ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.1, January 2006, pp.96−107
DOI: 10.1360/jos170096

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

Peer-to-Peer 环境下多粒度 Trust 模型构造*

张 骞[1+], 张 霞[1], 文学志[1], 刘积仁[1], Ting Shan[2]

[1](东北大学 计算机软件国家工程研究中心,辽宁 沈阳 110179)

[2](School of Information Technology and Electronical Engineering, University of Queensland, Australia)

Construction of Peer-to-Peer Multiple-Grain Trust Model

ZHANG Qian[1+], ZHANG Xia[1], WEN Xue-Zhi[1], LIU Ji-Ren[1], Ting Shan[2]

[1](National Engineering Research Center for Computer Software, Northeastern University, Shenyang 110179, China)

[2](School of Information Technology and Electronical Engineering, University of Queensland, Australia)

+ Corresponding author: Phn: +86-24-83661102, E-mail: zhangqian@neusoft.com, http://www.neu.edu.cn

Zhang Q, Zhang X, Wen XZ, Liu JR, Ting Shan. Construction of peer-to-peer multiple-grain trust model. *Journal of Software*, 2006,17(1):96−107. http://www.jos.org.cn/1000-9825/17/96.htm

Abstract: Trust is multi-faceted and the peer's needs are different in different situations. A peer may need to consider its trust in a specific domain of another peer's capability or in a combination of multiple domains. Current Peer-to-Peer trust model could not promise the trust computation of different domains. This paper presents a novel Peer-to-Peer multiple-grain trust model and gives a distributed implementation method, which considers the trust computation of different domains. Mathematical analyses and simulations show that, compared to the current trust model, the proposed model is more precise on trust computation of multiple domains and more robust on trust security problems.

Key words: peer-to-peer; ontology; trust tree; documentary point; multiple-grain trust model

摘 要: 信任是多方面的,在不同的应用场景中,同一节点在不同领域具有不同的可信度.现有信任模型粒度过于粗糙,不能很好地解决同一 Peer 节点在不同领域、不同方面的可信度计算问题.据此,提出一种新的 Peer-to-Peer 环境下的多粒度信任模型,并给出该模型的数值分析和分布式实现方法.分析及仿真结果表明,该模型与已有模型相比,在可信度计算的粒度、模型的安全性等方面有较大的提高.

关键词: 对等网络;本体;信任树;档案点;多粒度信任模型
中图法分类号: TP393 文献标识码: A

目前,有关 P2P 的应用日益广泛[1],但仍然缺乏有效的信任机制提高系统整体的可用性[2].因此,有必要建立一种新的分布式信任机制,这已经成为当前的研究热点之一[3−6].

在 P2P 网络环境中,节点间的信任来自于两个方面:一个是对节点本身所能提供的服务质量的信任;另一个是对节点对于其他节点的评价的信任.不同节点对其他节点的信任评价所依据的标准可能是不一样的,以 P2P

* Supported by the National Natural Science Foundation of China under Grant No.60473031 (国家自然科学基金); the National Natural Science Foundation of Jiangshu Province under Grant No.BK2004119 (江苏省自然科学基金)

Received 2005-03-16; Accepted 2005-07-11

环境下的文件共享应用为例,节点的服务质量体现在各个方面,比如下载速度、文件质量及所属领域等.一些节点可能希望从其他节点获取音乐文件,则其所关心的是提供音乐文件的服务质量;而另外一些节点可能更加关注与之交易节点的整体服务质量.

目前存在的信任模型的粒度过于粗糙,不能对用户针对某一具体领域的信任度进行量化.本文提出了一种新的 Peer-to-Peer 环境下的多粒度信任模型和全局部分可信度,将信任模型的粒度进一步细化,能够对用户针对某一具体领域的信任度进行量化.事实上,本文提出的模型是一个多粒度模型 MGM(multiple-grain model),可以针对信任树中任一层次、任一领域计算可信度.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,文献[3]讨论了基于 PKI 的信任模型、基于局部推荐的模型和基于数据签名的模型及其存在的问题,除此之外,还存在以下几类模型:

(1) 基于角色的信任模型.在这类系统中,节点依据其兴趣,加入不同的社区,社区是拥有共同兴趣的节点集合,同一个节点可以加入不同的社区.依据节点对于不同社区的隶属程度,决定其在不同方面的可信度.如 Dynamic Coalitions 中采用的信任模型,但该模型仍然存在一定的缺陷[4].

(2) 基于 Bayesian 的信任模型.这类系统的核心思想是:依据一定的参数(如文件质量、下载速率等),利用 Bayesian 概率的方法计算可信度,.如 Yao Wang 在文献[5]中讨论的信任模型.但其可信度的计算实质上是基于用户自身的主观判定,往往具有片面性.

(3) 全局可信度模型.为获取全局的节点可信度,该模型通过邻居节点间相互满意度的迭代来获取节点的全局可信度.如 Stanford 的 EigenRep[6]和文献[3]的全局信任模型.文献[3]中的模型与我们的模型相似,为便于描述,我们将文献[3]的单粒度模型简记为 SGM(single-grain model).

SGM 模型的核心思想是: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.即

$$T_i = \sum_k (R_{k,i} \times T_k) \tag{1}$$

对于任意节点,T_i 为节点 i 的全局可信度,$R_{i,j}$ 为节点 i 对节点 j 的推荐度.

$$R_{i,j} = \frac{S_{i,j} - F_{i,j}}{\sum_k S_{k,j}} \tag{2}$$

$S_{i,j}$ 和 $F_{i,j}$ 分别为 i 对 j 在历史交易中积累的满意次数和不满意次数.该模型与 MGM 模型有两点相似:一是都试图通过迭代方法计算节点的全局整体可信度(见第 2 节);二是通过分布 Hash 机制放置节点的全局整体可信度(见第 3 节).然而,该模型仍然存在以下问题:

(1) 模型的粒度很粗糙,它忽略了同一节点的可信度在不同领域、不同方面上的区别和联系.

比如,节点 i 在计算机领域比较擅长,但对音乐知识了解甚少.一些在音乐领域比较擅长的节点与 i 进行交易,显然对 i 的不满意程度会增加,导致 i 的可信度降低.因此,当其他节点在进行计算机领域的交易时,根据式(2)和式(1)计算出的 T_i 较小,从而拒绝与 i 交易,然而实际上,i 在计算机领域却是比较擅长的,这显然欠缺合理性.

