

对等网络中的搭便车行为分析与抑制机制综述

余一娇 金海

(华中科技大学服务计算技术与系统教育部重点实验室 武汉 430074)
(华中科技大学集群与网格计算湖北省重点实验室 武汉 430074)

摘 要 搭便车(free riding)行为在对等网络(Peer-to-Peer network)中日趋严重,它对网络健壮性、可用性和生命周期等造成了负面影响.设计、应用合理的搭便车行为抑制机制是当前对等网络研究的一个重要方向.文中首先描述搭便车行为的测量方法和发展趋势,然后分析搭便车行为的基本原理以及对对等网络性能的影响.依据数学建模和分析工具的差异,将已有搭便车行为抑制机制分为激励机制、博弈论方法、社会网络与经济模型三大类.文中归纳了三类抑制机制的建模方法,对其优、缺点进行了分析和比较,最后对搭便车行为抑制机制的未来研究方向进行了展望.

关键词 对等网络;搭便车;激励机制;博弈论;效用函数
中图法分类号 TP393

A Survey on Overcoming Free Riding in Peer-to-Peer Networks

YU Yi-Jiao JIN Hai

(Key Laboratory of Services Computing Technology and System of Ministry of Education,
Huazhong University of Science and Technology, Wuhan 430074)

(Key Laboratory of Cluster and Grid Computing of Hubei Province, Huazhong University of Science and Technology, Wuhan 430074)

Abstract Since the first report of free riding on Gnutella in 2000, large amounts of network measurement results show that free riding is prevalent in almost all Peer-to-Peer (P2P) networks, which reduces the robustness, availability and lifetime of P2P networks. Overcoming free riding becomes a hot research issue both in academic and industrial communities. The traffic measurement mechanisms in P2P networks are overviewed, which include active measurement and passive measurement. Some detailed traffic measurement results of Gnutella overlay network are illustrated, which indicate the development of free riding in P2P network in recent 6 years. The modeling of P2P network is introduced, and the impacts to system performance of free riding are discussed. Moreover, the authors focus on the mechanisms avoiding free riding and categorize the proposed approaches into incentive mechanisms, game theoretical approaches, and social network or economic models based methods. Each category is discussed in detail, and some typical work is introduced. Especially, the features of these three categories of solutions are analyzed and compared from the aspect of engineering implementation, which is helpful to both the P2P software developers and academic researchers. Finally, future works about the mechanisms discouraging free riding in P2P networks are outlined.

Keywords P2P network; free riding; incentive mechanism; game theory; utility function

收稿日期:2006-05-11;最终修改稿收到日期:2007-07-17. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2003CB317003)资助. 余一娇,男,1978年生,博士研究生,讲师,主要研究兴趣为计算机网络和计算机系统结构. E-mail: yjyu@mail.ccnu.edu.cn. 金海,男,1966年生,博士,教授,博士生导师,主要研究领域为计算机系统结构、网格计算、集群计算、并行与分布式计算、计算机网络、P2P计算、网络安全、语义网、网络存储、普适计算等.

1 引言

Napster 软件的出现给计算机网络技术领域带来了重要变革. 系统中每个节点既是服务器又是客户机的通信模式, 是对传统非对称客户机/服务器通信模式的重大改变. 尽量利用终端节点的计算、存储和通信资源的 P2P 通信模式已成为当前互联网应用的重要模式之一. 一些基于对等网络技术的文件共享系统, 如 Gnutella, e-Donkey, BitTorrent 等, 在全球范围内的在线用户数量有时达数百万之多. 对等网络技术自诞生伊始就一直本着“人人为我、我为人人”的信息共享与服务理念. 然而对等网络中的每个节点是否真正贯彻了对等提供服务、享受服务的信息共享理念, 有待通过对等网络流量测量和统计分析来证实.

2000 年 8 月 Adar 和 Huberman 对用户数量超过百万的大型对等网络 Gnutella 一天的流量进行了测量, 并作出了统计分析^[1]. 文献[1]首次指出 Gnutella 网络中各个节点无论在提供信息共享和网络维护方面都存在极大差异. 多数节点没有共享文件, 或共享了为数极少的文件, 或是共享几乎不被人访问的文件. 许多节点加入对等网络的初衷只是想从对等网络中获取其它节点提供的服务, 而不乐意为对等网络做贡献. 文献[1]把对等网络节点的只享受信息资源服务而不为系统作贡献的行为称为搭便车; 存在搭便车行为的节点称为搭便车者. 至今对等网络研究对搭便车行为尚缺乏形式化的定义, 但有如下共识: 搭便车是指对等网络中的节点仅从系统中获取其它节点提供的服务, 而不为对等网络作贡献的行为.

显然, 搭便车现象与 P2P 通信模式提倡的协作共享理念不一致, 因此文献[1]自发表之后立即引起了对等网络领域研究者的广泛关注. 至今, 搭便车现象测量验证与抑制机制研究大致经历了 3 个阶段. 在第一阶段, 人们试图测量不同对等网络中的搭便车现象, 分析其数字特征. 经过这一阶段的研究, 搭便车现象作为一种客观存在已成为研究者的共识, 并认为是对对等网络健康发展的一种挑战^[2]. 在第二阶段, 一些搭便车行为抑制机制被提出, 如激励机制、基于博弈论的方法等, 其中少数已应用到真实系统中. 在第三阶段, 研究者意识到还没有某种方法可以彻底解决对等网络中的搭便车行为, 因此应对抑制机制开展更深入地研究, 并且通过在真实系统

中实现抑制机制去验证方法的效率.

研究者和 P2P 应用软件开发人员面对已提出的数十种搭便车行为抑制机制, 关注的是以下问题: (1) 各类抑制机制的基本模型和分析方法; (2) 抑制机制对 P2P 应用系统的前提假设在真实网络环境中是否可满足; (3) 抑制方法的适用场景和控制效率; (4) 各类抑制机制在工程应用中面临的困难; (5) 未来就抑制机制应进一步研究哪些问题. 本文的目的是对已有抑制机制进行分类、介绍和评价, 明确未来对等网络中搭便车行为抑制机制研究的方向.

本文第 2 节介绍搭便车行为的测量方法和测量结果; 第 3 节对 P2P 文件系统和搭便车行为进行数学建模, 说明其基本原理, 分析搭便车行为对对等网络的影响; 第 4 节将已有搭便车行为抑制机制分为 3 类, 详细介绍其原理; 第 5 节讨论 3 类控制方法的特点以及存在的不足; 第 6 节归纳搭便车行为抑制机制应进一步研究的方向; 第 7 节对全文进行总结.

2 搭便车行为的测量

搭便车行为是通过长期观测真实对等网络流量而发现的一种客观现象. 为了全面理解搭便车行为及其影响, 我们首先介绍搭便车行为的测量方法和测量结果.

2.1 测量方法

对等网络中存在普遍的搭便车现象最初是通过测量 Gnutella 文件共享系统而发现的. 从某个具体的 P2P 应用系统测量得出的结论是否可以反映所有 P2P 系统的一般规律, 不同的测量方法得出的结论是否可靠等问题人们在不断地研究. 2000 年以来, 研究者用多种测量方法在不同对等网络环境中进行了长时间的测量和分析, 且不断深入. 例如, 在测量周期方面, 文献[1]仅是对 24 小时的 Gnutella 流量进行测量和分析; 而文献[3-6]中测量周期都长达一周以上, 部分实验测量时间长达几个月、流量数据记录条数高达几十亿条. 延长测量周期, 有利于排除网络突发事件对测量结果的干扰, 从而得出比较客观的结论.

对等网络中的流量测量技术可分为主动式测量和被动式测量两大类. 主动式测量方法被文献[1, 3, 6-8]所采用, 其基本方法是: 充分利用 Gnutella 等 P2P 软件的开源特性, 修改单个节点程序的源代码, 使其可以记录该节点所发出的查询请求报文、收到的查询反馈报文、转发报文等日志信息. 用于测量的

特殊节点,可以主动发出大量的资源查询请求,记录收到的响应信息.最后,对该测量节点保存的日志信息进行统计和分析.主动式测量方法灵活易实现,测量人员可以根据需要获得对等网络的数据包转发信息.然而,主动式测量存在的一个问题是:测量节点并不是正常的节点,其测量行为会干扰对等网络的正常行为,增加网络的额外开销,如数量众多的测量请求包等.通常主动式测量的规模较小,一般只对几万个节点的行为进行测量.但文献[7]中报道,在允许存在一定误差的前提下,其开发的快速测量节点可以在7分钟内测量节点数量超过百万的对等网络拓扑结构.

被动式测量方法被文献[4-5]采用,基本方法是收集物理网络边界路由器中的流量记录或转发报文记录.位于不同物理网络节点之间的通信,必定会经过边界路由器,因此在边界路由器上获取的流量数据可以客观地描述不同节点之间的通信特征.多数P2P应用系统都会在TCP或UDP报文头中加入本系统的特定标志字段(如GNUT(Gnutella),oxe3(eDonkey2000))和采用特定的端口号.根据特殊字段和特定的端口号可以判断边界路由器转发的数据包是否源于P2P应用.被动式测量的优点是:测量不产生额外的通信负载,不干扰正常的对等网络行为特征,且通常测量周期和测量范围都比较大.不足之处是:由于测量点局限于边界路由器,同一物理网络内部的对等网络流量没有被记录.另外有些P2P应用担心数据包会被防火墙过滤,采用了一些加密措施或将数据包包含在HTTP报文中传输,因此会出现漏测现象.从目前的研究报道来看,主动式测量技术和被动式测量方法均已比较成熟.

