ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.1, January 2007, pp.157−167
DOI: 10.1360/jos180157

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

基于相似度加权推荐的 P2P 环境下的信任模型*

李景涛+, 荆一楠, 肖晓春, 王雪平, 张根度

(复旦大学 计算机与信息技术系,上海 200433)

A Trust Model Based on Similarity-Weighted Recommendation for P2P Environments

LI Jing-Tao+, JING Yi-Nan, XIAO Xiao-Chun, WANG Xue-Ping, ZHANG Gen-Du

(Department of Computer and Information Technology, Fudan University, Shanghai 200433, China)

+ Corresponding author: Phn: +86-21-51355369 ext 809, E-mail: lijt@fudan.edu.cn, http://www.fudan.edu.cn

Li JT, Jing YN, Xiao XC, Wang XP, Zhang GD. A trust model based on similarity-weighted recommendation for P2P environments. *Journal of Software*, 2007,18(1):157−167. http://www.jos.org.cn/1000-9825/18/ 157.htm

Abstract: In decentralized peer-to-peer file-sharing networks, due to the anonymous and self-organization nature of peers, they have to manage the risk involved with the transactions without prior knowledge about each other's reputation. SWRTrust, a global trust model, is proposed to quantify and to evaluate the trustworthiness of peers, which includes a mathematical description and a distributed implementation. In SWRTrust, each peer is assigned a unique global trust value, computed by aggregating similarity-weighted recommendations of the peers who have interacted with it. Previous global trust models are based on the assumption that the peers with high trust value will give the honest recommendation. This paper argues that this assumption may not hold in all cases. Theoretical analyses and experimental results show that SWRTrust is still robust under more general conditions where malicious peers cooperate in an attempt to deliberately subvert the system, converges more quickly, and decreases the number of inauthentic files downloaded more effectively than the previous models.

Key words: peer-to-peer; trust; distributed hash table; similarity; anonymity; convergence speed

摘 要: 在诸如文件共享等无中心的 Peer-to-Peer 网络中,对等节点具有匿名性和高度自治的特点,并且由于缺乏对与之交互的节点的可信程度的知识,节点需应对交互过程中可能出现的威胁.提出了一种基于节点评分行为相似度加权推荐的 peer-to-peer 环境下的全局信任模型(简称 SWRTrust),用于量化和评估节点的可信程度,给出了模型的数学表述和分布式实现方法.已有的全局信任模型建立在信任值高的节点其推荐也更可信这个假设基础上,SWRTrust 对该假设的合理性提出了质疑.分析及仿真实验结果表明,SWRTrust 较已有模型适用于遏制更广泛类型的恶意节点攻击,在迭代的收敛速度和网络中的成功下载率等性能指标上有较大提高.

关键词: 对等网络;信任;分布式哈希表;相似度;匿名性;收敛速度
中图法分类号: TP393 文献标识码: A

类似 Gnutella[1]和 Kazaa[2]的文件共享 peer-to-peer(P2P)网络在 Internet 上的应用取得了成功.P2P 网络因其开放性以及节点(peer)的匿名性和自治等本质特征,为计算机病毒、垃圾数据、伪造文件等在 P2P 网络上的传

* Supported by the National Natural Science Foundation of China under Grant No.60373021 (国家自然科学基金)
Received 2005-07-11; Accepted 2006-02-23

播提供了有利的条件,如 VBS.Gnutella worm 蠕虫病毒[3]的流行.最近对 Kazaa[2]的研究表明,有超过 50%的音频文件是被污染的(polluted)[4].另外,由于缺乏激励机制,有 25%的节点是 free riders[5](只从其他节点下载文件,而不提供文件上载服务).文献[6]使用博弈论模型,从理论角度分析表明:引入依据节点的"可信程度"的高低对节点区分服务这种机制是解决以上两个问题的一个有效的办法.文献[7–13]均给出了各自的信任模型,并通过仿真实验说明 P2P 网络中引入信任机制可以抑制上述问题.

所谓信任模型,是指建立量化的评价体系,以信任值度量节点的"可信程度".通过节点间的交互历史所反映出的节点的"可信程度",本质上是节点的实际物理属性基于其参与策略的一个综合能力的"投影",既可以反映节点的物理能力,也同时体现了节点参与网络的主观态度[13].信任模型有不同的具体应用模式,比如 Kazaa[2]让信任值高的节点在下载时获得较多的带宽.

分布式、匿名性和自治是 P2P 网络的本质属性,也是其成功的主要原因.因此,在 P2P 环境下的信任模型要解决的核心问题是:在不损失这些本质属性的前提下,实现节点信任数据的计算、存储和分发的在网络中分布式进行,并且在上述各环节应占用很少的网络资源,同时保证规模的可扩展性.P2P 环境下的信任模型所面临的首要挑战是协同作弊问题[8,9]:在恶意节点形成作弊的团体、协同伪造信任值时,信任模型能否有效地识别,乃至遏制作弊行为是评价模型的重要指标.

本文旨在构造一种 P2P 环境下的基于相似度加权推荐的全局信任模型(简称 SWRTrust),第 2 节给出模型的数学表述.第 3 节给出信任值分布式求解算法.第 4 节的仿真实验表明,该模型在遏制协同作弊方面较文献[9,10]的模型有一定优势.第 5 节总结全文.

1 相关工作

引入 PKI 系统中一些成熟的机制来实现信任模型是一种可行方案[11].这类系统中,有一个或一组权威节点维护一个可信的节点集合.这些权威节点可以颁发证书给可信的新加入的节点,节点以证书作为其身份的凭证使用网络中的资源,这类系统往往是中心依赖的,与 P2P 的分布式属性不相符合,有单点失效问题.此外,对数字证书的信任是客观的,可用于验证节点的实际身份,却无法反映节点参与网络的主观态度.

不依赖于可信第三方的信任模型主要有两类:基于微支付的模型和基于社会信任网络的模型.在基于微支付的模型[12]中,节点接受服务需支付一定的虚拟货币,提供服务可以获得虚拟货币.然而,这需要一个完整的计费系统跟踪记录每一笔小额交易,因此不具有工程可行性[6].