(2) 模型及实现协议在放置节点可信度时,没有考虑档案点(可信度存放点)间的异构性.某些高可信节点的档案可能存放在低可信节点上,高可信度通常意味着高访问量,档案点需要面对的是频繁的可信度更新及计算,这对于计算能力、网络带宽等都非常有限的低可信档案点而言,可能成为严重的负载(见第 3.1 节).

(3) 该模型及实现协议应用场景有限.模型协议要求在获取可信度时,必须预先知道服务节点的 ID,即:假设用户能够对自己的需求进行精确的描述.然后,通过系统提供的功能(如搜索),来获得满足需求的 ID 列表,根据 ID 获取节点可信度,最后从中选择可信度高的节点交易.但在诸如聊天消息等应用中,该假设并不成立(见第 3.1 节).

(4) 模型的协议实现在应付节点间大规模交易的情况时,可能出现较高的消息开销.比如:交易模式不是 1 对 1 交易而是 N(N 为系统级规模)对 N 时,服务节点的可信度求解和更新所需的消息开销分别会达到 $O(N^2)$和 $O(N^3)$的规模(见第 3.2 节).

(5) 该模型没有充分考虑安全性问题.对于档案点本身的恶意行为,该模型及协议无能为力(见第 3.3 节).

本文针对 SGM 模型存在的问题,提出了一种新的全局信任度模型 MGM,并给出了该模型的分布式计算协议及实现方法,最后是仿真实验.

2 多粒度信任模型

首先给出本文对信任实体和信任本体的定义,然后给出信任树的定义,最后引出全局可信度的定义.

定义 **1**. 信任实体定义为一个五元组 E,E=(I,D,P,T,NE).其中:I 为该实体在 Peer 节点内的唯一标识;D 为实体属性集;P 为定义在 E 上的处理符集;T 为实体的全局部分可信度;NE 为与该 Peer 节点在 E 上有过交易的 Peer 节点集.一个信任实体代表了 Peer 节点的一种偏好(领域).

定义 **2**. 信任本体定义为一个七元组 O,O=(I,E,EO,DO,PO,T,NO).其中:I 为该本体在 Peer 节点内的唯一标识;E 为本体 O 的子实体集;EO 为 O 的子本体集;DO 为 O 的属性集;PO 为定义在 O 上的处理符集;T 为本体 O 的全局部分可信度;NO 为与该 Peer 节点在 O 上有过交易的节点集.

定义 **3**. Peer 节点的信任树是一棵多叉标签树.信任树 T_r 定义为一个七元组,T_r=(ID,$Root(T_r)$,V,E_r,TR,TN,P).其中:ID 为 Peer 节点在 P2P 网络全局范围内的唯一标识;$Root(T_r)$表示树的根节点;V 为信任树的树节点集;E_r 为边集,是 V 上的一个二元关系;TR 为节点的全局整体可信度;TN 为与该节点有过交易的 Peer 节点集;P 为 T_r 上的处理符集.

信任实体、信任本体与信任树的映射关系为:(1) 信任树的根节点对应为一个虚拟本体,记为 RO.RO 是非任何本体的子本体.(2) 信任本体映射为信任树中的节点,该本体的子本体或子实体映射为对应树节点的孩子节点,如图 1 所示.

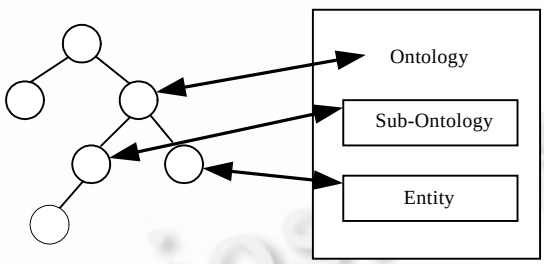


Fig.1 Relationship between trust tree and ontology

图 1 信任树与本体间的映射关系

以 P2P 环境下的文件共享应用为例,假设节点 R 有“音乐”和“计算机”两方面的偏好,可以构造本体 A 与 B 及 R 对应的信任树 T_r(如图 2 所示).

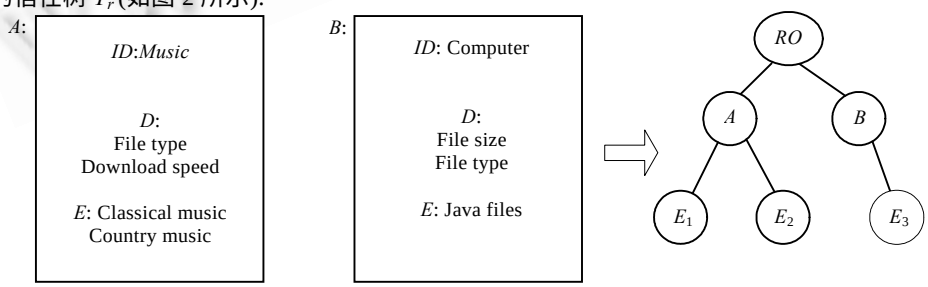


Fig.2 Trust tree of node R

图 2 节点 R 的信任树

定义 **4**. 全局信任树是一棵多叉标签树,定义为一个三元组 $QT_r=(V,E,DO)$.其中:V 为全局信任树的树节点集;E 为边集;DO 为全局范围内存在的领域集.任意 $u,v\in V$,若$(u,v)\in E$,则称 u 是 v 的父节点;若$(u,v)\in E^+$(E^+为 E 的传递闭包),则称 u 是 v 的祖先节点.存在映射 $f:V\to DO$,f 满足以下条件:

(1) $|V|=|DO|$,即 V 与 DO 的势相等;任意 $v\in V$,存在 $do\in DO$,满足 $do=f(v)$,$v=f^{-1}(do)$.

(2) 任意 $u,v\in V,u\neq v$,则 $f(u)\neq f(v)$;若$(u,v)\in E^+$,则 $f(v)$是 $f(u)$的子领域.

