

P2P 网络中基于 Gossip 算法的信任模型

王 平, 邱 劲, 邱玉辉

(西南大学计算机与信息科学学院, 重庆 400715)

摘 要: 为了解决 P2P 网络开放性和自治性带来的信任问题, 提出一种基于 Gossip 算法的信任模型。该模型在估计节点的信任关系时引入时间退化因子, 使用基于网络拓扑结构的 Gossip 算法传递消息、聚集信誉。模拟实验结果表明, 该模型能更有效地评估节点间的信任关系, 提高查询效率, 减少网络拥塞。

关键词: Gossip 算法; 信任; 信誉; P2P 网络

Trust Model Based on Gossip Algorithm in P2P Networks

WANG Ping, QIU Jin, QIU Yu-hui

(Institute of Computer and Information Science, Southwest University, Chongqing 400715)

【Abstract】 In order to resolve the trust problem in open and autonomous P2P network, a trust model based on Gossip algorithm is proposed in which peers can autonomously deal with the information and identify trust belief with aging factor in P2P network. Simulation experimental result shows that the model can identify the peers' trust with lower overhead and higher efficiency.

【Key words】 Gossip algorithm; trust; reputation; P2P networks

1 概述

随着互联网的飞速发展、各种信息资源的膨胀, 如何在大规模网络环境下可扩展地支持海量数据信息的共享、管理和应用, 是目前互联网面临的挑战。P2P 网络因其特性为解决这一问题提供了新的可行方案。然而, 由于开放和自治的本质, P2P 网络中充斥着大量不可靠的服务和欺诈行为, 如在 Gnutella 中节点极易收到遭受 Gnutella 病毒感染的文件。因此, 为了提高 P2P 网络的整体可用性和服务质量, 信任机制的研究成为 P2P 网络研究的一个热点。

针对 P2P 网络中的信任问题, 人们提出了大量模型。文献[1]利用 P-Grid 提供的信息估计节点在未来欺骗的可能性。文献[2]基于信任的传递性提出基于全局声誉的信任模型 EigenRep。文献[3]提出一种综合局部声誉和全局声誉的信任机制 PeerTrust。Gossiptrust^[4]则通过节点的并发执行计算出全局信誉。然而, 现有的模型依然存在一系列问题: 文献[1]的方法没有考虑全局信息; EigenRep 的收敛性较差且具有较高的通信代价; PeerTrust 没有给出信任因素以及置信因子的具体度量方法, 且难以对抗共谋行为。Gossiptrust 是一种新颖的信誉集成机制, 具有较好的适应性和鲁棒性, 但没有考虑信任的时效性和网络的拓扑结构。

此外, 非结构化 P2P 网络因其简单、健壮性良好而在商用 P2P 系统中占主导地位。因此, 本文针对“如何在非结构化 P2P 网络中可扩展、高效地收集信息以确定和维护信任关系”这一问题, 提出了一种基于 Gossip 算法的信任模型。

2 基本定义

模型以非结构化 P2P 资源共享网络为例。假设在时间点 t , 节点 a 从节点 b 对其资源 θ 进行下载, 交互的结果为 $O_{a,b}^t(\theta)$ 。若 a 从 b 获得可靠资源 θ , 则 $O_{a,b}^t(\theta)=1$; 否则, $O_{a,b}^t(\theta)=0$ 。那么, 在 $1\sim t$ 的这段时间内, a 与 b 基于资源 θ

进行的所有交互结果为 $O_{a,b}^t(\theta)=1$, 其中, $t \in Z, t > 0$, 且在某一时间点, 特定的节点间只能完成一次交互。

在任意时间点 t , a 和 b 之间基于 θ 的交互历史可表示为 $E = \langle s_{a,b}^t, f_{a,b}^t \rangle$, 其中, $s_{a,b}^t$ 为成功交互的次数; $f_{a,b}^t$ 为交互失败的次数。被信任者 b 提供可靠资源或者不提供可靠资源的倾向性受其行为支配, 可将其定义为 $B_{a,b}(\theta) \in [0,1]$, 其中, $B_{a,b}(\theta) = p(O_{a,b} = 1)$ 。

