

分散 P2P 网络中资源的自适应共享策略

徐毅, 王家伟

(重庆交通大学信息科学与工程学院, 重庆 400074)

摘要: 针对 P2P 网络中资源共享交换效率问题, 提出一种计算网络资源共享可能度的自适应策略, 以决定节点对资源的共享抉择。该策略考虑了 P2P 网络中资源共享存在受限存储空间、下载中断和冗余下载等问题, 以资源共享可能度作为资源共享抉择依据, 提出资源共享可能度在单个代理的计算方法。实验结果证明自适应策略能够较好地避免资源的冗余下载, 提高 P2P 网络资源共享效率。

关键词: P2P 网络; 资源管理; 自适应共享

Self-adaptive Sharing Policy for Resources in Decentralized P2P Network

XU Yi, WANG Jia-wei

(School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074)

【Abstract】 To solve the problem of resource sharing efficiency in P2P networks, this paper presents a self-adaptive sharing policy of computing resource sharing probability. The policy is used to decide whether a resource can be shared. It considers some key issues about resource sharing in P2P networks, such as limited store memory space, interrupted or redundant download process. A local computing method is used to compute the resource sharing probability in single agent. Experimental results show that the policy can effective decrease redundant download process numbers of sharing resources.

【Key words】 P2P network; resource management; self-adaptive sharing

1 概述

P2P 网络由分散分布、相互间能够共享交换资源的一组网络节点组成, 本质上是分散、无中心的, 从而使它非常适合在无须考虑扩展性、负载均衡的环境下实现对大量资源的共享。随着 Internet 的普及、端用户系统资源的丰富及网络带宽的快速增加, P2P 网络在协同工作、分布式信息共享、大规模并行计算等方面显示出独特的优势, 成为新的发展热点。现在非常流行的文件共享网络系统如 BT、eMule 都是 P2P 网络系统。

对 P2P 文件交换网络的应用调查表明^[1], 用户间共享交换的资源大多数是大型资源, 如音频、视频资源。这类资源的下载占 P2P 网络上数据传输量的 65% 以上。而 P2P 网络中用户共享的动力来自于用户间的资源交换。一个用户共享自己资源, 同时期待能从其他用户中获得感兴趣资源作为回报。在这种环境下, 用户选择共享其他用户很少或根本不请求的大型资源是不明智的, 共享这类资源不会有助于用户从其他用户得到交换资源回报, 同时这类资源还要消耗用户大量存储空间。

由于 P2P 网络分散的本质, 没有中心节点实施统一的阻塞控制算法确保资源下载, 导致对资源的可靠下载需求难以得到满足。当节点从另一节点下载大型资源时, 必须考虑该资源下载还未完成时资源就已被不被共享的风险。这种情形重复出现会极大影响 P2P 网络的效率。因此, 研究资源共享策略可帮助用户在共享大型资源时尽量减少低效资源或无效资源(不被其他用户请求的资源)带来的空间浪费和其他开销。

目前有很多研究考虑解决 P2P 网络中资源共享的问题, 如以 Napster 为代表的集中式资源管理模型、以 Gnutella 为代

表的纯分布式资源管理模型及流量控制模型^[2]等方法。但 these 方法都有其不足之处, 虽然在一定程度上解决了资源共享问题, 但由于对资源共享抉择缺少启发式选择机制, 无法利用先前历史资源搜索经验, 因此降低了网络中资源利用率^[3]。本文对资源共享采用了一种更好的策略, 根据其他节点对共享资源的请求量和该共享资源在 P2P 网络上当前时刻的可获取性, 计算资源共享可能度, 自适应决定是否在当前节点共享该资源。通过实验验证, 该策略可以有效地减少数据冗余传输, 同时 P2P 网络的效率也不会受到影响。

2 P2P 网络中资源共享的关键问题

假设研究的 P2P 网络是一种无结构、无中心的分散体系结构, 也就是通常所说的纯 P2P 网络。在纯 P2P 网络中, 不存在维护管理网络现有节点和资源信息的中心管理节点, 对资源的一次搜索请求可传播到整个网络。这保证了网络中每个节点都能收到任意节点发起的搜索请求。在研究中利用代理完成资源共享度计算及共享抉择, 假设 P2P 网络的每个节点上都有一个代理, 帮助用户确定他与其他节点的资源共享决策。后文讨论中不再特别区分节点和代理。

P2P 网络中资源共享主要面临如下 3 个关键问题:

(1) 受限存储空间。P2P 网络的每一个节点物理上都位于一台存储容量有限的计算机上。节点的存储容量限制了该节点共享资源的数量和体积。在 P2P 网络中, 节点提供资源共