在被测量对象方面,文献[4]中明确提出了信令流和数据流的分类.信令流主要包括节点加入、退出、资源查询请求、查询反馈响应、路由维护报文等.数据流报文则是资源查询定位后,在服务请求节点与服务提供者之间的数据传输报文.例如:节点A发现节点B上有所需视频文件,节点A直接从节点B节点下载视频文件产生的对等网络流量称为数据流.通常,主动式测量主要是针对对等网络中的信令流数据包;而被动式测量对数据流报文的测量更为精确.在已开展的试验中,主动式测量技术采用得比较多,因此得到的结果关于信令流的分析比较充分,而数据流较少.

2.2 测量结果

为了描述对等网络中搭便车现象在近年来的变

化规律,以下简要介绍2000年至2006年期间部分对等网络流量测量的统计结果.表1示出了2000年和2005年不同研究者对Gnutella系统的测量结果对比.其中2000年的测量结果来源于文献[1],2005年的测量数据来源于文献[6].通过表1中数值对比可见,五年来搭便车现象在Gnutella系统中不仅无缓解,反而越来越严重.

表1 Gnutella中搭便车现象测量结果

| 时间 | 不共享文件节点所占比率/% | 共享文件数小于10的节点所占比率/% | 1%热心节点提供的查询响应所占比率/% |
|-------|---------------|--------------------|---------------------|
| 2000年 | 66 | 73 | 47 |
| 2005年 | 85 | 86 | 50 |

文献[3]中报告了2001年对Gnutella系统的网络流量测量结果.与文献[1]相比,本次测量周期更长,更关注对等网络的拓扑结构特征,更侧重对流量数据的定量分析.该文指出,在2001年期间Gnutella对等网络中节点保持连接数平均值为3.4,但少数热心节点保持的连接数达到数千,甚至几万;节点连接数分布近似呈现幂律(Power-law)分布规律(由于部分节点的连接数太低,使得度数的实际分布曲线和幂律分布略有差异).网络连接数代表节点在同一时刻为其它节点提供资源查询或下载服务的数量,节点保持连接数的巨大差异与文献[1]中报告的搭便车现象具有一致性.

文献[4]中报告了2001年对用户数量众多的P2P网络(Gnutella, Fast-Track和DirectConnect)的测量结果.该文认为3个不同的对等网络尽管在用户数量和数据流量特性上存有一些差异,但在节点保持的连接数分布方面均是极少数热心节点维持了大量的连接,搭便车现象在3个对等网络中广泛存在.特别是文献[4]中不仅测量了对等网络中的信令流数据,还测量分析了数据流,发现二者都存在严重的搭便车现象.

文献[8]中描述了2005年对Gnutella对等网络的测量和统计结果,依然是多数节点不愿意提供文件共享服务.

总结以上2000年至2006年期间关于对等网络流量测量结果,可得以下3个结论:(1)搭便车现象随着时间推移,变得日趋严重;(2)搭便车现象在所有P2P应用系统中均普遍存在;(3)搭便车行为是一种跨地域、跨不同物理网络的普遍现象,尽管在不同地域和子网中的严重程度存在差异.

3 搭便车行为的建模和分析

搭便车行为是对等网络节点用户具有自私心理作用下的一种结果. 如何通过数学建模, 分析搭便车行为的原理, 定量讨论它对对等网络性能的影响是研究有效搭便车行为抑制机制的前提.

3.1 搭便车行为的建模

对对等网络性能和用户行为进行数学建模的研究始于 2003 年, 其基本的工具是排队论和随机过程^[9-12]. 由于搭便车行为是对等网络中的一种普遍行为, 在关于对等网络性能建模和分析的研究中, 一般都会考虑节点类型的分类、搭便车者比率对网络

性能的影响.

文献[9]采用排队论对对等网络的服务能力进行数学建模, 并对搭便车行为对性能影响进行了分析. 对基于 P2P 技术的共享文件系统建模如图 1 所示^[9]. 其中思考期(think time)是指节点刚加入到对等网络, 处于既不发出查询请求, 也没有下载文件的阶段. 每个节点在加入到对等网络之初, 都会经过该阶段, 因此被认为是一个必经阶段. 当节点发出第一个查询请求后, 则将进入到在线(on-line)阶段, 它包括发出请求、下载文件等活动. 当节点获取到全部所需要文件资源后, 可能会选择离开对等网络, 从而进入离线阶段(off-line). 图 1 描述了节点在 3 个不同阶段的状态转换关系以及请求等待队列.

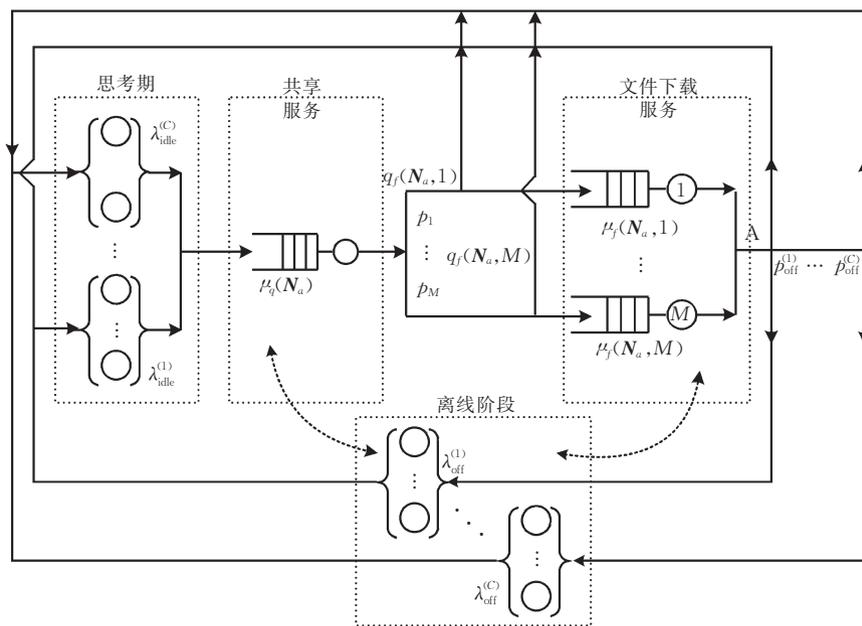


图 1 利用排队论对 P2P 文件共享系统建模

决定对等网络服务能力的关键参数包括对等网络提供服务队列数量、单个文件在网络中的备份数量、节点的类型数量等. \mathbf{N}_a 是一个向量, 表示所有在线节点; $N_a^{(C)}$ 是表示第 C 类型的节点的数量. μ_q 为对等网络提供资源搜索和下载服务的队列数量. 每一个提供信息资源服务的节点都将拥有一个提供服务的队列, 提供信息资源服务的节点越多, 则服务队列的数量就越大, 网络服务能力就越强. 文献[9]把对等网络中的节点分为非搭便车者和搭便车节点两类, 分别标注为类型 1 和 2, 整个对等网络中所拥有的服务队列计算方法如式(1)所示:

$$\mu_q(\mathbf{N}_a) = \mu_q(N_a^{(1)} + N_a^{(2)}) \quad (1)$$

根据本文第 2 节描述的对等网络实际测量结果, 多数搭便车节点完全不提供文件资源共享服务, 因

而它们不提供服务队列. 依据式(1)可得: $N_a^{(2)}$ 越大, 则 $\mu_q(\mathbf{N}_a)$ 就越小, 对等网络提供服务的能力越弱.

某个具体文件在对等网络中的备份数量(有多个节点提供同一文件下载服务)也是决定对等网络服务能力大小的关键指标. 若某文件在网络中仅有一个备份, 则所有搜索和下载请求都由该节点处理, 因而该服务队列较长, 排队等待时间长, 查询响应和下载速度慢, 网络服务性能差. 如果某个文件在对等网络中有多个备份, 如同时有多个下载请求, 对等网络可提供并行下载服务, 服务响应速度快. 不妨假设对等网络中存在 M 个不同文件, 并将所有文件按备份数量从大到小依次排序, 则第 i 个文件的服务能力定义为 $\mu_f(\mathbf{N}_a, i)$ ^[9].

$$\mu_f(\mathbf{N}_a, i) = \mu_f(N_a^{(1)}, i) = \frac{N_a^{(1)} \times H \times K}{i^{\circ}} \quad (2)$$

其中, H 表示单个节点对整个系统的服务能力; K 等于 $1/\sum_{j=1}^M 1/j^\alpha$; α 是表示文件分布的参数. 对一个确定的对等网络系统, 式(2)中的 H, K 和 α 均为常量. 因此由式(2)可得, 在保持对等网络系统中节点总数不变的前提下, 如果搭便车节点数量越多, 则 $N_a^{(1)}$ 就越小, μ_f 则较小, 网络服务能力较低. 如果 i 越大, 单个文件在对等网络中的备份数量越少, 则服务能力越低.