基于社会信任网络的模型借鉴社会学有关信任的研究成果,又可分为局部信任模型和全局信任模型.在局部信任模型系统中,节点通过询问有限的其他节点以获取它们对某个节点的推荐度,再综合自己和该节点交互的历史经验,确定节点的信任值.在这类系统中,往往采取简单的局部广播的手段,其获取的节点信任值也往往是局部的和片面的[10].Xiong Li[8]给出了一个适用于 P2P 电子社区的局部信任模型,节点的可信度是对以往该节点向其他节点提供服务的水平的综合评价.模型考虑全面,引入了节点对交互的反馈、反馈的可信度、节点参与交互的次数、交易的属性和节点所在社区 5 个因素度量节点的可信程度.

全局信任模型引入的目的在于削弱恶意节点协同作弊的效果,节点具有唯一的全局信任值,它综合了整个网络对该节点的信任评价,因此恶意节点不能仅通过"相识"的若干个同伙给出不实评价而获得高的信任值.Eigentrust[9]和窦文的模型[10]是已知的全局信任模型.然而,它们均仅使用节点的全局信任值本身作为推荐度的权重,即假设具有高全局信任值的节点其推荐也更加可信.但这个假设并不总是成立的,因此,在文献[9]提到的伪装的恶意节点一类攻击下(见第 4.5 节实验),这两个模型的性能较差.

针对上述诸问题,本文提出 SWRTrust 模型,做出的主要贡献有:

1) 提出了一个基于相似度加权推荐的全局信任模型(SWRTrust),放宽了上述假设条件,以节点评分行为的相似度加权其推荐度计算全局信任值,仿真实验中(尤其在伪装的恶意节点模型下),模型的性能明显优于 eigentrust 和窦文的模型.
2) 引入一个简单算法优化相似度计算,以解决相似度计算中的稀疏性问题,这个问题在节点数量庞大

的 P2P 环境下尤为重要,已知的信任模型中缺乏对这个问题的讨论.

3) 使用改进的 Chord 方案实现信任数据的分布式存储与查询,首次明确提出了防止伪造信任数据的 3 个对节点匿名性的要求.

2 SWRTrust 全局信任模型

2.1 全局信任值的度量

全局信任值是从整个网络角度观察得到的节点 i 的信任值,它综合了网络中所有节点的对 i 的评价.在 n 个节点的网络中,任意节点 i 具有全局唯一的信任值,记作 T_i.

对于任意节点 j ($j\neq i$),其个人对 i 的信任评价可以为 $bZ_{ji}+(1-b)T_i$,$b\in[0,1]$,其中,Z_{ji} 表示 j 与 i 的直接交互经验.即节点 j 在对 i 信任评价时,可以调节 b 的大小,在直接经验和全局信任值之间有所取舍.

定义 **1**. 局部满意度 $S_{ij}=G_{ij}-F_{ij}$.公式中:G_{ij} 为在节点 i 看来与节点 j 交互成功的次数;F_{ij} 为在节点 i 看来与节点 j 交互失败的次数.由定义可知:S_{ij} 是节点 i 对以往与节点 j 交互经验的总结,$S_{ij}>0$ 说明 i 对 j 有正面的评价;$S_{ij}<0$ 说明 i 对 j 有负面的评价.

定义 **2**. 局部信任值,即归一化的局部满意度,其定义为 $L_{ij}=\dfrac{\max(S_{ij},0)}{\sum_j \max(S_{ij},0)}$,若 $\sum_j \max(S_{ij},0)=0$,则令 $L_{ij}=T_i/n$,n 为网络规模.L_{ij} 是节点 i 根据直接交易历史对节点 j 作出的信任评分,也是 i 对 j 的推荐度.L_{ii} 无实际意义,令其为 0,否则,i 可以伪造 L_{ii},为自己作出很高的信任评分.

定义 **3**. 相似度 C_{ij} 刻画了节点 i 和节点 j 的评分行为的相似程度.j 与 i 的相似程度越高,说明 j 与 i 对网络中其他节点的看法越一致.相似度可以用多种函数刻画[14],本文采用余弦相似度函数,用两个向量的夹角余弦表示两个向量的相似度,即

$$C_{ij}=\frac{\sum_k B_{ik}B_{jk}}{\sqrt{\sum_k B_{ik}^2}\cdot\sqrt{\sum_k B_{jk}^2}}.$$

用作比较的是成功交易率向量 $B_i=[B_{i1},B_{i2},\ldots,B_{in}]$,其中,$B_{ik}$ 是 i 对 k 的成功交易率评分,是节点 i 看来与 k 以往交互历史中,成功或失败交互数占其与节点 k 交互总次数的比率.

当 $i\neq k$,$G_{ik}+F_{ik}=0$ 时,$B_{ik}=0$;

当 $i\neq k$,$G_{ik}+F_{ik}>0$ 时,$B_{ik}=\begin{cases}G_{ik}/(G_{ik}+F_{ik}), & G_{ik}\geq F_{ik}\\ -F_{ik}/(G_{ik}+F_{ik}), & G_{ik}<F_{ik}\end{cases}$.

对任意 i,令 $B_{ii}=1+\varepsilon$,其中,ε是任意小的正数*,即所有节点均认为自己是最值得信赖的.

定义 **4**. 推荐矩阵 $R=(R_{ij})$,R 的 i 行 j 列的元素为 $R_{ij}=L_{ij}\cdot C_{ij}$.不同于其他全局信任模型,本文的推荐矩阵元素是节点 i 和节点 j 的相似度 C_{ij} 与 i 对 j 的局部信任评分 L_{ij} 的乘积.这样做的原因将在定义 6 中给出.