树中节点编码的基数为 d,根节点的编号为 1,对于其余任意节点,设其编号为 A,则其第 i 个($i\in[0,d-1]$)个子节点的编号为 $A\times d+i$.对于任一领域 k,使用 $f^{-1}(k)$的标识作为 k 的全局唯一标识,记为 L_k.

定义 **5**. 称三元组$(S_{i,j,k},F_{i,j,k},L_k)$为节点 i 对 j 在 k 领域的评价(comment),记为 $E_{i,j,k}$.其中,$S_{i,j,k}$ 与 $F_{i,j,k}$ 分别代表 i 与 j 在 k 领域交易成功和失败的次数,L_k 为领域 k 的标识.

任意 Peer 节点 i 对节点 $j(j\neq i)$在 k 领域的局部看法,称为节点 j 在 k 上的局部可信度.局部可信度可由 i 与 j 在 k 领域及其相关领域交往的历史计算得出.下面重点讨论节点的全局可信度.

考虑社会网络分析中的基于节点入度(in-degree)的中心测量(centrality-measurement)方法[7,8],即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.这里,我们对此进行了扩展以求解全局领域可信度,即任意节点在某一领域的全局领域可信度,由与之在该领域或其相关领域发生过交易行为的其他节点对它的推荐,以及这些节点在该领域的全局领域可信度来决定.因此,首先给出推荐度的定义.

定义 **6**. 称 $R_{i,j,k}$ 为节点 i 对 j 在 k 领域的推荐度(recommend degree).

$$R_{i,j,k}=\begin{cases}\dfrac{1}{2}\times\left(\dfrac{\sum_{m=1}^{N_{j,k}}w_{k,m}R_{i,j,m}}{N_{j,k}}+\dfrac{S_{i,j,k}-F_{i,j,k}}{\sum_{l}S_{l,j,k}}\right), & N_{j,k}>0\\ \dfrac{S_{i,j,k}-F_{i,j,k}}{\sum_{l}S_{l,j,k}}, & N_{j,k}=0\end{cases}\tag{3}$$

其中:$N_{j,k}$ 为 j 信任树中 k 所对应树节点的孩子节点数目;$\sum_{l}S_{l,j,k}$ 为所有与 j 在 k 领域有过交易且交易成功的次数.$w_{k,m}$ 表示 k 的分支领域对 k 的重要程度.考虑 $\sum_{l}S_{l,j,k}$ 或 $S_{i,j,k}-F_{i,j,k}<0$ 的情况:此时,若 $N_{j,k}=0$,则 $R_{i,j,k}=0$;否则有 $R_{i,j,k}=\dfrac{\sum_{m=1}^{N_{j,k}}R_{i,j,m}}{N_{j,k}}$.意味着节点 i 与 j 在 k 领域没有交易,但在 k 的分支领域有过交易,则 i 仍然可以在 k 领域对 j 作出推荐.

定义 **7**. P2P 网络中任意节点 i 在 k 领域的全局可信度 $T_{i,k}$.设

$$T_{i,k}=\sum_{j}(T_{j,k}\times R_{j,i,k})\tag{4}$$

其中,j 为与 i 在 k 或其相关领域有过交易的节点.设 k 领域全局可信度向量为 $T_k=(T_{1,k},T_{2,k},\ldots,T_{n,k})$,则称

$$T_k=T_k R_k\tag{5}$$

为 P2P 网络关于信任关系矩阵 R_k 的信任方程.其中,$R_k=|R_{i,j,k}|_n$,元素 $R_{i,j,k}$ 为节点 i 对节点 j 在 k 领域的推荐度.若 k 为根领域,则 $T_{i,k}$ 为全局整体可信度,否则 $T_{i,k}$ 为全局部分可信度.方程(4)的解可由下述定理求得.

定理 **1**. P2P 网络在 k 领域关于其信任关系矩阵 R_k 的信任方程的 Jacobi 和 Gauss-Sedei 迭代收敛.

证明:由式(5)易得 $T_k^T=R_k^T T_k^T$,即$(I_k^n-R_k^T)T_k^T=0$,令 $H_k=R_k^T$,则收敛的充分条件是

$$\max_{1\le i\le n}\sum_{j=1}^{n}|H_{i,j,k}|<1\tag{6}$$

易知 $I_k^n-H_k$ 为严格对角占优矩阵,于是定理得证.限于篇幅,详细证明不再给出. □

至此,我们构造了一个基于信任树的多粒度信任模型,该模型可以针对信任树中任何领域计算可信度.随着信任树的高度和树中节点数目的增加,领域划分得越细,计算出的全局可信度也就越能逼近真实情况.

值得一提的是,SGM 模型也采用了迭代的方法,但因为其模型本身很粗糙,因而迭代相对简单得多.而且,通过迭代计算出的可信度并不能准确反映节点的真实情况.另外,在 SGM 的迭代中,比如计算 i 的可信度,任意与 i 发生交易的节点的档案点都要参与迭代(意味着较多的消息开销).事实上,这是没有必要的,在本文的模型中,只有在某一具体领域和其相关领域有过交易的节点的档案点才参与迭代,而且参与迭代的档案点数目与领域细化的程度成反比,从而进一步减少了参与可信度计算的节点数目,因而具有更好的工程可行性.

3 可信度的分布求解协议

本文通过 Hash[9]表放置节点的全局可信度.SGM 模型也采用了基于 Hash 的放置方法,但其方法及实现协议仍然存在一些问题(见第 1 节问题(2)、问题(3)及问题(4)).针对这些问题及 MGM 模型的特点,本文对其方法进行了改进,并重新设计了实现协议.

3.1 可信度的放置

首先给出本文对档案点的定义及分析,然后讨论档案点维护的数据结构.

定义 **8**. 令 $L_{i,j}$ 为网络中任意节点 i 与 j 在 Terrace[8]拓扑中所处层次之差,即 $L_{i,j}=layer(j)-layer(i)$,其中 $layer(i)$ 表示 i 所在层次,设 HTD 为任意均匀的 Hash 函数.

(1) 节点 i 在 Terrace 中的投影记为 GD_i,即 $GD_i=HTD(ID_i)$,ID_i 为全局唯一的节点标识符.设节点 i 与 GD_i 的层次之差为 $L_{i,g}$,若 $L_{i,g}>0$,则将沿 GD_i 到 *d-tree* 根节点路径上的第 $L_{i,g}$ 个祖先节点(若该节点为 ID_i,则选第 $L_{i,g}-1$ 个祖先节点)作为 i 的整体档案点,记为 ZD_i;否则将 GD_i 作为 i 的整体档案点.整体档案点存放节点的全局(整体和部分)可信度.