不同结构的 P2P 应用系统, 容忍搭便车行为的能力也有所区别. 通常对等网络可分为 3 类: CIA (具有集中索引服务器的对等网络, 如 Napster)、DIFA (分布式洪泛路由机制对等网络, 如 Gnutella)、DIHA (分布式哈希表索引对等网络, 如 CAN, CHORD). 文献[9]针对 3 类对等网络, 考察了搭便车节点在整个对等网络节点中所占的不同比率对网络吞吐率的影响. 图 2 示出了利用上述模型分析搭便车节点比率与网络性能之间的关系^[9].

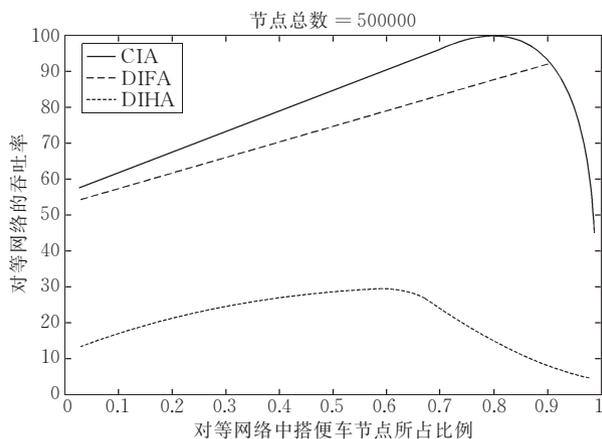


图 2 搭便车者比率对网络吞吐率的影响

当搭便车节点比率较小时, 网络中的节点多数是热心提供服务的节点, 网络信息资源丰富, 但下载请求少, 网络通信资源大量闲置. 如果此时有搭便车节点加入对等网络, 提出下载请求, 网络服务能力得到充分利用, 网络吞吐率因此线性提高.

当网络中搭便车节点超过一定比率后, 对等网络吞吐率随着搭便车节点增多而急剧下降. 其理由是: 搭便车节点持续增多, 服务请求数量不断增大, 而提供服务的节点没有增加, 甚至会降低. 网络服务队列数量降低, 单个服务队列的长度增加, 完成单个服务所需要的排队等待时间变长. 另外, 过多数据报在网络中传输, 可能导致网络拥塞和数据重传, 网络有效吞吐率降低.

3 种不同类型的对等网络支持搭便车节点的能

力也存在一定差异. 从图 2 中可见, 当搭便车节点占 60% 时, 分布式哈希表索引机制对等网络的吞吐率就开始降低. 而分布式洪泛机制和集中服务器式的对等网络则能容忍更高的搭便车节点比率. 当搭便车节点所占百分比高达 90%、80% 时网络吞吐率才开始降低. 从图 2 的分析可见, 对等网络具有较强的支持搭便车节点的能力. 但真实网络测量结果表明, 搭便车节点所占比率已超过 90%, 过多的搭便车节点已降低了对等网络的性能.

BitTorrent 采用了激励机制来抑制搭便车行为, 其方法是各个节点以较大优先权选择向其提供较大下载带宽的节点提供数据上传服务. 系统鼓励节点担任种子节点, 节点提供的上传服务越多, 能享受的下载服务优先级就越大. 文献[10]针对该搭便车抑制机制, 分析得出单个搭便车节点能获取的最大带宽, 如式(3)所示. 其中 N 是整个系统中节点的数量; n_u 是单个节点能提供的最大并发下载连接数, 在 BitTorrent 中设定为 4; μ 为假设所有节点具有的相同的上传带宽.

$$N \frac{1}{N - n_u} \frac{\mu}{n_u + 1} \approx \frac{\mu}{n_u + 1} \quad (3)$$

当 N 足够大时, 搭便车者能获取的最大下载带宽可近似地认为是评价下载带宽的 $1/(n_u + 1)$. 直接提高 n_u 的值, 即可降低搭便车者享受的服务, 从而迫使搭便车者不得不为对等网络做贡献. 然而提高 n_u 会导致节点需支持的并发 TCP 连接数增多, 对等网络易出现超时重传或拥塞现象, 网络有效吞吐率会下降.

值得指出的是, 由于对对等网络建模方法和假设的差异, 关于搭便车行为对整个对等网络性能影响的分析结果存在一些差异, 但基本的分析结论比较接近.

3.2 搭便车行为对对等网络的影响

广泛存在的搭便车行为会降低对等网络的性能, 增加系统的脆弱性; 如果搭便车现象进一步加剧可能会导致整个 P2P 应用系统最终崩溃. 以下具体分析搭便车现象给对等网络带来的诸多不良影响.

(1) 对等网络中在线节点越多, 热心节点所承担的资源查询与下载服务任务就越重, 热心节点可能因长期过载而宕机或主动退出网络. 很多热心节点是无结构对等网络中的关键路由节点, 一个热心节点宕机极大可能会导致整个对等网络的连通性发生巨大变化. 根据复杂网络理论, 网络服务过多地依赖少数热心节点, 则系统抵御外界协同攻击能力比

较低.

(2) 多数节点的搭便车行为会降低对等网络的生命周期. 文献[2]认为节点加入到某个特定的对等网络就是为了从中获得自己需要的信息资源(假设信息资源都是通过文件形式保存、查询或下载). 节点不断从对等网络中查询、收集自己需要的文件, 当节点收集到的文件数等于或者接近对等网络中所有共享文件总数时, 该节点会退出对等网络. 如果共享文件多集中在热心节点上, 则热心节点需要从对等网络中下载的文件数量较少, 退出系统的时间就越快. 一个热心节点的退出, 将导致整个对等网络中共享文件总数急剧下降, 从而致使更多的节点退出, 最终对等网络由于缺乏丰富的共享资源而快速崩溃.

(3) 如果搭便车现象过于严重, 对等网络将趋近客户机/服务器通信模式, P2P 通信模式具有的健壮性、可扩展性都将不复存在, 甚至其可用性比传统客户机/服务器通信模式还差. 因为传统的服务器在设计 and 部署过程中, 多在机器硬件配置、网络接入环境等方面做了特定处理, 而对等网络中的热心节点多数是由普通的个人计算机来承担, 在软、硬件环境配置方面不如传统的服务器可靠、健壮.

尽管搭便车现象对对等网络运营存在以上负面影响, 但至今多数对等网络软件开发者依然容忍搭便车现象存在, 以下 3 个因素值得考虑:

(1) 搭便车现象使对等网络抵御随机选择节点攻击的能力较强. 由于多数节点是搭便车节点, 随机选择节点则容易得到搭便车节点, 对大量搭便车节点进行攻击, 尽管被攻击节点服务终止或退出对等网络, 但该攻击对整个对等网络的拓扑结构和服务影响甚小.

(2) 从对等网络维护和管理角度来看, 搭便车现象使得对等网络中节点的重要性不再均衡. 多数搭便车节点由于为对等网络承担的任务比较少, 其加入或退出对整个网络的影响较小, 因此对等网络管理者可以重点管理数量不多的热心节点. 从工程的角度来看, 被管理节点数量越少, 越有利于精细管理.

(3) 允许一定程度的搭便车现象, 有利于吸引自私的 P2P 节点用户加入到对等网络, 从而提高对等网络的用户数量. 大量的用户对开展对等网络中的商业应用服务有较大帮助.

因此, 为了确保对等网络高效、安全、可靠运行, 有必要采用适当的措施对过于严重的搭便车行为进行抑制.

4 搭便车行为抑制机制

自 2003 年以来, 研究者提出了一系列的搭便车行为抑制机制. 考察已有对等网络中搭便车行为抑制机制研究成果, 发现它们有一个共同的指导原则: 根据节点已做出的贡献, 确定它能够享受服务的能力. 贡献越小的节点, 从网络中查询、下载信息资源的能力就越弱.

已提出的抑制机制在数学建模和分析工具、测量对象和测量点位置等方面存在较大差异. 为了清晰地描述已提出的搭便车行为抑制机制, 本文从数学建模和分析工具角度, 将已提出的抑制机制分为基于节点行为的激励机制、基于博弈论的方法、采用社会网络或经济模型的控制策略 3 大类. 本文先通述 3 类抑制机制设计与分析的基本思路, 然后对其中有代表性的方法举例详细描述.

4.1 激励机制

激励机制 (incentive mechanisms) 是最早提出的搭便车行为抑制方法, 也是应用最广泛的方法. 关于激励机制研究中的一些子问题, 如定义和计算节点的信誉度, 已成为 P2P 研究中的热点. 不同激励机制之间的差异主要体现在效用函数 (节点享受服务能力与节点为系统已做贡献的关系) 定义、测量点选择等方面. 图 3 给出了激励机制的基本算法流程. 关于效用函数的定义与合理性分析, 将在 4.1.1 中作深入讨论.

1. 提出一个效用函数;
2. 每个节点定期或事件触发地计算节点的效用函数;
3. 节点请求发起信息查询或下载服务;
4. 如果节点效用函数值较高, 则进行下载操作, 下载后转步 6;
5. 如果节点效用函数值低, 拒绝下载服务, 转步 7;
6. 节点重新计算效用函数;
7. 系统等待事件触发, 有服务请求则跳转到步 3; 如果节点为其它节点提供了服务, 则跳转到步 6.

