

P2P 网络可用性关键技术研究进展^{*})

张三峰 赵生慧 方群 于坤 吴国新

(东南大学计算机网络和信息集成教育部重点实验室 南京 210096)

摘要 P2P 网络是典型的由应用需求推动学术研究的网络技术,提高可用性是 P2P 网络的一个研究主线。在概括 P2P 网络应用的类型,总结 P2P 网络的性质模型的基础上分析了影响 P2P 网络可用性的若干因素,包括结构化 P2P 网络拓扑参数优化、覆盖网络拓扑感知和优化、网络异构和负载均衡、抗波动性能优化、容错路由以及激励机制。对各个关键技术问题的解决方案进行分类对比,并根据比较的结果对今后的发展趋势进行了展望。

关键词 对等网络,拓扑参数,可扩展性,拓扑感知,波动,容错,负载均衡

Research and Evolvement of Key Technologies Enhancing Practicability of P2P Network

ZHANG San-feng ZHAO Sheng-hui FANG Qun YU Kun WU Guo-xin

(Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract P2P network is a new technology specially promoted by needs of application. Enhancing practicability is a main thread of P2P research. Surveys types of P2P application, summarizes the property model of P2P network and analyzes key technologies affecting practicability of P2P network including topology parameter optimizing of structured P2P, topology awareness of overly, heterogeneous and load balancing, enhancing churn resistibility, fault tolerance and incentive mechanism. Classifies and compares solutions of key problems and discusses directions of research based on the comparing result.

Keywords Peer-to-peer, Topology parameters, Scalability, Topology awareness, Churn, Fault tolerance, Load balancing

1 前言

Peer-To-Peer 网络(在本文中 P2P、对等网络同义)是一种强调节点之间逻辑对等的新型计算模式。随着计算机硬件成本的降低和计算机网络及其应用的普及,用户端的资源越来越丰富。P2P 网络技术应运而生;利用 P2P 技术可以提升用户端资源(处理能力、存储能力、应用软件等)的利用率。在互联网研究重点从互联互通向应用和服务进行跃迁的背景下,P2P 技术凭借其部署灵活、廉价等特点将在新一代的网络服务体系结构中占据重要位置,可以显著提高互联网服务的效益规模。一方面,无处不在、无时不在的互联网及其蕴含的价值潜力和巨大的海量资源为 P2P 技术的发展提供了巨大的动力;另一方面,互联网的匿名、动态性、异构性特征又给 P2P 网络的可用性带来极大的挑战。

P2P 网络是典型的由应用需求推动学术研究的网络技术,提高可用性是 P2P 网络的一个研究主线。已有的研究从结构化 P2P 网络拓扑参数优化、无结构 P2P 网络可扩展性优化、覆盖网络拓扑感知和优化、网络异构和负载均衡、抗波动性能优化、容错路由以及激励机制等方面提高 P2P 网络可用性。本文目标是总结 P2P 网络应用发展、性质模型和面临的挑战,以此为基础对影响 P2P 网络可用性的各个关键技术进行深入分析,并对相关解决方案分类对比,根据比较的结果对

今后的发展趋势进行展望。

2 P2P 网络应用和特性

P2P 网络技术已被广泛应用于各个领域(文件共享、对等计算、网络存储、应用层组播、信息检索、P2P 平台、P2P 协同)并还在不断扩展中。

一般来讲这些应用具有消除中心节点瓶颈,带宽对称,规模大,多路并行下载速度快,吞吐量高,自组织,不依赖服务器 DNS 等基础设施,易于部署,充分利用 Internet 边缘的海量闲置计算资源,成本低,可实现高性能计算等优势。

这些应用体现出 P2P 网络在体系结构、网络组织和管理、资源利用等方面的一些共同的本质特征。图 1 总结了 P2P 网络的性质模型,其中箭头表示因果关系。

(1) 对等、无中心特性

P2P 节点间是逻辑上对等的互连关系,每个节点既是客户端、也是服务器端,所有资源和服务的交换都是在节点间完成。这带来很多好处:调动了传统上只能处于客户端位置的用户积极性;可以发掘网络边缘的数据、计算、存储等资源;避免了服务器性能瓶颈和单点失效问题;网络流量分布更加分散,更好地实现了网络负载均衡;多点并行下载提高内容分发网络的吞吐量;分布式搜索引擎可以实现深度搜索。

(2) 自组织特性

^{*} 国家 863 计划(编号 2007AA01Z422),国家自然科学基金重大研究计划项目,高可用的新一代网络服务体系结构及其关键技术的研究(编号 90604003),安徽省教育厅自然科学基金项目(编号 2005kj088),安徽省高校教师青年基金(编号 2007jq1061)。张三峰 博士研究生,主要研究领域为 P2P 网络架构、激励,信息检索;赵生慧 副教授,主要研究领域为内容分发、缓存和备份;方群 副教授,主要研究领域为 P2P 网络安全和信任;于坤 博士研究生,主要研究领域为 P2P 网络激励和信任;吴国新 教授,博士生导师,主要研究领域为分布式计算、网络安全、信息化关键支撑技术、网络性能评价。