(2) 领域 k 在 Terrace 中的投影记为 D_k,即 $D_k=HTD(L_k)$,L_k 为全局唯一的领域标识符.设 L_k 与 D_k 的层次之差为 $L_{k,D}$,若 $L_{k,D}>0$,则将沿 D_k 到 *d-tree* 根节点路径上的第 $L_{k,D}$ 个祖先节点作为 k 的领域档案点,记为 LD_k;否则将 D_k 作为 L_k 的领域档案点.领域档案点存放该领域内节点的全局部分可信度.

Terrace 是基于 *d-tree* 的非对称树型 DHTs 拓扑结构,拓扑中节点所处层次与其可信度成反比,同层节点的可信度相近,上述定义能够以较高的概率保证任意节点 i 的档案点所在层次不会低于 i 所在层次.因此,任意节点的档案能够以较高的概率保存到与之可信度相近的节点上,避免了因档案点的异构性所引起的低可信档案点负载过重的问题,同时使得系统具有更好的容错能力.

图 3 描述了改进前后可信度放置的一个例子,节点 D 希望将对 C 的评价写入 C 的档案点.改进前的方法(如图 3(a)所示):hash 得到 C 的逻辑地址为 111,则将档案数据写入 E.改进后的方法(如图 3(b)所示):hash 得到 C 的逻辑地址为 111,$L_{C,E}=layer(E)-layer(C)=1$,则将档案数据写入 B;设交易领域为 k(computer),得到 C 的逻辑地址为 121,假设 $L_{k,B}=0$,则由 B 将档案数据写入领域档案点 H 中.

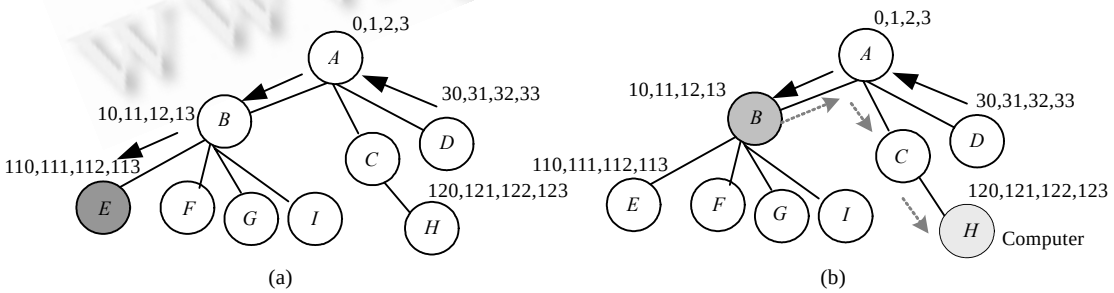


Fig.3 The method of placing the file dates improved before and after

图 3 改进前后的档案数据放置方法

每个整体档案点 d 至少需要包含一个如图 4(a)所示的数据结构.节点 d 是网络中节点 i 的档案点,ID_i 是 i 的标识;$L_{k1},\ldots,L_{kp}$ 为领域标识;$ID_{j11,1},\ldots,ID_{j1m,0},\ldots,ID_{j2h,k1},\ldots,ID_{jnd,kp}$ 为与 i 发生交易的节点标识,如 $ID_{jnd,kp}$ 为与 i 在 k_p 领域有交易的节点标识,$ID_{j1m,1}$ 为与 i 在根领域有过交易的节点标识;$S_{j11,i,1},\ldots,S_{j1m,i,1},\ldots,S_{j2h,i,k1},\ldots,S_{jnd,i,kp}$ 与 $F_{j11,i,1},\ldots,F_{j1m,i,1},\ldots,F_{j2h,i,k1},\ldots,F_{jnd,i,kp}$ 分别为与 i 发生交易的节点汇报的交易成功与失败的次数,如 $S_{jnd,i,kp}$ 表示以 $ID_{jnd,kp}$ 为标识的节点汇报的在 k_p 领域交易成功的次数.

$T_{j11,0}^{(m)},\ldots,T_{j2h,k1}^{(m)},\ldots,T_{j2h,kp}^{(m)}$ 为推荐节点目前的全局可信度;$T_i^{(m+1)}$ 为由 d 计算出的 i 的全局整体可信度,$T_{i,k1}^{(m+1)},\ldots,T_{i,kp}^{(m+1)}$ 为由 d 计算出的 i 的全局部分可信度,如 $T_{i,kp}^{(m+1)}$ 为 i 在 k_p 领域的全局部分可信度.

$\boldsymbol{ID_i}$	$T_i^{(m+1)}$		
$S_{j11,i,1}$	$F_{j11,i,1}$	$ID_{j11,1}$	$T_{j11,1}^{(m)}$
$S_{j12,i,1}$	$F_{j12,i,1}$	$ID_{j12,1}$	$T_{j12,1}^{(m)}$
…	…	…	…
$S_{j1m,i,1}$	$F_{j1m,i,1}$	$ID_{j1m,1}$	$T_{j1m,1}^{(m)}$
$\boldsymbol{L_{k1}}$	$T_{i,k1}^{(m+1)}$		
$S_{j21,i,k1}$	$F_{j21,i,k1}$	$ID_{j21,k1}$	$T_{j21,k1}^{(m)}$
…	…	…	…
$S_{j2h,i,k1}$	$F_{j2h,i,k1}$	$ID_{j2h,k1}$	$T_{j2h,k1}^{(m)}$
…	…		
$\boldsymbol{L_{kp}}$	$T_{i,kp}^{(m+1)}$		
$S_{jn1,i,kp}$	$F_{jn1,i,kp}$	$ID_{jn1,kp}$	$T_{jn1,kp}^{(m)}$
…	…	…	…
$S_{jnd,i,kp}$	$F_{jnd,i,kp}$	$ID_{jnd,kp}$	$T_{jnd,kp}^{(m)}$