图 3 激励机制的算法流程

测量点位置选择是激励机制设计的关键. 目前关于测量点的选择主要有两种方法: (1) 节点对自身行为做自我测量, 评价节点在网络中的贡献大小; (2) 每个节点对邻接节点进行测量和监督, 各个节点的贡献大小由邻接节点评价. 第一种测量是节点把自身每次提供服务或享受服务的行为记录下来, 工程上容易实现, 且开销比较小. 但它很难对付少数节点的恶意欺骗, 如虚报为其它节点提供信息服务的次数. 节点互相测量与监督的方式能有效地防止

少数节点的恶意欺骗,但分布式测量的系统开销比较大. 以下根据测量点位置差异,将激励机制分类讨论.

4.1.1 节点自我测量的激励机制

效用函数 (utility function) 是搭便车抑制机制研究中的一个关键概念^[2]. 效用函数可能涉及以下自变量: 节点共享文件的数量、节点已下载文件的数量、节点已上传文件数量、节点已下载数据的大小、节点已上传数据的大小等. 定义计算复杂性小,却能客观地反映搭便车控制中关键问题的效用函数是激励机制设计的核心. 以下通过介绍不同效用函数定义,分析激励机制的一般研究方法. 式(4)~(6)是文献[2]中定义的 3 个效用函数.

$$U(P_i, T) = \alpha \cdot |UList(P_i, T)| \quad (4)$$

式(4)左端的 $U(P_i, T)$ 表示在 T 时刻, 节点 i 的效用函数. 右端的 $UList(P_i, T)$ 表示在 T 时刻, 节点 i 所提供的共享文件数; α 是一个规范化系数, 是个常量. 采用式(4)作为效用函数, 节点能享受的服务质量正比于节点共享的文件数量. 效用函数(4)十分简单, 节点所共享文件数量决定节点能享受的服务质量.

$$U(P_i, T) = \beta \cdot \sum_{(\forall F_j \in UList(P_i, T))} size(F_j) \quad (5)$$

式(5)从文件大小角度来计算节点的效用函数. 式(5)先把一个节点提供的共享文件大小求和, 然后乘以规范化系数 β . 对比式(4)和式(5), 两者差异是在大文件和小文件的处理方面有区别. 采用式(4), 有利于共享多个小文件的节点享受高服务质量; 而式(5)则有利于共享大文件的节点. 式(4)和式(5)定义的效用函数均是静态效用函数, 没有反映节点所提供的文件被其它节点下载次数动态信息.

$$U(P_i, T) = \sum_{(\forall F_j \in UList(P_i, T))} \gamma(k + DCount(F_j, T)) \cdot size(F_j) - \sum_{\forall F_l \in DList(P_i, T)} size(F_l) \quad (6)$$

效用函数(6)既考虑了节点提供的共享文件的大小, 又考虑了提供的共享文件在对等网络中的受欢迎程度.

表示节点 i 在时刻 T 的奖励值, $\sum_{\forall F_l \in DList(P_i, T)} size(F_l)$

表示节点 i 在时刻 T 的惩罚值. 奖励值中包括节点为其它节点提供下载文件大小之和, 惩罚值是节点从网络中下载数据大小之和, 已下载信息量越多则惩罚值越大. 效用函数(6)可以有效区分表 1 中那些

提供不被访问信息资源的搭便车节点. 然而式(6)在增加合理性的同时也增加了计算复杂度, 它揭示了效用函数设计中的一个矛盾: 复杂度与合理性的冲突. 计算比较简单的效用函数, 如式(4)、式(5), 合理性较弱; 而合理性强的定义则计算复杂度大. 采用式(4)~(6)效用函数, 可以有效限制节点的搭便车行为. 在节点搭便车行为突出时, 可让节点既无法发送查询请求, 也不能下载信息资源. 搭便车者若不能享受信息资源共享服务, 它将离开对等网络, 对等网络在线用户数量降低. 然而对对等网络运营者而言, 在线用户数量大小是衡量系统价值的重要指标. 例如利用 P2P 系统从事广告业务, 它可以允许不提供共享数据的节点存在, 因为系统可以主动向搭便车节点发送广告信息. 在线节点数量越多, 则广告效果越好. 如果严格执行上述效用函数, 尽管有效抑制了节点的搭便车行为, 却不利于提高对等网络的用户数量.

为了解决上述矛盾, 免费赠品 (freebie) 在效用函数设计中应该给予重视^[2]. 免费赠品的基本方法是: 鼓励搭便车者长时间呆在对等网络中, 允许它不对信息共享做贡献, 但可以根据在线时间长短给予它一定的享受网络信息服务的权力. 免费赠品实现比较简单, 如把节点在线时间作为自变量加入到效用函数. 式(6)中的奖励值计算公式中包含变量 k , 它可以是节点在线时间的一个度量. 节点在线时间越长, k 则越大, 从而奖励值越大, 效用函数值越高.

上述 3 个效用函数具有一个共同特点: 节点能享受的服务能力与其所做贡献的绝对大小 (如提供的文件数量、上传数据大小) 直接相关. 如果对等网络中所有节点的软硬件配置和物理网络接入环境具有一致性, 则比较贡献的绝对大小是合理的. 然而真实的对等网络中节点配置具有极大异质性, 包括服务器、个人电脑、甚至手持电脑设备. 一个有较大服务能力的设备与一个服务能力较弱的设备比较所做贡献的绝对大小, 显然有失公允. 特别是有些节点由于物理接入方式局限, 决定了它的下载和上传很难达到平衡. 如采用 ADSL 方式 (通常上行 128 kbps, 下行 1.5 Mbps) 连接到互联网中的节点, 如果直接根据上下下载数据大小计算效用函数, 一个竭尽全力为其它节点提供服务的节点可能还是被认为是搭便车者, 从而它的下载请求可能遭到拒绝或降低服务优先级, 显然这对通过 ADSL 接入的节点是不公平的. 因此定义效用函数, 应充分考虑对等网络节点在软硬件配置和物理网络接入环境方面存在的客观

差异.

文献[13]认为评价一个节点为对等网络所做贡献不能单纯地按绝对贡献大小评价,继而提出了如式(7)所示的效用函数值大小比较方式,其中 $C_i(t)$ 表示节点所作绝对贡献值, d_i 表示节点的最大可支持物理带宽. 式(7)中的效用函数值不是简单的绝对贡献值,而是服务贡献值与节点所能提供的最大带宽之比值. 它对物理上贡献能力低、但尽力做了贡献的节点比较有利.

$$\frac{C_i(t)}{d_i} \geq \frac{C_j(t)}{d_j} \rightarrow U_i(\theta_i, x_i(t)) \geq U_j(\theta_j, x_j(t)) \quad (7)$$

多数基于激励机制的搭便车行为控制方法是在节点提出查询或下载请求时计算该节点的效用函数,然后确定服务质量. 但也有研究者认为在对等网络负载不大的情况下,应该尽量为所有节点服务. 因此他们把激励机制的控制点选择在信息服务提供节点,而不是请求发起节点^[13-14]. 提供信息资源下载的节点根据当前网络状态,将其判定为拥塞和不拥塞两种类型. 如果服务提供节点不拥塞,则尽力满足同一时刻所有节点提出的服务请求. 但当服务提供节点发生拥塞时,它根据查询或下载请求发起节点的效用函数值对服务请求进行优先级分类. 先满足效用函数值高的节点发出的服务请求,后满足效用函数值小的节点服务请求;特别拥塞的情况下,可能拒绝搭便车者的服务请求^[13-14].

文献[15]中定义的效用函数如式(8)所示:

$$u(p, t) = \int_{t'=0}^t b(p, r(t'), t') a_l(p, t') - \sum_{r \in R} c(p, r) \cdot part(p, r, t) \quad (8)$$

式(8)与式(4)~(6)有两个区别:(1)在时间域上采用连续方式,用积分计算节点 p 在时刻 t 的效用函数;(2)评价收益的角度不同. 式(4)~(6)是从整个 P2P 系统的角度来计算各个节点的效用函数,节点为其它节点提供了服务,则是正收益;节点从网络中下载了数据,则是负收益. 式(8)却是从单个节点角度来评价效用函数,其中积分部分表示节点从网络下载的数据量. 从对等网络中下载资源越多,则节点自身收益越大;求和部分表示节点为其它节点提供的下载服务总和,为系统贡献越大,则节点效用函数值越小.

根据式(8)的计算结果,文献[15]把节点分成理性主义者、搭便车者和贪婪者 3 类. 理性主义者期望效用函数值 $u(p, t)$ 较大;搭便车者则仅期望自身提

供的下载服务量最小;贪婪者则仅注意积分部分值较大,即节点从网络中获得下载数据总量最大. 文献认为无论是搭便车者还是贪婪者,都不值得提倡,对等网络提倡理性主义节点.