自组织的定义：“自组织是一个过程，在此过程中，一个系统的组织自然本能地增加，该过程不需要环境或其他外部系统的控制。”自组织管理模式下，由于服务是分散于各个节点的，部分节点或网络遭到破坏对其它部分的影响很小，网络一般能够自动重构、自动维护、自动修复、保持其功能，系统的鲁棒性好。自组织的管理模式还使得网络部署更加灵活方便，无须架设专门的服务器软硬件设施，也不需要专门的管理维护人员。

(3) 匿名、自治

匿名性体现在：P2P 网络不标识文件的作者或创建者、发行者以及读者或其他消费数据者，一般也不追踪用户的历史信息。自治性体现在：P2P 网络不是任何个人或组织专有的，每个节点对所拥有资源及自身行为具有完全的控制权，它可以选择资源开放的程度、加入和离开的时机。匿名、自治保护了用户的隐私，增加了用户的自由度，体现了自由软件的精神，受到多数用户的欢迎。

(4) 覆盖网络

P2P 网络屏蔽底层通信细节，是一种应用层虚拟网络，有效地屏蔽了不同网络、操作系统、计算机硬件、以及编程语言等所带来的异构性。P2P 网络以开放的接口把服务提供给用户，使得 P2P 网络易于扩充、易于部署。

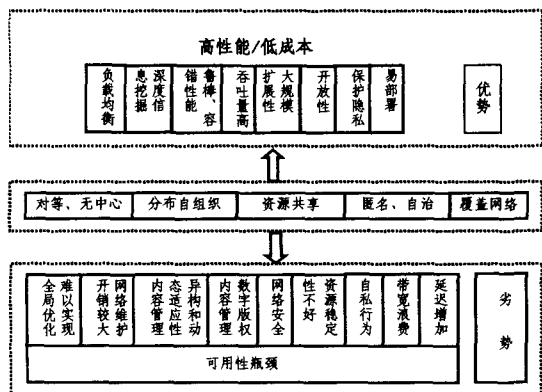


图1 P2P网络性质模型

(5) 资源共享

“我为人人，人人为我”的资源共享模式降低了整体成本。

总之，P2P 网络可以以低廉的成本完成以前需要昂贵的设备才能完成的任务：①使用现有的基础设施、不需要专门的网络、服务器设备；②利用闲置的 Internet 边缘资源，资源利用率高；③自组织管理模式的管理运营成本极低；④无中心瓶颈，即使在大规模网络中也不需要昂贵的高端设备。因此，P2P 网络的优势可以概括为高性能/低成本。

基于 Internet 的 P2P 网络具有一些内在的特性，这些特性在 P2P 网络从诞生走向成功的过程中发挥了重要作用。但是，随着网络环境的发展变化，P2P 网络在向更深层次、更广范围发展的时候，也面临新的挑战。

首先，P2P 网络的运行平台——Internet 已经从最初的规模较小、覆盖范围有限、参与群体单一的网络发展成为规模庞大的巨型网络。这些不同利益群体的目标、行为以及他们之间的矛盾和冲突严重影响 P2P 网络的可用性。例如：2000 年前后，用户通过 Napster 交换自己喜欢的娱乐内容并由此掀起 P2P 网络的流行浪潮，但是知识产权保护者的诉讼使得旧有的 Napster 服务对用户不再可用。P2P 网络极大地增加了用户和用户之间的直接交互，但是自治系统的管理者以及政

府部门却因为担心传播有害内容或者泄漏商业秘密而给 P2P 网络设置障碍，比如协议识别、封锁端口等，这些行为也降低了 P2P 网络的可用性。P2P 网络服务的可用性，依赖于其参与者贡献的带宽、计算、存储、内容等资源，网络用户的自私行为显然会降低可用性。由于 P2P 网络应用常常需要较高的带宽以及拓扑不匹配等因素，大量骨干网络带宽和接入网络带宽被 P2P 网络应用占用，ISP 没有从网络流量增加过程中获得更高的收益却必须为此增加管理和运营成本，因此 ISP 采取技术和经营手段限制 P2P 网络的可用性。

其次，Internet 表现出越来越强烈的动态性、异构性特征。随着互联网的发展，P2P 网络的参与者在计算资源和用户行为等方面表现出更多的差异性。这些异构特性影响了 P2P 网络在拓扑优化、连通性、资源稳定性等方面的性能。另外，Internet 上的 P2P 应用由于规模庞大、匿名、自治的特点还导致很强的动态性。动态节点给 P2P 网络可用性带来性能上的影响，比如：延迟增加、路由错误的概率增加、网络维护开销增加等等。

当前互联网中广泛存在的利益冲突给 P2P 网络应用带来的影响，动态、异构的 Internet 环境对结构化 P2P 网络的影响是制约 P2P 网络实用性的两个重要问题。如前所述，现有 P2P 网络技术在面对这两个问题的时候显得力不从心，而解决这些问题对于 P2P 网络实用化进程又非常迫切。这种矛盾正是 P2P 网络可用性研究的背景和动机。

