

P2P 网络中基于激励机制的副本策略

郭良敏^{1,2}, 杨寿保¹, 路卫娜¹, 张瑞¹, 武斌¹

(1. 中国科学技术大学计算机科学与技术学院, 安徽合肥 230027; 2. 安徽师范大学计算机科学与技术系, 安徽芜湖 241003)

摘要: 为了在 P2P 网络的副本技术应用中调动节点主动提供存储空间的积极性、增强节点间的相互合作, 借鉴信息经济学中委托—代理关系的相关知识以及激励的概念, 提出基于激励机制的副本策略, 从而提高 P2P 网络的性能。模拟实验表明, 所提出的基于激励机制的副本策略可以削弱由于节点间的不合作所引起的不能成功创建文件副本的影响, 可以在一定程度上提高网络的搜索性能。与目前几种经典的副本策略相比, 所提出的策略降低了网络的搜索响应时间, 提升了搜索的成功率, 并使得节点的负载相对达到均衡。

关键词: 副本; 激励机制; 委托节点; 代理节点; P2P 网络

中图分类号: TP393 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2010.10.014

Replication strategy based on incentive mechanism in peer-to-peer networks

GUO Liangmin^{1,2}, YANG Shoubao¹, LU Weina¹, ZHANG Rui¹, WU Bin¹

(1. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China;

2. Department of Computer Science and Technology, Anhui Normal University, Wuhu 241003, China)

Abstract: To encourage peers in P2P network to provide their storage space saving replications and strengthen cooperation between them, the replication strategy based on incentive mechanism was put forward using information economics about principal-agent relationship and the incentive concept for reference. The experimental results show that the proposed replication strategy can weaken the negative effect brought about by noncooperation between peers as far as possible to create replications successfully, and indeed improve search performance of P2P networks to a certain extent. Compared with classical replication strategies, the new method can reduce search response time, promote search success ratio and balance peers' load.

Key words: replication; incentive mechanism; principal-peer; agent-peer; P2P network

0 引言

P2P 网络^[1]是一种分布式网络, 该网络中的参与者(节点)共享他们所拥有的一部分硬件资源, 如

处理能力、存储能力、网络连接能力等, 节点既充当客户机又充当服务器。

P2P 系统按照发展历程可简单分为结构化和非结构化两种类型, 其中非结构化的 P2P 系统应用较

为广泛.但是非结构化 P2P 系统中的搜索机制一般采用洪泛^[2](flooding)方法,每次搜索会产生大量的网络流量,增加带宽消耗,影响网络性能.除此之外,无论是结构化还是非结构化 P2P 系统,若对节点中的某文件请求率过高,则会增加该节点的负载,使该节点成为热点,影响其工作效率,也给其他节点带来访问延迟.为了减少网络流量和访问延迟,缓解节点压力,并提高搜索效率和文件的可用性,需要在其他节点上放置文件的拷贝,即副本技术,它是 P2P 系统中常用的一种数据管理机制.

但是,在 P2P 系统中由于节点的自治自私性,每个节点在对外共享自己的资源时是完全出于自愿的.因此,网络中每个节点都希望尽可能多地使用其他节点的资源,而尽可能少地贡献自己的资源.这种自治自私性导致网络中节点存在着不可靠的自主行为,它严重影响了网络的性能.现有的副本策略基本没有考虑节点的自治自私性(为简单起见,本文假设自治自私性仅表现在节点是否愿意保存副本的问题上,并认为网络中的节点不存在欺诈行为)所引发的节点间的不合作问题,即节点未必愿意提供存储空间去保存其他节点中的文件的副本.因此,为了提高节点保存副本、提供服务的积极性,本文利用信息经济学中委托—代理关系的相关知识以及激励的概念,提出基于激励机制的副本策略(incentive replication,简称为 IR),让节点自愿地去保存相应文件的副本,主动为其他节点提供可靠服务.

本文其余部分安排如下:节 1 介绍相关工作;节 2 介绍信息经济学中的相关基础知识^[3-4];节 3 详述基于激励机制的副本策略;节 4 是模拟实验,对所提出的副本策略进行性能测试;最后是本文的结论.

1 相关工作

P2P 系统中最常采用的副本创建策略可分为三类,宿主复制(owner replication,简称为 OR)、随机复制(random replication,简称为 RR)和路径复制(path replication,简称为 PR)^[5]:

①宿主复制:若目标文件不在查询源节点中,并且查询成功,则仅在查询源节点上保存该文件的副本,这种情况下,系统中副本的数量与该文件被访问的频率基本成正比关系,并且该数量是有限的;

②随机复制:按照一定的概率随机选取节点部署文件副本.此方法需了解网络的结构,不太容易实现;

③路径复制^[6-7]:若目标文件不在查询源节点中,并且查询成功,则在查询源节点到目标文件节点的路径节点上(包括查询源节点)保存该文件的副本.由于高度数的节点会比低度数节点转发更多的请求,所以该策略会使大量副本存在于高度数的节点上.

