第41卷 第8期 西 安 交 通 大 学 学 报 Vol.41 №8
2007年8月 JOURNAL OF XI′AN JIAOTONG UNIVERSITY Aug. 2007

# 一种可扩展的反馈信任信息聚合算法

李小勇，桂小林，赵娟，冯大鹏
（西安交通大学电子与信息工程学院，710049，西安）

**摘要**：针对现有的动态信任聚合算法中利用基于信任链的广播方式进行反馈信任信息搜索而导致的系统运算收敛慢、可扩展性差等问题，建立了直接信任树(DTT)的概念，并基于DTT提出了一种新的可扩展的反馈信任信息聚合算法.根据节点之间的直接信任关系构建DTT，然后利用DTT进行反馈信任信息搜索，同时引入质量因子和距离因子两个参数来自动调节聚合计算的规模.仿真实验表明，算法能够显著提高反馈信任信息聚合计算的收敛性，具有较好的恶意反馈行为检测能力，在恶意节点比率增大时，算法也表现出较强的稳健性.

**关键词**：信任信息搜索；聚合算法；直接信任树；可扩展性
**中图分类号**：TP311 **文献标识码**：A **文章编号**：0253-987X(2007)08-0879-05

# Novel Scalable Aggregation Algorithm of Feedback Trust Information

Li Xiaoyong, Gui Xiaolin, Zhao Juan, Feng Dapeng
(School of Electronics and Information Engineering, Xi′an Jiaotong University, Xi′an 710049, China)

**Abstract**: Focusing on the problem that current dynamic trust aggregation algorithms searching for feedback trust information is based on trust chain in broadcast way, leading to slow in convergence and bad in scalability, a new scalable feedback trust information aggregation algorithm is proposed, in which the concept and definition of direct trust tree (DTT) are presented. Based on DTT a novel aggregating algorithm of scalable feedback trust information is proposed, and the DTT is constructed according to node's direct trust relation. Then, the feedback trust information is searched by using DTT instead of in broadcast manner. Two new parameters, quality factor and distance factor, are introduced to adjust the scale of aggregating computation automatically. Simulation results show that the algorithm remarkably enhances the convergence of feedback trust information aggregating computation. Moreover, it has better detecting capability against malicious feedback actions and has stronger robustness when the ratio of malicious nodes is increased.

**Keywords**: trust information search; aggregating algorithm; direct trust tree; scalability

随着对网格计算、普适计算、P2P计算、ad hoc等大规模的分布式应用系统的深入研究，系统表现为由多个软件服务组成的动态协作模型，系统形态正从面向封闭的、熟识用户群体的和相对静态的形式向开放的、公共可访问的和动态协作的服务模式转变.在这种动态的和不确定的环境里，公钥基础设施中基于证书机构的静态信任机制已不能适应这种需求，由此产生了动态信任管理问题[1-4].

动态信任管理研究中的一个关键问题是如何通过有效的方式聚合反馈信任信息[5].虽然文献[6-8]采用不同的数学方法进行反馈信任信息的聚合，但他们都通过基于信任链的广播方式在整个系统中进

收稿日期：2007-01-08. 作者简介：李小勇(1975～)，男，博士生；桂小林(联系人)，男，教授，博士生导师. 基金项目：国家自然科学基金资助项目(60273085)；国家高技术研究发展计划资助项目(2002AA104310)；教育部新世纪优秀人才计划资助项目(NCET-05-0829).

行反馈信任的搜索，导致了在大规模的分布式环境下系统运算的慢收敛性和巨大的时空开销，进而影响了系统的可扩展性.

在大规模分布式系统中，对于任何信任管理机制的设计，可扩展性是首要考虑的因素[8]. 本文就此建立了一种基于直接信任树(Direct Trust Tree, DTT)的反馈信任信息聚合机制，以此解决反馈信任信息搜索和聚合计算的收敛性问题，增强算法在大规模分布式系统中的可扩展性.

# 1 算法设计

## 1.1 DTT的构建

首先给出主体的定义和分类，然后进行直接信任关系的定义和DTT的构建.