定义 **5**. 集合 U_i 为与节点 i 交互过的上游节点的集合,其定义为

$$U_i=\{j|\text{节点 } j \text{ 与节点 } i \text{ 曾经交互过,并对 } I \text{ 作出了信任评分 } L_{ji}\}.$$

定义 **6**. 全局信任值向量 $T=[T_1,T_2,\ldots,T_n]^T$,上标“T”表示转置.在单次迭代中,节点 i 的全局信任值 T_i 是 U_i 中的所有节点对 i 的局部信任评分的加权和.在 eigentrust 和窦文的模型中,都仅以节点的全局信任值作为权重,即 $T_i=\sum_{k\in U_i}(L_{ki}\cdot T_k)$.这意味着节点的全局信任值越高,其推荐意见也越重要.直觉上这点似乎是成立的,然而,节点的全局信任值高低和节点推荐度的重要程度并不是完全一致的,例如在伪装的恶意节点攻击下(见第 4.5 节),部分恶意节点通过伪装获得较高的全局信任值,同时给其同伙较高的推荐度.针对这一点,本文提出基于相

* $i\neq j$ 时,B_{ij} 最大为 1.

似度加权推荐的全局信任值度量:

$$T_i = \sum_{k \in U_i} (L_{ki} \cdot C_{ki} \cdot T_k) \tag{1}$$

用 k 和 i 的评分行为的相似度 C_{ki} 作为 k 的推荐度的权重.该模型能有效遏制伪装的恶意节点攻击这类协同作弊,仿真实验印证了这一点.在其他类型的攻击下,该模型比 eigentrust 和窦文的模型也有更好的效果.其矩阵形式为

$$T=R^T \cdot T \tag{2}$$

2.2 全局信任值迭代计算的收敛性

基于式(2)的迭代收敛与否,决定了全局信任值向量 T 的解的存在性.下面用迭代矩阵 R^T 的范数小于 1 证明此简单迭代法的收敛性.

定理 **1**. 对于任意初始向量 $T^{(0)}$,基于式(2)的简单迭代法 $T^{(k+1)}=R^T \cdot T^{(k)}$收敛.

证明:上述迭代收敛的充分条件是矩阵 R^T 的范数$\|R^T\|<1$.[7]

因为$\|R^T\|_1=\max_i \sum_j |L_{ij} \cdot C_{ij}| \le \max_{i,j} |C_{ij}| \cdot \max_i \sum_j L_{ij} \le \max_{i,j} |C_{ij}|$.

当 $i=j$ 时,有 $L_{ij}=L_{ii}=0$,所以, $L_{ij} \cdot C_{ij} = L_{ii} \cdot C_{ii} =0$,亦即上式 $\max_{i,j} |C_{ij}|$ 中 $i \ne j$.

又因为 C_{ij} 是向量 B_i 和向量 B_j 的夹角的余弦,

并且 B_i 和 B_j 中至少有一个分量 $B_{ii}(B_{ii}=1+\varepsilon)$和 $B_{ji}(|B_{ji}| \le 1)$不相等,

所以对任意 i 和任意 j,$i \ne j$,必有$|C_{ij}|<1$,即 $\max_{i,j} |C_{ij}| <1$.也即$\|R^T\|_1<1$.命题得证. □

2.3 全局信任值的进一步讨论

如果允许节点 i 计算和提交自己的全局信任值,则 i 可以随意伪造 T_i.因此,本文第 3 节实现 SWRTrust 时,任意节点 i 均被指派其他节点作为其信任值管理节点,记作 M_i,用于存放、提交和计算 i 的信任值.

对 SWRTrust 模型一类可能的攻击是:如果恶意节点 i 和 j 的评分行为非常相似,此时,C_{ij} 和 C_{ji} 的值都比较大,且 i 和 j 之间互相给出不真实的高评分,试图获得较高的全局信任值(后文第 4 节中的 IM(individual malicious peers)类、CM(collective malicious peers)类和 DM(disguised malicious peers)类恶意节点具有该特点).注意式(1),全局信任值是综合所有和节点 i 交互过的节点对其的推荐度,并经过多次迭代后得到的,恶意节点 i 很难通过有限个相识的恶意节点来获得高的信任值,除非 i 仅和这几个节点交互,而从不向其他节点提供恶意服务.后文第 4 节仿真实验的结果说明:只要恶意节点的数量不超过正常节点,那么上述协同作弊行为是可以被识别的.信任值的评估者用自己和推荐者的评分行为的相似度作为是否采纳其推荐意见的依据,这具有一定的合理性.因为类似 SWRTrust 的信任模型[8−10]是基于推荐的模型,对这类模型直接、有效的攻击就是由“相识”的若干恶意节点伪造推荐度来互相“吹捧”,以获得虚高的信任值.节点相信和自己评分行为相似的节点给出的推荐意见,可以有效地遏制这类攻击:因为如果恶意节点“不按规则办事”,给出虚假的评分,那么,由于其评分行为和正常节点相似度低,正常节点不会采纳其推荐意见,恶意节点的“吹捧”目的无法达到;而如果恶意节点给出正确的评分,虽然其评分行为和正常节点相似,但其给出的正确推荐意见可以被用来作正确的信任值计算.

如何准确地估计两个向量之间的相似度,本身就是一个难的问题,特别是在系统规模扩大以及用于比较的向量的非零分量个数很少的情形下(稀疏性问题).本文第 2.4 节简要地给出了一种改进的相似度度量算法,用于解决相似度计算中的稀疏性问题.需要强调的是:本文的核心内容是提出一个度量节点全局信任值的模型,节点间的相似度仅作为一个度量的因素引入模型,其他适用的相似度度量方法均可以运用到 SWRTrust 模型中来.