(a) Whole documentary point

(a) 整体档案点

$\boldsymbol{L_{k1}}$	
ID_{i1}	$T_{i1,k1}^{(m+1)}$
ID_{i2}	$T_{i2,k1}^{(m+1)}$
…	…
ID_{im}	$T_{im,k1}^{(m+1)}$
$\boldsymbol{L_{k2}}$	
…	
$\boldsymbol{L_{kn}}$	
ID_{j1}	$T_{j1,kn}^{(m+1)}$
…	…
ID_{j1}	$T_{j1,kn}^{(m+1)}$

(b) Domain documentary point

(b) 领域档案点

Fig.4 The structure of whole and domain cumentary point

图 4 整体与领域档案点的数据结构

每个领域档案点 q 至少包含一个如图 4(b)所示的数据结构.其中:$L_{k1},\ldots,L_{kn}$ 为领域标识,$T_{i1,k1}^{(m+1)},\ldots,T_{j1,kn}^{(m+1)}$ 为节点的全局部分可信度,如 $T_{j1,kn}^{(m+1)}$ 为标识为 ID_{j1} 的节点在 $L_{\boldsymbol{kn}}$ 领域的全局部分可信度.

3.2 分布求解协议

首先给出协议的几个主要原语及其语义.

Put_K(ID_v,ID_u,$E_{u,v,k}$,L_k,f):Peer 节点 u 将对节点 v 在领域 k 的评价 $E_{u,v,k}$ 写入 v 的整体档案点,并触发 f 过程,L_k 为领域 k 的标识.

Put_K(ID_v,$T_{v,k}$,L_k,):将 ID_v 在 L_k 领域的全局可信度 $T_{v,k}$ 写入(更新)到 L_k 的领域档案点.

Get_K(ID_v,$T_{v,k}$,L_k):从 ID_v 的整体档案点结构中获取节点 v 在 k 领域的全局可信度,并写入本地变量 $T_{v,k}$ 中.

Get_K(L_k,$T_{v,k}$,ID_v):从 L_k 的领域档案点结构中获取节点 v 在 k 领域的全局可信度,并写入本地变量 $T_{v,k}$ 中.

Get_K(L_k,T_k):从 L_k 的领域档案点结构中获取 k 领域内全部节点在该领域的全局可信度,并写入本地变量 T_k 中,T_k 为全局可信度的列表.

Get_K(L_k,T_k,δ):从 L_k 的领域档案点结构中获取领域 k 内可信度值不小于 δ 的全局可信度,并写入本地变量 T_k 中,T_k 为全局可信度的列表.

Update_K(ID_v,L_k):重新计算 ID_v 在领域 L_k 的全局可信度,并将结果更新到 ID_v 的整体档案点和 L_k 领域档案点.

任意节点作为一般用户节点的算法如下:

Procedure *Evaluate_K*(ID_v,L_k,*FT*)//在 L_k 领域,Peer 节点 *u* 执行对 *v* 的评估过程

If (*FT*=*true*) then

$S_{u,v,k}=S_{u,v,k}+1$

Else

$F_{u,v,k}=F_{u,v,k}+1$

Endif

Put_K(ID_v,ID_u,$E_{u,v,k}$,L_k,*FTrust*)

End

任意节点作为整体档案点的算法如下:

Procedure *FTrust*(ID_v,$E_{u,v,k}$)

将 ID_v,$E_{u,v,k}$ 存入档案点数据结构;

Get_K(L_k,T_k)

For each $T_{j,k}$ in T_k

if $(j \neq v)$ and $\left(\sum_l (S_{j,v,l}+F_{j,v,l}) \neq 0\right)$ then //l 为 *v* 的信任树中 *k* 对应的树节点及其孩子节点

Calculate($R_{j,v,k}$) //计算 $R_{j,v,k}$,$R_{j,v,k}$ 为节点 *j* 对 *v* 在 L_k 领域的推荐度

Endif

Endfor

$T_{v,k}=\sum_j (T_{j,k}\times R_{j,v,k})$

Put_K(ID_v,$T_{v,k}$,L_k,) //将 $T_{v,k}$ 和 ID_v 写入到 L_k 领域的档案点.

Update_K(ID_v, L'_k) //更新 ID_v 在领域 L'_k 的全局可信度, L'_k 为 L_k 父领域.

End

领域档案点在收到整体档案点 *Put_K*(ID_v,$T_{v,k}$,L_k,)请求后,将 $T_{v,k}$ 和 ID_v 写入本地数据结构.

由以上分析可知,节点 *u* 与 *v* 发生交易后,通过 *Put_K*(ID_v,ID_u,$E_{u,v,k}$,L_k,*f*)将评价 $E_{u,v,k}$ 写入 *v* 的整体档案点,引发对 *v* 在 L_k 领域的可信度的重新计算,重新计算过程中需要通过 *Get_K*(L_k,T_k)获取与 *v* 在 L_k 领域有过交易的其他节点在该领域的全局可信度,然后通过 *Put_K*(ID_v,$T_{v,k}$,L_k,)将 $T_{v,k}$ 和 ID_v 写入到 L_k 领域的档案点,并更新 ID_v 在 L_k 父领域的全局可信度,因此,消息复杂度为 $O(h)$,其中 *h* 为 ID_v 对应的信任树的深度,和与节点 *v* 发生交易的节点数目无关.

在 SGM 中,节点 *v* 的可信度的重新计算需要获取全部与 *v* 有过交易的其他节点的全局可信度,因为这些可信度存放在不同的档案点上,故其消息复杂度为 $O(N)$,其中 *N* 可能为系统级规模.另外,对于 SGM,在应付节点间大规模交易的情况时,如 *N* 对 *N*(*N* 为系统级规模)的交易模式时,可信度求解所需的消息开销是 $O(N^2)$,可信度更新所需的消息开销是 $O(N^3)$.

在本文实现中,任意节点仅需一次消息开销就可获取 L_k 领域内全部(*Get_K*(L_k,T_k))或部分(*Get_K*(L_k,T_k,δ))节点在该领域的全局可信度,在应付如 *N* 对 *N* 的交易模式时,可信度求解所需的消息开销是 $O(N)$,更新所需的消息开销是 $O(N^2)$.原语 *Get_K*(L_k,T_k,δ)主要基于以下考虑:用户通常只希望了解某一领域的高可信节点,因此没有必要获得该领域全部节点的可信度;与 *Get_K*(L_k,T_k)相比,通过设置参数δ,*Get_K*(L_k,T_k,δ)仅传输可信度值不小于δ的全局可信度,因而降低了网络传输的开销.