Yamamoto 等^[8]提出了基于随机路径复制(path random replication)和自适应的路径复制(path adaptive replication)的副本策略,它只是将传统的路径复制的副本放置算法进行优化,让负载更加均衡.Tewari 等^[9]提出了基于文件的请求概率制作相应的副本,将下载文件所经过的链路数作为评判性能的尺度.Cohen 等^[10]提出了 square-root 副本策略,该策略重点在于文件副本的数量问题.Lamehamedi 等^[11]是通过比较副本的维护代价,选择合适的位置进行副本创建.Sun 等^[12]提出了基于副本访问代价的动态副本策略,主要以缩短响应延迟为目的.

但是,上述的这些副本策略以及国内外已有的绝大多数关于 P2P 网络副本的研究^[13-16]基本都是假设节点是充分合作的,并没有考虑节点的自治自私性所导致的不合作问题以及这种不合作给副本创建所带来的阻力.本文主要针对节点间的不合作问题,提出基于激励机制的副本策略,利用激励机制增强节点间的相互合作,最终使得副本策略能够更好地发挥其作用.

2 信息经济学基础

2.1 委托—代理关系的基本概念

只要在建立或签订合同前后,市场参与者双方掌握的信息不对称,这种经济关系都可以被认为委托—代理关系.构成委托—代理关系的基本条件有两个:①市场中存在两个以上相互独立的个体,且双方都是在约束条件下的效用最大化者.在这两个个体中,代理人必须在许多可供选择的行动中选择一项预定的行动,该行动既影响其自身的效益,也影响委托人的收益.②代理人和委托人都面临市场的不确定性和风险,且他们之间掌握的信息处于非对称状态.

信息经济学将达成委托—代理均衡合约的条件概括为两个:参与约束和激励相容.所谓参与约束就是代理人履行合约后所获得的收益不能低于某个预定收益额;激励相容就是代理人以行动效用最大化

原则选择具体的操作行动。

2.2 激励机制的目标

对于委托人来说,只有使代理人行动效用最大化,才有可能获得其自身效用最大化的收益。然而,要使代理人采取效用最大化行为,必须对代理人进行有效的刺激。这样,委托人与代理人之间的收益协调问题,转化为激励机制的设计问题。

设计激励机制的目标是:①针对代理人的隐蔽信息而面临的不利选择地位,激励的目标是如何使代理人“自觉地”显示他们的私人信息或真实偏好,即所谓的“如何让人说真话”;②针对代理人的隐蔽行动而可能面临的风险问题,激励的目标就是如何使代理人“自觉自动地”尽最大努力工作,诱使代理人不采取道德风险行动,即所谓的“如何让人不偷懒”。激励机制的核心就是“我怎样使某人为我谋事”。

3 基于激励机制的副本策略

借鉴上述信息经济学的思想,P2P 系统中某文件的源节点(即拥有该文件的初始节点,简称为 h_p)与保存文件副本的副本节点(简称为 r_p)之间近似地认为满足委托—代理关系。因此,本文假定某文件的源节点为委托人(又称委托节点),存放文件副本的节点为代理人(又称代理节点)。本节主要给出奖惩代理节点所设计的激励模型以及相应的副本策略。

3.1 激励模型

令 e 为代理节点所付出的努力程度, $c_i(e)$ 是代理节点为副本 i 付出努力 e 的成本(如存储空间、负载及带宽等)。委托节点成功创建副本 i 的效用函数 $U_h(\cdot)$ 用下式度量(这里将委托节点创建副本的成本忽略不计):

$$U_h(e) = b_i(e) - w_i(e) \quad (1)$$

式中, $b_i(e)$ 代表委托节点创建副本 i 后所取得的收益; $w_i(e)$ 是代理节点保存副本 i 所获得的报酬,该报酬由委托节点给予。

代理节点接受副本 i 的效用函数 $U_r(\cdot)$ 如下式所示,即所获得的报酬与自己所付出努力的成本之间的差值:

$$U_r(e) = w_i(e) - c_i(e) \quad (2)$$

在网络中,无论是委托节点还是代理节点,他们都想最大化自己的效用,获得更多的收益。根据委托—代理关系中的参与约束原则,代理节点从合约

中得到的效用不能小于不接受该合约所可能得到的效用 μ_i (可称为保留效用),参与约束条件可以写为

$$w_i(e) - c_i(e) \geq \mu_i \quad (3)$$

委托节点希望代理节点选择在既定约束条件下能产生最多的效用,即 $U_h(e)$ 最大。若要使 $U_h(e)$ 取得尽可能的最大值,则需差值 $(b_i(e) - w_i(e))$ 取得最大值,对于委托节点希望能最小化 $w_i(e)$ 来得到最大值。同时,为了满足式(3)的要求,一般情况下,委托节点希望代理节点所选择的努力恰好可以满足约束条件,即

$$w_i(e) - c_i(e) = \mu_i \quad (4)$$

由于对于具体的某副本 i , μ_i 和 $c_i(e)$ 都可唯一确定,因此满足式(4)的 $w_i(e)$ 相对达到最小值。

经济学理论认为,若所选择的努力程度 e^* 使边际收益等于边际成本,则此时的收益最大。边际收益一般指每增加一个单位产品所增加的收入,即保存一个副本 i 所增加的收入;边际成本一般指每增加一单位产品所增加的成本,即保存一个副本 i 所增加的成本。显然, e^* 为代理节点使委托节点收益最大化所付出的努力程度。