**定义1** 根据担任角色的不同，主体分为3种类型：① 服务提供者(Service Provider, SP)；② 服务请求者(Service Requester, SR)；③ 反馈者(Feedback Rater, FR).

SP可以提供不同质量的服务，设某主体 $P_i$ 可提供 $m$ 个级别的服务 $S=\{s_1,s_2,\cdots,s_m\}$，SP向SP请求某种服务，SP根据需要请求FR提供反馈信任，最后SP根据评估算法进行信任评估，进行是否提供服务的决策，以保障系统的安全.

直接信任表示在给定的上下文中，一个主体根据直接接触行为的历史纪录而得出的对另外一个主体的信任程度. 主体 $P_i$ 和 $P_j$ 之间的直接信任值用 $E(P_i,P_j)$ 表示，并规定 $E(P_i,P_j)=\{e_{ij}\mid e_{ij}\in[0,1]\}$，由 $P_i$ 根据自己与 $P_j$ 的直接交互经验计算得到，主体交互之后彼此提交满意度的评价，我们采用概率可能性的方法来区分主体提供的不同服务质量，0表示完全不满意，1表示完全的满意，值越大表示满意度越高.

**定义2** 设主体 $P_i$ 与 $P_j$ 在最近的 $h$ 个交互中产生的信任满意度评价为集合 $E_{ij}=\{e_{ij}^{(1)},e_{ij}^{(2)},\cdots e_{ij}^{(h)}\}$，其中 $0\leqslant e_{ij}^{(k)}\leqslant 1$，$h<H$，$H$ 为系统允许的最大的历史记录数，$E_{ij}$ 中的元素按照交互的时间顺序排列，$e_{ij}^{(1)}$ 表示离现在较久的一次交互，$e_{ij}^{(h)}$ 表示离现在最近的一次交互，则 $P_i$ 对 $P_j$ 的直接信任值为

$$E(P_i,P_j)=\begin{cases}\sum_{k=1}^{h}e_{ij}^{(h)}\gamma(k)/h, & h\neq 0\\ 0, & h=0\end{cases} \tag{1}$$

式中：$\gamma(k)\in[0,1]$ 是衰减加权因子，用来对发生在不同时刻的直接信任信息进行合理的加权. 根据人们的行为习惯，对于新发生的交互行为应该给予更大的权重，这也反映了信任关系随时间的变化而衰减的属性. 定义衰减因子为

$$\gamma(k)=\begin{cases}1, & k=h\\ \gamma(k-1)=\gamma(k)-1/h, & 1\leqslant k\leqslant h\end{cases} \tag{2}$$

**定义3** 一个主体的邻居节点指与自己有直接交互行为的主体.

**定义4** 以主体为父节点，所有的邻居为子节点，邻居也有邻居节点，这样就可以构造成一棵多叉多层的带权有向树，称为DTT，节点 $P_i$ 的DTT表示为

$$\mathrm{dtt}(P_i)=(<P,C>,E) \tag{3}$$

式中：$P$ 是节点(主体)的集合；有向边 $C$ 表示父节点和子节点的直接信任关系；权值 $E$ 就是直接信任值. 在DTT中，节点所在的层数表示为 $l$，我们规定根节点的层数 $l=0$，根节点的直接邻居 $l=1$，邻居的邻居 $l=2$，依次类推.

图1为一个DTT的例子，表示第 $k$ 个交互后，节点 $P_0$ 的DTT(注意图1只画出其中的3层，在动态信任关系中DTT是一种高度动态变化的数据结构，每一个时刻，每一个节点都有一棵DTT).

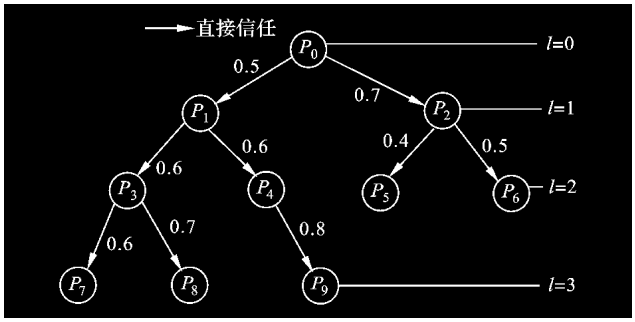


图1 根据直接信任关系构建的DTT

## 1.2 反馈信任的聚合

反馈信任，表示主体间通过第三者的间接反馈形成的信任度，也叫做声誉、反馈信任、间接信任和等级评分等.