2.4 相似度度量算法的改进

随着 P2P 网络中节点数量和资源总量的增加,节点 i 和节点 j 的成功交易率向量 B_i 和 B_j 会变得非常稀疏,两个向量中同时不为 0 的分量数目会很小,甚至为 0.受文献[15]启发,本文提出 C_Similarity(class-based similarity)算法度量节点评分行为的相似度.对任意节点 i,C_Similarity 先将节点 i 的成功交易率向量 B_i 的维数

降低,得到 i 的节点类评估向量,记作 A_i;然后,用 A_i 计算节点间的相似度.A_i 中等于 0 的分量个数很少,从而解决了用于相似度计算的向量稀疏性的问题.同时,由于 A_i 的维数远小于 B_i,相似度计算的消耗也会相应地降低.C_Similarity 算法可归结为 3 个步骤: (1) 将被评分节点分为 q 类,记作 $NC_1,\ldots,NC_q$;(2) 按公式 $A_{ik}=\sum_{u\in NC_k}B_{iu}$ 计算每个节点 i 对节点类 $NC_k(k=1,\ldots,q)$的评分 A_{ik},得到向量 $A_i,A_i=[A_{i1},A_{i2},\ldots,A_{iq}]$;(3) 计算 A_i 和 A_j 的夹角的余弦得到节点 i 和 j 的相似度.即

$$C_{ij}=\frac{\sum_{k=1}^{q}A_{ik}A_{jk}}{\sqrt{\sum_{k=1}^{q}A_{ik}^2}\cdot\sqrt{\sum_{k=1}^{q}A_{jk}^2}}.$$

有多种成熟的分类算法可以实现步骤(1)中对被评分节点的分类,本文实现时采用的是基于简单向量距离分类法的算法.C_Similarity 算法的设计原则出于以下考虑:SWRTrust 模型的应用背景是 P2P 网络环境,虽然相似度计算在节点本地进行,不会引起额外的网络通信量,但仍然要求计算的消耗和存储的开销尽可能地小.C_Similarity 算法能够有效地解决稀疏性问题,并且计算量小.算法实现时,仅需要在每个节点 i 的信任值管理节点存储一个对 i 的成功交易率评分向量即可.

3 全局信任值分布式计算的实现

3.1 信任数据的存储与查询

在 P2P 这样一种分布式的、节点高度自治的环境下,需要有一套非集中式的数据管理方案实现节点的信任值的分布式存放与查询.分布式哈希表(DHT)可以用于指派节点的信任值管理节点.本文实现中,基于 Chord 协议[16]为网络中每个节点设定信任值管理节点.网络中每个节点 i 有一个全局唯一的标识 ID_i,是一个 m 比特的二进制数.它是 i 在 Chord 环形逻辑空间中的逻辑地址,是 i 的物理地址的单向 Hash 值(例如在 Internet 环境下,可由节点的 IP 地址通过 SHA-1 计算得到).键值(key)也是 m 比特二进制数,它是由节点的标识(ID)通过单向 Hash 函数 H 计算得到的.如果 ID_i 是某个键值 k 最近的后继节点标识,那么节点 i 负责管理 k 所对应的节点的信任值.即有如下定义:

定义 7. 设 H 为任意均匀的单向 Hash 函数,网络中任意节点 i 的标识 ID_i 在 m 比特逻辑地址空间的投影 $H(ID_i)$的最近的后继节点标识所对应的节点称为 i 的信任值管理节点 M_i.即 ID_{M_i} 是 $H(ID_i)$最近的后继标识.

基于 Chord 协议,信任数据的定位和查询操作具有可以证明的消息复杂度上界[16].在 n 个节点的网络中,任意节点 i 可以在 $O(\log n)$的消息复杂度内将其对节点 j 的信任评分写入 M_i;同样,任意节点 i 可以 $O(\log n)$的消息复杂度从 M_j 获取节点 j 的信任数据(如 T_j).每个节点需要维护的路由表(finger table)的规模仅为 $O(\log n)$.

基于 Chord,网络中每个节点 i 均被指派了信任值管理节点,而同时,i 又是一个或若干个其他节点的信任值管理节点.作为信任值管理节点,M_i 至少具有 4 项功能:(1) 存储节点 i 的与信任值计算相关的数据;(2) 验证所存储的数据的合理性(比如:比较 i 最新提交的 G'_{ij} 和 M_i 当前所存储的 G_{ij},如果 $G'_{ij}-G_{ij}>1$,极有可能是节点 i 提交了不真实的数据);(3) 向其他节点提交 i 的全局信任值 T_i、成功交易率向量 B_i 等数据;(4) 计算它所管理的节点 i 的信任值.

3.2 信任值管理节点的匿名性

上述机制可以有效地指派任意节点 i 的信任值管理节点 M_i.然而,如果没有对节点匿名性的支持,节点 i 和 M_i 可以相互“勾结”,协作对信任数据做假.本文对信任数据分布式存放节点的匿名性的具体要求给出了定义:

(1) 任意节点 i 的信任值管理节点 M_i 是无法知道节点 i 的物理地址的;

(2) 任意节点 i 不能选择 i 自己的标识 ID_i,以便使 ID_i 正好是存储网络中某个节点 j 的信任值的逻辑地址;

(3) 任意节点 j 在与 M_i 通信时,如果 M_i 不是 j 的后继或前驱节点,且不在 j 的 Finger table 中,则 j 无法知道 M_i 的物理地址和确切的逻辑地址.

对于要求(1),由于 M_i 仅知道自己管理的键值和键值所对应的标识(ID_i),而 ID_i 是节点 i 的物理地址的单向 Hash 值,由于 Hash 函数的单向性,因此,M_i 无法知道 i 的物理地址;对于要求(2),由于节点 i 的标识 ID_i 是 i 的物理地址的单向 Hash 值,i 不能选择某个特定的物理地址(如 MAC 地址或 IP 地址),这一点也是满足的;实现要求(3),需要对 Chord 的部分 API 调用加以限制,本文采用 AChord 方案[17]中的 4 条限制规则,可以证明引入对要求(3)的支持后,并没有增加 Chord 的消息复杂度[17].改进后的 Chord 满足对信任值管理节点匿名性的要求,同时继承了 Chord 的性能和查询精确的特点.在 eigentrust 和窦文的设计中使用其他 DHT 算法,但均没有明确提出对节点匿名性的具体要求.

3.3 全局信任值分布式求解算法

引入信任值管理节点的全局信任值求解算法具有工程可行性,下面首先给出用到的两个原语及其语义:

Submit(ID_i,(ID_j,ID_k),Value1,Value2):将节点 j 对节点 k 的局部信任值等与信任计算相关的数据 Value1, Value2 提交到 i 的信任值管理节点 M_i.Value1 和 Value2 的具体含义由上下文决定.