激励机制的方法就是通过 $w_i(e)$ 使代理节点在选择该合约时获得的效用大于选择其他合约所获得的效用,即激励相容约束;对所有的选择,有 $w_i(e^*) - c_i(e^*)$ 的值最大,即 e^* 是代理节点对应的最优选择。

3.2 模型计算的简化

委托节点的“收益”主要体现在本身负载的减轻、相应处理能力的提升、带宽消耗的减少。代理节点的“收益”体现在因保存副本而获得的报酬,但同时会伴随着自身处理能力、共享存储空间以及带宽资源的减少。为了简化整个计算过程,本文做如下设置:将代理节点所采取的努力程度 e 的值简化为 0 或 1。

(I) 若 e 值为 0,则代表该代理节点不愿意接受此副本在本地的创建,相应的 $c_i(e)$, $w_i(e)$ 和 $b_i(e)$ 均为 0。

(II) 若 e 值为 1,则表示该代理节点同意在本本地创建副本。

①相应的 $c_i(e)$ 用代理节点负载增加的比例与处理能力、带宽、存储空间减少的比例的加权和来衡量,如下式:

$$c_i(e) = \alpha_r \times L_r + \beta_r \times P_r + \gamma_r \times W_r + \lambda_r \times S_r \quad (5)$$

②相应的 $b_i(e)$ 用委托节点负载减少的比例与

处理能力、带宽增加的比例加权和来衡量,如下式:

$$b_i(e) = \alpha_h \times L_h + \beta_h \times P_h + \gamma_h \times W_h \quad (6)$$

③相应的 $w_i(e)$ 根据式(4)计算而得,在初始时,任何节点的保留效用认为是 0,将式(5)代入式(4)就可计算出 $w_i(e)$.

其中式(5)中的 $L_r = (\text{代理节点 } r \text{ 所接收到的副本 } i \text{ 的请求数}) / (r \text{ 创建 } i \text{ 之后接收到的总请求数})$,代表负载增加的比例; $S_r = (\text{副本 } i \text{ 的大小}) / (r \text{ 在存放 } i \text{ 之前所拥有的空间大小})$,代表存储空间减少的比例; P_r 是处理能力减少的比例,为计算简单, P_r 近似用(节点 r 所接收到的副本 i 的请求数) / (r 最多可处理的请求数)代替; W_r 是带宽减少的比例,近似用 L_r 代替. 式(6)中的 $L_h = (\text{因创建了 } i \text{ 而减少的请求数}) / (\text{委托节点 } h \text{ 创建 } i \text{ 之前接收到的总请求数})$,代表负载减少的比例; P_h 是处理能力增加的比例,近似用(因创建了 i 而减少的请求数) / (h 最多可处理的请求数); W_h 是带宽增加的比例,近似用 L_h 代替. 为了统一化各个参数,引入 $\alpha_r, \beta_r, \gamma_r, \lambda, \alpha_h, \beta_h$ 和 γ_h 作为各参数的权重. 各参数的权重值是系统初始化时设置的,定义如下:①计算 $c_i(e)$ 时,本文认为对于代理节点可能更注重负载和存储空间的变化,因此可以假设 $\alpha_r = 0.4, \beta_r = 0.1, \gamma_r = 0.1$ 和 $\lambda_r = 0.4$;②计算 $b_i(e)$ 时,本文认为委托节点更注重的是负载和带宽的变化,因此可以假设 $\alpha_h = 0.4, \beta_h = 0.2, \gamma_h = 0.4$.

3.3 副本策略

系统的目标是通过增加副本数量以及在合适的节点上放置副本来尽量均衡系统内节点的负载、提高系统的搜索效率. 由于系统中节点的自治自私性,负载轻的节点未必愿意提供自己的共享存储空间给负载重的节点,本文通过上述的激励模型,增强节点间的相互合作,激励节点贡献自己的存储空间.

在系统中,每个节点可能会拥有双重身份,既是委托节点又是代理节点. 作为代理节点,它获得的总效用(记作 $\sum U_r$,为区分作为委托节点所获得的效用,本文称其为“合作效用”)可以近似地反映出该节点与其他节点的合作程度. 为了激励节点间相互合作,成功创建副本,系统规定:服务倾向被提供给拥有高合作效用的节点. 具体如下:

(I) 若节点的合作效用 $\sum U_r$ 高,则该节点可以享受系统中其他节点的服务机会就多、服务质量就高,主要体现在:对于合作效用高的节点的请求会

被优先处理,因此其搜索文件的效率、搜索成功率会相对较高,其创建副本的成功率也会较高等.