基金项目: 重庆市自然科学基金资助项目(CSTS, 2008BB2352)

作者简介: 徐毅(1973 -), 男, 讲师、硕士, 主研方向: 嵌入式系统软件, 数据采集, 分布处理; 王家伟, 副教授、硕士

收稿日期: 2009-12-26 **E-mail:** xuyitiger@sohu.com

享的动力来自于期望从其他节点获取自己需要的资源。一种合适的策略是在某个资源不再被其他节点请求共享时，能自适应从共享中移除以节省存储空间，因为该资源不会增加节点从其他节点获取资源的可能性。

(2)中断的下载过程。当一个大型资源被某节点请求获取时，提供共享节点会因提供该资源共享而付出相当的数据传输带宽和下载时间。而中途中断的下载过程对于提供共享节点是不利的，因为中断的下载过程也会耗费提供共享节点的数据传输带宽和下载时间，但未完成的资源共享不会帮助共享节点向请求节点获取资源回报。

(3)冗余下载问题。在 P2P 网络中，因为无法确保完整下载，通常一个节点会向多个提供相同资源的节点提出共享请求以提高获取资源的成功率。在这种情形下，当请求节点从其中一个资源提供节点获得共享资源后会停止其他下载过程，导致其他资源提供节点出现中断的下载过程。通过自适应策略，由共享节点根据网络其他节点提供相同资源的数量和该资源在网络上的需求做出共享决定，可减少不必要的冗余下载过程。

本文提出一种基于资源共享可能度计算的自适应共享策略，以解决在 P2P 网络中共享资源所面临的上述问题。这种自适应共享策略只需要每个节点代理进行本地计算就能决定具体资源的共享决策，可以有效地运行在以分散、无中心为本质特征的 P2P 网络中。

3 资源自适应共享策略

在 P2P 网络中，代理向其他代理提供某个共享资源，以期望从其他代理获取资源回报。一个代理所能获取的资源回报取决于希望获取该共享资源的节点数量^[4]。这也意味着该共享资源可能被共享的概率与希望获取该共享资源的代理数量是正比的，用资源共享可能度表示这个概率。如果共享可能度高，意味着代理共享该资源能够得到更大回报，在共享抉择时应先考虑。而资源共享后，在其他代理下载该资源时还请求共享的代理数量会减少。因此，资源的共享可能度在不同时刻是动态变化的，必须进行调整以反映请求共享代理数量的变化。

3.1 资源共享可能度计算

对 P2P 文件共享网络资源请求数据分析可知，对资源的请求随时间的增加而减少。通常情况下，对新共享的资源点击下载请求较频繁，对已经共享过一段时间的资源的下载请求则少一些。

基于 P2P 网络请求资源的这种特性，假设代理对一个资源按其共享效用评价资源价值。当一个新资源被 P2P 网络上的某代理引入时，其初始价值一定，但随着其他请求共享代理完成该资源的下载，其价值就会下降。在任意时刻，对某资源的下载请求数对于 P2P 网络上所有代理都是相同的。因为资源价值由资源下载请求数目决定，所以 P2P 网络上所有拥有某特定资源的代理都能用相同评价函数评价该资源价值。当网络上所有期望获得该资源的代理都得到该资源后，资源价值最终会下降到 0。为说明这种特征，代理的资源价值评价函数应是线性递减的。具体形式如下：

$$V_t = V_0 \left(1 - \frac{t \text{时刻共享资源代理数}}{\text{期望获取资源代理总数}}\right)$$

其中， V_t 是资源在时刻 t (从资源引入 P2P 网络开始计算) 的价值； V_0 是最初引入资源代理设定的初始价值。定义资源在时刻 t 的共享可能度为 P_t ：

$$P_t = V_t / V_0 = 1 - \frac{t \text{时刻共享资源代理数}}{\text{期望获取资源代理总数}} \quad (1)$$

因此，无论 V_0 是多少，都与计算出的资源共享可能度无关。

资源价值评价函数以资源的相对评估值和资源的生存时刻为参数计算出共享可能度。该值在 0~1 之间，当代理通过该函数计算得到的共享资源的共享可能度为 0 时，可以将该资源从共享存储中移出。一个代理得到其他代理返回的参数值后就能够算出当前时刻某资源的共享可能度。具体分析计算过程如下：