反馈信任值计算主要是根据信任的传递性计算信任值($a$ 信任 $b$，$b$ 信任 $c$，那么 $a$ 信任 $c$)，在以下情况下，需要计算反馈信任值：①2个主体发生交互行为前，而SP没有SR的相关交互记录；②SP有SR的记录，但比较陈旧，需要更新. 主体之间的反馈信任值是一个由第三方提供的反馈信任值，在某个交互过程中实体 $P_i$ 需要评估实体 $P_j$ 的反馈信任度. 设反馈者的集合为 $\{W_1,W_2,\cdots,W_L\}$，$E(W_k,P_j)$ 表示第 $k$ 个反馈者对 $P_j$ 的直接信任值.

**定义 5**　反馈信任值聚合函数定义为

$$R(P_i,P_j)=\begin{cases}\dfrac{\sum_{k=1}^{L}(\rho_k E(W_k,P_j))}{\sum_{k=1}^{L}\rho_k}, & L\neq 0\\ 0.5, & L=0\end{cases} \tag{4}$$

式中：$L$ 为反馈者的个数；$E(W_k,P_j)$为第 $k$ 个反馈者对 $P_j$ 的直接信任值；$\rho_k$ 为反馈者加权因子.

反馈信任不能采取简单平均的办法，不同的反馈者所在的 $l$ 不同，有些反馈者是 $P_i$ 的邻居($l=0$)，而有些不是($l\neq 0$). 定义 5 中我们用 $\rho_k$ 对每一个反馈信息进行加权. 当没有反馈信息时，根据文献[9]的结论，取默认值 $R(P_i,P_j)=0.5$. 根据每一个反馈者所在的 $l$，给出 $\rho_k$ 的定义

$$\rho_k=\begin{cases}1, & l=0\\ \prod_{m=0}^{l}E(P_m,P_{\text{succ}}), & l>0\end{cases} \tag{5}$$

式中：$E(P_m,P_{\text{succ}})$表示从 $P_0$ 到 $P_k$ 信任路径上实体 $P_m$ 对它的后继节点 $P_{\text{succ}}$的直接信任值. 例如在图 2 中：$l=0$ 时，$\rho_0=1$ 表示节点的自信任；对于直接邻居节点，也就是 $l=1$ 时，$\rho_2=0.7$；当 $l=2$ 时，$\rho_6=0.6\times 0.7=0.42$；当 $l=3$ 时，$\rho_9=0.5\times 0.6\times 0.8=0.24$.

在传统的通过信任链或者广播方式计算反馈者的信任值时，节点将反馈信息搜索请求直接发送给所有的邻居节点，邻居再将该请求发送给自己的邻居节点，依次类推. 以这种方式计算反馈信任信息时，可扩展性和收敛性是最大的挑战. 例如，对于有成千上万个节点的系统，要向系统所有的主体通过广播的方式查询反馈信任度，系统的开销将会非常巨大，另外当信任链的跳数较多时，运算的收敛性速度会非常慢. 我们的主要目标是找到一种合理的方式来控制反馈信任信息聚合计算的规模，以提高系统的效率. 传统网络系统中利用 TTL(Time to Live)控制反馈信息搜索请求消息的生存周期是可能的方案之一，但该方法并不适合于高度动态变化的分布式系统，TTL 是一种被动的控制模式，当 TTL 的值设定好之后，系统就被动的等待 TTL 周期的结束，不能主动的改变聚合运算的结束时间，所以需要探索新的机制.

通过 $\rho_k$ 的定义可以看出，随着 $l$ 的增大，$\rho_k$ 的值逐渐减小. 另外，当反馈者的直接信任值较小时，说明它的可信度较低，进而说明它的反馈信息也具有较低的可信度. 根据以上 2 个性质，为了提高在大规模分布式系统中反馈信任信息的聚合速度，我们引入质量因子和距离因子这 2 个参数来调节反馈信任聚合计算的规模.