(II) 若节点的合作效用 $\sum U_r$ 低,则其请求可能不被及时地处理或者不被处理,这样会降低搜索文件的搜索效率和成功率,并且其创建副本的成功率也可能会降低等.

3.3.1 副本的放置

当节点上某共享文件在一段时间 T 内的访问次数超过某个阈值 Ω 后,该节点(委托节点)就需要创建此文件的副本. 对于副本的存放位置,即代理节点的选择问题,本文考虑三个因素:第一个因素就是文件的传输代价,而节点间的距离(用消息的往返时间 RTT 来表示)又是影响传输代价的一个重要参数,因此委托节点会选择距离相近的节点作为代理节点;另外,选择距离相近的节点作为代理节点也会降低副本的更新代价. 第二个因素是节点的在线时间,在线时间长的节点所提供的服务会更加可靠,在此类节点上创建的副本会被网络中的其他节点充分利用,这样才能更好地减轻委托节点的负载. 关于在线时长的计算,本文通过最近一段时间内的在线时长(其中未有下线)与以往时间内节点的在线稳定因子的乘积而得,其中在线稳定因子 = (以往时间段内的总在线时长) / (以往时间段内的总在线时长 + 以往时间段内的总不在线时长). 第三个因素就是节点的合作效用,若有多个可被选择的节点,则从中选择合作效用高的节点,如此可以激励节点间相互合作. 为了管理的方便,委托节点对于自己的共享文件需要维护如图 1 所示的数据结构.

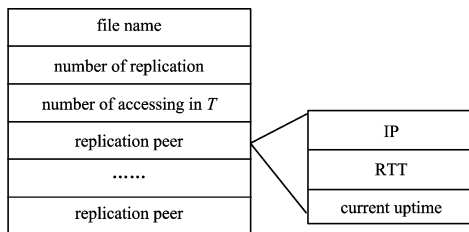


图 1 文件数据结构

Fig. 1 Data structure of file

副本放置的具体过程如下:

(I) 委托节点通过邻居表中的节点信息向周围节点发送副本创建请求 request(请求中含有节点的合作效用值);

(II) 接收到该请求的节点,若已有该文件的副本,则将请求转发给自己的下行邻居节点(即非

request 的发送方节点),若没有该文件副本,则根据得到的效用值决定是否接受该副本.若在某时刻接收到的请求数多于一个,则节点会优先处理具有高合作效用的节点请求;

(III) 若代理节点接受该副本,则其会发送一个同意接受的反馈消息(包含其合作效用值和在线时长)给委托节点.委托节点收到该反馈后,根据消息的 RTT 值和合作效用值选择相对较近的和高合作效用的节点作为代理节点的后备,然后再从后备中选择在线时间长的节点作为代理节点.接着,发送副本给代理节点,并向网络宣布该节点上有某文件的副本,同时记录副本节点的相关信息.对于不合适的节点暂时不处理;

(IV) 若代理节点不接受该副本,则会将请求转发给自己的下行邻居节点,并发送拒绝接受的消息给委托节点.委托节点收到消息后,根据返回的消息决定是否发送新的请求给其他节点,若需要,则转(I),若不需要,则过程结束.

在这里,某个节点可能多次收到同样的请求,为缩减节点处理请求的时间,本文对委托节点发出的副本创建请求中加一个标签 tag,用以区别不同的副本创建请求.当节点收到相同 tag 的创建请求时,直接将请求转发,不作其他处理.所以在上述的步骤(II),接收到请求的节点通过请求中的 tag 值来判断该请求是否需要处理.在步骤(III)所做的修改是,节点只要发送了同意接受的反馈消息,它就应将相应请求的 tag 值存储起来,以备步骤(II)中的判断.

副本放置算法的伪代码如图 2 所示.在此算法中, $sort_{RTT-\sum U_r}()$ 的时间复杂度约是 $O(N * \lg N)$, N 是委托节点收到的同意请求数目, $sort_{uptime-\sum U_r}()$ 的时间复杂度约是 $O(k * \lg k)$, $sort()$ 的时间复杂度是 $O(t * \lg t)$, t 是代理节点收到的副本创建请求的个数.因此算法的复杂度是 $O(N * \lg N + k * \lg k + |S''| + t * \lg t + |Q|)$.

