构造基于推荐的 Peer-to-Peer 环境下的 Trust 模型*

窦 文+, 王怀民, 贾 焰, 邹 鹏

(国防科学技术大学 计算机学院,湖南 长沙 410073)

A Recommendation-Based Peer-to-Peer Trust Model

DOU Wen+, WANG Huai-Min, JIA Yan, ZOU Peng

(School of Computer Science, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: Phn: +86-731-4512852, E-mail: douwen@vip.sina.com, http://www.nudt.edu.cn

Received 2003-04-28; Accepted 2003-06-30

Dou W, Wang HM, Jia Y, Zou P. A recommendation-based peer-to-peer trust model. *Journal of Software*, 2004,15(4):571~583.

http://www.jos.org.cn/1000-9825/15/571.htm

Abstract: For most peer-to-peer file-swapping applications, sharing is a volunteer action, and peers are not responsible for their irresponsible bartering history. This situation indicates the trust between participants can not be set up simply on the traditional trust mechanism. A reasonable trust construction approach comes from the social network analysis, in which trust relations between individuals are set up upon recommendations of other individuals. Current p2p trust model could not promise the convergence of iteration for trust computation, and takes no consideration for model security problems, such as sybil attack and slandering. This paper presents a novel recommendation-based global trust model and gives a distributed implementation method. Mathematic analyses and simulations show that, compared to the current global trust model, the proposed model is more robust on trust security problems and more complete on iteration for computing peer trust.

Key words: peer-to-peer; trust; distributed hash table; sybil attack; slandering

摘 要: 在诸如文件共享等无中心的 Peer-to-Peer 环境中,资源共享是用户自愿的行为.在这类系统中,由于用户不为自身的行为担负(法律)责任,因而节点间的信任关系往往很难通过传统的信任机制建立.一种更合理的考虑是参考人际网络中基于推荐的信任关系建立方法.现有的模型不能很好地解决模型的迭代收敛性问题,同时缺乏对诸如冒名、诋毁等安全性问题的考虑.针对上述问题,在节点推荐的基础上提出了一种基于 Peer-to-Peer 环境的信任模型,并给出了该模型的数学分析和分布式实现方法.分析及仿真表明,该信任模型较已有模型在迭代的收敛性、模型的安全性等问题上有较大改进.

关键词: 对等网络;信任;分布 Hash 表;冒名;诋毁

* Supported by the National Grand Fundamental Research 973 Program of China under Grant No.G1999032703 (国家重点基础研究发展规划(973)).

作者简介: **窦文**(1968－),男,重庆人,博士,讲师,主要研究领域为分布计算;**王怀民**(1962－),男,博士,教授,博士生导师,主要研究领域为分布式计算,人工智能,Agent 计算;**贾焰**(1960－),女,教授,博士生导师,主要研究领域为分布式计算,超大规模数据库,并行数据库;**邹鹏**(1957－),男,教授,博士生导师,主要研究领域为分布操作系统,分布式计算.

中图法分类号: TP393 文献标识码: A

虽然目前的 Peer-to-Peer(简称 P2P)应用日益广泛,但仍然缺乏有效的机制以提高系统整体的可用性,这非常显著地表现为应用中大量欺诈行为的存在以及不可靠的服务质量[1].以众多的文件共享应用为例,25%的文件是伪造文件(faked files),同时,不负责任的用户随意地中止(文件上载)服务,使得服务质量无法得以较好的保证.一种可能的办法是对用户评定信任等级,在多个同样服务可选的情况下,信任等级高的节点成为首选.

在传统的网络环境(如 Internet)和应用(如电子商务)中,信任关系的建立依赖于可信的第三方,比如认证中心(CA).只要个体持有该 CA 所颁发的证书即被认为是可信的,同时,恶意用户必须承担(法律)责任.然而,在目前广泛存在的 Peer-to-Peer 环境中(如文件共享应用),有一种排斥 CA 的倾向,这主要基于以下几点考虑:(1) 集中式的认证往往伴随着额外的费用和开销,而 P2P 环境通常追求零开销[2](zero-dollar-cost certificates),用户自愿参与网络、自由交易并且不准备为自己的行为负(法律)责任;(2) 对单点失效的顾虑,这里指的单点失效有两方面的含义:(a) 物理上的,即可信(认证)服务器的崩溃导致整个 P2P 系统的崩溃;(b) 社会或法律意义上的单点失效,即由于政治、法律等原因导致可信(认证)服务器无法正常工作从而致使 P2P 系统崩溃(Napster 的崩溃即是一个此类例子).因此,对于目前日益广泛的诸多 Peer-to-Peer 环境,建立一种新的分布式信任机制是十分必要的.这种必要性不仅体现在用户对 P2P 网络的有效使用上,也体现在有利于网络的良性发展上.

在社会网络中,信任关系是人际关系的核心,个体间的信任度往往取决于其他个体的推荐,同时,作为推荐者的可信度也决定其推荐个体的可信度.实际上,这种互相依赖的信任关系组成了一个所谓的信任网络[3](Web of trust).在这样的信任网络中,任何个体的可信度都不是绝对可靠的,但可以作为其他个体决定其交互行为的依据.基于信任网络的 Peer-to-Peer 系统与人际网络有很大的相似性[4],这表现在:(1) 网络中的个体在与其他个体的交互中会留下零星的“信用”信息;(2) 个体对于交互对象具有充分的选择权;(3) 个体往往不看重绝对的可靠性或服务质量,即个体可以忍受少量错误的选择带来的损失,比如文件共享应用;(4) 个体有义务为网络中的其他个体提供推荐信息.因此,这为借鉴社会学研究的某些结论提供了可能.

本文旨在构造一个用于 Peer-to-Peer 环境的全局信任模型,并给出该模型数学表述和分布式实现方法,从而使任意节点可以随时地、较方便地获取其他节点的全局可信度.分析和仿真说明,本文提出的模型不仅克服了已有模型的部分局限性(见第 1 节),而且具有较好的工程可行性.

1 相关工作

目前存在若干基于 Peer-to-Peer 环境的信任模型,可以归为以下几类:

(1) 基于 PKI[5]的信任模型,在这类系统中,存在少数领袖节点(leader peers),领袖节点负责整个网络的监督,定期通告违规的节点.这些领袖节点的合法性通过 CA 颁发的证书加以保证.正如我们前面所述,这类系统往往是中心依赖的,具有可扩展性、单点失效等问题.如 eDonkey 的诸多 server[6].

