

# 基于随机漫步的信任路径搜索算法

刘智勇, 郑 滔, 伍伟绩

(南京大学软件学院, 南京 210093)

**摘 要:** 传统的局部信任模型采用简单洪泛的方法获得信任信息, 针对该方法效率较低且对网络资源消耗较大的问题, 提出一种基于随机漫步的搜索信任路径的算法。通过以往遗留的路径信息改进搜索, 可有效减少多余信息的数目和信任回路的出现。该算法适用于对网络资源占用比较敏感的环境。

**关键词:** 随机漫步; 信任管理模型; 信任路径搜索

## Trust Path Search Algorithm Based on Random Walk

LIU Zhi-yong, ZHENG Tao, WU Wei-ji

(Software Institute, Nanjing University, Nanjing 210093)

**【Abstract】** The value of recommendation trust is collected by flooding in traditional local trust model, but the method has low efficiency and takes up a lot of network resources. This paper presents a trust path search algorithm based on random walk, which is able to improve the search by using the path information in the past. The algorithm is efficient and valuable in network environment which is sensitive to resources occupying because it reduces the number of unnecessary messages and trust circuits effectively.

**【Key words】** random walk; trust management model; trust path search

### 1 概述

在开放的分布式环境中, 主体往往要和不了解甚至完全陌生的主体进行交互。由于缺少中心化的管理权威, 许多基于传统软件系统形态的安全机制如访问控制列表、公钥证书体系等已不能满足安全需求。信任管理就是为了适应这种开放性的动态系统而提出的新的安全机制。

目前主要有全局信任模型和局部信任模型 2 种不依赖于可信第 3 方的信任模型。在局部信任模型中, 共享信息的获取有 2 种途径: 一是通过向其他节点洪泛信任请求获得, 该方法可扩展性差, 时间复杂度和对网络带宽造成的影响也都不能忽略。另一种是通过采用 DHT 机制的 P2P 存储系统如 Chord 等获得, 利用杂凑(hashing)的方式, 将数据和结点运算成一个键值(key), 利用键值来完成数据的放置与维护。这种方法并不适合于节点频繁加入和离开的系统。

本文提出一个基于随机漫步(random walk)的局部信任路径搜索算法, 它不仅减少了网络带宽的消耗, 同时也能有效减少回路的出现。

### 2 相关工作

信任管理的思想最早由 Blaze 等人提出, 是采用一种统一的方法描述和解释安全策略, 安全凭证以及用于直接授权关键性安全操作的信任关系。它能在整个系统中保持统一的安全策略, 同时还具有很强的表达能力和可扩展性, 很好地适应了分布式系统的需要。Beth 等人提出信任定量化的概念和方法, 将信任分为直接信任和推荐信任, 根据肯定和否定经验数计算实体完成任务的概率, 以此表示信任, 并给出了信任合成的方法。Rahman 等人提出的信任度评估模型给出了信任度的传递协议和计算公式。

根据收集推荐信任的范围, 信任模型可分为全局信任模

型和局部信任模型。在全局信任模型中, 每个节点具有唯一的全局信任值, 这个信任值通过对网络中所有交易反馈进行分析得到。全局信任模型 EigenTrust<sup>[1]</sup>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。在 EigenTrust 模型中, 通过节点间信任度的迭代来实现信任的传播, 从而为每个节点计算全局信任值。每次交易都会导致在全网络范围内的迭代, 在网络节点数目增加的情况下, 通信量和计算量的增加将导致网络性能急剧下降。这也是大部分全局信任模型共同的问题所在。

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。通过限制获得反馈和评价信息的范围, 使得由此带来的通讯开销不致过大, 但其获取的节点可信度也往往是局部的和片面的。PeerTrust<sup>[2]</sup>是已知的局部信任模型, 该模型通过节点的交易反馈信息来量化、比较节点的可信度, 并基于结构化 P2P 提出了分布式的回馈信息保存和信誉值计算方案。但基于 DHT 的回馈信息的管理开销不容忽视。文献[3]提出在 P2P 环境下基于贝叶斯网络的信任模型。该信任模型主要关注于描述信任的不同方面, 使得节点可以根据不同的场景来按需获取节点不同方面的性能。该信任模型能够适应于规模较小的 Gnutella 网络, 或节点交互集中的 Gnutella 网络。文献[4]运用模糊集合理论来度量并推导信任关系, 对信任管理问题进行了建模, 给出了信任类型的定义机制和信任的评价机制, 构造了一个完整的主