Query(ID_j,T_j,L_{ji},B_j):基于 Chord 协议,查询节点 j 的全局信任值 T_j,j 对 i 的局部信任评分 L_{ji} 和 j 的成功交易率向量 B_j.

算法 **1**. 全局信任值求解算法.

网络中任意节点 i 同时具有两个角色:它既是用户节点,同时也是若干个用户节点的信任值管理节点.任意节点 i 作为一般用户节点和信任值管理节点的算法分别如下所示.

(1) 节点 i 作为一般用户节点的算法:

```
UpdateAndSubmitTrustdata( )   //节点 i 与 j 每次交互后更新并向 M_i 提交 G_ij 和 F_ij
{
    If (成功交易) G_ij←G_ij+1;
    else F_ij←F_ij+1;
    Submit(ID_i,(ID_i,ID_j),G_ij,F_ij);   //向 M_i 提交 G_ij 和 F_ij,并触发 M_i 的 UpdateLocaltrust( )过程
}
```

(2) 节点 i 作为节点 u 的信任值管理节点的算法:

```
UpdateLocaltrust( )   //i 收到 Submit(ID_u,(ID_u,ID_v),G_uv,F_uv)后,触发更新 L_uv,B_uv 的过程
  {
      验证 G_uv,F_uv 的合理性;
      依定义 1~定义 3 计算 S_uv,L_uv,B_uv;
    Submit(ID_v,(ID_u,ID_v),L_uv,B_uv);   //向 M_v 提交 u 对 v 的评分 L_uv 和 B_uv,触发 M_v 将 ID_u 加入集合 U_v 的过
                                          程,并可选地触发 M_v 的 CalcGlobaltrust( )过程
  }
  CalcGlobaltrust( )   //迭代计算 i 所管理的节点 u 的全局信任值
  {
       for (every j∈U_u ( j≠u))
       {
           Query (ID_j,T_j,L_ju,B_j);
           计算节点 j 和 i 的评分行为的相似度 C_ji;
           T_u←T_u+L_ju·C_ju·T_j;
       }
      return T_u;
  }
```

有两种模式触发 CalcGlobaltrust( )过程:每当节点 i 收到其他节点对 u 的局部信任评分时,触发 CalcGlobaltrust( )更新 u 的全局信任值;或者每当网络运行一定的时间后触发(例如,设置门限值 p,当 i 收到的对 u 的局部信任评分的数量大于 p 时触发).

先考虑每当收到其他节点对 u 的局部信任评分时就触发的情形:CalcGlobaltrust( )过程仅需查询那些与 u

交互过的节点,如果按照 Xiong Li[8]所给出的方法,则仅需查询与 u 最近一段时间窗口内交互过的节点来获得它们对 u 的评分,例如仅记录最近的 50 个与 u 所交互过的节点对 u 的评分,因为 u 较早交互行为对节点最近的表现的说服力较弱.这种方案也易于实现,只要 u 的信任值管理节点维护一个先进先出队列淘汰那些对较早时间进行的交互的评分即可.在下面的讨论中,我们设时间窗口为 f,即仅记录与 u 最近交互过的 f 个节点.在 eigentrust[9]中,求解节点 i 的全局信任值的迭代通过与 i 的交互过的节点在全网络范围扩散,直到所有节点的全局信任值的连续两次迭代的结果之差小于某个系统指定的极小常量,其消息复杂度为 $O(f^2)$.消息开销造成该协议仅仅适用于小规模网络.注意算法 1,在 CalcGlobaltrust( )中,仅需要以 Query(ID_j,T_j,L_{ju},B_j)原语查询一轮与 u 交互过的节点 j 的 T_j,L_{ju} 和 B_j,因此,我们的算法消息复杂度为 $O(f)$.窦文[10]方法的消息复杂度和我们的相同,后文第 4 节的实验说明,我们的算法与 eigentrust 和窦文的相比具有更好的迭代收敛特性.

引入门限 p 的目的正是想通过减少 CalcGlobaltrust( )过程的运行次数来减少网络中的不必要的数据通信量.因为信任值管理节点要等到收到 p 个对 u 的信任评分时才触发 CalcGlobaltrust( )更新 u 的全局信任值,而不是每次收到对 u 新的评分就更新.p 的选取与信任值的“新鲜度”有关,依具体的 P2P 应用而定为宜.若 p 过大,那么 u 的信任值得不到及时更新,会不够“新鲜”;若 p 过小,则会造成较大的数据通信量.

4 仿真及其结果分析

本文使用 Stanford 大学的 Query Cycle Simulator 软件包[18],该软件包仿真典型的文件共享式 P2P 网络.在该软件包中,我们加入了代码,实现了 SWRTrust 模型,同时也实现了 eigentrust 和窦文的模型作为对照.若无特别声明,SWRTrust 模型中的相似度度量均采用定义 3 给出的方法.

在该软件包中,每次仿真由若干个周期(cycle)组成,每个周期内,网络中的每个节点随机地发出一个文件下载请求后等待,收到拥有文件的节点对它的响应后,从所有响应的节点中选择全局信任值最高的节点下载文件.如果下载不成功,则从响应节点列表中删除该节点,再从列表中选择信任值最高的节点下载,直到成功下载为止.没有发出下载请求的节点要么应答或转发请求,要么处在不活跃状态.我们在每个周期结束时收集数据,然后仿真进入下一个周期.每个仿真进行多次,实验分析中所使用的数据均为多次实验结果的平均值.系统初起时,令每个节点的信任值均为 $1/n$,即 $T^{(0)}=[1/n,1/n,...,1/n]^T$.$n$ 为仿真的 P2P 网络的规模,取其为 500 个节点,恶意节点的邻居节点个数为 6 个,而正常节点仅有 3 个邻居节点,拥有的邻居节点越多意味着越有可能被查询到,可以有更多的机会提供恶意的上载服务;并且,恶意节点均谎称拥有热门类别的文件,这样做也会提高被选中作为下载源的可能性,而正常节点无此假设.对于 eigentrust,有 10 个高可信节点,高可信节点拥有 6 个邻居.网络的其他设置参见文献[19].