信任是一个非常复杂的主观概念, 目前并没有一致的定定义。本文把信任 $T_{a,b}(\theta)$ 定义为: 信任者 a 对被信任者 b 在未来交互提供可靠资源 θ 的概率。

3 信任模型

3.1 直接信任

假设 a 对 b 提供可靠资源 θ 的概率为 $B_{a,b}(\theta)$, 则根据贝叶斯概率统计方法, 可利用 a 与 b 已有的交互历史事件估计 $B_{a,b}(\theta)$ 的期望值。因此, a 关于 b 在时间点 t 提供可靠资源 θ 的信任 $T_{a,b}^t(\theta)$ 可定义为: 在时间点 t , $B_{a,b}(\theta)$ 基于 $O_{a,b}^t(\theta)=1$ 的期望值为

$$T_{a,b}^t(\theta) = E[B_{a,b}(\theta) | O_{a,b}^t(\theta)] \quad (1)$$

使用 beta PDF(beta Probability Density Function)计算, 得

$$E[B_{a,b}(\theta) | O_{a,b}^t(\theta)] = \frac{\alpha}{(\alpha + \beta)} \quad (2)$$

根据标准 beta PDF, 在未来的交互中, b 提供可靠资源

基金项目: 国家“973”计划基金资助项目(2003CB317008)

作者简介: 王 平(1979-), 女, 博士研究生, 主研方向: 分布式人工智能, 信息安全; 邱 劲, 博士研究生; 邱玉辉, 教授、博士生导师

收稿日期: 2009-05-11 **E-mail:** wangping@swu.edu.cn

$\theta : O'_{a,b}(\theta) = 1$ 的分布函数可以表示为如下参数设置的 beta PDF 函数：

$$\begin{aligned}\alpha &= s_{a,b}^{lr} + 1 \\ \beta &= f_{a,b}^{lr} + 1\end{aligned}\quad (3)$$

在信任的估计中，两者的交互历史决定了 beta PDF 函数中 α 和 β 的值。在标准贝叶斯算法中，每一次交互结果都被赋予相同的权重，并没有考虑交互发生的时间。然而，信任具有动态性，是信任者对被信任者未来行为的预测。直观上，两者之间发生较早的事件对信任值的影响应小于发生较晚的事件。为此，模型中引入时间退化因子 μ 以反映信任的这一特性。

假设 a 与 b 进行交互，如果 b 提供了可靠资源 θ ，则有 $o = 1$ ；否则， $o = 0$ ，那么时间退化更新为

$$\begin{aligned}s_{a,b} &= \mu s_{a,b} + o \\ f_{a,b} &= \mu f_{a,b} + (1 - o)\end{aligned}\quad (4)$$

假设 $\{o_1, o_2, \dots, o_n\}$ 为交互事件的结果序列，则经过 n 次直接交互后：

$$s_{a,b}^n = o_n + \mu o_{n-1} + \dots + \mu^{n-1} o + \mu^n \quad (5)$$

如果 $T_{a,b}$ 估算正确， $T_{a,b}$ 将趋于一个常量。这样，经过较长时间即 n 具有较大值时，有

$$\begin{aligned}s_{a,b} &\approx n T_{a,b} \\ f_{a,b} &\approx n(1 - T_{a,b})\end{aligned}\quad (6)$$

且 $Beta(s_{a,b} + 1, f_{a,b} + 1)$ 无限趋近于 $T_{a,b}$ ，即

$$E(s_{a,b}^t) \approx \frac{T_{a,b}}{1 - \mu} \quad (7)$$

假设 $m = \frac{1}{1 - \mu}$ 为整数，根据标准贝叶斯方法， m 次交互后与经过无限次交互所得的期望值相同。综上所述，时间退化因子为

$$\mu = 1 - \frac{1}{m} \quad (8)$$

其中， m 为模型中设置的节点行为趋于稳定的交互次数。此外，只要交互的时间不终止，则

$$\begin{aligned}s_{a,b} &= \mu s_{a,b} \\ f_{a,b} &= \mu f_{a,b}\end{aligned}$$

依然成立。

3.2 信誉

非结构化 P2P 网络的开放性和动态性使信誉^[5]在进行信任决策中具有重要作用。现有的模型主要采用洪泛算法聚集信誉，洪泛算法虽具有良好的鲁棒性和扩展性，但易产生大量冗余消息，严重消耗资源，造成网络拥塞。Gossip 算法作为一种典型的消息扩散算法而被广泛应用于分布式系统^[4]。为此，模型采用 Gossip 算法传递消息、聚集信誉。