通过分析 P2P 网络模型，设定如下参数：

N ：网络中期望获取某资源的代理数目总数。

u ：网络中代理的平均上传能力。

s ：网络中代理共享比率，即网络中共享资源代理的数量与已获取资源代理数的比率。

g_t ：资源下载周期 t 中获得资源代理数。

C_t ：资源下载周期为 t 时共享资源代理数。

R_t ：资源下载周期为 t 时的下载请求平均数。

P_t ：资源下载周期为 t 时资源共享可能度。

以下载一个资源的平均时间作为一个下载周期，以下载周期作为时间计量单位。一个下载周期从一个共享资源代理获取资源的平均代理数量等于共享代理的平均上传能力 u 。

当代理共享比率 s 不等于 0 时，请求下载代理在下载完成后有一定概率共享刚获得的资源。因为代理平均上传能力为 u ，所以每个提供共享资源的代理都能同时支持 u 个资源下载，在 t 时刻获取资源的数量为

$$g_t = u C_t \quad (2)$$

其中， C_t 是在 t 周期共享资源的代理数， C_t 包括前一个周期 $t-1$ 共享资源的代理数 C_{t-1} ，还包括在周期 $t-1$ 中获取资源的代理中愿意提供共享该资源的代理数，因此：

$$C_t = C_{t-1} + s g_{t-1} \quad (3)$$

将式(2)代入式(3)得到：

$$C_t = C_{t-1} (1 + su) \quad (4)$$

因为在第 1 个周期只有引入资源代理提供共享，所以设定迭代初始条件 $C_1=1$ ，得到：

$$C_t = (1 + su)^{t-1} \quad (5)$$

将 C_t 代入式(2)得到：

$$g_t = u(1 + su)^{t-1}$$

在周期 t 后获得资源的代理总数为

$$\sum_t C_t = \sum_t u(1 + su)^{t-1} = \frac{(su + 1)^t - 1}{s} \quad (6)$$

将式(6)和资源请求总数 N 代入式(1)可知， t 周期后资源共享可能度为

$$P_t = 1 - \frac{(su + 1)^t - 1}{Ns} \quad (7)$$

完成所有请求获取资源代理的资源下载所需的下载周期可由 $Ns = (su + 1)^t - 1$ 求出：

$$t_{\max} = 1 + \log_{su+1}^{(Ns+1)} \quad (8)$$

3.2 网络全局参数 s 、 u 和 N 的本地计算策略

通过上面的分析可知，代理可以用式(7)确定资源的共享可能度。但在分散的 P2P 网络中，对单个代理来说，要得到与整个网络相关的全局参数非常困难，而式(7)中的 s 、 u 和 N 都是全局参数。下面着重讨论这些参数的本地计算策略。

(1)确定网络中代理平均上传能力 u

为计算 u , 网络中每个代理都要向网络其他代理报告(可通过广播等方式)自身的上传能力 u_i 。现有 P2P 网络大多具有这种能力, 它们采用的节点发现协议要求新加入网络节点报告其上传能力。假设代理上传能力不因时间发生变化, 每个代理只需在加入网络时报告一次上传能力。代理可以通过下式在本地单独计算出整个网络的代理平均上传能力:

$$u = \sum_{i=1}^{i=N} u_i / X$$

其中, X 是代理收到报告上传能力的代理总数。

(2)确定网络中代理共享比率 s

可以假设不共享资源代理是那些报告上传能力为 0 的代理。因此, 代理共享比率 s 可由下式计算得到:

$$s = 1 - (u_i = 0 \text{ 的代理数}) / X$$

因此, 根据从其他代理提供的上传能力 u_i 值, 任一代理都能在本地计算出 s 值。

(3)确定网络中期望获取某资源代理数目总数 N

如上所述, 在纯 P2P 网络中, 对资源的一次搜索请求可以传播到整个网络, 这保证了网络中每个节点都能收到任意节点发起的搜索请求。可假设传播搜索资源请求所需时间远小于大型资源下载时间, 代理在一个下载周期内就应收到所有期望获取共享资源但还未获取该资源代理的请求。因此, 在周期 t 时, 代理收到的资源请求总数与在周期 t 之前获取资源的代理总数之和就是期望获取某资源的代理总数 N :

$$N = R_t + \sum_{i=1}^{t-1} C_i \quad (9)$$

其中, t 可以取值 $1, 2, \dots, t_{\max}$ (t_{\max} 是所有期望获取资源的代理都完成下载的周期数)。

3.3 共享可能度的本地计算策略

将式(6)代入式(9)可以得到:

$$N = R_t + \frac{(su+1)^{t-1} - 1}{s}$$

将 N 代入式(7)可以得到:

$$P_t = 1 - \frac{(su+1)^t - 1}{sR_t + (su+1)^{t-1} - 1} \quad (10)$$

在式(10)中, 任一代理能够知道当前时刻收到的资源请求数量 R_t , t 是资源从最初引入网络后的生存时间, 以下载周期为单位计量。以当前时间减去资源生成时间就能算出 t 。因此, 将 R_t, t, u, s 代入式(10), 任一代理就能在本地计算出某资源的共享可能度 P_t 。