3.3 安全机制

SGM 中的讨论是在假定档案点可信的条件下提出来的,事实上,每个节点都可以作为其他节点的档案点,档案点并不一定可信.在多粒度模型环境中,档案点的恶意行为包括以下 4 种情况:

(1) 在交易双方提交评价时,私自修改评价,起到诋毁(抬高)其他节点的作用;

(2) 在交易双方提交评价时,多个档案点协同将正面的评价修改为负面评价,或反之;

(3) 私自篡改档案结构中的数据;

(4) 档案点互相合作,协同篡改档案结构中的数据.

针对上述问题,本文采用如下方法对其进行抑制:交易双方交易完成后,将新评价同时提交给交易领域档案点、交易领域父领域档案点(根领域除外)和整体档案点.下面分别就上述 4 种情况进行讨论:

(1) 情况 1:交易评价在提交时,各个档案点分别进行可信度计算,每个档案点获取其余档案点的计算结果,通过表决来判定恶意的档案点;对于情况 3,因为是单个档案点私自修改数据,因此,也可通过档案点之间表决的方式使恶意档案点暴露.

(2) 情况 2:只有当 3 个档案点全部是恶意节点、并且多次协同修改针对某个节点的评价时,其恶意行为才可能成功.然而,频繁地修改评价会引起交易双方的警觉,从而使恶意档案点暴露.事实上,全局信任树的高度与 Terrace 树相比很小,领域档案点一般位于可信度较高的节点上.因此,协同诋毁(抬高)也就更加困难.

(3) 情况 4:不妨设被篡改的数据是 Peer 节点 i 在 k 领域的评价,则 i 在 k'领域(k 的父领域)的评价也必须被篡改,且 k'的领域档案点必须是恶意节点.否则,可以从 k'的领域档案点获得 i 在 k'领域的全局可信度,并根据 k'的领域档案点中的数据(或 k 的领域档案点中的数据)重新计算 i 在 k'领域的全局可信度,若两个全局可信度相差过大,则可以判定存在恶意档案点.因此,至少需要保证 k 领域的全部祖先领域(i 节点信任树中对应的祖先领域及信任树根节点的父领域)的档案点都参与协同,作弊行为才可能成功.即至少需要有 $L+2$(L 为 k 对应的树节点在 i 信任树中的层次)个档案点参与作弊.然而,若 L 的值很小,由定义 8 可知,k 领域及其祖先领域的档案点可信度会很高;若 L 很大,显然作弊成功的可能性更小.

上述方法需要对领域档案点中的档案结构进行修改,限于篇幅,在此不再详述.在 SGM 中,任意档案点都可以私自篡改档案结构中的数据,而不受任何其他节点的制约.另外,根据 SGM 的可信度放置方法,高可信节点的档案可能保存在低可信节点上,而低可信档案点更容易作弊.

另外,对于一些更为狡猾的恶意节点,SGM 中的安全机制无能为力.比如对任意节点 u 和 v,节点 u 在音乐方面比较擅长,v 在计算机领域比较擅长.u 频繁(可能是恶意的)与 v 从事音乐方面的交易,显然,u 对 v 的评价多是负面的,根据 SGM 中的可信度计算方法,v 的全局可信度将急剧下降,进而影响到 v 在计算机领域的交易.而对于多粒度模型,可信度能够针对不同粒度的领域层次进行计算,某个领域可信度的降低,对不相关领域的可信度影响不大.如上例,v 在受到的恶意攻击时,仍然可以在计算机领域保证有较高的可信度.另外,v 可以在交易发生前获取 u 在音乐领域的全局可信度,从而拒绝与之交易.

4 模型的改进

4.1 信任树的更新

随着时间的推移,用户的兴趣可能发生变化,这就需要对信任树进行动态更新.这里提出基于聚集的更新算法,将全局信任树中的各个领域 D_j 看作关键词的集合,记为 dk_j.令 O 表示用户当前偏好的领域集合,即当前信任树对应的领域;U 表示用户新增加的偏好领域集,R 表示用户不再感兴趣的领域集,R 及 U 初始为空.算法如下:

(1) 对用户提出的任意查询 q,根据文献[10]中的上下文语义查询扩展方法,获得该查询的查询扩展关键词,连同查询包含的关键词一起,组成关键词集合 kw_i.

(2) 在一定时期之后,利用文献[11]中的方法对关键词集合 $kw_i(i=1,2,\ldots,m)$进行聚类,得到聚类集合 $C_p(p=1,2,\ldots,z)$.

(3) 利用文献[12]中的相似度计算方法,计算每个聚类 C_p 与领域关键词集合 $dk_j(j=1,2,\ldots,n)$的语义相似度 $SI_j(j=1,2,\ldots,n)$,令 $SI_k=\max(SI_j,j=1,2,\ldots,n)$,若 $SI_k>\lambda$(λ为相似度阈值)且 $SI_k\notin O$,则将 D_k 添加到集合 U 中.否则,若 $SI_k<\lambda$且 $SI_k\in O$,则添加 D_k 到集合 R 中.

(4) 对集合 U 及 R 中的每一项,最终由用户判定是否添加到本地信任树中或从信任树中删除.

上述步骤(1)主要考虑到在通常情况下,用户并不能准确地描述自己的需求.对于步骤(4),我们认为只有用户才能判定自己真正偏好的领域.

4.2 存储空间问题

可用空间在这里指的是 Peer 节点在某一时刻剩余的可被利用的存储空间.根据定义 8,领域档案点可能保存有多个领域的评价,尤其对于一些热门领域,领域内的交易量很大,因此需要较大的存储空间.虽然领域档案点一般位于可信度较高的节点上,然而,这些档案点的可用存储空间仍然有限,我们提出如下解决方法:

方法 1 修改了全局部分可信度的放置方法,记任意 Peer 节点 h 的可用空间为 h_s,给定空间大小阈值为 e,算法步骤如下:

(1) 根据定义 8 选择领域档案点,记为 i.若 $i_s<e$,则转向(2).否则,将档案保存在 i 中,算法结束;

(2) 从 i 的孩子节点中选择可用空间最大的节点 j,若 $j_s\geq e$,则将档案数据保存在 j 中,并在 i 中保存指向 j 的指针.否则考察 i 的其他子孙节点,若所有子孙节点的可用空间都小于 e,则转向(3);

(3) 将放置档案数据的任务交给 i 的父节点,重复上述过程,直到找到合适的档案点 m,将档案数据保存在 m 中,并在 i 中保存指向 m 的指针,算法结束.