3.2.1 基于网络拓扑结构的查询策略

在 Gossip 算法中，每个收到消息的节点随机地选择 $fanout$ 个邻居节点进行传播。研究表明，Gnutella 等非结构化 P2P 网络中节点分布满足幂率分布，节点间链接差异巨大：有的链接丰富，在消息传播过程中收到消息的概率较大；有的节点链接相对稀疏，在消息传递过程中有可能被孤立，成为信息孤岛。本文采用文献[6]提出的方法，在查询目标节点的推荐信息过程中，根据链接结构的差异赋予不同链接不同的权值，然后根据不同的权值控制本地信息流的传递，从而

提高整个路由效率。为了便于讨论，给出如下定义：
 $outView(i)$ ：节点 i 的出度节点集； $inView(i)$ ：节点 i 的入度节点集； $link_{ij}$ ：节点 i 指向节点 j 的有向边，表示 i 从 j 中进行资源共享； $\varphi(link_{ij})$ ：链接 $link_{ij}$ 的权重，有：

当 $|inView(j)| > 0, |outView(j)| > 1$ 时：

$$\varphi(link_{ij}) = \frac{\alpha_1(\alpha_2 - \alpha_2 / |outView(j)|)}{|inView(j)|}$$

当 $|inView(j)| > 0, |outView(j)| = 1$ 时：

$$\varphi(link_{ij}) = \frac{\beta_1}{|inView(j)|}$$

当 $|inView(j)| > 0, |outView(j)| = 0$ 时：

$$\varphi(link_{ij}) = \frac{\beta_2}{|inView(j)|}$$

其中，常量 $\alpha_1 > 0, \alpha_2 > 0, 0 < \beta_2 < \beta_1 < \frac{\alpha_2}{2}$ ； $\|$ 表示节点的出入度。

每个节点 i 维护的信息为

$$\begin{aligned}Inf(i) &= \{ID_i, IP_i, \theta_i, outView(i), inView(i), neighbor(i), R_{i,x}^t(\theta)\} \\ neighbor(i) &= \{ID_{j_k}, IP_{j_k}, outView(j_k), inView(j_k), T_{i,j_k}^t(\theta), \\ &\quad tr_{j_k \rightarrow x}^t(i, j_k)\}\end{aligned}$$

其中， $neighbor(i)$ 为节点 i 的邻居表； $T_{i,j_k}^t(\theta)$ 为邻居节点的信任值； $tr_{j_k \rightarrow x}^t(i, j_k)$ 为邻居节点的推荐可信度； $R_{i,x}^t(\theta)$ 为节点 i 对网络中非邻节点的全局信誉。

综上所述，基于拓扑结构的 Gossip 算法具体步骤如下：

- (1) 源节点在查询目标节点关于资源 θ 的推荐信息的过程中，首先生成查询消息，初始化生存周期 TTL (信息传播迭代的跳数，根据小世界网络模型，设 $TTL = 6$)。
- (2) 检查 TTL ，若 $TTL = 0$ ，消息无效，丢弃；若 $TTL > 0$ ，则根据邻居节点表中节点的出入度计算邻居节点的链接权重 $\varphi(link_{ij})$ ，然后根据链接权重计算邻居节点 j 作为下一个消息传递节点的概率：

$$P_{ij_k} = \frac{\varphi(link_{ij_k})}{\sum_{j_k \in outView(i), k=1}^{|outView(i)|} \varphi(link_{ij_k})} \quad (9)$$

