

基于 Chord 的层次化结构信誉模型

张国华, 黄烟波

(中南大学信息科学与工程学院, 长沙 410083)

摘要: 针对现有信任模型流量过大、节点不能很好被激励的问题, 设计一种基于 Chord 的 P2P 层次化结构信誉模型 TMCL。在高效获取信息的同时, 减轻基于推荐的信任模型中普遍存在的流量冲击问题。在交易中采用双向选择机制, 遏制节点作恶, 激励节点积极主动地提供服务。实验证明, TMCL 较已有的信任机制在降低网络性能开销和提高网络服务质量方面都取得了较好的效果。

关键词: P2P 网络; 信誉模型; 层次化结构; 双向选择

Trust Model Based on Chord Layer Construction

ZHANG Guo-hua, HUANG Yan-bo

(School of Information Science and Engineering, Central South University, Changsha 410083)

【Abstract】 Aiming at large flow and inefficiently to incent peer in the current trust model, this paper proposes a Trust Model base on Chord Layer construction(TMCL). It can access the information efficiently and alleviate flow impact which widely exists in trust model based on recommendation. At the same time it adopts bi-direction select mechanisms in transaction which prevents the malicious peers and incentive peer provides better services. Experiment shows that compared with the existing trust model, TMCL has better performance in reducing network overhead and improves the QoS in network.

【Key words】 P2P network; trust model; layer construction; bi-direction selection

由于 P2P 的开放性、匿名性等特点, 节点之间的交流和服务存在着很大的安全隐患, 信任模型作为一种解决方案, 得到了广泛的研究与发展, 但信誉机制引入后, 为了保证信誉信息的客观性与有效性, 耗费了大量的系统资源收集信息, 计算信誉值, 给网络带来了巨大的流量负载。因此, 本文设计了一种 P2P 系统中基于 Chord 的层次化结构信誉模型 (Trust Model in P2P systems based on Chord Layer construction, TMCL), 在保证高效获取信息的同时, 建立激励与惩罚机制, 以求在较小的网络流量开销下提供良好的网络服务质量。

1 相关研究

近年来, 为抑制网络中的不良行为, 提高网络服务质量, 国内外在信任机制方面进行了大量的研究并提出多种信任模型, 但依然存在很多问题亟待解决, 具体表现在以下几方面:

(1) 信誉信息计算带来的巨大流量: 文献[1]提出的基于局部声誉的信任模型 HBDTM, 简单且通信代价小, 但抗攻击能力差, 信任度量具有一定的片面性。文献[2-3]提出的基于全局声誉的信任模型 EigenRep, Trust 能更准确地估计节点行为, 但随之产生的交互信息给网络带来了巨大的流量冲击。文献[4]提出的基于置信因子综合局部声誉和全局声誉的信任机制 PeerTrust, 虽兼备上述 2 种信任机制的优点, 但依旧带来了大量流量。

(2) 信誉信息获取方面: 为保证信誉信息的真实性与可信性, 避免节点修改自己的信誉信息, 信誉信息一般分散存放于网络中其他节点上。如何高效地发现推荐信息是信誉研究的一个难题, 建立如 P2Grid 那样的分布式存储系统, 虽然具有较快的搜索速度, 发现的准确性得到了保证, 但是以信誉信息更新频繁为代价, 且节点性能的差异会严重影响整个系统的效率。目前普遍采取的局部 Flooding 方式, 虽然模式简

单, 拥有高鲁棒性、高可扩展性、简便易行等优点, 但查询结果可能不完全, 且会产生大量冗余消息。建立基于超级节点或群组的层次化信任模型, 如 GroupRep RASSA^[5]等, 上层结构都是利用 Flooding 方式进行查找的, 同样存在查找不完全、速度慢等问题。