3 P2P 网络可用性定义

通俗地讲，当一个用户向 P2P 网络发出一个资源查询请求或者资源获取请求的时候，如果此时 P2P 网络因为网络连通不畅、可用资源较少等因素而不能给予用户满意的查询响应速度，或者要求用户长时间排队等待下载，甚至无法找到用户所需的资源，那么用户认为该 P2P 网络服务“不好用”，如果这种现象经常发生，用户则认为“不可用”。由此可见，“可用性”是从使用者的角度对一个系统做出的评价（比如：快速、方便灵活、可靠、安全），因此它是一个综合多维属性的概念。

借鉴可信 P2P 体系结构^[1]以及互联网软件可信机理^[2]的相关研究，本文对“可用性”概念作如下定义：

定义 1(可用性, Availability) 是评价一个系统在给定环境下、给定时间段内，实现给定目标的能力的尺度。它包括以下几个属性：

响应能力 (Responsiveness): 不仅包括延迟、抖动等性能指标，还强调用户对系统性能感受的一致性。

可靠性 (Reliability): 系统在某时间点上可操作并能完成服务请求的概率。

安全性 (Security): 系统确保能够防范意外的或者恶意的的外部攻击。

下面结合该定义解释 P2P 网络的可用性问题。

P2P 网络可用性是有特定环境的，在本文中即是 Internet 环境。这个环境包括的因素有：实现连通功能的物理网络、自治匿名的用户、网络服务提供商 (ISP, Internet Service Provider)、互联网内容提供商 (ICP, Internet Content Provider)。因此底层网络的连通能力、自治系统的管理策略、ISP 的经营和计费策略、用户的自治行为、ICP 的经营策略都会影响 P2P 系统的可用性。

P2P 网络可用性是指一个持续时间段上的可用性，高可用的 P2P 网络不仅要在网络稳定阶段保证网络服务的响应

能力与安全可靠性,还要能在成长阶段和衰退阶段保证网络服务同样可用。

P2P 网络可用性目标包含了功能指标和性能指标,高可用性不仅要实现给定的功能目标(资源定位、网络传输、内容分发、网络协同等功能),而且要实现给定的性能指标(用户操作响应速度、网络延迟、抖动等),由此可见,可用性也包括了对于服务质量保障(QoS)的要求。

4 P2P 网络可用性关键技术研究进展

4.1 结构化 P2P 网络的拓扑参数优化

自 2000 年以来,从图论的角度分析、设计基于 DHT 的结构化方法,一直是热点问题,每年在 SIGCOMM, INFO-COM, IPTPS 等重要学术会议上都有更优化的 DHT 方法提出。这些 DHT 方法^[22-36]是对静态拓扑(如多机互联网络拓扑)的扩展:静态拓扑结构映射到 P2P 网络中可以产生新的 DHT 方法,已有的理论分析方法仍然可以用于分析 DHT 方法的特性。可以从五个方面分析对比这些 DHT 方法的优劣:节点度数、网络直径(或者平均跳数)、容错性、负载均衡和维护开销。度数-网络直径的折衷(状态-效率的折衷)是结构化 P2P 系统研究发展的主线。网络直径越小,延迟就越小,路由效率就越高。度数越高,节点状态越复杂,网络维护开销越大。Ratnasamy Sylvia^[37]等提出:Chord, Pastry 等 DHT 方案是以 $O(\log_2 N)$ 度数去实现 $O(\log_2 N)$ 网络直径, CAN 方法是以 $O(d)$ 度数去实现 $O(N^{1/d})$ 网络直径,那么是否存在以 $O(d)$ 度数去实现 $O(\log_2 N)$ 网络直径的 DHT 方法? 或者说“对于节点度为 $O(\log_2 N)$ 和 $O(d)$ 的 P2P 网络,网络直径下界是否是 $\Omega(\log_2 N)$ 和 $\Omega(N^{1/d})$?”,其中 N 表示节点个数, d 表示最大出度。Jun Xu^[38]等以图论方法分析了这个问题,通过研究目前各种 DHT 路由策略,他们发现平均节点度和平均路由长度存在一种渐进优化曲线的关系。他们认为在路由表分布一致性(Uniform)和维持常量拥塞的条件下, $O(\log_2 N)$ 和 $O(N^{1/d})$ 其实就是节点度 $O(\log_2 N)$ 和 $O(d)$ 网络的直径下界。

然而 Uniform 并不是构建结构化网络的必要条件,几个基于 Butterfly, de-Bruijn 图和 Kautz 图的常数度的 DHT 方法都超出了渐进优化曲线的界限,而具有更好的“度-直径”折衷关系。有向图的 Moore 界说明:在最大出度为 d , 直径为 D 的有向图中最大节点个数 N 满足如下约束: $N \leq 1 + d + d^2 + \dots + d^D = N_{\max}$ 。由此界限可以得出:节点度为 d 和 $\log_2 N$ 的规则网络的直径下界分别是 $\Omega(\log_d N)$ 及 $\Omega(\frac{\log_2 N}{\log_2(\log_2 N)})$ 。de-Bruijn 图 $B(d, D)$ 的节点数为 $N = d^D$, Kautz 图 $K(d, D)$ 的节点数为 $N = d^D + d^{D-1}$, 它们已经非常接近 N_{\max} , 所以对常数度的规则图已经很难进一步优化。

