在已有模型的基础上改进了本地信用的度量，并根据节点信用度自适应调整邻居节点。接下来将对模型涉及到的主要内容进行介绍。

3.1.1 本地信用

本地信用是指节点自身对目标节点的信誉认同。在本文用 l_{ij} 表示节点 i 对节点 j 的本地信用。两节点在共享文件过程中的行为称为交易。对交易结果的评价采用类似文[7]的方法(见表 1)：

表 1 交易评价情况表

交易评价	交易说明	满意度
好	交易所得与服务描述一致，交易速度快	+1
一般	交易所得与服务描述大体一致，交易速度一般	+0.5
不好	交易所得与服务描述有相当差别	−0.5
极差	交易所得与服务描述完全不同，甚至可能含有病毒	−1

在这里用 A_{ij}、B_{ij}、C_{ij}、D_{ij} 分别表示交易结束后节点 i 对节点 j 评价为“好”、“一般”、“不好”、“极差”的次数。

满意度影响因子定义为：

$$\theta=1-(C_{ij}\times 0.5+D_{ij}\times 1)/(A_{ij}+B_{ij}+C_{ij}+D_{ij}) \quad (1)$$

因此：

$$l_{ij}=\theta\times(A_{ij}\times 1+B_{ij}\times 0.5)/(A_{ij}+B_{ij}+C_{ij}+D_{ij}) \quad (2)$$

观察可得：在总交易次数一定的情况下，θ 是关于 C_{ij} 与 D_{ij} 的减函数，当节点 j 提供的恶意服务次数增加时，θ 值减小，因此 l_{ij} 数值也减小。这种本地信用计算方法可以把节点的多种交易评价结果加权综合考虑，能更合理地体现本地节点对目标节点的信用评估。另外为了防止节点蓄意地抬高某些节点的本地信用值，本文定义了规格化的本地信用值(normalized local trust value) L_{ij} (在本文余下部分如无特殊说明均采用规格化的本地信用值)：

$$L_{ij}=\frac{l_{ij}}{\sum_j l_{ij}} \quad (3)$$

其中，$0<L_{ij}<1$ 且 $\sum_j L_{ij}=1$

3.1.2 节点内容相似度

P2P 网络中每个节点感兴趣的内容不尽相同。兴趣相同的节点由于有共同的爱好，会更重视和关注相互间的推荐。这便需要设计一种描述节点间相似度的数学模型。本文假设 P2P 网络中总共提供 m 种内容分类(比如音乐、计算机等)，分别为 C_1，C_2，…，C_m。每个节点可对任意的分类感兴趣。这里用分类向量 $\vec{V}_i=[a_1, a_2, \cdots, a_m]$ 表示节点 i 所感兴趣的内容分类情况，其中 $a_k=0$ 表示节点对 C_k 分类不感兴趣，$a_k=1$ 表示节点对 C_k 分类感兴趣($k\in[1,m]$)。

用节点分类向量 $\vec{V}$ 的相似度来表示节点的内容相似度。节点 i 的分类向量 $\vec{V}_i$ 与节点 j 的分类向量 $\vec{V}_j$ 的相似度可以用 $\vec{V}_i$ 与 $\vec{V}_j$ 夹角的余弦表示。因此，节点 A 与节点 B 的内容相似度定义为：

$$Simi_{ij}=\frac{\vec{V}_i\cdot\vec{V}_j}{\|\vec{V}_i\|_2\times\|\vec{V}_j\|_2}=\frac{\sum_{k=1}^{m}a_k\times b_k}{\sqrt{\sum_{k=1}^{m}a_k^2}\times\sqrt{\sum_{k=1}^{m}b_k^2}} \quad (4)$$

其中 $Simi$ 值将在 0 与 1 之间变化，当两个节点感兴趣的分类完全相同时，$Simi$ 值为 1；当两个节点感兴趣的分类完全不同时，$Simi$ 值为 0。