4.1 节点类型定义

网络中的节点分为两大类:正常节点和恶意节点.正常节点无论在提供服务上(上载)还是在对其他节点的评价上(提交对其他节点的信任评分)都是真实的.恶意节点不提供真实的(上载)服务,并且对其他节点的信任评分不真实,甚至诋毁正常节点.我们构造了多种恶意节点类型来评估 SWRTrust 模型:

(1) 孤立的恶意节点,这类节点没有形成协同作弊的团体,我们称其为 IM 类.除了提供不真实的(上载)服务,IM 诋毁与其交互过的正常节点,并夸大与其交互过的恶意节点,即该类恶意节点令 $S_{ij}=F_{ij}-G_{ij}$.

(2) 协同的恶意节点,该类节点形成了协同作弊的团体,我们称其为 CM 类.除了实现 IM 类的所有功能,每个节点还极力夸大团体中的同伙,即对任意节点 j,若有 j 和 i 属于同一 CM 团体,则 i 令其对 j 的局部信任评分 $L_{ij}=1/||CM||$,$||CM||$是 CM 节点的规模.

(3) 伪装的恶意节点,考虑了恶意节点的组合,IM 类和 DM 类同时出现,且互有分工.IM 类同上,DM 类节点提供真实的热门文件的上载服务(可以很快获得高的全局信任值),同时极力夸大 IM 团体中的节点,给 IM 类的节点很高的局部信任评分.即,若有节点 u 属于 DM 类,则对任意属于 IM 类的节点 v,有局部信任值 $L_{uv}=1/||IM||$.

事实上,某几种恶意节点的组合均可以形成新的恶意节点类型,以上 3 类恶意节点是其中具有代表性的,并且类型(2)和类型(3)中的恶意节点已经非常“狡猾”了.对其他一些恶意节点类型我们也进行了仿真,下面给出的

结论同样是成立的.

4.2 性能评价指标

在实验中,Chord 协议被简化,因为我们重在评估 SWRTrust 算法对抗恶意节点攻击的性能及其与同类算法的比较.性能评估的重要指标是算法的收敛速度和网络中的成功下载率.

成功下载率记作α,是正常节点的成功下载的次数占正常节点所有下载次数的比率,α直观地反映了信任模型应用的效果.每个仿真结束后,统计所有正常节点 i 所观察到的 G_{ij} 和 F_{ij} 来计算α,$\alpha=\dfrac{\sum_i\sum_j G_{ij}}{\sum_i\sum_j (G_{ij}+F_{ij})}$.

收敛速度体现了信任模型见效的速度.系统初起时,所有节点的全局信任值相等,节点选择下载源具有随机性.经过一定次数的下载,正常节点获得高的信任值,被选作下载源的可能性增大,网络的成功下载率维持在较高的水平.在后面的实验分析中,收敛速度是由网络中所有正常节点总的失败下载(inauthentic downloads)的次数随仿真周期的变化规律来反映的(因为如果节点的全局信任值以我们所期望的方式收敛,即任取正常节点 i 和恶意节点 j,有 i 的全局信任值大于 j 的全局信任值以很高的概率成立.由于节点总是选取信任值高的节点下载,那么网络中失败下载的次数接近于 0).经过较少的周期,网络中失败下载的次数便接近于 0,说明算法收敛较快.失败下载次数记作β,每个仿真周期结束时统计所有正常节点 i 所观察到的 F_{ij} 来计算β,$\beta=\sum_i\sum_j F_{ij}$.

4.3 IM类仿真及讨论

IM 类仿真是指网络中的恶意节点均为 IM 类,调整网络中 IM 类节点的数量以观测实施 SWRTrust, eigentrust 和窦文的模型的效果.

图 1 给出了网络中 IM 节点数占 40%时,分别使用 4 种算法选择下载源时,失败下载次数β趋于 0 的速度.可以看到:使用 SWRTrust 模型从第 6 个周期开始,几乎完全杜绝了失败下载的发生,这说明 SWRTrust 完全遏制了恶意节点,使其无法获得较高的信任值.图中的 Random 是指不使用任何信任模型,节点每次随机选取下载源下载.实际上,Random 给出了最坏情况下β的变化情况.在后面的实验中,我们同样用 Random 算法作为参照.在其他的 IM 节点规模(占节点总数的 10%~50%)下,实验结果与图 1 类似.图 2 给出了在 IM 节点的规模发生变化时(分别占节点总数的 10%,20%,30%,40%,50%),成功下载率α的变化情况.可以看出:使用 SWRTrust 模型,在 IM 节点数接近 50%时,α仍能维持在 80%以上.

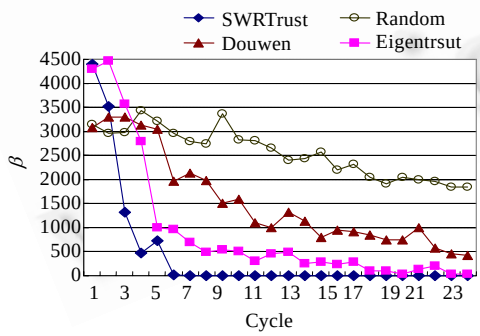

Fig.1 β of each cycle when the simulation proceeds

图 1 β随仿真周期的变化规律

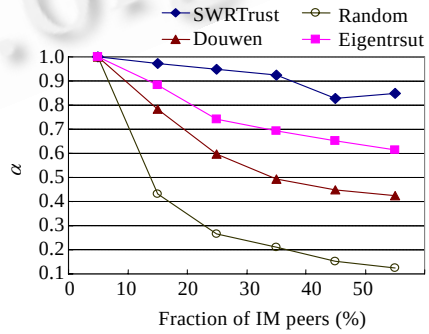

Fig.2 α vs. different percentage of IM peers

图 2 不同规模 IM 类节点存在时的α

4.4 CM类仿真及讨论

CM 类恶意节点彼此“相识”,它们具有更强的协同作弊能力.SWRTrust 中节点相信与自己行为相似的节点的推荐,相当于正常节点也形成了协作的团体.因此在 CM 类仿真中,SWRTrust 比其他模型更有优势.图 3 给出了

CM 节点数分别占总节点数 10%,20%,30%,40%,50%时α的变化情况.当 CM 节点数占 40%时,β随仿真周期的变化规律与图 1 相似,由于篇幅所限,这里不再绘图说明.