**作者简介:** 刘智勇(1983 - ), 男, 硕士, 主研方向: P2P 网络, 信任计算; 郑 滔, 教授; 伍伟绩, 硕士研究生

**收稿日期:** 2009-01-20 **E-mail:** mg0632036@software.nju.edu.cn

观信任管理模型。

各种信任模型都需要通过一定的算法得到多条信任路径。并根据所得到的多条信任路径进行综合计算得到一个综合信任值。文献[5]提出 PGP 中信任机制的 BDP 近似算法，得出了较多的不相交路径，但未考虑如何使各条路径的信任度进行优化。文献[6]利用有向图对信任网络进行建模，提出了基于主观逻辑的信任搜索算法，能对多条信任路径的信息进行合成。

### 3 基于随机漫步的信任路径搜索算法

#### 3.1 算法思想

在复杂的大规模网络拓扑图中，要找到所有的信任路径会花费很多时间，带来的网络带宽开销也难以忍受。通过在搜索信任路径时进行筛选，使得到的结果尽可能为较优的多条路径。本文用以下 2 个标准来判断一条信任路径的优劣：(1)推荐信任值。推荐信任值越高的路径，在合成最终结果时所占的比重越大。(2)路径长度。信任路径越短，在信任传递的过程中误差越少，被欺骗的可能性也越小。

假设节点  $a$  要对节点  $b$  进行信任评估， $a$  通过洪泛的方式寻找到达  $b$  的信任路径。当一条信任路径被发现后，信任值通过信任路径反向传递到发出请求的节点。现有的算法在收集信任值时，只收集信任值，信任路径本身的信息被直接丢弃。本文提出的算法通过保存这部分信息，在以后其他节点提出相同的查询请求时进行了优化。

##### 定义 信任路径信息

假定  $i$  和  $j$  是网络中 2 个不同的节点， $L$  为一条从节点  $i$  到节点  $j$  的信任路径， $L$  的长度为  $length$ ， $W_L$  代表  $L$  的长度信息，则令  $W_L = p^{(length-1)}$ 。这条信任路径越短， $W_L$  值越大。当  $length=1$  时， $L$  是节点  $i$  到节点  $j$  的直接信任， $W_L=1$ 。设  $P_{i,j}$  代表从节点  $i$  到节点  $j$  的信任路径信息的总和， $P_{i,j} = \sum W_L$ 。其中， $p(0 < p < 1)$  为长度衰减因子， $p$  越小，则在计算  $p_{i,j}$  的值时越重视短链。若  $p=1$ ，那么在计算  $p_{i,j}$  的值时短链与长链的价值相同。

当节点  $a$  要与节点  $b$  发生交互时，节点  $a$  的信任评估模块检查其邻居的推荐信任值。向推荐信任值高于阈值  $\theta$  的邻居发出查询节点  $b$  信任度的请求。邻居节点  $c$  在收到  $a$  的信任查询请求后，根据保存在本地的关于  $b$  的信任路径信息决定是否进一步转发请求。 $c$  计算转发信任查询请求的概率  $K$ ：

$$K = P_{c,b} / P_{a,b} + F \quad (1)$$

$F(0 < F < 1)$  为基础转发概率，保证即使  $c$  完全没有关于  $b$  的路径信息也会以一定的概率转发。当网络中节点频繁变化时  $F$  可以设为较高值以保证得到足够的搜索结果，而当网络比较稳定时  $F$  可以设为较低值减少带宽消耗。

对于一个新加入的节点或改变在网络拓扑中位置的节点，由于参与的交互行为数量有限，其拥有的信任路径信息的可靠性和完整性得不到保证，此时应设一个较大的  $F$  值。随着节点参与的交互行为逐渐增多，可以逐渐减小  $F$  值。