3.3.2 副本的删除

由于节点的共享存储空间是有限的,因此代理节点为了充分利用自己的共享存储空间,需要将某些副本删除.系统中的代理节点会统计自己共享存储空间内所存放的副本在最近的时间段 τ 内被访问的次数 $C_{Access}(\tau)$,利用 $C_{Access}(\tau)$ 、副本的大小 F_{size} 以及与源节点的距离 D 综合出副本的删除代价 $cost$ (其值等于 $C_{Access}(\tau) * F_{size} * D$).按这个 $cost$ 从低到高将各个副本排序,根据需要选择代价较低的副

Replication-placement algorithm

- 1: Principal-peer send requests to neighbor peers;
- 2: Agent-peer received requests;
 - 2.1: $Q = sort(\sum U_r)$; /* Q 是发送创建请求节点的集合,并按合作效用降序排列 */
 - 2.2: Send agreeable message to node $Q[i]$, goto 3; /* 节点 $Q[i]$ 的合作效用是代理节点满意的 */
 - 2.3: Forward the requests, and send refusal message to node $Q[j]$, goto 2; /* 节点 $Q[j]$ 的合作效用是代理节点不满意的 */
- 3: $S = sort_{RTT-\sum U_r}(RTT_s)$; /* S 是委托节点依据距离和合作效用降序排列的节点集合 */
 - 3.1: $(k, S') = satisfy(S)$; /* 选择距离较近的节点构成集合 S' 并获得节点数目 $k = |S'|$ */
 - 3.2: $S'' = sort_{uptime-\sum U_r}(S')$; /* 依据在线时长对集合 S' 降序排列 */
 - 3.3: Send replication to agent-peer $S''[i](i=0,1,2,\dots,k-1)$;
 - 3.4: Principal-peer declare these agent-peers to network, and save their information and pay reward;
- 4: if $(K < r)$ then Send request to another peers, goto 2; /* K 是委托节点已有的代理节点数, r 是需要创建的副本数 */

图 2 副本放置算法

Fig. 2 Replication placement algorithm

本删除.副本删除的同时,代理节点要向该副本的源节点发送已将某文件副本删除的消息.代理节点所管理的副本相关信息如图 3 所示.

replication name
source peer IP
number of total accessing
number of current accessing (τ)

图 3 副本数据结构

Fig. 3 Data structure of replication

为了防止删除的副本将来又会被多个节点请求,该节点会根据副本在之前的热门程度和访问规律对以后流行程度进行预测,做出不同的反应:①存放该副本源节点的相关信息,以便提高将来搜索该文件的效率;②将副本暂时存放到私有空间,以备需要时直接放到共享空间中.对于②,仅对只读文件有效,对于可修改的文件,还得从源节点获得最新的文件副本.

4 模拟实验与分析

本文选取 P2P 文件共享系统作为模拟设定场景,并以 flooding 算法作为底层的查询协议,运用 C 和 MATLAB 做模拟实验来验证所提出基于激励机

制的副本策略的性能.

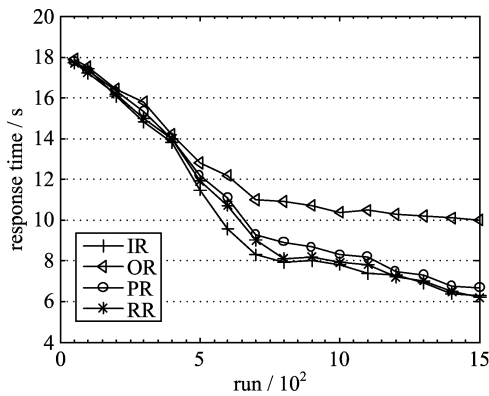
4.1 实验参数设定

在实验中,假定该文件共享网络包含1 000个节点(其中关于节点是否合作,我们考虑表1列出的几种情形,将网络中的节点分为合作节点(cooperative peer)、非合作节点(disoperative peer)以及有可能表现为合作也有可能表现为非合作的不确定节点(uncertain peer)),网络中初始文件数是200个,并且每个初始文件是相异的,每个节点的共享存储空间是5个副本的大小,同一个节点内不存在相同的两个或两个以上的文件或副本.查询模拟按轮(run)进行,每轮随机选择200个节点发出请求.具体的参数设置及其缺省值如表2所示.为了简化实验,本文

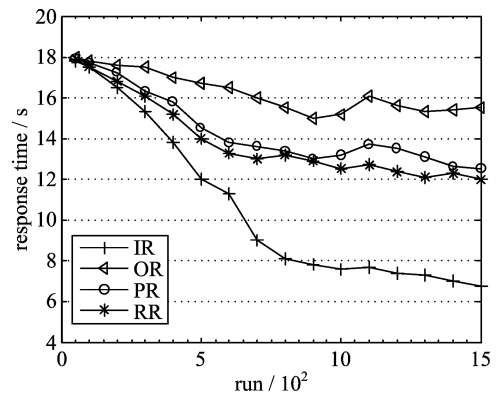
表1 各类节点所占比例

Tab.1 Proportion of different peers

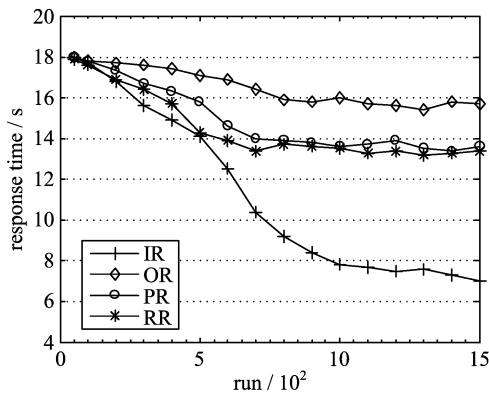
scenarios	cooperative peer	disoperative peer	uncertain peer
1	100%	0	0
2	50%	30%	20%
3	20%	50%	30%
4	0	80%	20%