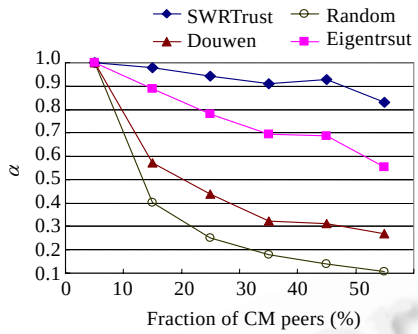

Fig.3 α vs. different percentage of CM peers

图 3 不同规模 CM 类节点存在时的α

4.5 IM&DM类仿真及讨论

在 eigentrust[9]中,针对 IM&DM 类恶意节点类型做了仿真实验,但效果不理想,文献[9]也提到这种类型的恶意节点能破坏 eigentrust 模型.窦文[10]没有提到这种类型的恶意节点模型.如图 4 所示,本文的仿真实验表明:在 IM&DM 类型节点存在的网络中,即便只有少量的 DM 类节点,eigentrust 和窦文的模型均不收敛(即不能通过全局信任值高低区分出正常节点和恶意节点),这印证了 eigentrust 的实验结论.不收敛的根本原因在于:eigentrust 和窦文均仅使用节点的全局信任值作为推荐度的权重,即假设具有高信任值的节点其推荐意见也更可靠.然而,DM 类节点虽然具有高的信任值,但其推荐意见却是虚假的.SWRTrust 使用节点的相似度加权其推荐度,可以甄别出 DM 类节点,因此仍然具有很好的性能.

图 4 给出了网络中 IM 和 DM 节点数占 40%,其中,DM 节点占恶意节点总数 10%时,分别使用 4 种算法选择下载源时β的变化规律.可以发现:SWRTrust 的性能甚至优于图 1 所示的只有 IM 类的情形,这是由于 DM 类节点贡献了一部分成功的上载服务.即在 IM&DM 恶意攻击下,IM 类节点提供恶意的上载服务是需要 DM 类节点支付成本的.当 DM 类节点占恶意节点数的比例大于 10%时,对 eigentrust 和窦文的模型的干扰更大,两者均不收敛.下一个实验的目的在于了解 DM 类节点所占的比例变化对 SWRTrust 性能的影响.图 5 给出了使用 SWRTrust 模型,在网络中 IM 和 DM 类节点占 40%时,DM 类节点占恶意节点总数的比例从 10%~45%变化,β随仿真周期变化的情况.

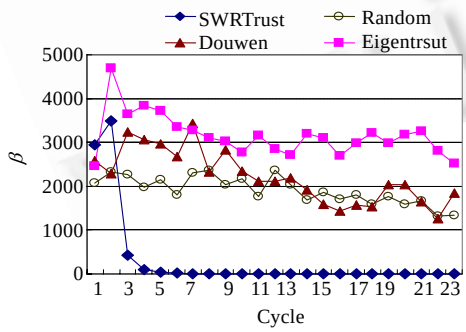

Fig.4 β of each cycle when the simulation proceeds

图 4 β随仿真周期的变化规律

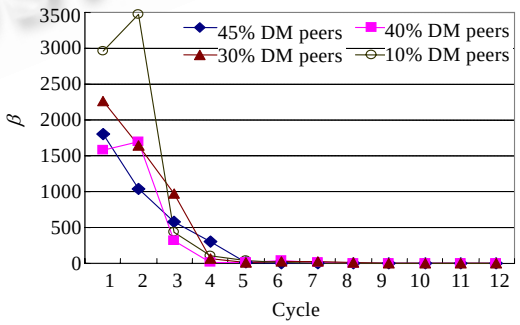

Fig.5 β vs. different percentage of DM peers

图 5 DM 类节点所占比例变化时的β

可以看到一个有趣的现象:随着 DM 类节点数量的增加,β反而减少.这说明 IM&DM 类型恶意节点对 SWRTrust 没有影响,DM 类节点越多,则为网络贡献的正常的上载服务也越多,而 DM 类对 IM 类的夸大作用却被 SWRTrust 遏制了.当 DM 类节点超过 50%时,SWRTrust 出现不收敛的情况.但此时,恶意节点中有超过一半的

节点提供正常上载服务,这对恶意节点来说支付的成本太高,因此不具有使用该模式提供恶意服务的动机.

4.6 在节点给出的评分数量稀少的情况下的算法性能

随着 P2P 网络中节点数量和资源总量的增加,节点 i 的成功交易率向量 B_i 会变得稀疏,很难正确估计两个节点的评分行为的相似度.本节讨论此情形下分别实施 SWRTrust 和 C_SWRTrust(基于 C_Similarity 的 SWRTrust 模型)的效果.网络中 CM 节点数占 40%时,分别进行了两组实验:第 1 组限定节点仅对 1%与其交互过的节点给出评分;第 2 组将这个比率提高到 5%.图 6 给出分别使用上述两种算法选择下载源时失败下载次数β变化的规律,网络中其他类型的恶意节点存在时的实验结果类似,这里不再另绘图说明.

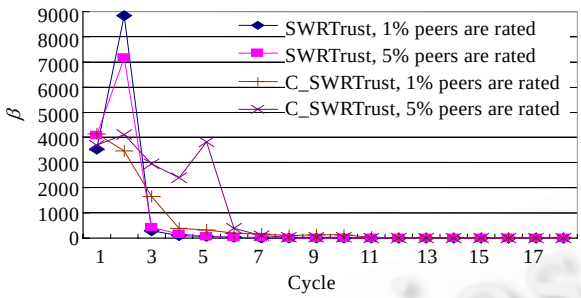

Fig.6 β measured when only sparse peers are rated

图 6 节点给出的评分数量稀少时的β

在仿真开始的几个周期内,使用 SWRTrust 模型时,网络中失败下载次数较多;而使用 C_SWRTrust 时曲线较平滑,失败下载次数较少.这是由于节点的成功交易率向量中非零分量个数稀少,节点间相似度评估不准确,因此节点选择下载源的依据——全局信任值不能反映节点实际的可信程度,造成了大量失败的下载.而 C_SWRTrust 模型较好地解决了稀疏性问题,因此没有出现失败下载次数急剧增大的情况,但算法收敛较慢.