**定义 6**　质量因子 $\eta\in[0,1]$是系统设定的一个常数，只有反馈者的直接信任值 $e\geqslant\eta$ 时，该反馈者的反馈信息才是可信的. 当 DTT 中某个节点的 $e<\eta$ 时，该节点和以该节点为根的子树上所有节点的反馈都是不可信的.

通过质量因子我们可以有效控制反馈信任聚合运算的规模，一方面可以增强系统的运算收敛性，另一方面也可以减少低信任值的节点的恶意反馈，提高系统的安全性.

**定义 7**　距离因子 $\lambda$ 是一个大于等于 1 的正常数，用来控制反馈信任请求他信息在 DTT 中的传播深度. 当 $l\leqslant\lambda$ 时，节点将请求把信息转发给自己的邻居节点(子节点)，否则停止传递.

通过距离因子可以减少信任链的长度，有效提高系统的聚合运算速度.

式(4)给出了当有 $k$ 个反馈节点时如何进行反馈信任的聚合计算，但并没有给出如何获得这 $k$ 个反馈节点，下面结合定义 5 和定义 6 给出搜索反馈节点的递归算法，算法中函数 level($P_i$)表示 $P_i$ 在 DTT 中的层次.

**算法 1**　RequestFeedbacks($P_i$,$P_j$,$\lambda$,$\eta$)

设集合 $N(P_i)$表示节点 $P_i$ 的所有邻居节点；

设集合 $F(P_j)$表示所有针对节点 $P_j$ 的反馈节点；

```
if (level(Pi)<λ)) then 结束;
  for(所有 Pk∈N(Pi)) do
      if (E(Pi,Pk)>η) then
        在 Pk 的本地数据库查询 E(Pk,Pj);
        if(E(Pk,Pj)存在) then
          F(Pj)=F(Pj)+Pk;
          RequestFeedbacks(Pk,Pj,λ,η);
        endif
      endif
    endfor
  endif
  返回 F(Pj);
  结束
```

### 1.3　总体实现算法

开放的分布式系统中，恶意的主体可以通过提交不诚实的反馈来抬高其他恶意主体的信誉或者诋毁其他正常主体，特别是协同作弊对系统的危害更

大.反馈信任评估的基本需求是:①信任更新方式可以更好地反映近期的反馈信任度;②有效地减少恶意欺骗方式对信任计算造成的影响.

对于第1个需求,可以通过减小定义2中参数$h$的值的办法实现.通过定义2可以看出,参数$h$值反映了算法对节点过去行为的遗忘程度,$h$的值越小,表示近期行为对直接信任$E$的影响越大,也就是系统对节点近期行为越敏感,而定义5中反馈者权重$\rho_k$的计算又是依赖于直接信任$E$的,所以通过对参数$h$的调整,就可以更好地反映近期的反馈信任度.

对于第2个需求,可以通过提高质量因子$\eta$的值的方式实现.例如,可以设定$\eta=0.61$,表示直接信任值小于0.61的节点的反馈将不被系统采纳,这样可以避免信任值较低的节点的恶意反馈问题.下面给出反馈信任信息评估的实现算法.

**算法2** ComputeFeedbackTrust($P_i$,$P_j$)

输入:$h$,$\lambda$,$\eta$;
根据等式(1),(2),(3)构建dtt($P_i$);
根据算法1计算$F(P_j)$;
for(所有$P_k \in F(P_j)$) do
  根据等式(1),(2)计算$E(P_k,P_j)$
  根据等式(5)计算$\rho_k$;
  根据等式(4)计算$R(P_i,P_j)$;
endfor
输出:$R(P_i,P_j)$.
结束

## 2 模拟实验结果及分析

为了评测算法的性能,通过一个模拟的P2P网络进行了测试,实现平台选择Net logo[10],一个基于JAVA语言实现的"复杂多主体系统"建模工具,可以实现大规模分布式系统中节点间的并行工作.作为参照,我们同时实现了P-T模型[6].表1为参数的设置.