- (3) 选取 p_{ij} 最高的 $fanout$ 个节点进行传播，并令 $TTL = TTL - 1$ 。

- (4) 邻居节点收到消息后，检查其邻居表中是否有与目标节点交互的历史：如果有，则返回相关信息给源节点；如果没有，则将消息传递给邻居节点的邻居节点，即重复步骤(2)~步骤(4)。这样反复递归直到找到目标节点推荐信息，或者 $TTL = 0$ ，消息无效被丢弃。

3.2.2 推荐可信度

推荐可信度反映了推荐信息的可靠性，由于目标节点的全局信誉由聚集目标节点的邻居节点的推荐信息而得，因此推荐信息的可信度将直接影响目标节点全局信誉的准确性。

这里，节点 a 对节点 b 关于 c 的交互历史的推荐信息可信度定义为

$$tr_{b \rightarrow c}^t(a, b) = \frac{s_{a,b \rightarrow c}^{lr} + 1}{s_{a,b \rightarrow c}^{lr} + f_{a,b \rightarrow c}^{lr} + 2} \quad (10)$$

其中， $s_{a,b \rightarrow c}^{lr}$ 为 a 对 b 关于 c 的推荐满意的次数； $f_{a,b \rightarrow c}^{lr}$ 为 a 对 b 关于 c 的推荐不满意的次数，其计算见式(4)。

3.2.3 全局信誉

客观上,参与评价的推荐节点数越多,全局信誉的期望值就越接近真实值。为了防止少数恶意节点通过共谋来抬升彼此的信誉,模型在聚集推荐信息的过程中,不仅考虑推荐节点的可信度、推荐节点对目标节点的信任,还考虑推荐节点数。

假设 a 为源节点, c 为目标节点, b_i 为目标节点的推荐节点, m 为推荐节点数,则 a 对 c 关于 θ 的全局信誉为

$$R_{a,c}^t(\theta) = e^{-m^{-1} \sum_{i=1}^m \frac{tr_{b_i \rightarrow c}^t(a,b)}{\sum_{j=1}^m tr_{b_j \rightarrow c}^t(a,b)}} \cdot T_{b_i,c}^t(\theta) \quad (11)$$

3.3 全局信任

在实际交互中,当交互历史较少时,节点更多地依赖于网络中的推荐信息;由于推荐信息存在不可靠性且获取需消耗一定的资源,因此当节点间的交互到达一定程度时,节点更多地依赖自身的交互。全局信任由此定义为

$$TR_{a,b}^t(\theta) = \xi T_{a,b}^t(\theta) + (1-\xi) R_{a,b}^t(\theta) \quad (12)$$

其中, $\xi = \min(1, \frac{C_{a,b}}{C_{\min}})$, $\xi > 1 - \xi$, C_{\min} 为 a, b 之间交互结果趋于稳定的最小交互次数,对不同的信任议题, C_{\min} 有不同的取值。

4 实验及结果分析

为了验证模型的准确性和有效性,本文进行了如下模拟实验。实验使用文献[6]的方法仿真网络环境,相关参数设置如下: $fanout = 3, TTL = 6, \alpha_1 = \alpha_2 = 1, \beta_1 = 0.2, \beta_2 = 0.1$ 。

4.1 查询代价

实验 1 比较了传统的 Gossip 算法(S-C-g)和基于拓扑结构的 Gossip 算法(S-C-tg)在推荐信息传递过程中的查询代价(一条消息访问过的节点数占整个网络节点数的百分比)和返回率(返回推荐数与网络中和目标节点有交互的节点数的比率)。理论上,在消息的传递过程中,消息访问的节点数越少,查询的响应时间越短,网络资源开销就越小。