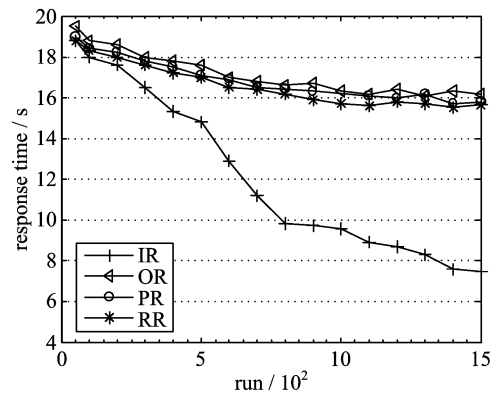
(a) Scenarios 1



(b) Scenarios 2



(c) Scenarios 3



(d) Scenarios 4

图4 搜索响应时间

Fig. 4 Response time of search

表2 实验参数及其缺省值

Tab.2 Parameters and default values

parameters	value
number of peers	1 000
number of files	200
searching algorithm	flooding
share storage size of each peer	5
max connection degree of each peer	10
TTL	7

未考虑文件的更新问题.

4.2 性能分析

本文主要与目前几种经典的副本策略(如宿主复制策略、随机复制策略、路径复制策略)进行比较来验证IR策略的优越性,即验证激励模型的有效性.下面就是分析表1所示的4种情形下各副本策略的性能.

4.2.1 搜索性能

实验中,首先比较了各副本策略下(IR,OR,PR和RR)的网络搜索性能:搜索效率(用搜索的响应时间来衡量)和搜索成功率.图4给出了表1所列

的 4 种情形下 IR 与其他几种副本策略的搜索效率。从图 4 中可以看出:在 Scenarios 1 里,即假设所有的节点都是充分合作的,各副本策略的性能差别不是很大;在 Scenarios 2,3 和 4 里,各副本策略的差距就比较明显,本文提出的 IR 策略在响应时间上相比其他副本策略有一定程度地减少;特别在 Scenarios 4 里,即 80% 以上的节点是不合作的,除 IR 策略仍然保持较好的性能外,OR,PR 和 RR 副本策略的性能都有很大程度地下降,并且在这种情形下,这 3 种策略的性能基本差不多。这充分说明了,当网络中绝大部分的节点表现为不合作时,没有考虑节点不合作性的副本策略就会很难发挥其应有的作用。因此,网络中某些节点存在的不合作性,使得 OR,PR,RR 策略未必可以成功地在节点上创建副本,从而导致其搜索效率低于基于 IR 的网络。

图 5 是对上述各副本策略搜索成功率的比较。图 5(a)反映的是所有节点都互相合作的情况,可看出 IR 策略与 PR,RR 策略差距很小,都比 OR 策略

的搜索成功率高些,原因在于 OR 策略所创建的副本数量是有限的;图 5(b),(c),(d)反映了当存在节点不合作情况时,IR 策略的性能就会优于其他 3 种策略,并且随着不合作节点比例的增加,IR 策略的优越性也会更加明显些。因此,从图 4,5 的结果可以看出,本文提出的激励模型能在一定程度上增强节点间的合作性,削弱了由于节点的不合作性对副本策略性能的干扰。

4.2.2 负载变化

实验假设在某一轮中某节点上某文件的请求数多于 20 时,该节点就得创建此文件的副本。初始时(即网络中不存在任何一个文件的副本),选择网络中的某个文件,该文件存储在节点 m 上,增加对其的请求量(每轮随机地选择网络中的若干节点给 m 发请求)。可以从图 6 看出,在开始的 100 轮之前节点 m 的负载(用 m 处理的请求数来表示)在不断升高。100 轮之后,不管网络中是否存在不合作的节点,IR 策略对于降低节点的负载始终是比较有效的,而对于其他 3 种策略,只有在节点合作时,才能