从多机互联网络等静态拓扑结构扩展新的 DHT 方法需要处理一些差异:前者由于硬件设计和实现的限制要求节点度数固定且较小,并且网络环境是相对静态、同构、可控的;而 P2P 网络对节点度没有严格要求,并且网络环境是相对动态、异构、不可控的。另外, Moore 界仅对规则网络有效,节点度数不对称的不规则网络中可以对节点度和网络直径的折衷进一步优化。

4.2 结构化 P2P 网络的路由容错相关研究

P2P 网络的本质特征决定节点失效并不是一个低概率事件,因此容错路由十分重要。节点失效原因包括节点的突然

崩溃,用户的非正常退出等。那么原先必须经过失效节点的那些消息是否能够恢复或者切换到替代路由? 恢复的时间是多少? 这被称为静态容错(static resilience)问题。Gummadu^[39]等人认为静态容错与路由方式和拓扑结构密切相关,并证明了环结构的静态容错有效性。环形拓扑的 Chord^[10]能够通过后继节点表应对静态容错问题。文献[40]给出了一种 $O(\log_2 N)$ 度数、 $O(\log_2 N / \log_2 \log_2 N)$ 网络直径的容错 DHT 结构,在给定较小的节点失效率下,网络所有活动节点都能在 $O(\log_2 N)$ 步内被定位,还能够容忍失效节点产生的随机错误数据。文献[41]提到了容错邻居和 Multiple out-edges 机制,但它们要么是只针对单点容错,要么是过多增加节点度数。这些思路都是首先考虑如何快速寻找替代路由,并尽可能低地控制替代路由的额外代价。在恢复原路由方面,一种考虑是由特定节点(如失效节点的邻居,通常是节点失效的发现者)接管失效节点所负责的路由环节,另一种考虑是由特定节点替代失效节点,Ratnasamy Sylvia 认为如果路由过程能自动通知替代节点,原路由路径将会很快被恢复^[37]。

已有的、基于字母表的 DHT 方法在追求“状态-效率”折衷关系的同时降低了节点的度数,减少了替代路由数目,使得网络的容错性降低。相关研究也没有考虑到实际系统中节点在实际度数、稳定性方面的异构特性对系统容错性的影响。

4.3 结构化 P2P 网络抗波动(Churn)性能研究

根据对部分已部署的 P2P 应用系统(Gnutella, BT, Kad, KaZaa 等)的测量,研究者发现 P2P 系统中存在网络波动(Churn)现象,而且除了少数节点的会话时间(从上线到线下的间隔)较长(一到数小时),多数节点的会话时间都很短(数分钟)。Churn 水平的衡量尺度包括会话时间(session time, 在线-下线)、生命时间(lifetime, 注册-注销)、可用率(availability)等等。简单的方法可以以会话时间作为衡量尺度。波动现象严重影响 P2P 网络的可用性:查询响应延迟增大;可能因为分区导致返回错误结果(查询消息没有被转发到真正的 root 节点);网络维护开销增加。

Bamboo 算法^[42]表明影响查询响应延迟的因素有:跳数、节点链路延迟、网络层的背景流量、排队、丢包、超时、可用带宽等等,因此它对网络层特性进行了模拟,并从三个方面提高 DHT 方法的抗波动性能:①由于带宽有限的时候,使用被动、按需的失效恢复方式会产生一个正反馈(失效恢复操作带来更大的维护开销加重了网络负担,带宽资源更加紧张),所以使用周期性的失效恢复方式;②在消息超时时间(节点在发送消息后等待多长时间认为节点已经失效,然后尝试其它邻居节点,这种时间占据了延迟的主要部分)的确定方法上,使用类似 TCP 协议的方法确定超时时间;③使用 PNS(临近邻居选择)的 DHT 方法改善延迟。Bamboo 方法没有从 DHT 拓扑结构的角度改进抗扰动性能,但是可以和该方面的算法互补。

Accordion^[43]算法是 MIT 的 PDOS 项目组提出的(Chord, Koorde 等著名的 DHT 方法都由该组提出)。它的思想和其它的 DHT 方法有明显的不同:一般的 DHT 方法使用尽量少的路由表项(常数个或者在 $O(\log n)$ 以内)实现 $O(\log n)$ 的网络直径,而 Accordion 专注于改善查询响应延迟。Accordion 维护一个尺寸可变的路由表。Accordion 使用预算内的所有网络带宽,多条路径并行发送查询请求,学习尽可能多的路由表项(选择性的把邻居的邻居设为自己的邻居)。各个节点的路由表尺寸可以不同:可用带宽大的节点带宽预算

大,路由表尺寸大(能者多劳)。在网络波动程度较高的时候,Accordion自适应的缩小路由表尺寸,减少无效路由表项数目,降低维护开销和带宽占用。Accordion的主要困难在于:根据网络状况(根据局部的信息判断网络整体的状况)自适应的调整带宽预算因而调整路由表尺寸;学习、选择路由表项的方法,剔除可能失效的路由表项的方法。