(3) 构建激励机制: 面对盛行的搭便车行为、恶意攻击, 以及病毒文件繁衍造成的不可靠服务, 如何构建及量化激励机制, 鼓励节点直接和间接互惠, 惩罚恶意节点以促进节点合作等都是目前亟待解决的问题。

2 基于 Chord 的层次化结构信誉模型

为了描述方便, 本文以共享资源为例来进行以下相关说明。

定义 1 如果节点 l 为节点 i 提供了资源, 称节点 l 为节点 i 的服务提供者。

定义 2 如果全局节点 l 计算且存储节点 i 信誉值, 称节点 l 为节点 i 的档案点。

定义 3 五元组 $R_i = (T_i, ID_i, ID_j, E, Q_i)$ 是节点 i 的档案点 l 的属性集, 其中:

- (1) T 是节点 i 的信任值;
- (2) ID_i 是提供服务节点的 ID ;
- (3) ID_j 是从 i 处下载了资源, 提交服务评价的节点 ID ;
- (4) E 是具体的服务评价;
- (5) Q_i 为节点 i 对本次交易的确认。

基金项目: 国家自然科学基金资助项目(60673164)

作者简介: 张国华(1984-), 女, 硕士研究生, 主研方向: P2P 算法, 信誉模型; 黄烟波, 教授

收稿日期: 2009-04-20 **E-mail:** zhangguohua@csu.edu.cn

定义 4 六元组 $P^i = (w^i, livetime^i, threshold^i, p_i^i, f^i, C(i))$ 是节点 i 的属性集, 其中:

(1) w^i 是节点 i 拥有的空闲带宽;

(2) $livetime^i$ 是节点 i 的在线时间。根据软件系统自动记录的在线时长与上线次数计算得出, 其计算公式如下:

$$livetime^i = \frac{uptime}{time} \quad (1)$$

(3) $threshold^i$ 是节点 i 愿意相信的服务提供者的信誉最低门限值;

(4) p_i^i 是节点 i 愿意给请求节点提供服务的请求节点信誉最低门限值;

(5) f^i 是节点 i 拥有的各类资源的总量;

(6) $C(i)$ 是节点 i 综合性能大小, 当节点刚加入时, 信誉值 T 占的比例较小, 随着节点交互的增加, 信誉值 T 的比例逐渐增大, 最高可达 0.5, 其计算公式如下:

$$C(i) = \alpha \times \frac{f^i}{F} + \beta \times \frac{livetime^i}{Time} + \varphi \times S(i) + \delta \times T_i$$

$$0 < \alpha, \beta, \varphi, \delta < 1, 0.5 \quad \alpha + \beta + \varphi < 1,$$

$$0 < \delta < 0.5, \quad \alpha + \beta + \varphi + \delta = 1 \quad (2)$$

其中, f 表示系统中所有资源的总量; $Time$ 表示系统创建以来的时间; $S(i)$ 是归一化后节点 i 的存储和计算能力大小。

2.1 模型结构

本文针对上述问题, 考虑到 P2P 网络中节点的异质性, 将网络中的节点根据其综合性能大小 $C(i)$ 分为全局节点(GN)和普通节点(ON)。GN 组成 Chord 的结构化网络, ON 组成无结构的局部化网络, GN 和 ON 组成层次化网络。ON 的信誉值和文件索引信息存储在所连接的 GN 中, 由 GN 负责管理, GN 的信誉值和自身拥有及维护的文件索引信息通过不同的 Hash 函数分布式存储到整个 DHT 网络中。