(2) 基于局部推荐.在这类系统中,节点通过询问有限的其他节点以获取某个节点的可信度.在这类系统中,往往采取简单的局部广播的手段,其获取的节点可信度也往往是局部的和片面的.如 Cornelli 对 Gnutella 的改进建议[7]中采用的就是这种方法.

(3) 数据签名.这种方法不追求节点的可信度,而是强调数据的可信度,以文件共享应用为例,在每次下载完成时,用户对数据的真实性进行判定,如果认可数据的真实性,则对该数据进行签名,获取签名越多的数据(文件),其真实性越高.然而,该方法仅针对数据共享应用(如文件共享),同时无法防止集体欺诈行为,即恶意的群体对某不真实的数据集体签名.目前流行的文件共享应用 Kazaa 采用的就是该方法(Sig2Dat[8]).

(4) 全局可信度模型.为获取全局的节点可信度,该类模型通过邻居节点间相互满意度的迭代,从而获取节点全局的可信度.据我们所知,Stanford 的 eigenRep[9]是目前已知的惟一与我们的模型相似的全局信任模型.因此我们这里会进行较为详细的讨论.

EigenRep 的核心思想是,当节点 i 需要了解任意节点 k 的全局可信度时,首先从 k 的交易伙伴(曾经与 k 发

生过交易的节点)获知节点 k 的可信度信息,然后根据这些交易伙伴自身的局部可信度(从 i 的眼光来看)综合出 k 的全局可信度.即

$$T_k=\sum_j (C_{ij}\times C_{jk}) \tag{1}$$

对于任意节点 I,j,C_{ij} 为节点 i 对节点 j 的局部信任度,T_i 为节点 i 全局的可信度.

$$C_{ij}=\frac{Sat_{ij}-UnSat_{ij}}{\sum_j (Sat_{ij}-UnSat_{ij})} \tag{2}$$

Sat_{ij} 和 $UnSat_{ij}$ 分别为节点 i 对 j 在历史交易中积累的满意次数和不满意次数.

EigenRep 与本文提出的模型在两点上具有相似性,一是都试图通过迭代方法计算节点的全局可信度,二是通过分布 Hash 机制放置节点的全局可信度(见第 3.1 节).然而,eigenRep 没有解决以下几个问题:

(1) 迭代的可收敛性问题.EigenRep 模型没有对迭代的收敛性作出确定性保证.即没有就矩阵 $C=\left|C_{ij}\right|_n$ 讨论迭代的收敛性.事实上,通过式(2)构造的线性方程组(1)不满足简单迭代的收敛充分条件[9].EigenRep 提出的一个补救策略是,假定网络中始终预先存在一个固定的亚可信的节点集合 P,P 中的节点拥有至少 $T_{i(i\in P)}>\psi$ 的全局可信度.在该假定的前提下,式(1)变为

$$T_k=(1-\alpha)\sum_j (C_{ij}\times C_{jk})+\alpha\, t_i \tag{3}$$

其中 $\begin{cases} t_i=0, & \text{if } t_i\notin P \\ t_i=\dfrac{\psi}{\alpha}, & \text{otherwise} \end{cases}$,$0<\alpha<1$,这保证了 C 的不可约性和非周期性,从而保证了式(3)计算的收敛性.我们认为,该假定的合理性值得商榷,因为这本质上使得这些节点拥有了“先天”的特权,同时,指定哪些节点组成集合 P 也是一个较难操作的问题.

(2) 该模型没有考虑惩罚因素,即模型没有对造成交易失败的节点在信任度上作出惩罚.考察式(2)和式(3),如果节点 i 对 j 的不满意程度增加,即 $Unsat_{ij}$ 增加,T_{ij} 甚至可能增加,这显然缺乏合理性.

(3) 该模型的协议实现没有考虑网络的性能开销,每次交易都会导致在全网络范围内的迭代,因此,该模型在大规模网络环境中缺乏工程上的可行性.

(4) 该模型及协议没有充分考虑安全性问题,例如对于冒名、诋毁以及协同欺诈防止(见第 3 节分析)等问题,eigenRep 没有加以讨论.

本文针对 eigenRep 存在的问题,提出了一种新的全局信任度模型,并给出了相关分析和分布计算协议,最后对模型进行了仿真检验.

2 全局信任模型

2.1 模型的定义和表示

首先给出本文对信任的定义.

定义 1. 设 P_{ij} 代表节点 i 对节点 j 的局部看法,即局部信任度.该看法来自于节点 i 与 j 的交互历史.这里设 $P_{ij}=\dfrac{S_{ij}}{I_{ij}}$,其中 I_{ij} 为节点 i 与 j 在最近某个固定时间 τ 内(τ 随具体应用而定,例如 1 个月)实际交互的次数,S_{ij} 为在节点 i 看来交易成功的次数,F_{ij} 为在节点 i 看来交易失败的次数.如果 $I_{ij}=0$,则设 $P_{ij}=0$.

引入 τ 表示我们的模型更注重节点近期的行为.S_{ij} 的更新需要用户自身的干预,因为只有用户自己才能决定某次交易是否成功.

定义 2. 信任 O_{ij} 代表了节点 i 对节点 $j(i\neq j)$(对于 $i=j$ 的情况,P_{ij} 无实际意义,设 $O_{ij}=1$)的信任程度在信任范围 λ 中的投影,设 $O_{ij}=[\alpha P_{ij}+(1-\alpha)T_j]\times\lambda$,这里,$T_j$ 为节点 j 的全局可信度,$T_j\leq 1$.其中 α 为常量且 $0<\alpha<1$,λ 为信任范围,$\lambda>1$.

对一个刚愎的用户而言,可设 $\alpha=1$,这时该用户只相信自己的交互历史,同时由于其交往有限,从而选择余地

较小.对于一个无主见的用户而言,可设α=0,这时该用户完全依靠其他用户的意见来决定取舍.一般地,可设α=0.5.

定义 3. 称二元组$\langle S_{uv}, F_{uv}\rangle$为节点 u 对节点 v 的评价,表示为 E_{uv}.

定义 4. 称 $R_{ij}=\dfrac{S_{ij}-F_{ij}}{\sum_k S_{kj}}$ 为节点 i 对节点 j 的推荐度(recommend degree).如果 $\sum_i S_{ij}=0$,或 $S_{ij}-F_{ij}<0$,则设 $R_{ij}=0$.

定义 5. 称矩阵 $R=\left|R_{ij}\right|_n$ 为网络的信任关系矩阵.