3.1.3 周围节点的推荐度

节点在评估目标节点的信用度时，不仅需要根据自身交易历史记录来判断，同时也应该广泛地征求其它节点，特别是与自己相似的节点的意见。因此，引入节点推荐度的定义：

$$R_{ij}=\sum_{k\in\delta(i)}L_{ik}\cdot Simi_{ik}\cdot L_{kj} \quad (5)$$

R_{ij} 表示节点 i 的周围节点对节点 j 信用度的综合评价。$\delta(i)$ 表示节点 i 的邻居域，邻居域不仅包括节点的直接邻居也包括邻居的邻居。在这个域内的节点把自己对节点 j 的本地信用提供给节点 i。而节点 i 在接受这个信用值之前需要衡量此邻居在节点 i 心目中的“份量”，这个“份量”通过 $L_{ik}\cdot Simi_{ik}$ 来衡量。从(5)式可以发现，当节点 i 对节点 k 的本地信用很低或两者的相似度很小时，无论节点 k 对节点 j 的评价如何，节点 i 都基本抛弃该值。需要说明的是本文后面介绍的自适应调整策略能使节点与其邻居的相似性尽可能高，从而在自身评价的基础上充分利用了邻居节点对目标节点的评价。

3.1.4 总体评价

$$T_{ij}=\eta\cdot L_{ij}+(1-\eta)\cdot R_{ij} \quad (6)$$

T_{ij} 表示节点 i 对节点 j 的总体评价。这里的 η 是节点的自信因子，当 η 为 1 时，表示该节点完全不在乎邻居的评价，只相信通过自己的历史交易得出的判断；当 η 为 0 时，表明该节点可能是新加入的节点，因为此时该节点缺少历史记录作为评价依据，只能完全相信邻居节点的推荐。当然，随着交易的进展，节点将从自己的交易结果中不断地积累经验，从而为将来交易做出更有价值的建议。

3.2 算法描述

在 P2P 文件共享网络中，当节点需要下载某个资源时，首先应该向其它节点发送查询请求，在获得资源提供者列表后根据 CSBTrust 模型选择高信用度节点交易。在交易过后，应该对交易结果按 3.1.1 节中的交易评价机制进行评价。另外，网络定期根据节点的历史交易情况自适应动态调整网络拓扑。主要算法如下。

3.2.1 服务查询

(1)查询原语：Query(SrcID, DestID, ReqServ, TTL)

(2)响应原语：Respond((SrcID, DestID, TTL)

文件共享系统中当节点需要请求某服务时，会产生一条查询消息，并向它的邻居转发。这里用原语 Query 表示，其中 SrcID 表示发起查询的源节点；DestID 为目标节点；ReqServ 表示所请求的服务；TTL 为该查询消息的转发跳数。收到该查询消息的节点一方面继续转发消息，另一方面如果自身存在该请求的资源，就回复一条应答消息，用原语 Respond 表示，其参数与 Query 类似，只不过这里的 DestID 为 Query 中的 SrcID，而这里的 SrcID 为应答节点。

3.2.2 选择交易对象

发出查询的节点将收到能提供服务的节点列表，然后需要对这些节点进行信用评估，从中选择信用度最高的节点下载资源。节点选择交易对象的过程如以下伪码所示(假设本地节点为 i)：

```
ChoosePeer(){
 评价队列置空;
for(每个候选节点 j){
  for(节点 i 的每个邻居){向邻居 k 请求 L_kj;计算 Simi_ik;}
  计算 R_ij, T_ij;
```

```
  评价队列 ←T_ij;
}
对评价队列排序;
返回节点 p(其中 T_ip = max_j(T_ij));
}
```

3.2.3 网络自适应策略

节点在文件共享过程中,将不断积累交易经验,在一段时间后,便能区分出哪些节点经常提供虚假资源,哪些节点提供真实资源。因此,网络定期根据 CSBTrust 模型得出的信用评价自适应动态调整网络拓扑。网络自适应策略如以下伪码所示(假设本地节点为 i):

```
Self-adapation(){
      disConnectNum=0;//断开的节点数
  for(节点 i 的每个邻居){
    if (L_ik<β){ //β是预先设置的一个阈值, k 是 i 的某个邻居
        断开与 k 的连接;
        disConnectNum++;
}
}
对 L_ip 降序排列(p 是与 i 交易过的节点),并把 p 放入队列 Q;
节点 j ← pop(Q);
while(disConnectNum>0){
   if(j 不是 i 的邻居 && Simi_ij>γ){//γ是预先设置的一个阈值
       把 j 设为 i 的邻居;
       disConnectNum--;
   }
 节点 j ←pop(Q);
 }
}
```