#### 3.2 信任路径信息的更新

在一个动态变化的网络拓扑中，不断有节点加入和退出，同时一个节点的可信程度也并不是固定的，而是经常随时间的变化而变动。因此信任路径也会随着时间推移逐渐失效。一个节点在近期得到的路径信息要比以前得到的路径信息更可信。设节点  $i$  对节点  $j$  的信任路径信息经过  $n$  单位个时间帧后将完全遗忘，在上一次得到路径信息后经过  $m$  个时间周期后又得到新的信任路径信息  $P'$ 。如之前的路径信息已被完

全遗忘，则直接令  $P_{i,j} = P'$ 。否则的话更新函数定义为

$$P_{i,j} = (1-\lambda) \frac{(n-m)P_{i,j} + nP'}{2n-m} + \lambda P_{i,j} \quad (2)$$

其中，

$$\lambda = \begin{cases} \alpha & P_{i,j} < P' \\ \beta & P_{i,j} > P' \end{cases}$$

由于每次得到的信任路径信息都是不完全的，即使网络拓扑和网络中节点信任值没有变化，所得到的信任路径信息也可能不同，因此设计  $\alpha$  和  $\beta$  分别为信任路径信息增加和减少学习因子。一般情况下取  $\alpha > \beta$ ，即信任路径信息增加的速度比降低的速度快。

#### 3.3 信任回路

在通过洪泛转发请求时，相同的请求消息可能被很多邻居节点发到同一个节点上，极易形成回路，产生多余消息，占用网络带宽。

在本算法中转发查询信任请求是基于信任路径信息的，邻居节点通过比较信任路径值得大小来决定是否转发信息，在转发的时候就有向信任路径值更高的节点发送的倾向。而要形成回路就必须有至少 1 步是从高信任路径值节点向低信任路径值节点发送请求，通过限制这个概率就可以有效地减少回路的出现。转发示例如图 1 所示。

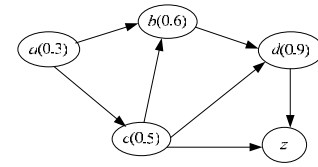


图 1 转发示例

有  $P_{az}=0.3$ ,  $P_{bz}=0.6$ ,  $P_{cz}=0.5$ ,  $P_{dz}=0.9$  基础转发概率  $F=0.1$ 。当节点  $a$  向邻居发送查询节点  $z$  信任度的请求时， $a$  向  $bc$ ,  $bc$  向  $d$  都会以高概率进行转发，而  $bc$  向  $a$ ,  $d$  向  $bc$  转发的概率就比较低。形成回路  $a-b-c-a$  的概率为 0.65，回路  $a-c-d-b-a$  的概率为 0.46。再加上信任度阈值的限制可以进一步降低形成回路的可能性。

#### 3.4 多级信任阈值

节点  $a$  在向邻居发出查询节点  $b$  信任度的请求时，会根据邻居的推荐信任值是否高于阈值  $\theta$  来进行第一次筛选。考虑到信任在传递过程中会不断丢失的，单一的阈值无法体现信任的这个特点。关于信任值得推荐合成，各种模型有不同的做法，这里以乘积式传递为例说明多级阈值。

以  $rt_{ij}$  代表节点  $i$  对节点  $j$  的推荐信任。如  $rt_{ij}=0.5$  则说明节点  $i$  对节点  $j$  的推荐信任值为 0.5。信任路径  $a-b-c-d-z$  的推荐信任值为  $rt_{ab} \times rt_{bc} \times rt_{cd} \times rt_{dz}$ 。

计算每跳的阈值：

$$\theta^t = 1 - \Theta^{t-1}, \quad 0 < \Theta < 1 \quad (3)$$

其中， $\Theta$  为阈值增加的幅度； $t$  为跳数， $\theta^t$  表示第  $t$  跳所采用的阈值。如  $\Theta = 0.8$ ，则第 1 跳的阈值为 0.2；第 2 跳的阈值为 0.36；第 6 跳的阈值为 0.74。通过多级阈值的限制，使得低信任值长链的数目大大减少。