任意节点的全局可信度由与之发生过交易行为的其他节点对它的局部看法以及这些节点的全局可信度来决定.如果以加权有向图 $G\langle V,E\rangle$来表示这种交互关系,设 $|G|=n$,$V=\{i|\exists j,$且 $i\xleftrightarrow{R_{ij}}j\}$,其中 $i\xleftrightarrow{R_{ij}}j$ 表示节点 i,j 发生过交易.

至此,我们实际上构造了一个以加权有向图形式表示的社会网络,在该网络中,个体间的关系(tie)为由交易带来的信任关系,关系的强度为个体间的局部信任度.这里,我们采用社会网络分析[10]中的基于节点入度(in-degree)的中心性测量(centrality measurement)[11]方法来求解网络中节点的全局可信度.Bonachi[12]提出的基于节点入度的中心性测量方法的核心是:(1) 所谓中心性(centrality)是指节点在网络中的重要性,重要性在不同的上下文中(即不同的关系网络中)具有不同的语义,在本文中,节点的中心性体现为节点的全局可信度,因为该网络中的关系为节点间交易带来的信任关系.(2) 通过节点的入度、相应的权值(R_{ij})以及(推荐)节点自身的重要性来判断目标节点的重要性(全局可信度).换句话说,越可信的节点给出的推荐越重要;同时,推荐者的数目越多,表示目标节点越重要(可信).这两点是相互关联的,少量的可信推荐者给出的评估可能比大量不可信推荐者(入度)给出的评估更重要,反之亦然,这需要根据具体的量而定.(3) 通过建立关系网络邻接矩阵 P 以及定义节点的中心性向量 X,可得线性方程组 $P^TX=X.P^T$ 为 P 的转置,X 为节点的全局中心性向量,在本文的上下文中,X 即为节点的全局可信度向量.

定义 6. 网络 N 中对于任意节点 i 的全局可信度 T_i,设

$$T_i=\sum_k (R_{ki}\times T_k) \tag{4}$$

其中 k 为与 i 发生过交易的节点.

设全局可信度向量 $T=\begin{bmatrix}T_1\\T_2\\\cdots\\T_n\end{bmatrix}$,

则称

$$R^TT=T \tag{5}$$

为网络 N 关于信任关系矩阵 R 的信任方程.由式(4)、定义 4 以及迭代收敛性可以保证 $0\le T_i\le 1$.其中,R 为信任关系矩阵,其矩阵元素 R_{ij} 为节点 i 对节点 j 的推荐度.

2.2 全局可信度的计算

暂不考虑分布的情况,对方程(5)解的计算可以通过传统的迭代方法来实现,例如,Jacobi 迭代、Gauss-Seidel 迭代等.迭代方法的选取主要看其是否适合于分布式环境.我们认为 Jacobi 迭代或 Gauss-Sediel 迭代都是可行的.这里我们仅考虑 Jacobi 迭代.

定理. 网络关于其信任关系矩阵 R 的信任方程的 Jacobi 和 Gauss-Sediel 迭代收敛.

证明:设 $H=R^T$,则 $HT=T$,即 $(I^n-H)T=0$ 收敛的充分条件为

$$\max_{1\le i\le n}\sum_{j=1}^{n}\left|H_{ij}\right|<1,$$

由式(2)得,对于矩阵 H 的每一列,

$$\sum_j H_{ij} = \sum_{\substack{j \\ i \neq j}} R_{ji} = \sum_{\substack{j \\ i \neq j}} \frac{\max(S_{ji} - F_{ji}, 0)}{\sum_k S_{kj}} < 1.$$

因此,$I^n - H$ 为严格对角占优矩阵.定理成立. □

因此,我们可得到对于任意节点全局可信度的 Jacobi 迭代:

$$T_1^{(k+1)} = R_{12}T_2^{(k)} + R_{13}T_3^{(k)} + \ldots + R_{1n}T_n^{(k)},$$
$$T_2^{(k+1)} = R_{21}T_1^{(k)} + R_{23}T_3^{(k)} + \ldots + R_{2n}T_n^{(k)},$$
$$\ldots,$$
$$T_n^{(k+1)} = R_{n1}T_1^{(k)} + R_{n2}T_2^{(k)} + \ldots + R_{n\,n-1}T_{n-1}^{(k)}.$$

3 全局可信度的分布求解协议

全局可信度的分布求解涉及 3 个主要问题,即:

(1) 安全性.这里,安全性包括冒名、诋毁以及协同作弊的防止.冒名者可能伪称高可信度的节点从而对整个系统的可用性造成破坏.协同作弊是指多个节点互相为对方提供过高的推荐值,从而破坏整个系统的可用性.诋毁是指恶意节点通过给予其他节点不真实的较低评价从而对系统造成间接损害(见第 3.3 节).

(2) 协议本身.即节点间如何协作完成迭代算法.

(3) 全局可信度的放置.即全局可信度在网络中如何放置使得任意节点 i 可以以有限的开销随时获取其他节点 j 当前的全局可信度 $T_j^{(k)}$.值得一提的是,$T_j^{(k)}$放置在 j 节点本身或者让节点 j 自己选择放置节点是不可取的,这可能导致欺诈行为.同时,正如上文所述,一个集中式的放置方案也不符合本文的初衷.

3.1 全局可信度的放置

本文通过分布 Hash 表(DHTs)[13]机制用于放置节点的全局可信度(值得一提的是,eigenRep 同样使用了 DHTs 放置策略,但没有就放置细节作细致的表述.eigenRep 使用 Berkeley 的 DHTs 协议 CAN[15]).在本文的实现中,我们引入 Terrace[14]P2P 网络构件,Terrace 是我们构造的基于 d-tree 结构的 DHTs 构件,通过 Terrace,网络中的所有节点投影到一个逻辑 d-tree 上,并赋予节点全局惟一的逻辑地址,逻辑地址的编码基数为 d-tree 的阶,例如,如果用八进制表示逻辑地址,则 d 为 8.树的最大空间由系统规模决定,例如,设逻辑空间为 IPv4 地址空间的投影,则逻辑空间的大小为 $0\sim2^{32}-1$.空间(树)中每个节点拥 d 个逻辑地址,逻辑地址为标量地址.以图 1 为例,图中地址由八进制表示,则除根节点最大扇出为 7 外,其余逻辑树节点的最大扇出为 8.从根节点起,每一个节点的子节点都是该节点逻辑地址在下一个基数位的列举,而节点自身包含当前基数位的所有列举(根节点例外,它没有第 0 位子树).通过均匀的 Hash,节点可以将对象(或对象索引)投影到同样的逻辑地址空间.