模型的逻辑结构如图 1 所示。

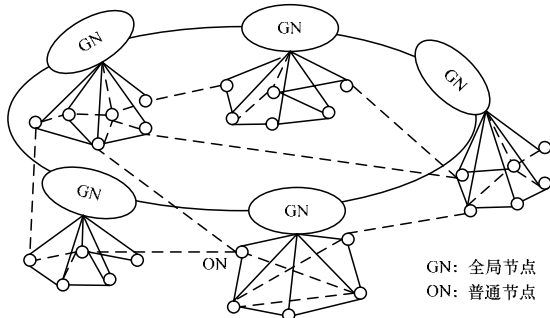


图 1 TMCL 基本结构

由式(2)可知, GN 具有较长的在线时间、较强的计算及存储能力、良好的信誉, 能提供大量的共享资源, 诚实地传递、响应其他节点的请求信息。与超级节点不同的是, $C(i)$ 的大小与信誉密切相关, 一旦作恶, 信誉值降低, 必将引起 $C(i)$ 值减少, 贬为普通节点。选取 $C(i)$ 值次之的节点为候选节点, 备份 GN 所存储的系统信息, 以应对 GN 降级或下线的情况, 增强系统的容错能力。

ON 根据物理位置远近选择合适的 GN 进行连接, 动态地形成以 GN 点为中心的组群。

在查找资源时, 首先在自己的本地信任库(存储以前交易过的节点信息及评价)查找相应节点; 若没有, 则转发给所连接的 GN。GN 得到查询请求, 首先在自身的信任库中查找; 若没有, 则通过哈希函数在 GN 所组成的 Chord 层查找。具

体见算法 1。

算法 1 资源查找算法

Input 请求节点的 ID 号, 要查找资源 R

Output 提供资源节点的 ID 号

```

1 Begin
2   If the request peer is ON
3     Search in local DB(R)
//在本地信任库中查找连接的节点是否有资源 R
4   If not find
5     sent the request to GN
//查询请求转发给 GN
6   Endif
7   Else the request peer is GN OR receive the request from ON
8     Search in GN's local DB(R)
//在全局节点的信任库中查找
9   If not find
10    find with hash(R) in GN.DHT
//如果没有, 则通过 Hash 算法在 DHT 层查找
11  If not find
12    return "the system didn't have R"
13  exit
14  Endif
15  Endif
16  Endif
17  Return the peer ID
18 End

```

由于 P2P 系统高动态的特性, 致使为每个节点都耗费大量的系统资源来计算全局信誉值必得不偿失, 因此在 TMCL 中, 为了减少信誉迭代产生的流量, 对在线时间短、数量大的 ON 只计算本地信誉值, 着重考察它们共享资源的能力。对在线时间长、数量少的 GN, 考虑到它们还肩负着跨区域查询的重担, 一旦将恶意节点选为 GN, 将给网络带来很大的负面影响。因此, 不能只计算衡量共享能力的本地信誉值, 还必须综合考虑推荐可信性, 以得出更能全面刻画节点行为的全局信任值。同时, 只给 GN 计算全局信誉值, 也是由全局信誉值的计算需要搜集大量历史反馈信息这一特性决定的, 只有节点经过一段时间活动, 具有了较长的存活期, 才能获得足够的评价依据, 使其信任度收敛到真实值。

2.2 基于双向选择的服务过程

在搜索到符合要求的服务后, 如何对其中的服务进行过滤以保证服务质量, 成为接下来需要解决的主要问题。在本模型中, 采用了基于信誉值的双向选择服务机制。一方面达到了选择更好服务的目的, 另一方面基于信誉给予节点不同的待遇, 对节点起到了激励与惩罚的作用。

一次典型服务过程如下:

(1)资源请求节点 A 通过算法 1, 得到提供资源的节点列表 B, 给 B 发送服务请求。

(2)服务提供者 B 中的节点, 如果 w^b 不为 0, 则向请求节点 A 的档案点提交信誉查询请求, 得到反馈后, 根据 A 的信任值大小决定是否响应服务请求。

(3)当响应的服务节点列表 B' 中节点不唯一时, 请求节点 A 向 B' 中各节点的档案点提交信誉查询请求, 接收反馈。

(4)A 根据反馈的信誉值, 选择信任值符合 $T_b \cdot threshold^A$ 的节点 b_1, b_2, \dots, b_n 发送连接请求, 获取资源。