4.4 结构化 P2P 网络适应异构环境的研究

有几个原因导致节点之间负荷与能力的比值严重扭曲,进而使得 P2P 网络负载不均衡:①底层网络异构,节点的存储空间、计算能力、接入带宽等的差异可能达到几个数量级;②网络资源的流行度非均匀分布,存储高流行度资源键值的节点可能过载;③结构化 P2P 网络节点标识符分配中的散列函数可能导致 $O(\log n)$ 的不平衡因子;④网络的动态性,节点的加入退出、资源的动态变化和迁移。所以,P2P 网络负载均衡需要解决如下几个问题:①异构参数测量,分布式的、自组织的、低开销的、动态自适应的异构参数测量,为负载均衡提供决策依据;②异构适应的拓扑结构;③节点和资源标识均匀分布算法;④缓存和副本复制策略;⑤动态负载调整算法。相关研究参见文献[44-52]。

在动态负载调整算法方面:P-Grid^[44]提出直接用新加入节点去分担重载节点的策略,省去了寻找和维护轻载节点及其负载状态的繁琐步骤。文献[45]从临近感知的角度讨论基于 DHT 的 P2P 系统的负载均衡问题,它使用一个构建在 DHT 之上的树结构,收集系统的负载信息,并进行负载重新分配负载以虚拟服务器为单位分配,通过重分配虚拟服务器对应的物理节点实现负载均衡在重定向虚拟服务器的时候考虑了临近信息,虚拟服务器在离过载节点最近的轻载节点之间传输。文献[46]以随机平衡二叉树的方式处理节点的动态加入和退出导致的负载不平衡。文献[51]把每个数据项哈希到多个 ID,选择负载轻的节点实际存储索引。

在节点和资源标识均匀分布算法方面:文献[47]研究了随机分配 ID 的负载均衡算法,使用分配-测量-分配的方法缩小随机分配导致的不平衡因子。文献[50]解决动态结构化环境下因为节点持续的加入和退出造成的节点和数据 ID 的不均匀分布问题。

4.5 拓扑优化相关研究

如前一章所述,由于拓扑不匹配而导致低效的 P2P 内容分发是造成带宽浪费、可用性降低的一个重要原因。因此,优化基于 P2P 的内容分发网络拓扑也受到广泛关注。

文献[20]提出通过实时的 RTT 测量来删除长连接、探测短连接的方案,可以较好地限制不必要的跨域通信。在结构化 P2P 网络方面,也有一些解决方案,如:临近路由 (PRS Proximity Routing Selection)、临近邻居选择 (PNS Proximity Neighbor Selection) 和拓扑感知的 ID 分配 (GLI Geographic Layout Identity)^[21]。PRS 指节点在路由选择时,处理标识空间距离和实际网络延迟之间的权衡,找出延迟最小或者是具有标识空间距离和实际距离最佳折衷点的邻居节点。PNS 的每个节点在路由表的建立和维护的过程中,选择在标识空间符合条件并且延迟较小的邻居作为路由表项。GLI 中节点 ID 标识与节点所在的物理位置相关,以标识为基础的覆盖网的路由就能够尽可能接近以 IP 为基础的物理网的路由。

在 P2P 技术已经发展成为 Internet 上最流行的应用、P2P 技术和互联网紧密交互的背景下,网络层的计费信息在一定程度上是拓扑信息的反映(跨域链路的成本和计费都高

于域内链路),那么能否结合运营商的计费信息,调整 P2P 网络拓扑,同时达到降低 P2P 运营成本和提高网络效率的目的,从而大幅提高 P2P 网络的可用性?

4.6 激励机制

由于 P2P 网络的匿名、自治本质特征,以及对等节点的个体利益最大化的行为,P2P 网络中类似 Free-riding 的自私节点现象严重影响服务的可用性。有关激励机制的研究已经成为新的热点问题。已有的激励机制可以概括为以下几类:基于经济关系的激励机制、基于声誉的激励机制、基于博弈论的激励机制、基于机制设计的激励方法。

基于经济关系的合作激励机制使用自由市场的等价交换方法实现激励,提供服务可以获得等价的回报。由于简单的物物交换减少了合作机会,通过一般等价物(声誉值或者虚拟货币等)增加合作机会。此类方法的设计分析中大量借鉴了经济学中的原理和模型。BitTorrent^[6]采用基于 2-way 交换的合作机制。它通过种子节点把对同一文件的请求聚集成一个群,同时每个文件分割成多个块以便多路并行传输。节点参与文件块交换时采用 TFT(以牙还牙)策略,也即在下载文件块的同时也为其它用户提供上传服务。

基于声誉的合作激励机制。这种方法中节点根据对方的声誉值决定合作程度,因此在节点的声誉值方面对不合作行为惩罚、对合作奖励可以形成基于间接互惠的合作机制。此类研究如 EigerTrust^[7], P2Prep^[8], P-Grid^[9], PRIDE^[10] 等。声誉系统一般要求用户具有唯一可信的身份标识,一般需要一个权威的第三方声誉管理机构。如何降低算法复杂度、开销,减少对 P2P 系统自组织性的破坏是这类方法改进的目标。