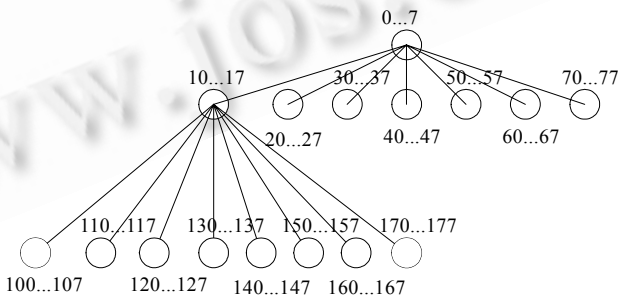

Fig.1 P2P nodes are projected into a d-tree based logic space, d is the upper bound of child trees a node could possess, and it is also the base of logic address

图 1 P2P 节点投影到一个由 d-tree 构成的逻辑空间上,其中 d 为树节点的最大子树个数,同时 d 也是逻辑地址基数

就本文而言,Terrace 提供了如下相关特征:

(1) 任意节点 i 通过 Terrace 在 $O(\log N)$的消息复杂度内将节点 j 的属性(如评价 E_{ij})写入 $H(\mathrm{ID}_j)$,同时保证节

点 i,j 难以获知 $H(\mathrm{ID}_j)$的具体位置(IP 地址).这是由于节点在 Terrace 中的逻辑地址与其物理地址无关.其中,H 为 Hash 函数,它将节点 j 的标识 ID_j Hash 到 Terrace 的逻辑空间.N 为网络规模,下同.如图 2(a)所示.其中,图 2(a)中节点 F 希望将其对节点 B 的评价 E_{FB} 写入 B 的档案点,由 Hash 得 B 的档案点为逻辑地址,为 121,则由 Terrace 的写入机制,E_{FB} 被写入节点 E.图 2(b)中节点 F 希望获取节点 B 的全局信任度 T_B,由 Hash 得 B 的档案点为逻辑地址,为 121,则由 Terrace 的路由机制,F 由 B 的档案点 E 获取节点 B 的全局信任度 T_B.

(2) 在不知 j 的 IP 地址的情况下,任意节点 i 可以按照 $O(\log N)$的消息复杂度从逻辑地址 $H(\mathrm{ID}_j)$获取节点 j 的有关数据(如全局可信度 T_j).其中,H 将节点 j 的标识 ID_j Hash 到 Terrace 的逻辑空间.如图 2(b)所示.

(3) Terrace 具有容错能力,即 Terrace 以较高的概率保证上面(1)、(2)操作的完成.限于篇幅,这里我们不对 Terrace 的容错机制作进一步阐述.建议有兴趣的读者参见文献[14].

(4) Terrace 具有较小的拓扑维护开销($O(d)$).

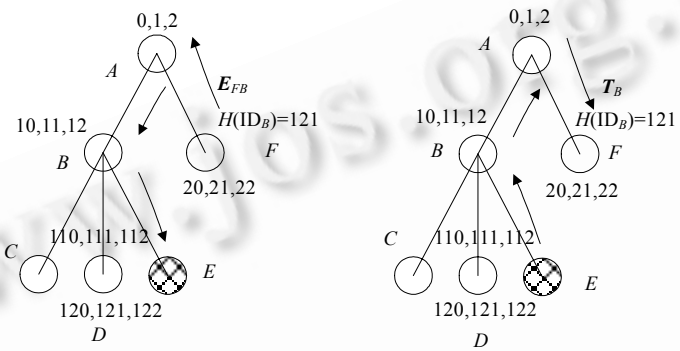

Fig.2 A part of logic space of 3-degree Terrace tree (Ternary)

图 2 三阶 Terrace 树(部分)逻辑空间(三进制)

定义 7. 设 HTD 为任意均匀的 Hash 函数(在仿真实现中我们采用 SHA-1),网络中任意节点 i 的标识在 Terrace 中的投影称为 i 的档案点 D_i,即 $D_i=HTD(\mathrm{ID}_i)$.ID_i 为全局惟一的节点标识符.

由于 Terrace 采用了单向 Hash 的方法,即节点加入拓扑时获取逻辑地址是随机的,无法根据节点的某个特征(如 IP 地址等)预先决定节点的逻辑地址,这为可信度的匿名放置的安全性带来一定优势.相对于 Terrace,CAN 和 Chord[16]的节点加入采用了双向 Hash 的方法,节点在拓扑中的逻辑地址由节点的某些属性预先决定(如 Chord 中节点的逻辑地址由其 IP 地址的 Hash 值决定,CAN 由节点的 ID 决定),因此,节点在拓扑中的逻辑位置固定,造成了任意节点和其档案点之间存在一一对应关系(忽略拓扑的动态因素),从而为节点与其档案点之间的协同作弊带来便利.

基于 Terrace,网络中的任意节点同时也是其他节点的档案点,事实上,全局信任度的迭代是通过档案点之间的协同完成的.因此,每个档案点 d 至少需要包含一个如图 3 所示的数据结构.

ID_r		$T_r^{(k+1)}$	P_k^r
$S_{k_1 r}$	$F_{k_1 r}$	$T_{k_1}^{(k)}$	ID_{k_1}
$S_{k_2 r}$	$F_{k_2 r}$	$T_{k_2}^{(k)}$	ID_{k_2}
…	…	…	…
$S_{k_T r}$	$F_{k_T r}$	$T_{k_T}^{(k)}$	ID_{k_T}

Fig.3 The structure of documentary point of peer d

图 3 档案点 d 的数据结构

如图 3 所示,节点 d 是网络中节点 r 的档案点,ID_r 为 r 的标识; $S_{k_1 r}\ldots S_{k_T r}$ 和 $F_{k_1 r}\ldots F_{k_T r}$ 为与 r 发生过交易的节点所报告的成功和失败次数; $\mathrm{ID}_{k_1}\ldots\mathrm{ID}_{k_T}$ 为 r 的推荐节点的标识, $T_{k_1}^{(k)}\ldots T_{k_T}^{(k)}$ 为推荐节点目前的全局可信度,$T_r^{(k+1)}$为由 d 计算出的节点 r 目前的全局可信度.P_k^r 为节点 r 的公钥(见第 3.3 节).