(5)服务结束后, A 根据对节点 b_1, b_2, \dots, b_n 提供服务的满意程度向 b_1, b_2, \dots, b_n 各自的档案点提交关于此次服务的评价

信息。

(6) 接收到评价信息的档案点向对应节点发送服务确认请求, 若返回 Q_i 值为真, 则根据评价信息更新信誉值。

交互过程如图 2 所示。

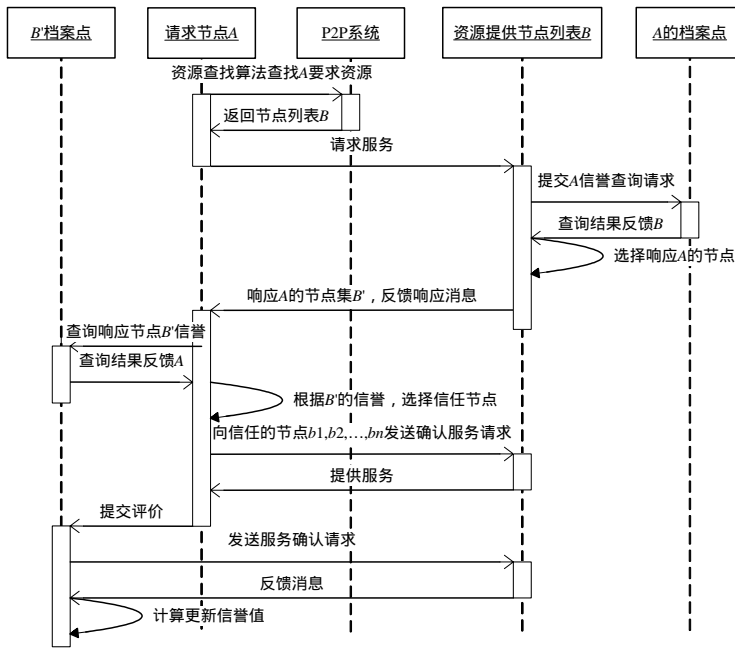


图 2 基于双向选择的服务过程

3 实验结果及分析

本文基于斯坦福大学开发的查询周期仿真器实现了信誉模型 TMCL。同时, 实现了 EigenRep^[2] 方案和 RASSA^[5] 方案。在实验中, 设定高性能节点、好节点、恶意节点的比例分别为 10%, 80%, 10%, 在联合欺诈的攻击模式下, 500 个仿真周期后对 EigenRep, RASSA 和 TMCL 模型产生的消息流量、节点的成功请求率进行了比较分析。

由图 3 可以看出, 3 种方案的网络流量差异很大。

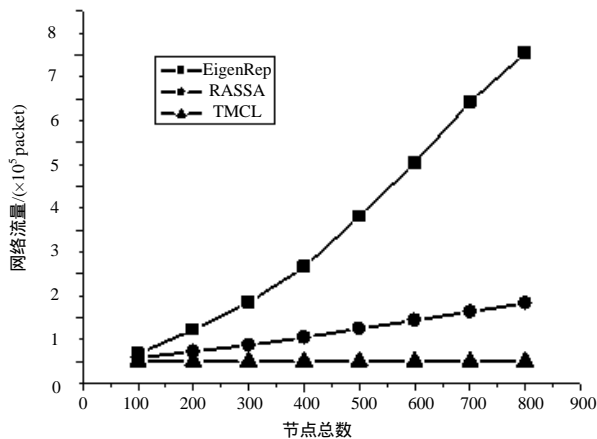


图 3 网络流量对比

在 100 个节点时, 流量比较接近, 随着节点数增多, 差异越来越明显, 到 800 个节点时, 由于 EigenRep 基于泛洪查找, 加之全局信誉计算消耗大量的流量, 高达 772 468 packet, 远远超过了另外 2 种。基于超级节点的 RASSA, 泛洪在超级节点层, 产生的流量较 EigenRep 小, 且超级节点数量相对增长慢, 因此流量增长速度比较稳定。TMCL 模型产生流量最小, 一直保持在十万以下。原因如下: (1) 架构在基于 Chord