#### 3.5 算法步骤

(1)源节点首先列出信任值高于阈值的邻居节点，向它们发送请求信息。同时发送参数信任路径信息和跳数  $t_{tl}$ 。

(2)邻居节点收到请求信息时，首先检查自己是否有目标节点的推荐信任值，若有则回应请求节点。

(3)检查参数  $t_{tl}$  是否为 0。若为 0，则停止发送请求信息。

若不为 0, 将  $t_{tl}$  的值减 1。然后计算转发信任查询请求的概率  $K$ , 以概率  $K$  转发请求。如果决定转发则转向步骤(2)。

#### 4 实验结果及分析

实验将本文算法与深度搜索算法进行比较, 模拟了一个 1 000 个节点, 节点间临近关系、推荐信任值与直接信任值随机分布的信任关系模型。每个节点平均与 10 个节点有直接信任关系。跳数限制为 5。以任意一个节点为源, 分别使用本文提出的算法和洪泛算法收集另一个任意节点的推荐信任值。为多次重复以上过程, 记录每 100 次查询的平均数据。记一个节点向另一个节点发送或转发查询请求为 1 个消息, 比较 2 种算法在网络中产生消息数的总量。图 2 表示每百次查询平均产生消息总数随交互的变化。

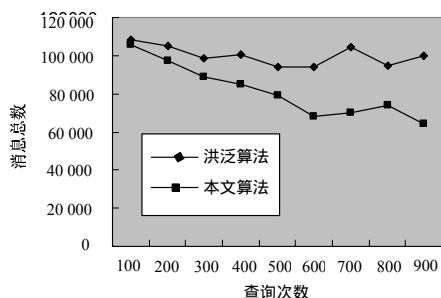


图2 每百次查询平均产生消息总数随着交互进行变化

##### 4.1 算法的有效性

从图 2 可以看出, 通过洪泛来进行查询所产生的消息总数大致保持在一个稳定的数量上。而通过本算法进行信任查询, 随着交互进行, 1 次查询在网络中产生并传播的消息总数在减少, 改进比率在增加。

##### 4.2 参数对算法的影响

图 3 是以某一点为源, 查询网络中所有其他节点的信任值, 其中采用不同的基础转发概率  $F$  所产生的消息总数的对比。从图中可以看出, 基础转发概率值越小, 产生的消息数就越少, 但所获得的信任值误差也越大。

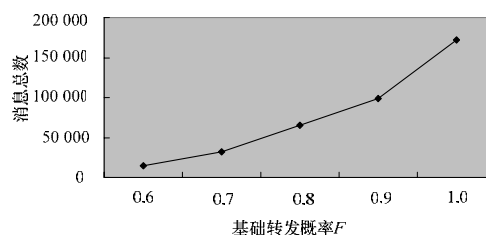


图3 采用不同  $F$  值时每百次查询平均产生消息总数

#### 5 结束语

在一个局部信任模型里, 收集推荐信任的过程中, 通过利用以往信任传递时遗留的信息, 使用定向随机漫步代替简单洪泛。根据信任阈值和信任路径信息来进行信息筛选, 可以有效减少多余消息的产生, 降低网络开销。

#### 参考文献

- [1] Kamvar S D, Schlosser M T, Molina G H. The EigenTrust Algorithm for Reputation Management in P2P Networks[C]//Proc. of the 12th Int'l World Wide Web Conf.. New York, USA: ACM Press, 2003.
- [2] Li Xiong, Ling Liu. A Reputation-based Trust Model for Peer-to-Peer E-commerce Communities[C]//Proc. of CEC'03. Newport Beach, California, USA: IEEE Press, 2003.
- [3] Yao Wang, Vassileva J. Trust and Reputation Model in Peer-to-Peer Networks[C]//Proc. of the 3rd International Conference on Peer-to-Peer Computing. [S. l.]: IEEE Press, 2003.
- [4] 唐文, 陈钟. 基于模糊集合理论的主观信任管理模型研究[J]. 软件学报, 2003, 14(9): 1401-1408.
- [5] Reiter M K, Stubblebine S G. Resilient Authentication Using Pathindependence[J]. IEEE Transactions on Computers, 1998, 47(12): 1351-1362.
- [6] 付江柳, 高承实, 戴青, 等. 基于主管逻辑的信任搜索算法[J]. 计算机工程, 2008, 34(3): 178-180.