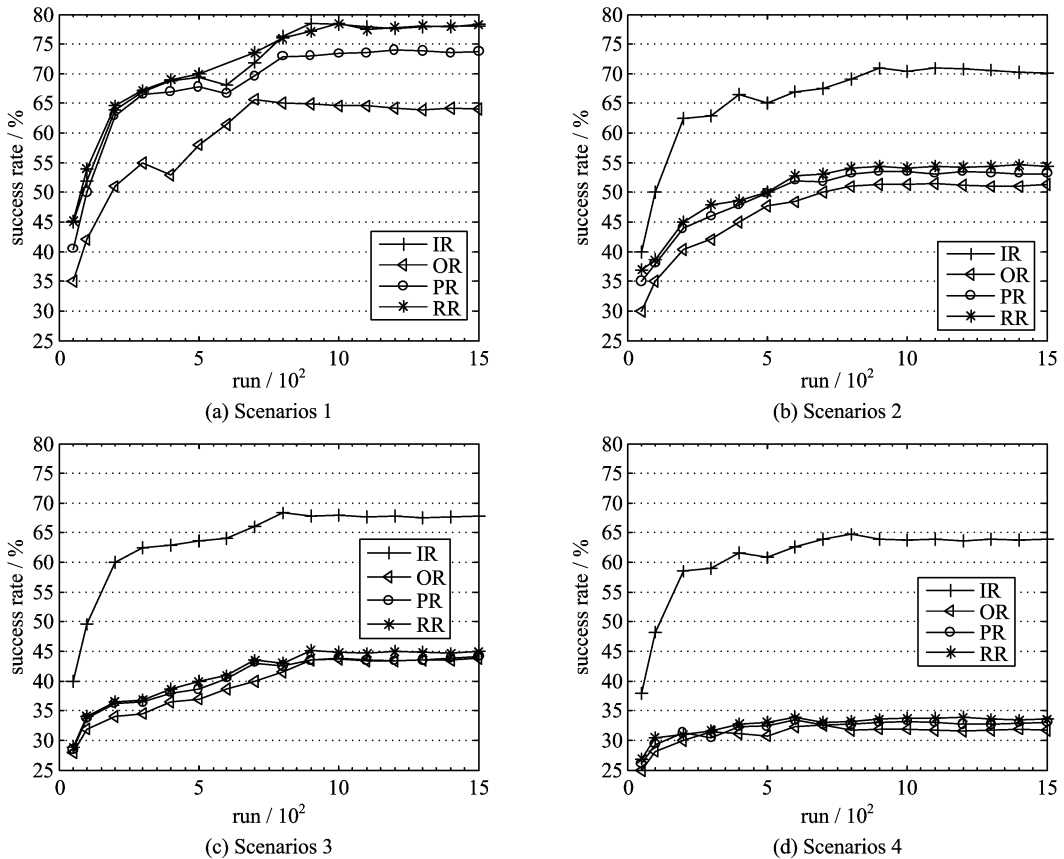


图 5 搜索成功率

Fig. 5 Success rate of search

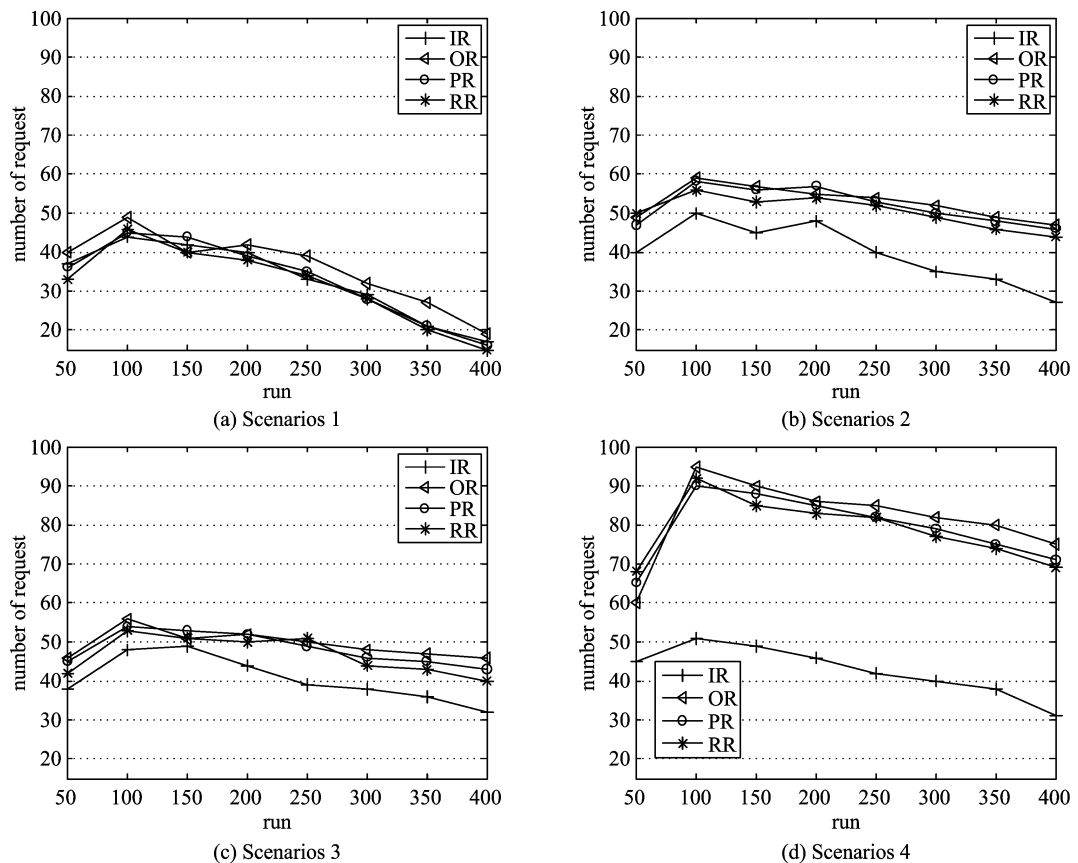


图 6 负载变化

Fig. 6 Load fluctuation

较好地降低负载,如图 6(a). 上述结果表明:即使有不合作的节点存在,IR 策略也能比较顺利地创建副本以降低某些节点的负载,从而达到负载的相对均衡,也就是说,基于激励模型的副本策略在消除热点瓶颈问题和提高系统可靠性方面是有一定效果的.

5 结论

针对 P2P 网络中节点的自治自私性,为提高节点贡献自己共享存储空间的积极性、增强节点间的相互合作,本文结合信息经济学中委托—代理关系的相关知识,提出了基于激励机制的副本策略. 实验表明,该副本策略可以改善由于节点的不合作所导致的副本不能成功创建的问题,能在一定程度上提高网络的性能.

但是本文的工作是以节点不存在欺诈行为这个假设为前提的,对于存在欺诈行为的网络,所提出的模型和副本策略应该如何做调整,以及节点的动态性所导致的副本不一致问题和如何根据文件现有的

访问情况预测其将来的使用率都是下一步所要研究的工作.

参考文献 (References)

- [1] Oram A. Peer-to-Peer: Harnessing the Power of Disruptive Technologies[M]. Sebastopol, CA, USA: O'Reilly & Associates, Inc, 2001.
- [2] Harchol-Balter M, Leighton T, Lewin D. Resource discovery in distributed networks[C]// Proceeding of 18th Annual ACM-SIGACT/SIGOPS Symposium on Principles of Distributed Computing. Atlanta, Georgia, USA: ACM, 1999: 229-237.
- [3] 乌家培. 信息经济学[M]. 北京: 高等教育出版社, 2002.
- [4] Macho-Stadler I, Pérez-Castrillo J D. 信息经济学引论: 激励与合约[M]. 管毅平, 译. 上海: 上海财经大学出版社, 2004.
- [5] Lv Q, Cao P, Cohen E, et al. Search and replication in unstructured Peer-to-Peer networks[C]// Proceeding of 16th ACM International Conference on Supercomputing. New York: ACM, 2002: 84-95.

- [6] Rowstron A, Druschel P. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility [C]// Proceeding of 18th ACM Symposium on Operating Systems Principles. Banff, Canada; ACM, 2001: 188-201.
- [7] Clarke I, Sandberg O, Wiley B, et al. Freenet: A distributed anonymous information storage and retrieval system [C]// Proceeding of International Workshop on Design Issues in Anonymity and Unobservability. Berkeley, CA, USA: Springer-Verlag, 2001: 46-66.