通过以上步骤可以逐渐地把恶意节点从节点的邻居中移去,并且把高信用度且与节点内容具有相似性的节点作为它的新邻居。这样做的好处是一方面节点在今后的请求查询中可以得到更可靠的邻居节点的推荐,另一方面可以增加请求查询的有效转发数及缩短资源的下载路径长度,从而提高下载质量和速度。

3.3 模型的评估

与 EigenTrust 相比,本模型不需要进行信用值的全局迭代。通过实验表明,算法收敛速度在同等条件下比 EigenTrust 快。另外,EigenTrust 在进行全局迭代时,为了获取网络中所有节点的信用值,需要通过邻居向全部节点发送信用请求,为了尽可能保证迭代效果,它需要把请求信用的消息 TTL 设置得足够大,才能覆盖全网络。因此它的消息复杂度为 $O(n^2)$。而本模型由于不需要全局迭代,只需向有限的邻居请求对其它节点的信用值,所以它的消息复杂度为 $O(n)$。需要指出的是本文的自适应调整策略使得节点的邻居是与之具有较高相似度的节点,从而在有限的信用请求范围内,仍可保证节点信用值计算的可靠性和合理性。

4 实验模拟结果与分析

为了检验 CSBTrust 模型的性能,本文以 Stanford 大学的 Query Cycle[8] 为基础搭建了一个 P2P 文件共享实验平台,并模拟实现了三种文件共享模型,分别是:没有引入信用机制的 NoTrust 模型;EigenTrust 模型;CSBTrust 模型。最后在不同的实验条件下,比较它们的性能。

4.1 实验基本参数与配置

默认情况下,系统中共有 100 个节点,2000 个文件,以及 10 个内容分类。2000 个文件平均分配给每个分类。初始状态下,每个节点随机选择自己感兴趣的分类和文件。模拟器执行若干个模拟周期(Cycle),节点在每个周期内向其它节点请求下载自己还未拥有的文件,这里的消息 TTL 设为 4。然后从返回的节点列表中选择信用度最高的节点下载。每个周期结束后,对交易结果进行汇总。本文对相同模拟过程进行多次模拟,取平均值分析。

4.2 文件下载质量的分析

系统中的节点分为两大类:一类是诚实节点,其提供真实文件上传服务,对其它节点进行公正的评价。一类是恶意节点,或提供虚假文件或对它节点进行不公正的评价。根据恶意节点的不同行为本文考虑了两种威胁模型(Threat Model)[6]:

4.2.1 个体威胁

这里的恶意节点收到其它节点发来的下载请求后,只上传虚假文件。为了检验不同数量规模的恶意节点对信用模型的影响,本文作了如图 1 所示的实验。当恶意节点数为 0 时,理想情况下节点下载到的都是真实文件。随着恶意节点所占比例的增加,总体上各系统虚假文件下载比例逐渐上升。NoTrust 模型的节点随机选择目标下载文件,所以其虚假文件下载率在三种模型中最高。而 CSBTrust 模型可以根据节点的交易经验来最大程度区分为恶意节点和诚实节点,从而丢弃恶意节点,达到惩罚恶意节点的目的。从图 1 可看到 CSBTrust 的效果要比 EigenTrust 好。