5 结论及下一步工作

本文提出了适用于 P2P 环境的基于相似度加权推荐的全局信任模型 SWRTrust;证明了全局信任值迭代求解算法的收敛性,并给出了求解算法的分布式实现;讨论了信任值存放节点的匿名性要求.仿真实验说明,SWRTrust 模型对几类典型的协同作弊行为有一定的识别和遏制能力.

其他类型的恶意节点攻击以及它们可能对 SWRTrust 产生的影响需进一步研究.AChord[17]仅提供一种“弱”的节点匿名性支持,更有效的保障信任数据安全性的机制有待研究.SWRTrust 具体实现时可以基于任何一种适用于 P2P 网络的安全机制,例如 Herbivore[19].信任值管理节点本身是否可信是一个问题,一个解决方案是大多数表决机制,为每个节点设定若干个信任值管理节点,均存储信任数据的副本,采纳多数信任值管理节点提交的信任数据.提供数据可用性和大多数表决机制的副本技术有待探索.

致谢 在此,我们向本文的审稿人表示感谢,感谢他们对本文提出的深入而有建设性的修改意见.

References:

[1] Gnutella. http://www.gnutella.com/

[2] Kazaa. http://www.kazaa.com/

[3] VBS.Gnutella Worm. http://securityresponse.symantec.com/avcenter/venc/data/vbs.gnutella.html

[4] Liang J, Kumar R, Xi Y, Ross K. Pollution in P2P file sharing systems. In: Makki K, Knightly E, eds. Proc. of the IEEE Infocom 2005, Vol.2. Miami: IEEE Press, 2005. 1174−1185.

[5] Saroiu S, Gummadi PK, Gribble SD. A measurement study of P2P file sharing systems. In: Kienzle MG, Shenoy PJ, eds. Proc. of the Multimedia Computing and Networking 2002 (MMCN 2002). SPIE Press, 2002. http://www.cs.washington.edu/homes/gribble/papers/mmcn.pdf

[6] Buragohain C, Agrawal D, Suri S. A game theoretic framework for incentives in P2P systems. In: Shahmehri N, Graham RL, Carroni G, eds. Proc. of the 3rd Int'l Conf. on Peer-to-Peer Computing (P2P 2003). Los Alamitos: IEEE Press, 2003. 48−56.

[7] Shi WM, Yang HF, Wu YS, Sun X. Numerical Analysis. 2nd ed., Beijing: Beijing Institute of Technology Press, 2004. 91−93 (in Chinese).

[8] Xiong L, Liu L. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. IEEE Trans. on Knowledge And Data Engineering, 2004,16(7):843−857.

[9] Kamvar SD, Schlosser MT, Garcia-Molina H. The eigentrust algorithm for reputation management in P2P networks. In: Bakonyi P, Hencsey G, *et al*., eds. Proc. of the 12th Int'l World Wide Web Conf. Budapest: ACM Press, 2003. 640−651.

[10] 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

[11] Altman J. PKI security for JXTA overlay networks. 2003. http://www.jxta.org/docs/pki-security-for-jxta.pdf

[12] Golle P, Leyton-Brown K, Mironov I. Incentives for sharing in peer-to-peer networks. In: Wellman MP, Shoham Y, eds. Proc. of the 3rd ACM Conf. on Electronic Commerce. New York: ACM Press, 2001. 264−267.

[13] 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).

[14] Deng AL, Zhu YY, Shi BL. A collaborative filtering recommendation algorithm based on item rating prediction. Journal of Software, 2003,14(9):1621−1628 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/1621.htm

[15] Zeng C, Xing CX, Zhou LZ. Similarity measure and instance selection for collaborative filtering. In: Bakonyi P, Hencsey G, *et al*., eds. Proc. of the 12th Int'l World Wide Web Conf. (WWW 2003). New York: ACM Press, 2003. 652−658.

[16] Stoica I, Morris R, Karger D. Chord: A scalable peer-to-peer lookup service for Internet applications. Technical Report, MIT, 2002. http://pdos.csail.mit.edu/chord/

[17] Hazel S, Wiley B. Achord: A variant of the chord lookup service for use in censorship resistant peer-to-peer publishing systems. In: Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems (IPTPS 2002). Cambridge, 2002. http://www.cs.rice.edu/Conferences/IPTPS02/

[18] Schlosser MT, Condie TE, Kamvar SD. Simulating a file-sharing P2P network. In: Proc. of the 1st Workshop on Semantics in P2P and Grid Computing (SemPGRID 2003). Budapest, 2003. 69−80. http://www.isi.edu/~stefan/SemPGRID/

[19] Goel S, Robson M, Polte M, Sirer EG. Herbivore: A scalable and efficient protocol for anonymous communication. Technical Report, TR2003-1890, Department of Computing and Information Science, Cornell University, 2003.

附中文参考文献:

[7] 史万明,杨骅飞,吴裕树,孙新.数值分析.第 2 版,北京:北京理工大学出版社,2004.91−93.

[10] 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

[13] 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.

[14] 邓爱林,朱扬勇,施伯乐.基于项目评分预测的协同过滤推荐算法.软件学报,2003,14(9):1621−1628. http://www.jos.org.cn/1000-9825/14/1621.htm

李景涛(1975－),男,甘肃兰州人,博士生,主要研究领域为计算机网络,分布式系统,信息安全.

王雪平(1974－),男,讲师,主要研究领域为计算机网络,网络安全,信息安全.

荆一楠(1978－),男,博士生,主要研究领域为计算机网络,信息安全,网络安全.

张根度(1937－),男,教授,博士生导师,主要研究领域为计算机网络,信息安全,数据通信,无线通信,信息工程.

肖晓春(1970－),女,博士生,主要研究领域为计算机网络,信息安全,网络安全.