文献[16]中提出了基于惩罚(payment-based)的激励机制. 该文中讨论节点是否遵从对等网络协议、转发来自其它节点的报文以及节点可用性. 如果节点频繁加入、退出对等网络,则会降低网络的服务效率. 文中仿真结果表明,基于惩罚的激励机制不仅能限制节点的搭便车行为,并且能提高节点在对等网络中的稳定性和可用性. 与其它论文不同,文献[16]并没有直接提出关于节点的效用函数,而是从整个系统的角度出发,利用博弈论的方法进行惩罚度计算.

从节点用户加入到对等网络中的心理动机分析讨论对等网络搭便车抑制机制,也是激励机制研究中的一种方法. 文献[17]把节点用户的心理动机分为个人因素和人与人之间的因素. 其中个人因素包括内因(如利他主义以及信誉等)和外因(如奖励、个人需求等). 人与人之间的因素包括爱好等. 该文认为应根据心理因素建立基于节点对系统贡献的奖励机制、独立标识,建立子社团以及节点之间的相互推荐等激励抑制机制. 由于该文作者不是从事纯粹的 P2P 技术研究,他们没有提出具体的激励机制实现方法. 但从节点用户心理分析的角度来讨论节点为什么存在搭便车行为值得关注,因为对等网络中节点的搭便车行为最终源于节点使用者的心理动机.

4.1.2 分布式测量机制

为了有效防止恶意节点在效用函数计算过程中的欺骗行为,可以采用不同节点之间的监督和测量方式判定节点是否为搭便车者. 文献[18]提出了一种基于分布式测量机制的搭便车行为抑制机制. 对等网络中的各节点首先观察其邻接节点发送、转发信令数据包和数据流报文行为;然后根据日志记录,计算效用函数发现对等网络中的搭便车者;最后采用强制手段迫使搭便车主义者必须为系统做贡献,否则将从系统中清除出去.

采用分布式测量方式,并不需要各个节点专门做测量操作,而仅需对通过自身转发的报文作日志记录. 如节点收到 P2P 报文时,节点记录下该报文来自哪个邻接节点,将转发到哪个邻接节点,然后定期统计各个邻接节点是为系统做贡献的报文数量较多,还是享受系统服务的报文数量较多. 文献[18]中定义了 3 类搭便车者:(1)发出查询响应报文次数

极低的节点,即该节点一般没有提供共享信息资源,或仅共享了不被其它用户访问的资源;(2)从邻接节点获得的资源数量远大于它为其它节点提供服务的报文数;(3)转发信令报文十分少的节点,该节点不为整个对等网络承担路由维护义务。

文献[6,18]中对搭便车者提出了三种不同程度的惩罚方案。第一个层次是快速降低搭便车者发出的查询请求消息包的 TTL 值。在无结构对等网络中,报文转发次数依赖于报文的 TTL 值。一般报文转发一次,其 TTL 值减 1。若采用以上惩罚机制,则搭便车者发出的查询报文 TTL 值可能一次减 2。报文在对等网络中转发次数降低,传播范围变小,获得响应报文数量会降低。该惩罚方法主要是针对不太自私节点的一种温和性惩罚。第二个层次是忽略自私节点的查询请求。节点收到邻接节点报文后,先根据邻接节点的类别判断是否转发报文。若报文源于搭便车者,则不转发该查询请求。第二个层次的惩罚方法将导致搭便车者无法从系统获得资源,并且它并不知道系统已经对它自私行为进行了惩罚。第三个层次的惩罚方法是直接取消与搭便车者的连接。如果搭便车者与邻接节点的连接中断,该节点则被从对等网络中清除,不再对系统带来不利影响。

4.2 博弈论方法

博弈论是分析社会群体和经济活动中,个体和群体选择最优行为策略的有效数学工具。对等网络中的节点行为,亦可以采用博弈论来进行分析。利用博弈论方法建模,一个节点的行为选择不再是自身的一个简单选择,它与同时在线的其它节点的行为选择紧密相关。节点在对等网络中的行为可以建模为长期博弈过程,节点可在包括热心为系统做贡献、贪婪下载、贡献收益平衡等多种行为策略集合中找到有利于最大化自身收益的策略选择。采用博弈论方法分析节点行为策略选择,实际上是从整个对等网络的角度来分析一个节点行为策略选择与其它节点行为策略选择的关系。

文献[19]采用博弈论作为工具分析和抑制对等网络中的节点搭便车行为。该文认为节点在对等网络中的行为选择是一个时间离散、非终止、不断重复的博弈过程,记为 G^∞ 。 G^∞ 中的参与者(player)集合是对等网络中全体在线节点。为了模型分析的简便,文献[19]假设节点在任一时间段都发出服务请求,且每个节点行为(actions)集合仅包括提供服务(serve)与不提供服务(don't)两个元素。

每个节点的偏好(preferences)用支付函数

(payoff function)的期望值来表示,主要涉及收益 U 和服务成本 C 。 U 与节点从对等网络中获得的服务数量直接相关, C 是节点为系统所做贡献自身的成本代价,如计算资源、网络带宽消耗等。任一节点 i 在时刻 t 的支付函数计算方法与节点信誉度 R 密切相关,其中 R 的计算方法如式(9)所示。若 t 时刻节点 i 提供服务,则 ω 为 1,否则为 0。鉴于 α 是 0~1 之间的非负实数,如果节点 i 在时刻 t 不为网络提供服务,则节点的信誉度会呈现下降趋势。若节点 i 完全不提供服务,则节点的信誉度下降,最终趋近于 0。文献[19]中定义 R_0 为 0, R_1 为 ω 。

$$R_t^i = R_{t-1}^i(1 - \alpha) + \omega \cdot \alpha, 0 \leq \alpha \leq 1, t \geq 2 \quad (9)$$

根据参与者在博弈过程能否获得其它参与者的行为选择信息,博弈分为完全信息博弈和不完全信息博弈两类。假设节点的行为选择信息通过洪泛机制快速在对等网络中传播,即一个节点选择某种行为策略后,立即将此选择信息广播到与之相邻的节点;然后邻接节点把该行为选择信息向其邻接节点递归传播。只要传播速度足够快,该博弈模型就是完全信息博弈^[19]。

不妨用 a^* 表示一次博弈过程中全体节点的行为组合,其中 a_i^* 表示节点 i 的行为选择。Nash 均衡是指具有以下特性的 a^* , 当且仅当其它任一节点 j 都采用 a_j^* 时,节点 i 必须选择行为策略 a_i^* 才能获得对自身有利的最大收益。也就是任何节点如果不选择 Nash 均衡中的策略,其所获收益均小于选择 Nash 均衡中的策略。基于博弈论的搭便车行为抑制机制在分析过程中的关键和难点是分析模型的 Nash 均衡。如果模型能计算出 Nash 均衡,则节点可以获知对自己最有利的行为选择策略。

求解纯策略均衡或混合式策略均衡是利用博弈论研究搭便车抑制机制的关键。纯策略是指对等网络中任一节点均采用确定性的行为策略,如提供服务与不提供服务。混合式策略均衡,则是一种非确定性的策略选择方法,它给出的结论是关于行为选择策略的一组概率组合。

以下讨论两个特殊的确定性行为组合实例。

① $a_i^* = \{\text{不提供服务}\}$; ② $a_i^* = \{\text{提供服务}\}$, 其中 i 为任一节点。当其它节点都选择不提供服务时,节点 i 如果选择提供服务,则需要消耗本节点的资源。既然其它节点都不提供服务,节点 i 就不可能从对等网络中获得收益。在付出没有回报的状态下,不如不付出,节点 i 选择不服务达到了最优的目标。又因 i

可代表对等网络中的任一节点,因此 $a_i^* = \{\text{不提供服务}\}$ 满足 Nash 均衡的定义,是一组 Nash 均衡.然而该 Nash 均衡对对等网络应用不具有指导意义.如果对等网络中所有节点都不提供服务,对等网络也就崩溃了.这与抑制搭便车行为是为了让对等网络更健康运行的目标相反.由此可见,并不是任得一组 Nash 均衡对克服搭便车现象都具有实际意义.

对任一节点 i , $a_i^* = \{\text{提供服务}\}$, a^* 不是 Nash 均衡.理由是当其它节点均提供服务的条件下,节点 i 选择不服务,它可从对等网络获得服务收益而自身无消耗,因此选择不提供服务的收益大于提供服务.

不妨对节点行为作两个假设:(1)在一个时间单元中,节点只进行一次行为选择;(2)同一对等网络中的节点行为选择具有相似性^[19].对称 Nash 均衡是行为组合 a^* 中任意两个节点在一次博弈过程中,它们选择一致的行为策略.由第二个假设,以下采用对称 Nash 均衡分析.既然是混合式策略均衡,那么不妨记节点在单位时间 t 内,选择提供服务的概率为 p ($0 < p < 1$),则选择不提供服务的概率为 $1 - p$.求解一组混合式策略 Nash 均衡,就是求解特定的概率 p .