基于博弈论的合作激励机制。重复博弈能够解释社会经济领域中的合作行为产生机理,基于这种博弈模型的激励机制有希望不借助附加管理机构而实现 P2P 网络节点间的合作。Mark^[11]首先在 Ad Hoc 网络中应用多人非对称的重复囚徒博弈考虑路由激励问题,提出一种简单的连续型 TFT 策略(每个节点根据从网络获得的吞吐量调整其合作级别)。相似研究还有 Vikram^[12], SLIC^[13] 等。简单的重复博弈模型在实际应用环境中遇到很大的困难:节点并非完全理性;节点难以了解全局信息,缺乏其它节点甚至自己的策略空间信息;P2P 网络本质上的动态性影响重复博弈的稳定性;合作只是多个均衡中的少部分,博弈的结果不能一定导致合作。

基于机制设计的激励方法和一般网络博弈的研究方法相反:它不是由类型组合求均衡解集合,而是寻求能把任何可能的节点类型配置映射成机制设计者所期望结果的方法。机制设计通过转移支付影响节点的收益以使设计者期望的结果成为均衡解^[14]。Vickery、Clarke 和 Groves 三人提出的 VCG 机制^[15]是一类重要的机制,它的奥妙之处在于能够激励参与者报告真实的私有信息,同时得到系统收益最大的博弈结果。

表 1 次价拍卖的报价净收益对比

报价	$c_1 < b_1 < c_2$	$b_1 < c_1$	$b_1 > c_2$	$b_1 = c_1$	$b_1 = c_1$	$b_1 = c_1$
	$b_2 = c_2$	$b_2 = c_2$	$b_2 = c_2$	$b_2 > c_2$	$c_1 < b_2 < c_2$	$b_2 < c_1$
收益	$v_1 = 5$	$v_1 = 5$	$v_1 = 0$	$v_1 > 5$	$0 < v_1 < 5$	$v_1 = 0$
	$v_2 = 0$	$v_2 = 0$	$v_2 > 0$	$v_2 = 0$	$v_2 = 0$	$v_2 < 0$

我们用一个 VCG 机制的简化形式——次价拍卖 (Second-Price-Auction) 说明它激励参与者报告真实信息的原理。

假设一个节点(机制的组织者)需要通过一个邻居节点转发报文,而且它希望选择成本最低的邻居转发。它有两个可能的邻居,它们的转发成本分别为 $c_1=10$ 和 $c_2=15$ (这里,转发成本也称作它们的类型)。两个邻居同时提供报价以决定谁负责转发。选择转发节点的规则是:选择报价最低的邻居,并且按次低报价付费,若两者报价相同,则随机选择一个。这样,诚实地报价(开出等于成本的报价,即 $b_1=c_1=10$, $b_2=c_2=15$)是两个邻居的最佳选择。结果是邻居 1 以价格 15 提供转发服务,净收益为 $v_1=15-10=5$,邻居 2 净收益为 $v_2=0$ 。从表 1 可以看出,在对方诚实的条件下,两个邻居都不可能通过开出比成本更高或者更低的报价而获得更高的净收益。两个邻居报告真实成本信息(防策略性),并且选择成本最低的邻居执行转发任务(有效性)是这种方法的成功之处。节点向邻居 1 的付费方式 $p=c_1+\Delta$, (其中, $\Delta=c_2-c_1$ 称为 Vickrey 折扣)是关键之处。节点付出了超过实际传输开销的费用(收支不平衡)是不足之处。

Ad hoc-VCG^[16]通过对 Ad-hoc 网络转发路径上节点的付费来激励中间节点转发报文,中间节点向目的节点汇报自己转发报文的开销,由目的节点计算开销最小的路径进行转发,VCG 算法计算转移支付以保证节点正确汇报自己的转发成本。由于 VCG 算法存在收支不平衡(BU, Budget Unbalance)弱点,源节点需要支付的虚拟货币大于实际传输的开销。COMMIT^[17]提供了一种 Ad hoc-VCG^[16]的改进:将总体支付分为分别由源节点和目的节点支付的两部分,并使用次价拍卖的方法决定传输路径,如果综合考虑源节点和目的节点的付出,系统实际也没有克服 BU 弱点。文献[18,19]讨论了在 P2P 或者 Ad-hoc 网络中应用 VCG 机制以分布式方式生成动态优化的拓扑的算法和协议,共同的缺点是假设存在能够补偿收支赤字的节点。总之,BU 的缺点制约了 VCG 机制在 P2P 系统的应用,因为自组织的网络中缺乏甘愿承担赤字的中心节点。

次价拍卖例子可以用于简单的环境。在 P2P 网络中,每个节点既是资源的提供者又是资源的使用者,所有的节点形成复杂的交易网络,在这种环境下,设计合适的供需匹配机制和结算方法并满足有效性和收支平衡的要求具有一定的挑战性。

结束语 本文在总结 P2P 网络性质模型、定义 P2P 网络可用性的基础上分析影响 P2P 网络可用性的关键技术和相关研究进展。我们认为在如下方面尚需关键技术的突破:①自组织的适应节点的个体理性特征的激励机制;②综合考虑接入服务提供商、内容服务提供商、用户的整体成本的拓扑感知和优化机制;③适应 Internet 的动态、异构特征,具有更好的容错性、抗波动性和负载均衡能力的 DHT 方法。