方法 2 的思想是:在领域档案点 h 的可用空间 $h_s<e$ 时,采用淘汰策略,使得 $h_s\geq e$.策略如下:

(1) 可信度最低策略.对所维护的可信度进行排序,依次删除最小的记录,直到 $h_s\geq e$.该策略以牺牲低可信节点为代价,使得领域档案点维护的信息向高可信节点集中.事实上,通过领域档案点获得领域内全部节点的可信度是没有必要的,用户实际想了解的只是该领域内的高可信节点,显然,该策略降低了原有的存储及通信开销.

(2) 最近最少更新策略.维护每个可信度最近更新次数的统计,依次删除更新次数最小的记录,直到 $h_s\geq e$.该策略以牺牲低可信节点及冷门领域为代价,使得领域档案点维护的信息向高可信节点及热门领域集中.

方法 3 要求领域档案点为维护的每个 Peer 节点设定一个生存周期,在当前周期结束之前,由节点发出继续存活的消息,则开始新的周期计时.否则,该节点的档案记录在当前周期结束时即被删除.

显然,方法 1 能够确保可信度得到放置,其消息开销为 $O(\log N)$;方法 2 本质上是被动淘汰策略,即完全由领域档案点决定所要淘汰的档案记录;方法 3 需要消耗较大的通信开销,但是该方法考虑到了新加入的节点,这类节点初始时一般具有较低的可信度.

5 仿真分析

我们通过仿真对本文提出的 MGM 模型及其实现协议进行了检验,内容包括可信度进化、容错能力、安全机制和存储策略.其中,可信度进化、容错能力及安全机制和 SGM 模型进行了对比.此外,我们还进行了负载均衡实验,负载均衡主要考察修改后的可信度放置方法对各层节点负载的影响.实验表明,随着节点间随机连接数目的增加,各层节点负载基本平衡,限于篇幅,在此不再给出具体结果.

实验 **1**. 可信度进化仿真.

该实验主要考察不同领域内节点的可信度在交易过程中的变化情况,从而观察 MGM 和 SGM 与真实情况的差异(如图 5 所示).对于规模为 200 个节点的仿真网络,我们将节点基于领域分为两大类,即计算机类节点(S_1)和音乐类节点(S_2),每类取 10 个节点作为观测点.计算机类节点在计算机领域比较擅长,而对音乐却知之甚少;音乐类节点恰好相反.

图 5 显示了 S_1 类和 S_2 类观测点在 MGM 模型和 SGM 模型下平均可信度的变化情况.显然,在 MGM 模型下,可信度的变化反映了两类节点的真实情况;而在 SGM 模型下,两类节点的界限比较模糊,因此不能准确反映节点的真实情况.这主要是由于 SGM 模型粒度的粗糙性造成的.

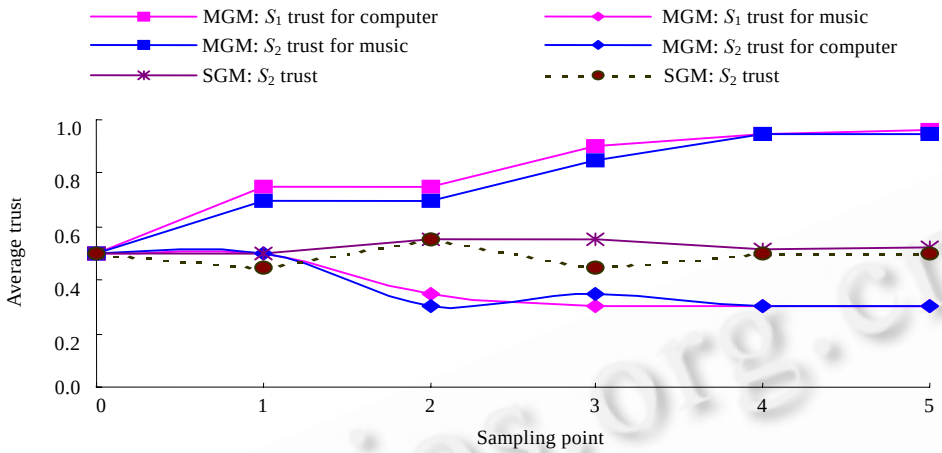


Fig.5 Trust evolution of two kinds of nodes

图 5 两类不同领域节点集合的平均可信度变化情况

实验 **2**. 容错仿真.

该实验在 10 000 个节点的规模下对 MGM 和 SGM 模型的容错机制进行了对比测试,即通过使不同规模的随机节点失效,观察在不同失效规模下,可信度求解失败的定位请求占定位总请求数的比率.

由结果可以看出(如图 6(a)所示),MGM 可以容忍比 SGM 更高规模的随机失效,这主要是由于 MGM 采用了与 SGM 不同的档案点选择方法与可信度放置策略.在考虑热点的情况下(如图 6(b)所示),MGM 表现出比 SGM 更好的容错能力,这主要是因为,在 MGM 中,高可信节点的档案点也是高可信节点,即使在整体档案点失效的情况下,仍然可以从领域档案点获取节点的可信度.与 MGM 不同,SGM 中的高可信节点的档案可能保存在低可信节点上,而低可信节点更容易失效.

Fig.6 Fault-Tolerance compare

图 6 容错能力对比测试

实验 **3**. 安全机制仿真.

该实验主要检验不同规模的作弊档案点对 MGM 模型以及 SGM 模型的影响,即观察在不同的作弊规模下,随机提交模拟交易请求,若交易双方的可信度均是真实的,则认为交易成功,否则为失败的交易.