3.2 分布求解协议

首先给出协议的几个原语及其语义:

Put(ID_v,ID_u,E_{uv},f): 节点 u 将对节点 v 的评价 E_{uv} 写入 Terrace 树中逻辑地址为 HDT(ID_v)的档案点,并触发该节点上的 f 过程,HDT 为 Hash 函数,下同.

Get(ID_v,T_v): 从 Terrace 树中逻辑地址为 HDT(ID_v)的档案点数据结构中读取 v 的全局可信度并写入本地 T_v 变量.用户通过该原语(获取)判断任意节点 v 当前的全局可信度.

Eval-Trans(ID_v): 节点每次与其他节点进行交易,最终通过该原语评估交易结果,ID_v 为交易对方,Ecal-Trans(ID_v)为 true 表示交易成功,否则表示交易失败.

如上文所述,任意节点 u 同时具有两个角色,即既是用户节点,同时也是某(几)个用户的档案点.任意节点 u 作为一般用户节点的算法如下.

Procedure Eval(ID_v) //Eval(ID_v)是每次用户节点 u 与节点 v 交易后必须执行的评估过程

if(Ecal-Trans(ID_v)=true) then

$S_{uv}\leftarrow S_{uv}+1$

else

$F_{uv}\leftarrow F_{uv}+1$

endif

Put(ID_v,ID_u,E_{uv},ReCalcTrust);

End

任意节点 u 作为档案点的算法如下:

Procedure ReCalcTrust(ID_v,ID_u,E_{uv})

将 ID_u,E_{uv} 存入档案点数据结构;

for (any $j\neq v$)

Get(ID_j,T_j);

If($S_{uj}+F_{uj}\neq 0$)

$$R_{jv}\leftarrow \frac{S_{jv}-F_{jv}}{\sum_j S_{jv}}$$

endif

endfor

$$T_v\leftarrow \sum_k R_{kv}\times T_k$$

end

从以上分布求解协议可以看出,当节点 u,v 进行完一次交易后(可能成功或失败),交易的结果会直接影响彼此的全局信任度 T_u,T_v,从而间接影响 u,v 作为推荐节点的可信度,最终影响其他节点的全局可信度(对于单向交易,如文件共享应用,交易结果只影响服务提供者的可信度).

在 eigenRep 中,任意节点 i 的任意一次交易都会引起迭代,迭代通过其交易伙伴在全网络范围扩散,直到所有节点的全局可信度在连续两次迭代的结果小于某个系统指定的极小常量,其消息复杂度为 $O(n^2)$.消息开销造成该协议仅仅适用于小规模网络.

通过上述协议描述可以看出,与 eigenRep 不同,在我们的协议中,任意节点 u 只需通过 Get(ID_v,T_v)即可获取节点 v 的全局可信度,交易结束后,通过 Put(ID_v,ID_u,E_{uv},f),将新的评价 E_{uv} 写入 v 的档案点,引发关于对 v 可信度的重新计算,在重新计算的过程中需要获取与 v 有过交易的其他节点的全局可信度(使用 Get 原语),因此,消息复杂度为 $O(n)$(实际上,节点 v 的交易伙伴往往远远小于网络的规模,我们甚至可以只把在最近一段时间 τ 以内与 v 交易的节点作为 v 的交易伙伴,从而进一步减少消息开销(参见定义 1)).可以看出,上述协议本质上是一种“交易

驱动”的“部分迭代”,网络中发生的交易越多、越频繁,就迭代得越完全.事实上,在一个交易稀少的网络中,如eigenRep 协议的“全面迭代”并没有实际意义,因为节点间交易稀少意味着绝大多数节点的可信度变化不大.第4.1 节的实验说明,我们的协议与 eigenRep 相比,具有相当的迭代收敛特性.

3.3 冒名、诋毁以及协同作弊的抑制

冒名是指全局可信度低的节点冒充全局可信度高的节点,并企图:(1) 通过欺骗其他节点的档案点来影响其全局可信度;(2) 通过欺骗其他节点使之与其进行(恶意)交易.

冒名现象会严重影响 P2P 网络的良性发展,恶意的个体可以通过冒充一个或多个高可信节点获取不应得的利益,同时随意作出伤害其他个体的行为.EigenRep 协议完全回避了冒名情况的处理,然而事实是,冒名可能对 eigenRep 协议产生严重的不良影响(见第 4.4 节).

冒名最直接的解决办法是采用认证机制.由于前文所述原因,本文不考虑引入基于可信第三方认证机制,从而使得问题相对复杂.例如,节点 i,j 都向节点 p 声称自己具有标识符 ID_k,p 如何判断其真伪?即使 i,j 都向 p 提供自己的 IP 地址,由于 p 事先并不知道 ID_k 所对应节点的 IP 地址,p 如何确定 ID_k 是否具有(或根本不具有)i,j 提供的 IP 地址?

这里,我们采用了 IP 地址与节点标识匹配的认证方法,称其为 $Cent_{IP\text{-}ID}$.

$Cent_{IP\text{-}ID}$ 的前提是:任意节点 u 具有各自的公钥 P_k^u 和私钥 S_k^u 对,同时要求任意节点 u 在提交对节点 v 的评估 E_{uv} 时,同时提交自身的 IP 地址 IP_u 和节点 v 的公钥 P_k^v(我们认为,具有交易意向的节点互相知道对方的 IP 地址是合理的).

对于(1),即节点 w 冒充节点 u 企图欺骗节点 v 的档案点以影响其全局可信度的情况.由于 $Cent_{IP\text{-}ID}$ 的前提从而使冒名者 w 面临两难的局面:如果 w 提交自身的 IP 地址 IP_w,v 的档案点 D_v 可以由 Terrace 网络通过 u 的档案点 D_u 获取 u 的公钥 P_k^u,然后用 P_k^u 和任意一个整数 r 构造一个对 w 的挑战(challenge),从而使 w 暴露;如果 w 提交 u 的 IP 地址 IP_u(这本身也具有一定难度),虽然 w 可以通过挑战,然而 D_v 可以通过与 u 交互发现 u 近期并没有提交对 v 的推荐,从而使 w 暴露.

对于(2),即节点 w 企图冒充节点 u 欺骗其他用户节点 v 使之与其进行交易的情况(这时假设 v 知道 w 的地址是合理的).由于节点 v 可以通过 Terrace 从 D_u 处获取 u 的公钥 P_k^u,并以之与整数 r 构造一个对 w 的挑战,从而使得 w 暴露.