- [8] Yamamoto H, Maruta D, Oie Y. Replication method for load balancing on distributed storages in P2P networks [J]. IEICE Transactions on Information and Systems, 2006, E89-D(1):171-180.
- [9] Tewari S, Kleinrock L. Proportional replication in Peer-to-Peer networks[C]// Proceeding of 25th IEEE International Conference on Computer Communications. Barcelona, Spain; IEEE, 2006: 556-567.
- [10] Cohen E, Shenker S. Replication strategies in unstructured Peer-to-Peer networks [J]. Computer Communication Review, 2002, 32(4):177-190.
- [11] Lamahmedi H, Shentu Z, Szymanski B, et al. Simulation of dynamic data replication strategies in data grids [C]//Proceedings of the 17th International Symposium On Parallel and Distributed Processing. Nice, France: IEEE Comput Soc, 2003: 10.
- [12] Sun Xin, Zheng Jun, Liu Qiongxin, et al. Dynamic data replication based on access cost in distributed systems[C]// Proceeding of the Fourth International Conference on Computer Sciences and Convergence Information Technology. Seoul, South Korea; IEEE, 2009: 829-834.
- [13] Kim K. Time-related replication for P2P storage system[C]// Proceeding of the Seventh International Conference on Networking. Cancun, MEXICO; IEEE Comput Soc, 2008: 351-356.
- [14] Shen Haiying, Zhu Yingwu. Plover: A proactive low-overhead file replication scheme for structured P2P systems [C]// Proceeding of IEEE International Conference on Communications. Beijing, China: IEEE, 2008:5 619-5 623.
- [15] Hassan O A H, Ramaswamy L. Message replication in unstructured peer-to-peer network[C]// Proceeding of International Conference on Collaborative Computing: Networking, Applications and Worksharing. White Plains, NY: IEEE, 2008: 337-344.
- [16] 冯国富, 张金城, 顾庆, 等. 一种基于覆盖网络拓扑的无结构 P2P 主动复制策略[J]. 软件学报, 2007, 18(9):2 226-2 234.