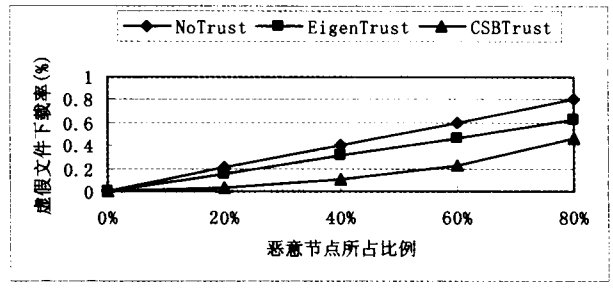

图 1 个体威胁下不同模型的虚假文件下载情况比较

4.2.2 诋毁与夸大

这类恶意节点不仅提供虚假文件上传,而且对与其它节点的交易结果进行和事实相反的评价,比如诋毁诚实节点的可信度,蓄意夸大其它恶意节点的可信度。由于 NoTrust 模型没有采用信用机制,恶意节点的任何评价都不影响节点的下载选择,所以本次实验只针对 EigenTrust 和 CSBTrust。为此,实验中总共执行 20 次查询,恶意节点占总节点的 20%。如图 2 所示,在刚开始的 3 个周期内,一方面节点选择下载比较盲目,另一方面 EigenTrust 进行的是全局迭代,而 CSBTrust 的内容相似性与自适应策略在运行初期无法很快体现出优势,所以虚假文件的下载数比 EigenTrust 多。但随着周期数的增加,节点的交易经验不断帮助其调整网络拓扑,断开与恶意节点的连接,建立与诚实节点的新连接,从而可以从可信邻居获得更多的诚实推荐,所以 CSBTrust 的虚假文件下载数越来越少。

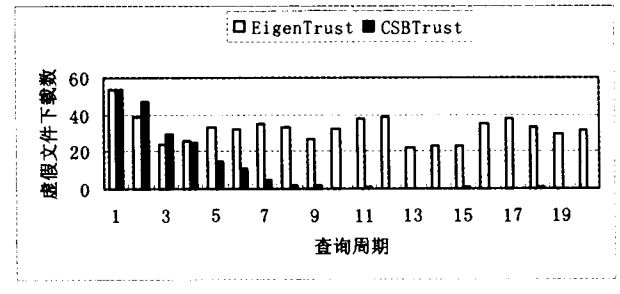

图 2 诋毁夸大情况下不同模型的虚假文件下载数比较

(下转第 133 页)

第三选择：S2-因为有精确的功能匹配和比较低的信誉匹配。

因此，通讯管理器将返回 S1 作为最佳的匹配。

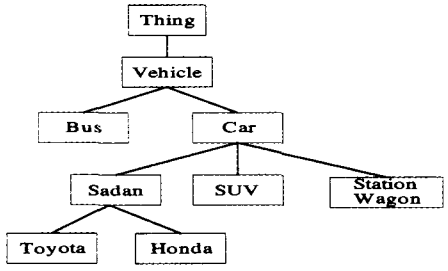

图 3 一个轿车本体

```
<profile:Profile rdf:ID="RequestToyotaSellService">
  <input>
    <profile:ParameterDescription rdf:ID="Price_Input">
      <profile:parameterName>Price</profile:parameterName">
      <profile:restrictedTo rdf:resource="Concepts.daml#Price"\>
    </profile:ParameterDescription>
  </input>
  <output>
    <profile:ParameterDescriotion rdf:ID="Car_Output">
      <profile:parameterName>ToyotaSedan</profile:parameterName">
      <profile:restrictedTo rdf:resource="Vehicle.daml#ToyotaSaloon"\>
    </profile:ParameterDescription>
  </output>
  <rns:ReputationAttributes>
    <rns:attribute>reliability</rns:attribute>
    <rns:attribute>availability</rns:attribute>
  </rns:ReputationAttributes>
</profile:Profile>
```

图 4 包含信誉参数的服务请求例子

结束语 本文用一个结构化的覆盖网络作为分布式的服务注册系统来提高其可扩展性和有效性。我们提出一个服务发现模型，利用语义 Web 技术促进基于信誉的服务发现和匹配。框架包括通讯管理模块、匹配模块、组合模块、信誉模块。文中详细描述了这些组件并提出了一个基于用户反馈的信誉评估算法。最后通过一个例子展示了此算法的使用。比较此领域已做的工作，该模型的优点是为服务信誉度加入了语义信息，并综合了第三方权威机构的评价，使得信誉度值更有效和准确。

本文的工作是我们正在进行研究工作的一部分，目的在于为服务组合机制提供一个开放的、可重用的基础结构。今后，我们计划进一步改进我们的服务发现模型，特别是在复杂环境下的信誉管理方案，并在真实的 Web 服务整合环境下配置这个模型，考察它的实用性。

参 考 文 献

1 Papazoglou M P, Georgakopoulos D. Service-oriented computing. Communications of the ACM, 2003,46

2 Bilgin A S, Singh M P. A DAML-Based Repository for QoS-Aware Semantic Web Service Selection. In: Proceedings of ICWS'04