$Cent_{IP\text{-}ID}$ 要求我们对原语 $Put(ID_v,ID_u,R_{uv},f)$进行如下扩展:

u:$(ID_v,ID_u,E_{uv},f,P_k^v) \xrightarrow{\text{Terrace}} D_v$

D_v:$(\text{Request on } P_k^u) \xrightarrow{\text{Terrace}} D_u$

D_u:$(P_k^u) \xrightarrow{\text{IP}} D_v$

$u \xleftarrow{\text{IP}} D_v$:$([r]P_k^u, ID_v)$ //构造挑战

u:$(r,ID_v) \xrightarrow{\text{IPorTerrace}} D_v$

... //Other local operation;

其中,u:$(x,y) \xrightarrow{\text{Protocol}} v$ 表示通过协议 protocol 将 x,y 由节点 u 发送到节点 v.

协同作弊是指节点 A,B 在互相知道对方 ID 的情况下,相互抬高对方的全局可信度.即双方反复通过原语 $Put(ID_v,ID_u,E_{uv},f)$为对方提供过高的推荐度 R_{AB} 或 R_{BA}.我们也将此行为称为夸大.与此相似的另一个行为是,节点 A 通过对节点 B 的档案点提交不真实的负面评价而试图降低节点 B 的全局可信度.我们将此行为称为诋毁.EigenRep 模型和协议对此未加讨论.

实际上,考虑 eigenRep 模型,由式(1)、式(2)可以看出,一个具有较高全局可信度的节点 A 可以通过无限增加对节点 B 的满意次数 Sat_{AB},从而对 B 的可信度加以夸大(如第 1 节所述,由于 eigenRep 不具备惩罚机制,从而也不存在诋毁问题).

对我们的模型来说,由定义 4 和定义 6 可以得出,对任意节点 v,

$$T_v^{(k+1)} = \sum_k (R_{kv} \times T_k^{(k)}) = \sum_k \left(\frac{S_{kv} - F_{kv}}{\sum_j S_{jv}} \times T_k^{(k)} \right) \tag{6}$$

因此,当恶意节点 u 试图诋毁或夸大 v 的全局可信度时,可以通过 Put 原语向 v 的档案点发送不真实的评价 E_{uv},并将 S_{uv} 设成一个大数.通过公式(6),可得

当 S_{uv} 足够大时,

$$T_v^{(k+1)} \approx \frac{S_{uv} - F_{uv}}{S_{uv}} \times T_u^{(k)} \tag{7}$$

因此,只要自身具有较高的全局可信度,并保证 $S_{uv} >> F_{uv}$,u 就可以对 v 进行夸大,反之,只要 $S_{uv} \approx F_{uv}$,u 即可对 v 进行诋毁.我们对档案点采取以下规则来抑制诋毁或夸大的效果:

(1) v 的档案点首先对评价 E_{uv} 中的 S_{uv} 和 F_{uv} 进行初步分析,对于 $S_{uv} - S'_{uv} > 1$,或 $F_{uv} - F'_{uv} > 1$,即认为该评价非法.其中$\langle S'_{uv}, F'_{uv} \rangle$为 u 对 v 上次提交的评价.因此,节点 u 不可能通过少量的诋毁或夸大操作就对 v 产生实质性影响.频繁的提交评价(例如秒级)则可能引起 D_v 的警觉而拒绝接受 u 的评价.另外,通过公式(7)可以看出,如果 u 本身为不可信节点(全局可信度较低),则其对 v 的夸大评价很难发生作用.

(2) 对每次交易,除要求客户方提交评价外,还要求服务方提交确认信息.以节点 u,v 为例,u,v 发生交易后(设 u 为客户),除 u 提交评价 E_{uv} 以外,也要求 v 在一定时间间隔 θ 内通过 Terrace 提交确认 A_{uv}.D_v 如果在间隔 θ 内收到 v 的确认(确认同样存在冒名问题,解决方法与我们前面讨论问题相似),则:

(a) 如果为正面评价,即 $S_{uv} > S'_{uv}$,D_v 以 $T_v^{(k)}$ 的概率接受该评价.

(b) 如果为负面评价,D_v 接受该评价.

如果在间隔 θ 内未收到 v 的确认,则:

(a) 如果为负面评价,D_v 以 $1-T_v^{(k)}$ 的概率接受该评价.

(b) 如果为正面评价,D_v 拒绝接受该评价.

我们从两个方面来看 u 对 v 的影响:如果 u 试图对可信度较高的节点 v 进行诋毁,则 D_v 接受的可能性为 $1-T_v^{(k)}$,因此效果不大.如果 u 试图夸大可信度较低的节点 v,则 D_v 接受的可能性为 $T_v^{(k)}$,效果同样有限.考虑到(1),夸大和诋毁行为可以被有效抑制.

上述规则带来的一个影响是,节点从低可信节点成长为高可信节点是一个非线性过程,其可信度越高,成长越快.节点从高可信节点变为低可信节点与此相似.一般地,对于新加入网络的节点,只要其用户确实可信,连续的成功交易可以加速其成为高可信节点的速度.对于高可信节点,由于可信度高,因此通常交易量较大,其连续"失信"也会加快其可信度的降低.

4 仿真及其结果分析

我们构造了多个仿真实验来检测我们的模型,作为参照,我们同时实现了基于 eigenRep 的模型的仿真.在我们的实现中,DHTs 机制被简化,即任意节点 v 的档案点 D_v 为随机指定的任意节点,这是因为在单机仿真环境中,基于 DHTs 的路由是不必要的.我们设想的应用场景为文件共享应用,即用户的目标是下载其所需的文件,下载文件的真实性是其判断本次交易成功与否的惟一标准.我们设计了以下几类节点(事实上,情况要复杂得多,几乎各种类型恶意节点的组合都可以产生一类新的恶意节点.考虑到实验的可操作因素,本文仅对以上类型进行了实验),即:

(1) 善意的节点.这类节点无论在提供服务上(上载)还是在对其他节点的评价上(提交对其他节点的评价),都是真实的.我们称这类节点为 S 类.

(2) 恶意节点.这类节点进一步分为以下几个子类:

(a) 单纯的恶意节点.这类节点只提供不真实的(上载)服务.我们称这类节点为 EE 类.

(b) 诋毁节点.这类节点诋毁所有与之交易过的 S 类节点,并提供不真实的(上载)服务.我们称这类节点为 ED 类.

(c) 夸大节点.这类节点夸大同伙的恶意节点.我们称这类节点为 EK 类.