在第 t 个时间单元内,节点 i 的支付函数计算方法如式(10),(11)所示^[19].若节点选择提供服务则用式(10)计算,否则用式(11).

$$Payoff_{\text{serve}} = p(-C + R_t^{\text{serve}} \cdot U) \quad (10)$$

$$Payoff_{\text{don't}} = (1 - p)(R_t^{\text{don't}} \cdot U) \quad (11)$$

将 ω 取值定义代入式(9),可得式(12),(13).

$$R_t^{\text{serve}} = R_{t-1}(1 - \alpha) + \alpha \quad (12)$$

$$R_t^{\text{don't}} = R_{t-1}(1 - \alpha) \quad (13)$$

根据混合式策略均衡的性质,无论节点在时刻 t 选择提供服务或者不提供服务,其支付函数应该相等,即

$$Payoff_{\text{serve}} = Payoff_{\text{don't}} \quad (14)$$

将式(10)~(13)分别代到式(14),求解可得 p , 如式(15)所示.

$$p = \frac{R_{t-1}U(1 - \alpha)}{-C + 2R_{t-1}U(1 - \alpha) + U\alpha} \quad (15)$$

式(15)中的 C, U, α 均是常量, R_{t-1} 是一个变量,因此 p 是随时间变化的变量.根据 Nash 均衡的定义,节点的行为策略是概率组合 $(p, 1 - p)$,以概率 p 提供服务,以概率 $1 - p$ 拒绝提供服务,其中 p 取值根据式(15)计算.

文献[20]主要是研究存在集中式服务器的对等网络中的搭便车行为抑制.该文假定 P2P 系统中存

在一个知道系统中所有节点行为记录的集中式服务器,其计算过程需要对等网络的全局知识.文献[21]也是采用博弈论的方法来抑制搭便车现象,但它采用区分服务的方法来实现.由于其建模过程更为复杂,在此不详细介绍.

4.3 社会网络和经济模型

对等网络中的节点存在的自私行为,不过是节点用户的个人心理因素的一种延伸.关于个人在社会行为中的自私性趋向,经济学、公共策略管理等领域已进行过深入研究^[22].文献[1]发表后,一些来自经济学、公共策略管理研究领域的学者发表论文讨论利用社会网络模型、经济模型、价格机制等方法来抑制对等网络中的搭便车现象.来自非计算机领域研究者的分析,得到了一些不同的结论,如文献[23]指出:对对等网络中的搭便车现象不必太担心,适度的搭便车现象是提高对等网络资源利用效率的一种有效方法.来自非计算机领域的一些研究结论,对在抑制搭便车现象与争取更多用户这一两难问题之间如何达到均衡具有启示意义.

文献[23]中采用社会网络的方法来研究对等网络中的搭便车现象.该文与其它文献的分析结论有一些差异,如它认为只要搭便车现象不是十分严重,搭便车行为是提高对等网络服务效率的一种有效手段,不应该抑制.社会网络模型是过去在社会公共政策管理领域研究应用比较广泛的理论模型,该文提出将此模型应用到对等网络中搭便车行为控制.但该文没有提出具体的搭便车行为抑制机制.

文献[24]介绍了 MMAPPs 系统中利用市场管理技术去促进节点之间的协作,降低搭便车现象的方法.图 4 是该文中提出的 P2P 网络中节点之间的关系.其中 Consumer 是网络服务消费节点,Provider 是提供服务的节点.在服务提供者和服务消费者之间,引入第三方节点(Other Peers)来监视服务发

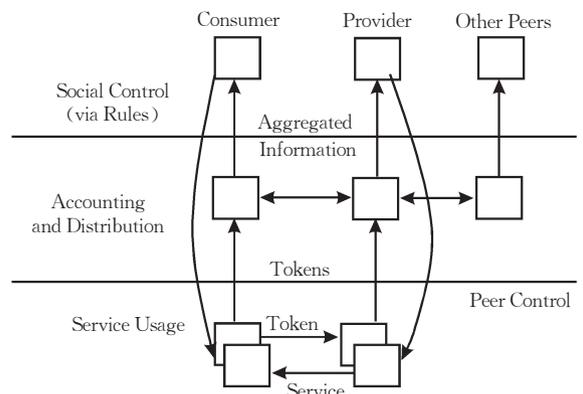


图 4 MMAPPs 的软件结构图

生过程.图4中包括3个分离的软件功能层,服务提供者和消费者通过Service Usage层实现服务.Accounting and Distribution层对节点的行为进行审计,依据这些信息可以实现抑制机制.Social Control层制定、保存对等网络中的消费和资源共享的规则.该文没有提出具体的控制方法,只对图4所示软件结构的工程实现作了介绍.

文献[25]中提出利用市场机制模型来控制虚拟组织中的搭便车现象.该文认为在对等网络等分布式应用领域,由于缺乏一个有效的中心控制者,不可能采用集中式的抑制机制,因此认为采用分布式的协同控制方法是一种必然.文中集中讨论了利用市场机制模型来抑制搭便车行为的工程实现方法,提出了一种基于代理的资源服务管理框架和非集中式联邦机制的资源管理策略.

文献[26]中提出了一种基于竞价机制的搭便车行为抑制机制.与其它论文相比,该文更关注节点对对等网络报文转发所做贡献的大小.其它论文讨论资源查询以及路由算法过程中,总是假设对等网络中的路由节点是无私的,路径中每个节点都会自觉为其它节点提供报文转发服务.该文认为既然节点在是否提供信息服务方面存在搭便车行为,这类自私节点也不可能很热心地为其它节点充当路由节点.因此,该文提出了关于转发报文的请求函数.节点如果为其它节点提供报文转发服务,则可以获得一定经济补偿;相反当节点需要其它节点为自身提供服务时,必须得用过去获得的经济补偿支付给为其服务的中间节点,从而迫使每个节点必须为其它节点服务.

此外,文献[27-31]中都提出了采用社会经济生活中审计、竞价等机制抑制搭便车行为.

5 抑制机制的评价

从仿真结果和理论分析结论来看,采用抑制机制后对等网络的可用性、生命周期、健壮性等都取得了显著提高,但尚未发现有论文将不同抑制机制在控制效率方面作比较.为了明确3类方法各自的特征和面临的挑战,有必要对3类方法的特性进行分析.

5.1 激励机制

基于激励的搭便车行为抑制机制在工程实现方面存在诸多优点,如效用函数根据单个节点行为的历史记录进行计算,不需要整个系统的全局信息,工

程实现比较简单等.但自我测量的激励机制面临以下挑战:

(1)效用函数严格程度的选择.过于严格的激励机制会导致搭便车者退出对等网络,从而用户数量急剧减少.特别对商用P2P系统,用户数量是影响整个系统存在的关键因素,太少的用户将导致商业运营失败.而不采用严格的激励机制,可能又难以达到抑制搭便车行为的初衷.在如何抑制搭便车行为与鼓励用户加入对等网络之间存在两难.

(2)文献[32]指出很多节点位于防火墙或地址转换服务器之内.防火墙可过滤外部网络对内部网络的服务请求;地址转换服务器内的节点没有外部IP地址,因此物理网络条件的局限导致它们不能为外网中的节点提供信息访问服务.2004年eDonkey对等网络中有25%~36%的节点由于防火墙和地址转换服务器的原因,不能为其它节点提供信息资源下载和路由服务^[32].如按照那些已提出的激励机制,这些节点将被认定为搭便车者,从网络中清除.失去如此多的节点,对对等网络自身发展是巨大打击.

分布式测量与监视机制可有效地防范那些不遵守协议的欺骗节点.然而在实际应用中,它面临以下两个挑战:

(1)节点之间互相监视开销大.例如:节点A与10个节点相邻,则节点A被其它10个节点监视.尽管来自10个邻接节点的数据可能存在一定差异,但多数关于A是否为搭便车者的判定结论应该是一致的.对一个节点有多个判定过程,这种重复判断是资源浪费.

(2)对连接数很高的节点,监视邻接节点会增加节点负担.连接数多的节点负载本来就比较重,还要监视众多的邻居节点,容易导致节点因负载过重而宕机.关键节点退出对等网络,将导致对等网络崩溃,这违背了抑制搭便车行为的初衷.

5.2 博弈论方法

采用博弈论作为数学工具分析搭便车行为抑制机制的效率是所有方法中定量化分析程度最高的一种.然而在实际P2P系统开发过程中,利用博弈论的方法面临以下问题:

(1)目前提出的抑制机制,一般都需要整个对等网络系统的全局知识.现有论文的仿真实验是在节点数量很少的情况下进行,可以满足该要求.然而真实的P2P系统节点数超过百万,且节点加入、退出网络行为频繁,该假设难以满足.

(2) 现有的基于博弈论分析方法中, 一些假设太严格, 与真实对等网络环境差异很大, 因此得到的 Nash 均衡未必具有真正的指导意义. 例如: 文献[19]中假设所有节点具有同质性. 而网络测量已经证明对等网络中是少数热心节点和多数搭便车者并存, 二者行为选择策略不具有一致性. 文献[19]中还提出利用洪泛法广播节点的行为选择信息. 当对等网络中的节点数量达到数万且节点地域分布广泛时, 完全信息博弈模型就不适合. 采用洪泛机制广播节点行为选择信息在大规模对等网络中不适合, 其理由如下: (a) 一旦节点数量巨大, 即使采用洪泛机制, 将一个信息广播到全网络的延时也会很大; (b) 每个节点各个时间段内的行为选择结果均在网络中广播, 则抑制机制的额外开销太大; (c) 各个节点需要容量巨大的存储空间去记忆其它节点的行为选择信息, 效率低下. 利用不完全信息博弈模型分析搭便车抑制机制应更适合对等网络应用环境, 但分析过程比较复杂.