3 Kalepu S, Krishnaswamy S, Loke S W. Reputation = f(User Ranking, Compliance, Verity). In: Proceedings of ICWS'04

4 Singh M P, Huhns M N. Service-Oriented Computing. Wiley, 2005

5 Liu Y, Ngu A, Yeng L. QoS Computation and Policing in Dynamic Web Service Selection. In: Proceedings of WWW, 2004

6 Vu Le-Hung, Hauswirth M, Aberer K. Towards P2P-based Semantic Web Service Discovery with QoS Support. In: Proceeding of Workshop on Business Processes and Services (BPS), Nancy, France, 2005

7 Vu Le-Hung, Hauswirth M, Aberer K. QoS-Based Service Selection and Ranking with Trust and Reputation Management. OTM Conferences,2005(1): 466～483

8 Day J, Deters R. Selecting the Best Web Service. In: the 14th Annual IBM Centers for Advanced Studies Conf., 2004

9 Paolucci M, Kawamura T, Payne T, Sycara K. Semantic matching of Web services capabilities. In: Proceedings of the 1st International Semantic Web Conference (ISWC)s, 2002

(上接第 94 页)

4.3 请求信用值的消息转发分析

EigenTrust 在计算全局信用值的时候需要全局迭代，严重的情况下将产生洪泛。而 CSBTrust 只在可信邻居节点范围内请求信用值，因此它的消息转发数是很小的。从图 3 中可知随着系统总节点数的增加，EigenTrust 的信用消息转发数将成指数级增长，很难在大规模网络实施。而 CSBTrust 模型的消息数基本保持在小范围内变化，因此，开销小，具有很好的可扩展性。

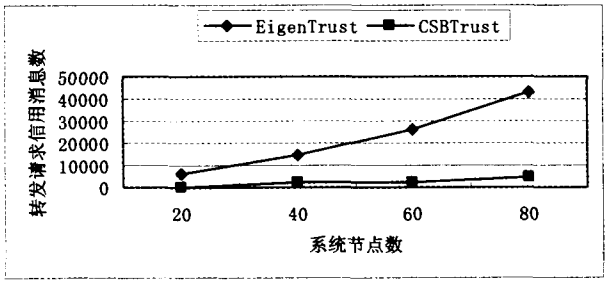

图 3 两种模型的请求信用消息转发数比较

结论与进一步工作 本文在已有模型的基础上构建了一个对等网络信用模型，该模型利用节点的内容相似性来评价节点信用度，并根据交易经验自适应动态调整邻居节点，这种方式能充分地利用可信节点的推荐来选择下载服务。实验表明该模型与同类模型相比，有效提高了下载服务的质量，并且具有良好的扩展性和较低的运行开销。

另外，本文主要从请求服务的节点角度出发评价节点信用，实际上提供服务的节点也需要对请求者进行信用评估，抵制 free-riding 行为，鼓励用户提供优质服务。因此，以此为出发点进一步完善整个信用模型将是作者接下去的主要研究目标。

参 考 文 献

1 Kazaa. http://www.kazaa.com

2 Liang J, Kumar R, Xi Y, Ross K. Pollution in P2P file sharing systems. In: Proceedings of IEEE INFOCOM 2005. IEEE Press, 2005,2:1174～1185

3 Xiong L, Liu L. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. IEEE Transactions on Knowledge and Data Engineering, 2004,16(7):843～857

4 Cornelli F. Choosing reputable servents in a P2P network. In: Lassner D, ed. Proc. of the 11th Int'l World Wide Web Conf. Hawaii: ACM Press, 2002. 441～449

5 窦文，王怀民，贾焰，等. 构造基于推荐 Peer-to-Peer 环境下的 Trust 模型. 软件学报,2004,15(4):571～583

6 Kamvar S D, Schlosser M T. EigenRep: Reputation management in P2P networks. In: Lawrence S, ed. Proc. of the 12th Int'l World Wide Web Conf. Budapest: ACM Press,123～134

7 袁巍，李津生，洪佩琳. 一种 P2P 网络分布式信任模型及仿真. 系统仿真学报,2006,18(4):938～942

8 Query Cycle. http://p2p.stanford.edu/www/qcsim.htm