(d) 冒名节点.这类节点冒充与其交互过的可信度高于自身的节点.提供不真实的(上载)服务.我们称这类节点为 EM 类.

在这里,我们假定文件共享网络是理想的,即任意用户可以找到任意文件及其声称为该文件拥有者的所有节点(文件可能并不真实).用户的行为较为简单,即从所有声称拥有其所需文件的节点中选择可信度最高的节点,并与之交易(下载).

对于规模为 1 000 个节点的仿真网络,我们设定的文件总数为 10 000.我们将 10 000 个文件随机分配到所有 1 000 个节点,并保证每个文件至少被一个 S 节点拥有(虽然这个假设不尽合理,但为了避免文件的初始分布影响最终与理想情况的比较(例如,如果文件仅仅被非 S 类节点拥有),我们作此假定).每个用户在整个仿真过程中必须完成 100 次交易(下载 100 次),每次交易目标为从其不曾拥有的文件中随机选择一个并试图进行下载.交易的成功使得该用户拥有该文件,失败的交易不会增加该用户拥有的文件.

在我们的仿真实验中,最终的评估标准是整个网络成功交易(下载)的次数.显然,在理想情况下(所有的节点都是 S 类节点),成功交易的次数为 10^5.

实验的仿真环境为 PIII 650MHZ,256MB.仿真基于 Java 实现.

4.1 EE类仿真

EE 类仿真是指网络中的恶意节点都为 EE 类.该实验主要为检验不同规模的 EE 类节点对我们的模型以及 eigenRep 模型的影响.结果如图 4 所示.

可以看出,与 eigenRep 模型相比,我们的模型在仅有 EE 类恶意节点的情况下,效果基本相当.在网络中 50%都为 EE 类节点的情况下,与理想网络相比,仍然可以达到较高的成功下载量(接近 79%).实际上,在这种情况下,由于节点开始是盲目的(所有节点起始可信度为 0),因此 EE 类节点的规模越大,系统初起时,这种盲目性造成的交易失败的可能性就越大.EigenRep 模型假设了一个亚可信节点集合 P(如第 1 节所述,eigenRep 通过该假定确保迭代的收敛.本实验将 P 的规模设为 100,其可信度始终保持为 0.5),因而系统初起时,节点具有较小的盲目性,因此交易成功率较我们的稍高些.然而,在网络中固定部分节点具有先天较高的可信度,这一假定本身在实际中是不合理的和较难操作的.同时,由于 eigenRep 的模型缺乏对恶意节点的惩罚(见第 1 节),如果 EE 类节点采取更灵活的策略,比如,以一定概率提供真实的服务(或只对高可信度节点提供真实服务),eigenRep 模型将很难对其进行惩罚.我们对此进行了实验.EE 类节点在该实验中改变策略为随机提供真实服务(对每个服务请求,以 0.5 的概率提供真实服务).从如图 5 所示的结果可以看出,面对更为“狡猾的”EE 节点,我们的模型较 eigenRep 具有明显的优势.

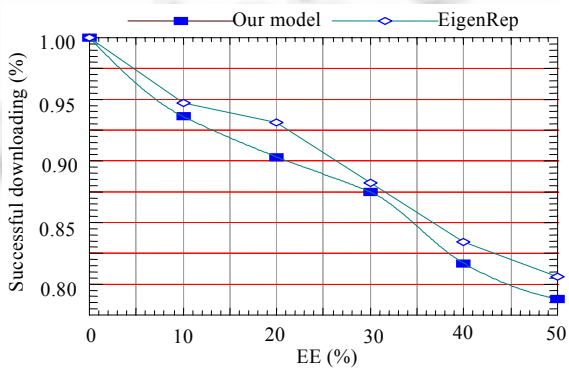

Fig.4 At different scale of EE nodes, the ratio to ideal case on successful downloading

图 4 不同规模 EE 类恶意节点下,与理想情况的成功交易比

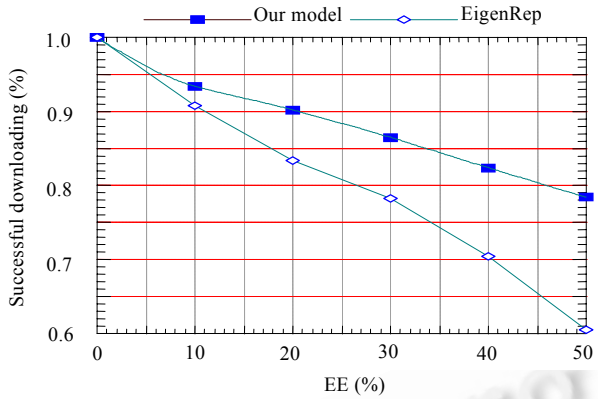

Fig.5 At different scale of EE peers (strategy changed), the ratio to ideal case on successful downloading

图 5 不同规模 EE 类节点下(改变策略后),与理想情况的成功交易比

4.2 ED类仿真

ED 类节点诋毁所有与之交易过的 S 类节点,并提供不真实的(上载)服务,这类节点试图通过降低可信节点的可信度来破坏网络的有效性.对于不同规模的 ED 类节点,我们的实验结果如图 6 所示.

由于 eigenRep 模型欠缺惩罚机制,因此诋毁机制对其影响不大.对于我们的模型,该实验检验了我们在第 3.3 节讨论的反诋毁方法.通过如图 6 所示的结果和对比可以看出,我们的模型较为有效地抑制了诋毁的影响,在系统节点中 50%都为诋毁恶意节点的情况下,仍然具有 80%左右的交易成功率.

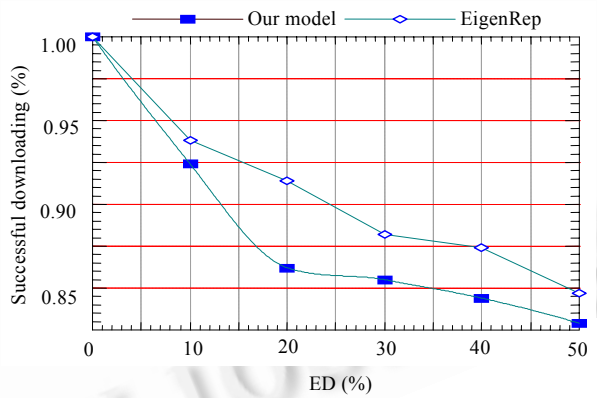

Fig.6 At different scale of ED peers, the ratio to ideal case on successful downloading

图 6 不同规模 ED 类节点下,与理想情况的成功交易比

4.3 EK类仿真

在实验中,我们令每个 EK 节点夸大(N_k–1)个其他 EK 类节点(N_k 为 EK 类节点的规模),这实际上是一种较为严重的协同作弊.EigenRep 模型对此未作任何处理,因此,随着 EK 类节点的增加(为了方便计,我们在实验中将 EK 类节点同时也设定为 EE 类节点),恶意节点很容易获取较高的可信度,同时由于 eigenRep 缺乏惩罚机制,造成系统的有效交易(下载)明显下降.与之相反,由于在我们的模型中对此作了处理(见第 3.3 节的反夸大机制),因此,夸大被明显抑制.