表1 模拟实验参数说明

| | 参数 | 单位 | 缺省值 |
|---|---|---|---|
| 运行参数 | 节点总数$N$ | 个 | 1000 |
| | 模拟计算次数$S$ | 次 | 2000 |
| 算法参数 | 历史交互记录数$h$ | 个 | 10 |
| | 距离因子$\lambda$ | | 4 |
| | 质量因子$\eta$ | | 0.61 |

### 2.1 收敛性和可扩展性评估

通过2个参数来考察算法的收敛性和扩展性.

**定义8** 聚合计算时间$t_c$定义为在各种不同的网络主体规模下,反馈信任聚合计算的时间开销.

**定义9** 平均存储开销$m_c$表示模型在信任评估计算中各种数据结构所占的平均存储空间的大小.设$m_{total}$为总存储空间的大小,$N$为节点总数,$m_c$定义为

$$m_c = \frac{m_{total}}{N} \tag{6}$$

将本文算法与P-T算法[6]进行比较,当其中一个算法的$t_c$和$m_c$的值较小时,本文算法具有较好的可扩展性.图2为在不同网络规模下2个算法的聚合计算时间$t_c$的比较,可以看出本文算法需要较少的聚合计算时间$t_c$,而P-T需要较多的聚合计算时间,说明本文有较好的运算收敛速度.同时也可以看出,随着网络规模的增大,本文算法的聚合计算时间缓慢增加,而P-T算法的聚合计算时间迅速增加,表明随着网络规模的增大,本文算法具有更好的可扩展性.

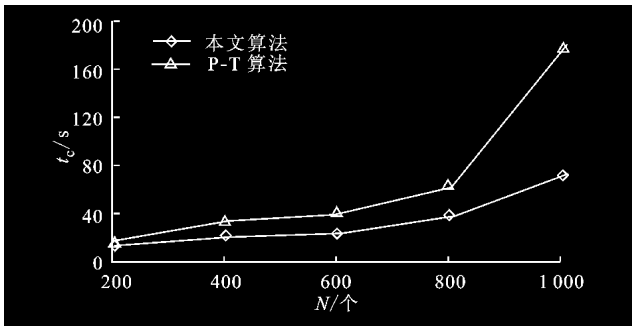


图2 模型聚合计算的时间开销比较

图3为不同网络规模下平均存储开销$m_c$的比较,可以看出本文算法需要较少的平均网络存储开销,2个模型的$m_c$值随着$N$的增加都缓慢减少,说明随着网络规模的增加,网络存储平均分布在各个主体之中.

### 2.2 模型的安全性能评估

动态信任管理的一个主要功能是进行节点恶意行为检测.信任机制应该具有较强的恶意行为检测能力.我们用恶意反馈行为的检测率来反映算法的安全能力.

**定义10** 设算法在交互过程中某个时间段共检测到$t_g$个诚实的反馈行为,检测到$t_b$个恶意的反馈行为,而系统中实际的恶意反馈者所占的比率为$\alpha$.恶意反馈行为的检测率$\theta$定义为

$$\theta = \frac{t_b/(t_b + t_g)}{a} = \frac{t_b}{a(t_b + t_g)} \quad (7)$$

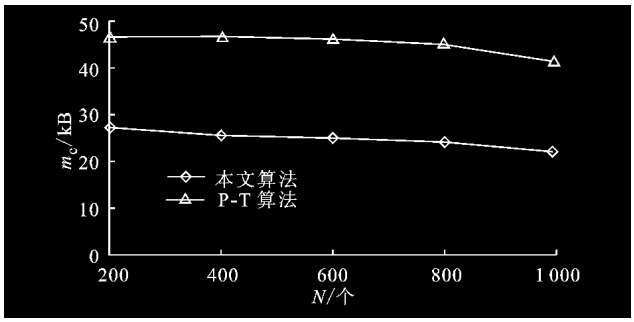


图 3　平均存储开销比较

图 4 为算法在各种不同比率的恶意节点环境下对恶意反馈行为的检测率比较. 实验中，所设定的恶意反馈节点的比率分别为 20%、50%、80%. 从图 4 可以看出，在 3 种情况下，算法对恶意节点的检测率都介于 80%～100%之间，说明本文算法对恶意反馈行为具有较强的检测能力，能有效地抵御恶意反馈行为的产生. 而且，随着恶意反馈节点所占的比率的增加，系统性能并没有下降，也说明本文算法具有较好的稳健性.