实验结果如图 1 所示。从中可以看出,在推荐信息的聚集过程中,本模型的查询开销远小于传统的 Gossip 算法,较好地解决了消息传递中的网络拥塞问题。

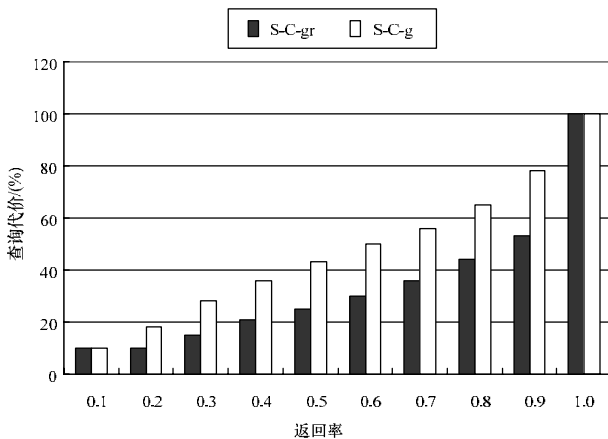


图 1 查询代价比较结果

4.2 时间退化因子

实验 2 模拟了节点的 150 次交互,交互过程设定为: $0 \leq t < 50$,交互成功; $51 \leq t < 100$,交互失败; $101 \leq t < 150$,交互成功,其中, t 为交互次数。为了简化模拟,这里只分析 $\mu = 0.9, \mu = 0.98, \mu = 1$ (不考虑时间退化)3 种情况。

实验结果如图 2 所示。从中可以看出,当节点行为发生变化时,引入时间退化因子的模型都能较快地做出反应,没有引入时间退化因子的模型则反应较慢,即模型在信任估计过程中对节点的突变行为具有较好的适应能力。

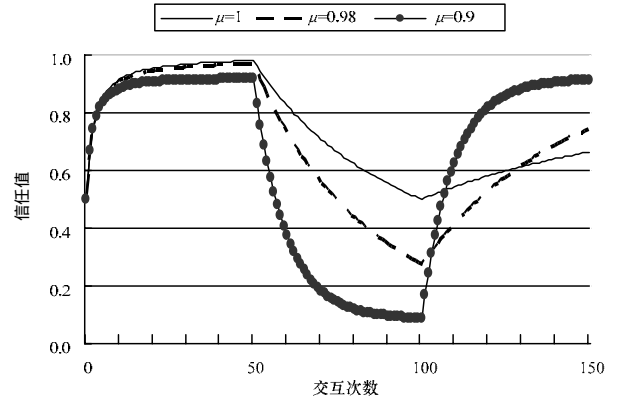


图 2 信任的时间退化

5 结束语

本文的信任模型与现有模型相比,由于使用了基于网络拓扑结构的 Gossip 算法来传递消息,因此降低了网络负载;同时在估计信任关系时引入时间退化因子,能更有效地评估节点间的信任关系。在 P2P 网络中,每个节点管理维护不同的资源,根据节点所维护的资源进行消息传递可有效降低负载。如何建立基于语义的信任网视图是下一步的研究重点。

参考文献

- [1] Aberer K, Despotovic Z. Managing Trust in a Peer-to-Peer Information System[C]//Proceedings of the 10th International Conference on Information and Knowledge Management. New York, USA: [s. n.], 2001: 310-317.
- [2] Kamvar S D, Schlosser M T. EigenRep: Reputation Management in P2P Networks[C]//Proc. of the 12th Int'l World Wide Web Conference. [S. l.]: ACM Press, 2003: 123-134.
- [3] Li Xiong, Ling Liu. PeerTrust: Supporting Reputation-based Trust for Peer-to-Peer Electronic Communities[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(7): 843-857.
- [4] Zhou Runfang, Hwang Kai. Gossip-based Reputation Aggregation in Unstructured P2P Networks[C]//Proc. of IEEE International Parallel and Distributed Processing Symposium. Xi'an, China: IEEE Press, 2007: 1-10.
- [5] 郭成, 明楚, 姚红岩. P2P 网络下基于推荐的信任模型[J]. 计算机工程, 2008, 34(24): 157-159.
- [6] Hai Zhuge, Xiang Li. RSM-based Hierarchical Gossip on P2P[C]//Proc. of International Conference on Algorithms and Architectures for Parallel Processing. [S. l.]: IEEE Press, 2007: 11-12.

编辑 张帆