如图 7 所示,相对于我们的模型,EigenRep 缺乏夸大抑制机制,恶意节点相互夸大可信度,从而吸引大量的(下载)交易,并使得这些交易成为无效交易.

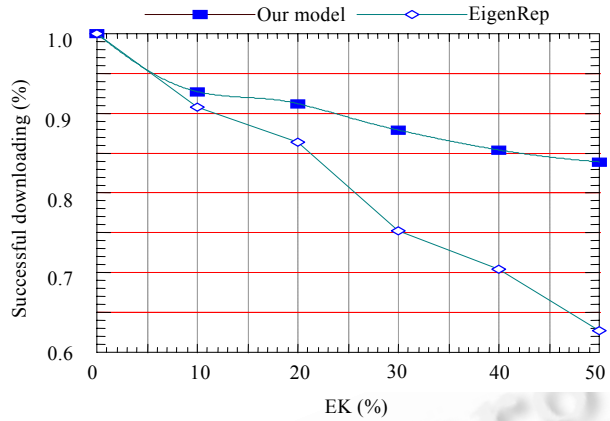

Fig.7 At different scale of EK peers, the ratio to ideal case on successful downloading

图 7 不同规模 EK 类节点下,与理想情况的成功交易比

4.4 EM类仿真

EigenRep 同样未对冒名作任何处理,反之,从第 3.3 节的讨论我们可以看出,通过使用反冒名机制,我们的模型可以完全不受冒名影响.对 eigenRep 而言,随着冒名节点的增多,恶意节点可以吸引相当数量的用户,从而产生大量无效交易,同时,冒名节点也可以冒充可信节点进行协同作弊(包括夸大和诋毁),从而更好地达到其(夸大和诋毁)目的.我们的实验给出了前一种情况的测试结果,其中,EM 节点在交易中冒充与其交易过的 S 类节点中可信度比它高的节点.如图 8 所示,从结果可以看出,随着冒名节点规模的增加,造成与冒名节点的交易大量增加,由于节点交易次数有限,因而有效交易明显减少(图中仅仅给出 eigenRep 在冒名情况下的测试结果,对我们的模型而言,由于冒名不起作用,因此结果实际上与第 4.1 节的实验效果相同,故此没有列出).

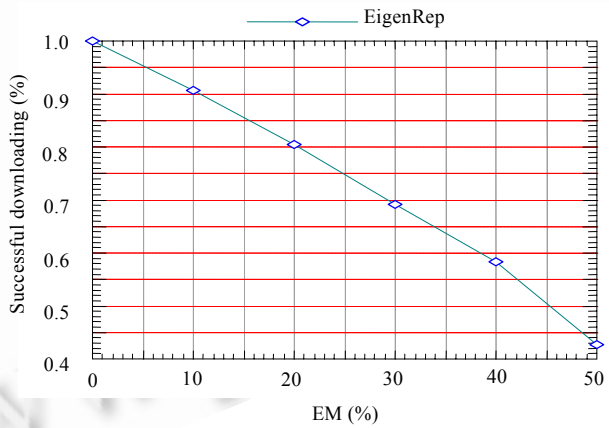

Fig.8 At different scale of ED peers, the ratio of eigenRep to ideal case on successful downloading

图 8 不同规模 EM 类节点下,eigenRep 与理想情况的成功交易比

5 结 论

本文借鉴社会网络分析中的中心性测量方法,构造了一个基于推荐机制的 Peer-to-Peer 环境下的全局信任模型.该模型克服了已有模型的若干局限性.通过分析和仿真说明,该模型比已有模型在多个指标上具有较大的提高.

References:

[1] Adarand E, Huberman B. Free riding on Gnutella. Technical Report, CSL-00-3, Palo Alto: Xerox PARC, 2000.

[2] Chen R, Yeager W. Poblano: A distributed trust model for P2P networks. Technical Report, TR-I4-02-08, Palo Alto: Sun Microsystem, 2002.

[3] Caronni G. Walking the Web of trust. In: Sriram RD, ed. Proc. of the IEEE 9th Int'l Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. IEEE Press, 2000. 153~159.

[4] Oram A. Peer-to-Peer: Harnessing the Power of Disruptive Technology. Sebastopol: O'Reilly Press, 2001. 222~238.

[5] Altman J. PKI Security for JXTA overlay networks. Technical Report, TR-I2-03-06, Palo Alto: Sun Microsystem, 2003.

[6] Albrecht K, Ruedi AR. Clippee: A large-scale client/peer system. Technical Report, TR-410, Swiss Federal Institute of Technology, 2003.

[7] 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.

[8] Sig2dat specification. 2002. http://www.geocities.com/vlaibb/

[9] Kamvar SD, Schlosser MT. 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.

[10] Scott J. Social Network Analysis: A Handbook. 2th ed, SAGE Press, 2000. 87~236.

[11] Wasserman S. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994. 173~196.

[12] Bonacich P. Eigenvector-Like measures of centrality for asymmetric relations. Social Networks, 2001,23(4):191~201.

[13] Ratnasamy S. Routing algorithms for DHTs: Some open questions. In: Kaashoek F, ed. Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems. Cambridge: Springer-Verlag, 2002. 45~52.

[14] Dou W. The research on trust-aware P2P topologies and constructing technologies [Ph.D. Thesis]. Changsha: National University of Defense Technology, 2003 (in Chinese with English abstract).

[15] Ratnasamy S. A scalable content-addressable network [Ph.D. Thesis]. University of Berkeley, 2002.

[16] Stoica I, Morris R, Karger D. Chord: A scalable peer-to-peer lookup service for Internet applications. Technical Report, TR-819, MIT Press, 2001.

附中文参考文献:

[14] 窦文.信任敏感的 P2P 拓扑构造及其相关技术研究[博士学位论文].长沙:国防科学技术大学,2003.