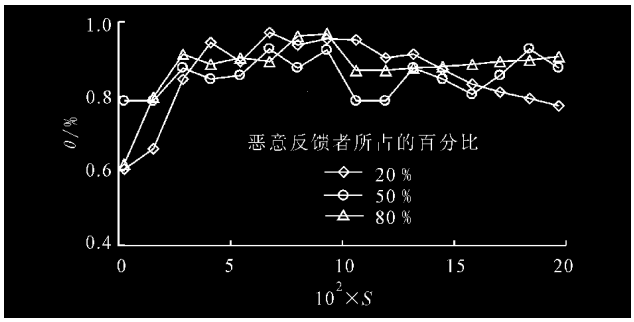


图 4　恶意反馈行为的检测率

## 3　结　论

通过本文的工作，我们可以得出结论：基于 DTT 的反馈信任信息聚合算法不但具有较好的可扩展性，而且具有较强的恶意行为检测能力. 同时，在网格计算、普适计算、P2P 计算等大规模的分布式系统中，对于任何信任评估算法的设计，可扩展性应该作为首要考虑的因素.

**参考文献：**

[1] Weeks S. Understanding trust management systems [C]// 2001 IEEE Symposium on Security and Privacy. Los Alamitos, USA: IEEE Computer Society Press, 2001:94-105.

[2] Viljanen L. Lecture notes in computer science: towards an ontology of trust [M]. Berlin: Springer-Verglag Press, 2005: 175-184.

[3] Chang E, Thomson P, Dillon T, et al. Lecture notes in computer science: the fuzzy and dynamic nature of trust [M]. Berlin: Springer-Verglag Press, 2005: 161-174.

[4] 李小勇，桂小林. 大规模分布式环境下动态信任模型研究[J]. 软件学报，2007，18(6)：1510-1521.
Li Xiaoyong, Gui Xiaolin. Research on dynamic trust model for large scale distributed environment [J]. Journal of Software, 2007, 18(6): 1510-1521.

[5] Yu B, Singh M. Developing trust in large-scale peer-to-peer systems[C]// Proceedings of the First IEEE Symposium on Multi-Agent Security and Survivability. Los Alamitos, USA: IEEE Computer Society Press, 2004:1-10.

[6] Xiong L, Liu L. Peer-trust: supporting reputation-based trust for peer-to-peer electronic communities [J]. IEEE Trans on Knowledge and Data Engineering, 2004, 16(7):843-857.

[7] Liang Z, Shi S. PET: a personalized trust model with reputation and risk evaluation for P2P resource sharing [C]// Proceedings of the 38th Hawaii International Conference on System Sciences. Los Alamitos, USA: IEEE Computer Society Press, 2005: 201-210.

[8] 常俊胜，王怀民，尹刚. DyTrust：一种 P2P 系统中基于时间帧的动态信任模型[J]. 计算机学报，2006，29(8):1301-1307.
Chang Junsheng, Wang Huaimin, Yin Gang. DyTrust: a time-frame based dynamic trust model for P2P Systems [J]. Chinese Journal of Computers, 2006, 29(8):1301-1307.

[9] Ziegler C N, Lausen G. Spreading activation models for trust propagation [C]// Proceedings of IEEE International Conference on e-Technology, e-Commerce, and e-Service. Los Alamitos, USA: IEEE Computer Society Press, 2004: 83-97.

[10] Tisue S, Wilensky U. NetLogo: design and implementation of a multi-agent modeling environment [EB/OL]. [2006-12-24]. http://ccl.northwestern.edu/papers/netlogo-swarmfest2004.pdf.

（编辑　刘杨　苗凌）