参 考 文 献

- [1] Melville L, Walkerdine J, Sommerville I. Method definition for dependability specification for P2P systems. TR0306F03_Method_Def. 2003
- [2] 王怀民,唐扬斌,尹刚,等. 互联网软件的可信机理. 中国科学 E 辑信息科学, 2006, 36(10): 1156-1169
- [3] Clark D D, Wroclawski J, Sollins K R. Tussle in Cyberspace, Defining Tomorrow's Internet // Proc. ACM SIGCOMM. 2002; 347-356
- [4] 纽约时报. <http://www.nytimes.com/2007/10/03/business/worldbusiness/03broadband.html>, 2007
- [5] 刘业. 适应自组织管理模式的 P2P 网络技术的研究[D]. 东南大

学博士学位论文. 2006

- [6] BitTorrent. <http://bitconjurer.org/BitTorrent/>
- [7] Kamvar S D. The eigenTrust algorithm for reputation management in P2P networks // Proceedings of the Twelfth International WWW Conference. Budapest, Hungary, May 2003
- [8] Cornelli F, Damiani E. Choosing reputable servers in a P2P network // Proc. of the Eleventh International WWW Conference. Honolulu, Hawaii, May 2002
- [9] Aberer K, Despotovic Z. Managing trust in a peer-to-peer information system // Proceedings the 10th international Conference on Information and Knowledge Management. Atlanta, Georgia, USA, 2002
- [10] Prashant D, Partha D. Securing reputation data in peer-to-peer networks // PDCS 2004. MIT Cambridge, USA, 2004
- [11] Félégyházi M, Buttyán L, Hubaux J-P. Equilibrium Analysis of Packet Forwarding Strategies in Wireless Ad Hoc Networks - The Static Case. PWC, 2003
- [12] Srinivasan V, Nugehalli P, Chiasserini C-F, et al. Cooperation in wireless ad hoc networks // Proceedings of IEEE Infocom. 2003
- [13] Sun Qixiang, Garcia-holina H. SLIC: a selfish link-based incentive mechanism for unstructured peer-to-peer net, works // Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS). 2004
- [14] Feigenbaum J, Shenker S. Distributed algorithmic mechanism design: recent results and future directions. Distributed Computing Column. Bulletin of the EATCS, 2003; 101-121
- [15] Milgrom P, et al. 拍卖理论与实务. 杜黎, 胡其英, 等译. 清华大学出版社, ISBN-7-302-12971, 2006
- [16] Anderegg L, Eidenbenz S. Ad hoc-VCG: A truthful and cost efficient routing protocol for mobile ad hoc networks with selfish agents [C]. ACM MobiCom, Annapolis, USA, 2003
- [17] Eidenbenz S, Santi P, Resta G. COMMIT: A sender centric truthful and energy efficient routing protocol for ad hoc networks [C] // The 5th IEEE Int'l Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (WMAN 05). Denver, Colorado, USA, 2005
- [18] Yuen S, Li Baochun. Strategyproof Mechanisms for Dynamic Multicast Tree Formation in Overlay Networks // Proceedings of IEEE INFOCOM 2005. Miami, Florida, March 2005
- [19] Yuen S, Li Baochun. Strategyproof mechanisms towards dynamic topology formation in autonomous networks. Mobile Networks and Applications archive, 2005, 10(6): 961-970
- [20] Liu Y H, Xiao L, Liu X M. Location Awareness in Unstructured Peer-to-Peer Systems. IEEE Transactions on Parallel and Distributed Systems, 2004, 16 (2)
- [21] Castro M. Exploiting Network Proximity in Distributed Hash Tables. Microsoft Research. Technical Report. 2002
- [22] The Chord Project. <http://pdos.csail.mit.edu/chord/>
- [23] Zhao B, Kubiawicz J, Joseph A. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report UCB/CSD-01-1141. Computer Science Division, U. C. Berkeley, 2001; 106-115
- [24] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems // Int'l Conf. on Distributed Systems Platforms (Middleware 2001). 2001; 135-141

(下转第 71 页)