编辑 金胡考

(上接第 155 页)

证明: 集合  $A$  中元素要构成  $F_{2^{2n}}$  在  $F_2$  上的一组基, 其中的元素必须线性无关。而  $A$  中元素的任意线性组合都可写成  $f_1(\alpha) + \alpha^k f_2(\alpha)$ , 其中,  $f_1(x)$  和  $f_2(x)$  为  $F_2$  上小于  $n$  次的多项式, 因此,  $A$  中的元素线性相关, 当且仅当存在适当的  $f_1(x)$  和  $f_2(x)$  使得  $\alpha^k = f_1(\alpha) / f_2(\alpha)$ , 其中,  $f_2(\alpha) \neq 0$ 。当  $k$  遍历  $[0, 2^{2n}-2]$  时,  $\alpha^k$  遍历  $F_{2^{2n}}$  的非零元。于是,  $A$  中元素能构成基的个数即为元素线性无关的个数, 即  $F_{2^{2n}}$  中所有非零元的个数  $(2^{2n}-1)$  减去  $f_1(\alpha) / f_2(\alpha)$  可表示出的非零元个数。

$F_{2^{2n}}$  中非零元若能够表示成  $f_1(\alpha) / f_2(\alpha)$ , 必然要求  $(f_1(x), f_2(x)) = 1$ 。而由定理 4, 任意 2 个互素对所成元素必不相同, 所以,  $f_1(\alpha) / f_2(\alpha)$  可表示出的非零元个数等于  $F_2$  上小于  $n$  次的互素多项式的对数减 2 (除互素对  $(0,1)$  和  $(1,0)$ )。

由定理 3,  $F_2$  上小于  $n$  次的互素多项式对数为  $2^{2n-1} + 1$ , 去掉  $(0,1)$ 、 $(1,0)$ , 共有  $2^{2n-1} - 1$  个, 所以,  $A$  能构成的基的个数为  $(2^{2n} - 1) - (2^{2n-1} - 1) = 2^{2n-1}$ 。于是,  $F_4$  上每个  $n$  级本原  $g(x)$ -类所含元素的个数为  $2^{2n-1}$ , 而  $F_2$  上的  $2n$  次本原多项式个数为  $\frac{\phi(2^{2n}-1)}{2n}$ 。因此,  $F_4$  上  $n$  级本原  $\sigma$ -LFSR 个数为  $\frac{\phi(2^{2n}-1)}{2n} 2^{2n-1}$ 。

显然, 定理 5 所得  $F_4$  上的计数公式与猜想的公式相符。但是, 对于有限域  $F_{2^m}$ , 目前还没有很好的解决办法。

#### 5 结束语

本文将本原  $\sigma$ -LFSR 的计数问题转化为线性空间  $F_{2^m} / F_2$  基的问题, 并在  $F_4$  上将其彻底解决。随着字 LFSR 的流行与处理器的发展, 相信  $\sigma$ -LFSR 序列会在密码设计中得到越来越广泛的应用。

#### 参考文献

- [1] Tsaban B, Vishne U. Efficient Linear Feedback Shift Registers with Maximal Period[J]. Finite Fields Application, 2002, 8(2): 256-267.
- [2] Zeng Guang, Han Wenbao, He Kaicheng. High Efficiency Feedback Shift Register:  $\sigma$ -LFSR[Z]. [2008-11-10]. <http://eprint.iacr.org/>.
- [3] 张猛, 曾光, 韩文报, 等. 本原  $\sigma$ -LFSR 序列的迹表示及其应用[J]. 电子与信息学报, 2009, 31(4): 942-945.
- [4] 张猛, 韩文报.  $\sigma$ -LFSR 的分类研究[C]//中国密码学会 2007 年论文集. 成都: 西南交通大学出版社, 2007: 27-35.
- [5] Benjamin A T, Bennett C D. The Probability of Relatively Prime Polynomials[J]. Mathematics Magazine, 2007, 80(3): 309-310.

编辑 张帆