5.3 社会网络和经济模型

利用社会网络和经济模型抑制搭便车行为, 一般都会借鉴其它学科领域比较成熟的数学模型. 该类模型种类繁多, 且已经有一些应用, 因此能提供解决问题的新思路^[33]. 鉴于经济学、公共政策管理等领域对节点用户的心理特征研究比较透彻, 在搭便车行为抑制机制研究中应借鉴该领域过去的一些结论.

然而该类方法大多假设对等网络软件具有对节点的行为审计和微支付(Micropayment)功能(能以极小的粒度, 如十分之一美分, 进行计费)^[33]. 因为多数可信的审计功能实现, 需要有一个集中式的审计中心. 集中式的审计中心不仅在可扩展性方面制约对等网络的发展, 且在信誉方面也难以获得所有节点的认同. 总之, 在节点行为动态性极强的大规模对等网络中, 该基本假设较难以满足. 因此这类方法尽管提出了一些模型, 但定量化分析和工程实现方面研究较少.

6 进一步的研究方向

如何抑制对等网络中的节点搭便车行为, 未来的研究将会更长期深入. 纵观已有研究成果, 未来可在以下方面作进一步研究.

(1) 在激励机制研究中, 如何确定新加入对等网络的节点所能享受的服务质量. 新节点因为在线

时间短, 如按式(6)所示效用函数来计算其能获取服务的质量, 则服务质量较低. 过低的服务质量, 不利于吸引新用户长期参与该对等网络. 相反, 如果新节点享受的服务质量很高, 则又面临着部分恶意节点的 whitewashing 行为.

Whitewashing 是指搭便车者在网络中受到惩罚后, 立即退出系统, 然后以新用户名或匿名再次登录到对等网络^[34]. 对等网络中的用户名通常是免费申请, 一个恶意用户可以申请多个用户名. 对新登录节点, 对等网络很难确认该节点是否是搭便车者. 对等网络一旦确定一个节点是搭便车者, 会立即限制其服务质量. 搭便车者一旦发现自身的服务受限制, 便立即退出, 然后用其它用户名加入对等网络, 则总能逃避惩罚. 总之, 如果不对新登录者提供一定行为限制, 则搭便车者总能贪婪地享受系统资源服务; 如果严格限制新登录节点的下载能力, 则会伤害真正的新登录节点. 因此, 如何在对等网络中判定搭便车者的 whitewashing 行为是抑制机制研究的关键^[34].

(2) 在研究抑制机制时, 对对等网络的建模和假设应更接近真实的对等网络环境. 目前的研究对对等网络的建模比较简单, 不足以描述对等网络的真实特性, 在过于简单的模型上分析得出的结论可能与真实的应用结果存在较大偏差. 如文献[19]中基于博弈论的抑制方法, 假设全网内节点的行为具有高度相似性, 对所有节点同一对待. 但事实上对等网络中存在自私节点和热心节点, 它们的行为特征具有显著差异, 而这些客观存在的特性在现有的建模中没有反映. 文献[11-12]在建模对等网络文件共享系统, 分析其可扩展性时, 假设所有节点对同一个文件都具有下载兴趣. 建模计算得出一个文件在系统中以 $O(\log_2(n))$ 时间复杂度在对等网络中快速传播, 其中 n 为对等网络中在线节点数量. 显然, 在真实对等网络环境中几乎不可能每个节点都对一个文件感兴趣, 它们既不会去下载文件, 也不会去提供服务, 因此真实的文件传播速度会低于该分析结果.

在对节点行为特征和对等网络建模时应深入考虑节点行为的动态特征、节点机器的硬件配置异质性、网络接入环境差异、节点用户心理动机多样性. 将热心节点(可能是真正的心用户, 也可能是对等网络服务运营商特意部署的特殊节点)与普通节点区分考虑等. 至少, 热心节点的存在应该在未来的控制模型和仿真实验中得到体现. BitTorrent 是较早采用种子节点方式来抑制节点用户自私行为的 P2P

软件,且在全球范围内的用户数量巨大.文献[10, 36]针对 BitTorrent 中的搭便车抑制机制和抑制效果进行了定量分析.文献[37]对在国内有广泛应用的 P2P 文件共享系统 Maze 中的搭便车现象也进行了深入的分析和统计.这类源于真实系统中的搭便车现象研究,相对单一的理论建模分析和仿真实验具有更大的指导价值.

(3) 抑制机制的效率应力争在真实对等网络中测试,并作定量分析.目前的分析主要依赖于实验仿真,但通常仿真实验的网络规模小,不超过 2000 个节点^[2].成功的对等网络应用,其同时在线用户的数量远大于仿真节点数.由于节点数量太少,仿真得出的结果不能充分说服 P2P 软件开发人员采用新的搭便车行为抑制方法.未来应尽快把已提出抑制机制在用户数量较多的 P2P 应用系统上测试.一旦某种抑制机制在用户规模达数万的大型对等网络中应用,并具有实际的控制结果,则对未来的研究将会有很大的启示作用.

由于对等网络中搭便车现象固有的复杂性,其数学建模难度大.复杂的建模有利于描述真实的对等网络,但定量分析却相对困难.过于简单的数学建模,尽管能得出一些定量的分析结论,但其工程指导性欠佳.文献[38]指出:过于复杂的抑制模型其实际所能取得的效果与比较简单的模型很接近.复杂的抑制方法其无论实现过程和计算复杂度都大于简单的模型,因此在工程应用中采用相对简单的建模更容易得到定量分析的结论.如何在算法的效率和可行性方面找到一种合理的平衡,是抑制机制设计和实现的关键.

(4) 应在搭便车行为抑制与鼓励信息共享之间找到合理的平衡点.搭便车现象是客观存在,并且一批 P2P 应用软件的设计者了解该现象将对对等网络可用性、健壮性等产生不良后果,但以 Gnutella 为代表的 P2P 应用软件系统在不断升级过程中仍未采用搭便车行为抑制机制^[6].一些已提出的基于激励的抑制机制在工程实现上并不困难,而 P2P 软件开发人员却不采用,可能是担心抑制搭便车现象会降低系统的用户数量.只有通过实际的运营比较,证明适当的抑制搭便车现象对整个 P2P 系统运营是利大于弊,才能让 P2P 软件开发人员采用搭便车抑制机制.如图 2 所示,适度的搭便车现象不仅不会降低对等网络的吞吐率,反而可以提高其吞吐率,因此惩罚搭便车者方面不宜采用过于严厉的方式.

(5) 节点负载的有效性与整个系统的公平性之

间的两难问题值得进一步讨论.从系统的观点出发,要让各个节点承担的义务保持公平,那么各个节点承担的工作负载越接近越好.然而,节点之间无论是在物理网络的接入还是在软件配置,节点自身的工作负载等方面都存在差异^[39].如何在异质的节点群体的行为控制间保持相对公平,值得深入研究.如何定义合理的效用函数,需要根据具体的对等网络应用的特征来判断.

(6) 如何利用对等网络拓扑结构的灵活性和可重构性,抑制搭便车行为值得深入研究.无论是物理网络的拓扑结构,还是对等网络的虚拟拓扑结构,都对网络的可用性、效率、抗协同攻击能力有直接影响^[3,40].对等网络的拓扑结构具有灵活的可重构性.文献[41]已在节点的接入过程阶段对此进行了考虑,以后还可以通过重构对等网络拓扑结构,强行转移热心节点上的信息资源到搭便车者上,抑制节点的搭便车行为.

7 结束语

本文系统介绍了对等网络中存在的搭便车现象,并讨论多种 2003 年以来提出的搭便车行为抑制机制.文中根据不同的数学建模和分析方法将已有控制方法分成 3 类,对各类抑制机制的基本方法、复杂性、控制效率和不足之处都进行了定性分析和比较.通过这些分析和比较,可以更清晰地了解搭便车行为研究的一般方法,有利于工程开发中选择合适的抑制机制.

本文是一篇全面地介绍对等网络中搭便车现象的测量方法、测量结果、对整个系统性能的影响的论文.特别是文中对搭便车行为抑制机制进行了系统的分类和比较.本文对了解对等网络中的搭便车现象及其抑制机制的研究背景、研究方法和进一步研究的问题,具有较大帮助.今后我们将对具体的抑制机制进行深入研究,提出工程实践性强且对用户透明的搭便车行为抑制机制.