4 实验结果

在实验网络中模拟了请求共享资源代理总数 $N=100$ 、代理上传能力 u_i 在 1~10 之间、网络中代理共享比率 s 在 0.1~1 之间的多种情形。为模拟真实环境下资源大小规模的差别, 下载时间周期设定在 10 s~100 s 之间随机选择。一个代理的共享缓存中可保存 3 个共享资源, 如果代理共享缓存中资源已满后还要下载其他资源, 必须从缓存中移除已有资源。实验中选取 2 种资源价值决定策略进行比较, 一种是随机移除某一资源, 另一种采用了本文基于共享可能度的自适应策略计算资源共享可能度, 移除资源共享可能度最低的资源。

图 1 是代理上传能力从 1 增加到 10 时资源平均下载周期变化的实验结果。当代理上传能力增加, 代理能够同时满足更多下载要求。 u_i 变大时, 下载一个资源需要的时间周期减少。从图中可以看出, 采用自适应共享策略使资源平均下载

周期相对对照组减少了 10%~30%, 资源的平均下载周期减少说明资源冗余下载和无效下载出现频率降低, 表明自适应共享策略能有效地减少资源冗余下载。

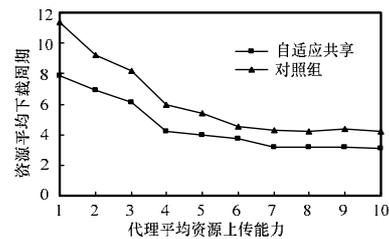


图 1 代理上传能力对资源下载的影响

图 2 是网络中代理共享比率 s 增加时资源平均下载周期变化的实验结果。当 s 变小(更多代理不愿共享资源)时, 整个网络中提供资源共享的代理数目减少, 资源更难被下载, 资源平均下载周期变大。当 s 减小到 0.1 时, 已不能保证资源下载请求能够完成, 因此, 很少有代理能获得请求资源, 资源平均下载周期急剧增加。这说明 P2P 网络中自私代理数目太多会严重影响整个网络效率。从自适应策略和对照组结果比较可以看出, 由于自适应策略能够减少资源冗余无效下载, 在较低代理共享比率时, 下载平均周期仍比对照组少, 即在网络中自私代理相对较多情形下, 自适应策略能更好保证资源共享下载效率。

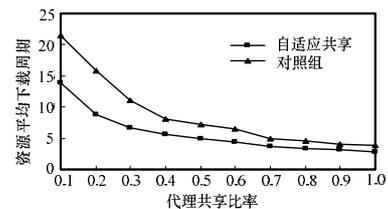


图 2 代理共享比率对资源下载的影响

实验结果证实, 自适应策略能够很好地避免资源的冗余下载, 提高 P2P 网络资源共享效率。

5 结束语

在分散 P2P 网络中, 大型共享资源的有效利用是 P2P 网络效率的关键。本文提出的自适应共享策略对于目前 P2P 网络中大型资源共享面临的问题提出了较好的解决方法, 单个代理能够根据自身获取的信息, 按照自适应策略计算出当前时刻资源的共享可能度, 为代理决定共享资源价值提供依据, 并能适时地根据网络当前时刻资源需求变化对共享资源价值进行调整, 保证共享资源的有效利用。

参考文献

- [1] Gummadi K P, Dunn R, Saroiu S, et al. Measurement, Modeling, and Analysis of a Peer-to-Peer File-sharing Workload[C]//Proc. of the 19th ACM Symposium on Operating Systems Principles. [S. l.]: ACM Press, 2003: 314-329.
- [2] Costa L, Amorim M, Fdida S. Reducing Latency and Overhead of Route Repair with Controlled Flooding[J]. Wireless Networks, 2004, 10(4): 347-358.
- [3] 基于 Agent 与分布式缓存的 P2P 资源发现策略[J]. 计算机工程, 2008, 34(21): 175-177.
- [4] Golle P, Leyton-Brown K, Miranov I. Incentives for Sharing in P2P Networks[C]//Proc. of the 3rd ACM Conference on Electronic Commerce. [S. l.]: ACM Press, 2001: 264-267.

编辑 张正兴