的层次化网络拓扑上, 查找资源时充分利用了 DHT 查找的高效性及网络节点邻近性优势, 提高查询速度的同时降低了网络流量; (2) 加入了节点异质性的考虑, 只为全局节点计算全局信任值的机制, 也大大减少了信任值计算时因全局迭代带来的网络流量。

成功请求率 RSQ : 假设合作节点共发起请求数为 Q , 其中有 q 个请求下载到了有效文件, 则 $RSQ=q/Q$ 。

由图 4 可知, EigenRep 的 RSQ 最低, RASSA 次之, TMCL 最高。这是因为 EigenRep 没有考虑恶意节点问题, 在 100 个节点时 RSQ 就只能达到 0.76, 随着恶意节点的增多, 在其联合欺诈的攻击下, RSQ 急速下降。基于超级节点(群组)的 RASSA, 由于超级节点的存在, 提供资源相对比较准确, 但同样会因恶意节点数目的增加, RSQ 下降, 增加到 500 后, 随着节点增多, 交互增加, 恶意节点被排斥出超级节点层, 影响减少, 因而 RSQ 趋向稳定。TMCL 由于一开始节点响应交互, 就具有一定的选择性, 信誉值相对低的节点在选择期间就被丢弃, 因此大大提高了下载到好文件的概率, 节点数为 100 时, 就具有较高的 RSQ , 达 0.842 41, 同时也减少了流量。但随着恶意节点数目的增加且未被识别出来之前, 它们总是不断地发送查询请求并给出响应, 因而在恶意节点联合欺骗下, RSQ 下降, 到 300 个节点时,

恶意节点猖獗到顶峰, 导致 RSQ 产生了一个低的峰值, 但同时因欺骗影响范围大, 信誉迅速降低, 对网络的破坏力减小, 因而当节点数增加 400 时, RSQ 值达到了一个高峰 0.846 15。随后, 由于恶意节点通过慢慢积累, 信誉增加, 在增加到 500 个节点时, RSQ 有所下降, 但随着系统规模的进一步扩大, 恶意行为被抑制, RSQ 逐渐稳步上升。

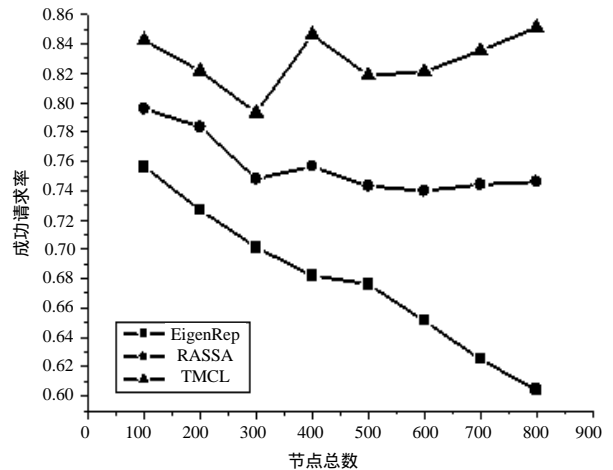


图 4 成功请求率对比

4 结束语

信誉模型 TMCL 充分考虑节点的异质性, 给予不同的节点不同的权利和义务, 同时通过双向选择机制对搜索到的服务进行过滤, 一方面保证了服务质量, 另一方面也对节点起到了很好的激励作用。但因篇幅限制, 本文没有深入研究信誉值计算问题, 下一步将进一步完善 TMCL 模型, 对推荐信息进行评估过滤, 保证信誉值的真实性, 同时对信誉值计算方法做进一步的探讨。

(下转第 186 页)