参 考 文 献

- [1] Adar E, Huberman B. Free riding on Gnutella. *First Monday*, 2000, 5(10): 32-35
- [2] Ramaswamy L, Liu L. Free riding: A new challenge to Peer-to-Peer file sharing systems//*Proceedings of the 36th Hawaii International Conference on System Sciences*. Hawaii, 2003: 220-229

- [3] Ripeanu M, Iamnitchi A, Foster I. Mapping the Gnutella network. *IEEE Internet Computing*, 2002, 6(1): 50-57
- [4] Sen S, Wang J. Analyzing Peer-to-Peer traffic across large networks. *IEEE/ACM Transactions on Networking*, 2004, 12(2): 219-232
- [5] Karagiannis T, Broido A, Brownlee N, Claffy K C, Faloutsos M. Is P2P dying or just hiding?//*Proceedings of the Globecom 2004*. Dallas, 2004; 1532-1537
- [6] Hughes D, Coulson G, Walkerdine J. Free riding on Gnutella revisited: The bell tolls? *IEEE Distributed Systems Online*, 2005, 6(6): 1-18
- [7] Stutzbach D, Rejaie R. Capturing accurate snapshots of the Gnutella network//*Proceedings of the 8th Global Internet Symposium*. Miami, 2005; 2825-2830
- [8] Meng S C, Shi C, Han D Y, Zhu X, Yu Y. A statistical study of today's Gnutella//*Proceedings of the 8th Asia-Pacific Web Conference*. Harbin, 2006;189-200
- [9] Ge Z H, Figueiredo D R, Jaiswall S, Kurose J, Towsley D. Modeling Peer-Peer File sharing system//*Proceedings of the IEEE INFOCOM 2003*. San Francisco, 2003; 2188-2198
- [10] Qiu D Y, Srikant R. Modeling and performance analysis of BitTorrent-like Peer-to-Peer networks//*Proceedings of the SIGCOMM 2004*. Portland, 2004; 367-378
- [11] Veciana G d, Yang X Y. Fairness, incentives and performance in Peer-to-Peer networks//*Proceedings of the Forty-First Annual Allerton Conference on Communication, Control and Computing*. Monticello, 2003
- [12] Yang X Y, Veciana G d. Service capacity of Peer-to-Peer networks//*Proceedings of the IEEE INFOCOM 2004*. Hong Kong, 2004; 2242-2252
- [13] Ma R T B, Lee C M, Lui J C S, Yau D K Y. Incentive P2P networks: A protocol to encourage information sharing and contribution. *ACM SIGMETRICS Performance Evaluation Review*, 2003, 31(2): 23-25
- [14] Ma R T B, Lee S C M, Lui J C S, Yau D K Y. An incentive mechanism for P2P networks//*Proceedings of the 24th International Conference on Distributed Computing Systems*. Tokyo, 2004; 516-523
- [15] Anceaume E, Gradinariu M, Ravoajia A. Incentive for P2P fair resource sharing//*Proceedings of the 5th IEEE International Conference on Peer-to-Peer Computing*. Konstanz, 2005; 253-260
- [16] Figueiredo D, Shapiro J, Towsley D. Incentives to promote availability in Peer-to-Peer anonymity systems//*Proceedings of the 13th IEEE International Conference on Network Protocols*. Boston, 2005; 110-121
- [17] Lui S M, Lang K R, Kwok S H. Participation incentive mechanisms in Peer-to-Peer subscription systems//*Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. Hawaii, 2002; 3925-3931
- [18] Karakaya M, Korpeoglu I, Ulusoy O. A distributed and measurement-based framework against free riding in Peer-to-Peer networks//*Proceedings of the 4th International Conference on Peer-to-Peer Computing*. Zurich, 2004; 276-277
- [19] Gupta R, Somani A K. Game theory as a tool to strategize as well as predict nodes' behavior in Peer-to-Peer networks//*Proceedings of the 11th International Conference on Parallel and Distributed Systems*. Fukuoka, 2005; 244-249
- [20] Golle P, Kevin L B, Mironov I. Incentive for sharing in Peer-to-Peer networks//*Proceedings of the 3rd ACM Conference on Electronic Commerce*. Tampa, 2001; 264-267
- [21] Ma R T B, Lee C M, Lui J C S, Yau D K Y. Incentive and service differentiation in P2P networks: A game theoretic approach. *IEEE/ACM Transactions on Networking*, 2006, 14(5): 978-991
- [22] Hardin G. The tragedy of the commons. *Science*, 1968, 162(3859): 1243-1248
- [23] Krishnan R, Smith M D, Tang Z L, Telang R. The impact of free-riding on Peer-to-Peer networks//*Proceedings of the 37th Hawaii International Conference on System Sciences*. Hawaii, 2004; 199-208
- [24] Strulo B, Smith A, Farr J. An architecture for peer-to-peer economies//*Proceedings of the 3rd International Conference on Peer-to-Peer Computing*. Linkoping, 2003; 208-209
- [25] Xavier P, Cai W T, Lee B S. Employing economics to achieve fairness in usage policing of cooperatively shared computing resources//*Proceedings of the 2005 IEEE International Symposium on Cluster Computing and Grid*. Cardiff, 2005; 326-333
- [26] Gupta R, Somani A K. Pricing strategy for incentive selfish nodes to share resources in Peer-to-Peer (P2P) networks//*Proceedings of the 12th IEEE International Conference on Networks*. Singapore, 2004; 624-629
- [27] Lang K R, Vragov R. A pricing mechanism for digital content distribution over Peer-to-Peer networks//*Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Hawaii, 2005; 211-221
- [28] Hales D, Edmonds B. Applying a socially inspired technique (tags) to improve cooperation in P2P networks. *IEEE Transactions on Systems, Man and Cybernetics-Part A: System and Humans*, 2005, 35(3): 385-395
- [29] Hausheer D, Liebau N C, Mauthe A, Steinmetz R, Stiller B. Token-based accounting and distributed pricing to introduce market mechanisms in a peer-to-peer file sharing scenario//*Proceedings of the 3rd International Conference on Peer-to-Peer Computing*. Linkoping, 2003; 200-201
- [30] Ham M J, Agha G. ARA: A robust audit to prevent free-riding in P2P networks//*Proceedings of the 5th IEEE International Conference on Peer-to-Peer Computing*. Konstanz, 2005; 125-132
- [31] Andrade N, Brasileiro F, Cime W. Discouraging free riding in a peer-to-peer CPU-sharing grid//*Proceedings of the 13th IEEE International Symposium on High Performance Distributed Computing*. Hawaii, 2004, 2004; 129-137
- [32] Wang W J, Chang H, Zeitoun A, Jamin S. Characterizing guarded hosts in Peer-to-Peer file sharing systems//*Proceedings of the Globecom 2004*. Dallas, 2004; 1539-1543

- [33] Feldman M, Chuang J. Overcoming free-riding behavior in Peer-to-Peer systems. *ACM SIGecom Exchanges*, 2005, 5(4): 41-50
- [34] Feldman I, Papadimitriou C, Chuang J, Stoica I. Free-riding and whitewashing in Peer-to-Peer systems. *IEEE Journal of Selected Areas in Communications*, 2006, 24(5): 1010-1019
- [35] Andrade N, Mowbray M, Lima A, Wagner G, Ripeanu M. Influences on cooperation in BitTorrent communities//*Proceedings of the 2005 ACM SIGCOMM workshop on Economics of Peer-to-Peer Systems*. Philadelphia, 2005: 111-115
- [36] Jun S, Ahamad M. Incentives in BitTorrent induce free riding//*Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems*. Philadelphia, 2005: 116-121
- [37] Yang M, Zhang Z, Li X M, Dai Y F. An empirical study of free-riding behavior in the Maze P2P file-sharing system//*Proceedings of the 4th International Workshop on Peer-to-Peer Systems*. Ithaca, 2005: 182-192
- [38] Courcoubetis C, Weber R. Incentive for large Peer-to-Peer systems. *IEEE Journal of Selected Areas in Communications*, 2006, 24(5): 1034-1050
- [39] Yu Y J, Jin H. A workload balancing based approach to discourage free riding in Peer-to-Peer network//*Proceedings of the APWEB/WAIM 2007 Workshop on Database Management and Application over Networks*. Huangshan, 2007: 144-155
- [40] Castro M, Costa M, Rowstron A. Should we build Gnutella on a structured overlay? *ACM Sigcomm Computer Communications Review*, 2004, 34(1): 131-136
- [41] Moscibroda T, Schmid S, Wattenhofer R. On the topologies formed by selfish peers//*Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing*. Denver, 2006: 133-142



YU Yi-Jiao, born in 1978, Ph. D. candidate, lecturer. His research interests focus on computer network and computer architecture.

JIN Hai, born in 1966, Ph. D. , professor, Ph. D. supervisor. His research interests include computer architecture, grid computing, cluster computing, parallel and distributed computing, semantic Web, peer-to-peer computing, network storage, network security, and pervasive computing.

Background

This work is supported by National Basic Research Program of China (973 Program) under grant No. 2003CB317003, which focuses on semantic storage technologies and high performance of services in grid and Peer-to-Peer (P2P) network environments. P2P applications are popular in recent years, in which peers are not only eager to enjoy services, but also expected to contribute to the P2P network. Some traffic measurement experiments are carried out since 2000 and the results, however, show that free riding is prevalent in almost all P2P networks. Free riding is a new challenge for P2P network, and a lot of overcoming mechanisms have been proposed. Unfortunately, most of proposed mechanisms are not

employed in real P2P networks, due to the complexities or unreasonable assumptions in modeling. The main objectives of this paper are to introduce the measurement of free riding, its impacts on P2P networks, and the proposed typical overcoming approaches. The authors categorize the proposed solutions, discuss the key ideas and features of solutions in each category, and summarize the future research issues. To the best of our knowledge, this paper is the first survey on the mechanisms avoiding free riding. In the past years, the authors have proposed a workload balancing based approach to discourage free riding in P2P network, and submitted a related national patent.