- ad hoc sensor radio networks[C]//Proc. Internet Performance Symp IEEE Global Telecommunications Conference (Globecom'01). San Antonio, TX, USA;IEEE, 2001:2944-2948
- [4] Sohrabi K, Gao J, Ailawadhi V, et al. Protocols for self-organization of a wireless sensor network[J]. IEEE Personal Communications Magazine, 2000, 7(5): 16-27
- [5] Mao Jianlin, Wu Xing, Wu Zhiming, et al. A novel energy-aware TDMA scheduling algorithm for wireless sensor networks[C]//First International Conference on Wireless Algorithms, Systems, and Applications (WASA2006). Xi'an, China; Springer-Verlag, 2006: 319-328
- [6] Fan Yong, Zhang Xiaotong, Duan Shihong, et al. Energy Consumption Distribution-aware Node Placement in Wireless Sensor Networks (WSNs) [C] // 2006 International Conference on Wireless Communications, Networking, and Mobile Computing. Wuhan, China; IEEE Comput. Soc, 2007: 4145-4149
- [7] Heinzelman W B, Chandrakasan A P, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks[J]. IEEE Transactions on Wireless Communications, Oct. 2002: 660-670
- [8] Santi P, Simon J. Silence is golden with high probability: maintaining a connected backbone in wireless sensor networks[C]//First European Workshop on Wireless Sensor Networks (EWSN 2004). Berlin, Germany; Springer-Verlag, 2004: 106-121
-
- (上接第 38 页)
- [25] Ratnasamy S, Francis P, Handley M, et al. A Scalable Content-Addressable Network//Proceedings of the ACM SIGCOMM. 2001: 168-175
- [26] Maymounkov P, Mazières D. Kademlia: A peer-to-peer information system based on the XOR metric//Proceedings of IPTPS' 2002. Mar. 2002: 53-65
- [27] Malkhi D, Naor M, Ratajczak D. Viceroy: A scalable and dynamic lookup network//Proceedings of the ACM PODC 2002. ACM Press, 2002: 183-192
- [28] Xu J, Kumar A, Yu X X. On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks. IEEE Journal on Selected Areas in Communications (JSAC), 2004, 22(1)
- [29] Harvey N J, Jones M B, Saroiu S, et al. Skipnet: A scalable overlay network with practical locality properties//Proceedings of USITS' 2003. Mar. 2003: 113-126
- [30] Aberer K, Mauroux P C, Datta A, et al. P-Grid: A Self-organizing Structured P2P System // Proceedings of the ACM SIGMOD2003. ACM Press, 2003: 29-33
- [31] Jagadish H V, Ooi B C, Vu Q H. BATON: A Balanced Tree Structure for Peer-to-Peer Networks//Proceedings of the 31st VLDB Conference. 2005: 661-672
- [32] Kaashoek F, Karger D R. Koorde: A Simple Degree-optimal Hash Table//Proceedings of the IPTPS2003. Feb. 2003
- [33] Fraigniaud P, Gauron P. An overview of the content-addressable networks D2B//Proceedings of the ACM PODC 2003. 2003
- [34] Loguinov D, Kumar A, Rai V, et al. Graph Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault Resilience//Proceedings of the ACM SIGCOMM2003, 2003
- [35] Shen H, Xu C, Chen G. Cycloid: A constant degree and lookup-efficient p2p overlay network//Proceedings of the IPDPS 2004. Santa Fe, New Mexico, Apr. 2004
- [36] Li D S, Lu X C, Wu J. FISSIONE: a scalable constant degree and low congestion DHT scheme based on Kautz graphs//Proceedings of the IEEE INFOCOM2005. 2005
- [37] Ratnasamy S, Shenker S, Stoica I. Routing Algorithms for DHTs: Some Open Questions//Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). 2002
- [38] Xu J, Kumar A, Yu X X. On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks. IEEE Journal on Selected Areas in Communications (JSAC), 2004, 22(1)
- [39] Gummadi K, Gummadi R, Gribble S, et al. The impact of DHT routing geometry on resilience and proximity // Proceedings of ACM SIGCOMM2003. Karlsruhe, Germany. Sep. 2003: 381-394
- [40] Naor M, Wieder U. A Simple Fault Tolerant Distributed Hash Table//Proceedings of the IPTPS2003. Feb. 2003
- [41] Skype. <http://www.skype.com/>.
- [42] Rhea S, Geels D, Roscoe T, et al. Handling churn in a DHT//Proceedings of the USENIX Annual Technical Conference 2004. Boston, USA, 2004
- [43] Li Jinyang, Stribling J, Morris R, et al. Bandwidth-efficient management of DHT routing tables // The Proceedings of the 2nd Symposium on Networked System Design and Implementation (NSDI'05). May 2005
- [44] Aberer K, Mauroux P C, Datta A, et al. P-Grid: A Self-organizing Structured P2P System // Proceedings of the ACM SIGMOD2003. ACM Press, 2003: 29-33
- [45] Zhu Y, Hu Y. Efficient. Proximity-aware load balancing for DHT-based P2P systems[J]. IEEE Trans on Parallel and Distributed Systems, 2005, 16(4): 349-361
- [46] Manku G S. Canada Balanced binary trees for ID management and load balance in distributed hash tables//Proceedings of the twenty-third annual ACM symposium on principles of distributed computing. New York, NY, USA, 2004
- [47] Mitzenmacher M. The Power of Two Choices in Randomized Load Balancing. IEEE Transactions on Parallel and Distributed Systems, 2001, 12(10): 1094-1104
- [48] Dabek F, Kaashoek M F, Karger D, et al. Wide-Area Cooperative Storage with CFS//Proc. 18th ACM Symp. Operating Systems Principles (SOSP). Oct. 2001: 202-215
- [49] Rao A, Lakshminarayanan K, Surana S, et al. Load Balancing in Structured P2P Systems//Proc. Second Int'l Workshop Peer-to-Peer Systems (IPTPS). Feb. 2003: 68-79
- [50] Godfrey B, Lakshminarayanan K, Surana S, et al. Load Balancing in Dynamic Structured P2P Systems//Proc. IEEE INFOCOM. Mar. 2004
- [51] Byers J W, Considine J, Mitzenmacher M. Simple Load Balancing for Distributed Hash Tables // Proc. Second Int'l Workshop Peer-to-Peer Systems (IPTPS). Feb. 2003: 80-87
- [52] Karger D R, Ruhl M. Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems // Proc. Third Int'l Workshop Peer-to-Peer Systems (IPTPS). Feb. 2004