由图 7 可以看出,MGM 模型在网络中 50%都为作弊档案点的情况下,与理想网络相比,仍然可以达到将近 81%的交易成功率.对于 SGM 模型,因为其安全机制没有考虑到档案点作弊的行为,在 50%都为作弊档案点的情况下,只有 27%的交易成功率.

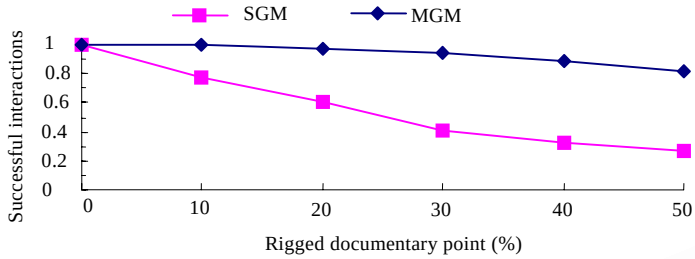


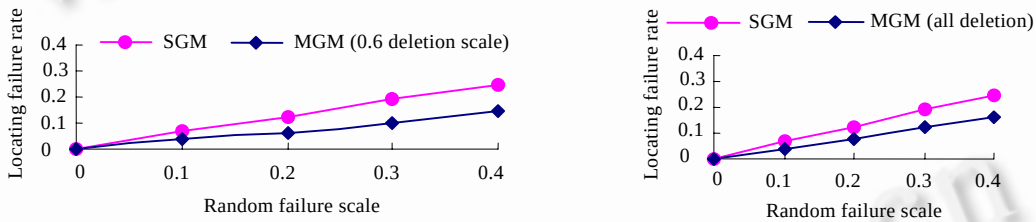
Fig.7 Test results of security mechanism simulation

图 7 安全机制对比测试结果

实验 **4**. 存储策略仿真.

存储策略仿真主要考察 MGM 模型改进后的存储方法对 MGM 模型的影响.

由第 4.2 节可知,领域档案点在可用存储空间已满时,会依据一定的策略删除其本身维护的部分档案信息,这会影响到 MGM 模型的容错能力及安全机制,如在整体档案点失效,同时领域档案点已满的情况下,某些低可信节点的档案可能会丢失,进而可能影响到可信度求解定位请求的失败.在删除规模分别为 0.6(如图 8(a)所示)和 1(如图 8(b)所示)的情况下,为整体档案点根据其可信度赋予不同的失效概率,观察不同的失效规模对可信度求解定位请求的影响.由结果可以看出,即使在删除规模为 1,不考虑领域档案点的情况下,MGM 模型的容错能力仍然比 SGM 模型具有明显的优势,这主要是因为在 MGM 模型中节点的档案点具有与节点相近或更高的可信度.



(a) With 0.6 deletion scale

(a) 0.6 删除规模下

(b) With all deletion

(b) 全部删除规模下

Fig.8 Fault-Tolerance compare

图 8 容错能力比较

6 结 论

本文提出了一种 P2P 环境下的多粒度信任模型,能够对用户针对某一具体领域的信任度进行量化,同时给出了该模型的数学分析及分布式实现方法.分析和仿真表明,该模型较已有模型在可信度计算的粒度、迭代的收敛性、容错能力、安全性等方面有较大的提高,具有更广泛的应用场景及较好的工程可行性.

致谢 在此,我们向对本文的工作给予支持的同行,尤其是国防科学技术大学计算机学院的窦文博士表示感谢.

References:

[1] Zhang Q, Sun Y, Liu Z, Zhang X, Wen XZ. Design of a distributed P2P-based grid content management architecture. In: Ilow J, ed. Proc. of the 3rd Communication Networks and Services Research Conf. New York: IEEE Press, 2005. 339−344.

[2] Adar E, Huberman BA. Free riding on Gnutella. Technical Report, CSL-00-3. Palo Alto: Xerox PARC, 2000.

[3] Dou W, Wang HM, Jia Y, Zou P. A recommendation-based peer-to-peer trust model. Journal of Software, 2004,15(4):571−583 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/571.htm

[4] Khambatti M, Dasgupta P, Ryu KD. A role-based trust model for Peer-to-Peer communities and dynamic coalitions. In: Cole JL, Wolthusen SD, eds. Proc. of the 2nd IEEE Int'l Information Assurance Workshop. New York: IEEE Press, 2004. 141−154.

[5] Wang Y, Vassileva J. Bayesian network trust model in peer-to-peer networks. In: Moro G, ed. Proc. of the 2nd Int'l Workshop on Agents and Peer-to-Peer Computing. Berlin: Springer-Verlag, 2004. 23−34.

[6] 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.

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

[8] 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).

[9] Ratnasamy S, Shenker S, Stoica I. Routing algorithms for DHTs: Some open questions. In: Druschel P, ed. Proc. of the 1st Int'l Workshop on P2P Systems. Berlin: Springer-Verlag, 2002. 45−52.

[10] Ogilvie P, Callan J. The effectiveness of query expansion for distributed information retrieval. In: Paques H, Liu L, Grossman D, eds. Proc. of the 10th Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2001. 183−190.

[11] Ankerst M, Breunig M, Kriegel HP, Sander J. OPTICS: Ordering points to identify the clustering structure. In: Delis A, Faloutsos C, Ghandeharizadeh S, eds. Proc. of the 1999 ACM SIGMOD Int'1 Conf. on Management of Data. New York: ACM Press, 1999. 49−60.

[12] Qiu YG, Frei HP. Concept based query expansion. In: Korfhage R, Rasmussen EM, Willett P, eds. Proc. of the 16th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 1993. 160−169.

附中文参考文献:

[3] 窦文,王怀民,贾焰,邹鹏.构造基于推荐的 Peer-to-Peer 环境下的 Trust 模型.软件学报,2004,15(4):571−583. http://www.jos.org.cn/1000-9825/15/571.htm

[8] 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.

张骞(1979－),男,山东金乡人,博士生,主要研究领域为 P2P 计算.

张霞(1965－),女,博士,教授,CCF 高级会员,主要研究领域为 P2P 内容管理,数据库技术.

文学志(1970－),男,博士生,主要研究领域为 P2P 安全.

刘积仁(1955－),男,博士,教授,博士生导师,主要研究领域为计算机网络技术.

Ting Shan (1978－),男,博士生,主要研究领域为 P2P 安全.