小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems

2015 年 3 月 第 3 期
Vol. 36 No. 3 2015

演化博弈框架下带奖惩的 P2P 激励机制研究

陆 坤,王世宇,谢 玲,李明楚

(大连理工大学 软件学院,大连 116620)
E-mail:mingchul@ dlut. edu. cn

摘 要: P2P 系统中引入激励机制,主要为了促进节点之间相互服务,从而提升系统性能. 现有的基于演化博弈的 P2P 激励机制,采用演化博弈的方法对 P2P 网络中的个体及交易建模,可以更真实反映个体行为变化的规律. 然而完全按照交易历史的服务方式对经常服务的个体不够公平,对经常不提供服务的个体也过于宽容. 本文在当前的激励机制中引入奖惩,通过仿真实验证明,这种带有奖惩的机制可以更好促进节点合作,也可以更快的使得系统达到最优平均收益.

关 键 词: P2P;激励机制;演化博弈;奖惩机制
中图分类号: TP393 **文献标识码:** A **文 章 编 号:** 1000-1220(2015)03-0488-05

Study of Incentive Mechanism with Punish-reward Under Evolutionary Game Framework

LU Kun, WANG Shi-yu, XIE Ling, LI Ming-chu
(School of Software, Dalian University of Technology, Dalian 116620, China)

Abstract: P2P incentive mechanism is aiming at encouraging peers to serve each other, so that the performance of system can be improved. Evolutionary game based incentive mechanism using evolutionary game to model peers and transactions in P2P network. It can have a better description of peers' behaviors changing. However, precisely provide service according to transaction history is unfair to those who always provide service and far too tolerant to those who seldom provide service. In this paper, we bring in punish and reward and prove it through experiments that this mechanism can better help facilitate cooperation and faster achieve best average payoff of the system.

Key words: P2P; incentive mechanism; evolutionary game; punish-reward mechanism

1 引 言

P2P(Peer-to-Peer)因其高度开放性、节点的自治性和匿名性等特性,得以广泛应用于文件共享,流媒体,搜索引擎等多个领域.

P2P 网络的稳定存在严重依赖于节点之间的相互合作. 然而在现实网络中大量存在着的"搭便车(Free-Riding)"行为,它们不遵循"相互协作,服务共享,互惠互利"的 P2P 技术理念,表现出自私,消极,恶意等特征,对系统的性能产生恶劣的影响. Saroiu S 等 2002 年的研究[1]表明,在 Guntella 文件共享系统中,70% 以上的资源仅由 5% 左右的用户提供. 这使得频繁提供服务的节点负载过大,网络的服务质量无法得到保障,造成了"公共品悲剧". 因此,如何有效地促进网络中的节点间相互合作,维护网络的可持续性成为当前 P2P 系统研究的一大挑战.

激励机制是促进系统中节点相互合作,应用最广泛的一种方法. 然而,在当前的研究中,绝大多数成果均按照交易历史或交易行为,采取某个特定概率为节点提供服务. 但是,有文章提出,适当的奖惩可以促进群体中的合作[12]. 因此,本文在激励机制中引入奖惩,建立模型,探讨基于奖惩的激励机制对系统中的合作是如何影响的.

2 相关研究

激励机制借鉴了在管理学中的相关概念,并应用于 P2P 系统中. 它的本质在于为网络中所有的节点提供信息,帮助节点了解向其发起请求者的属性,从而更好的决定是否向其提供服务,间接促使节点为了能够得到别人的服务而向他人提供服务. 从基于微支付的激励机制,到基于信誉系统的激励机制,再到基于经典博弈和演化博弈的激励机制,都在不同程度上实现了 P2P 网络中的合作促进,也都存在着不同的问题,例如违背 P2P 原则,单点失效等. 此外还有一些基于遗传算法的激励机制[2],基于全局信任[3],市场机制[4],基于社会规范[5]等的激励机制也是很多学者关注的研究方向.

目前的 P2P 系统激励研究中主要分为几类:基于微支付的激励机制,基于确定贡献度的激励机制,基于互惠的激励机制,基于演化博弈的激励机制等.

基于微支付的激励机制的核心观点是获取资源的有偿性. 系统要求节点在获取资源时要付给提供者虚拟的货币[6]. 然而这种机制严重依赖于中心节点对于节点的监管,容易导致系统的单点故障. 另外,定价的不公平会导致系统中出现通货膨胀或通货紧缩的现象[7]. 基于确定贡献度的机制要求节点在进入

收稿日期:2014-07-15 收修改稿日期:2014-12-23 基金项目:国家自然科学基金项目(61100194,61272173)资助. 作者简介:**陆 坤**,男,1980 年生,博士研究生,讲师,研究方向为激励机制、演化博弈等;**王世宇**,男,1991 年生,硕士研究生,研究方向为 P2P 激励机制;**谢 玲**,女,1982 年生,硕士,工程师,研究方向为图像处理、信息安全;**李明楚**(通信作者),男,1963 年生,博士,教授,研究方向为图论、信息安全及信任信誉系统.

系统后,必须先贡献一定数量的服务后才能向其他节点请求资源.然而节点在获得了服务使用权后,无法保证节点继续保持合作的行为,如引文[8].基于互惠的激励机制又被称为基于信誉的激励机制.它要求每个节点要知道自己邻居的历史行为,记录包括在一定时间段内邻居对其他节点的合作状况,并根据这些历史记录计算信誉值,再根据这些信任值决定自己对该邻居的行为.例如 Eigen Trust[9] 等算法.

基于演化博弈论的激励机制,采用演化博弈对节点及其之间的交互建模,更真实地反映了个体行为的变化规律.它主要研究理性行为由什么组成;如何将用户由完全理性个体过渡到有限理性个体;如何建立一个通用的模型,分析这种含有激励机制的系统的鲁棒性;如何优化算法,减轻系统中的计算复杂度[10].在每轮博弈时,每个个体会根据对方的行为和自己的策略决定是否提供服务,在每轮博弈结束时,每个个体会根据自己本轮的收益和整个网络的收益状况决定自己下一轮的行为.这种方式可以更好的模拟 P2P 网络中个体行为的变化规律.然而当前的研究中,一般都是以个体交易历史的数据为依据,即历史合作次数和总交易次数的比例作为概率对其提供服务.即当前个体得到服务的概率和其历史合作比例成线性关系.而研究表明,适当的奖惩对群体的合作有促进作用[12],本文给出了一种基于奖惩的激励机制.当用户交易历史中,以 50% 以上的比例进行合作的时候,对其当前得到服务的概率进行放大(奖励),而当历史是以小于 50% 的比例进行合作的时候,对其当前得到服务的概率进行缩小(惩罚).这样可以鼓励合作,促进整个 P2P 系统向着健康的方向发展.

3　奖惩与激励

3.1　激励机制

本文主要研究基于演化博弈的激励机制,考量请求者的交易历史,并以此作为是否为其提供服务的依据.本文以两种激励机制为例,即镜像激励机制(Mirror)和比例激励机制(Prop).

镜像激励机制(Mirror)

这种互惠策略来源于 Feldman 在 2003 年发表的文章[11].当一个镜像互惠策略的个体收到另外一个个体的请求时,它为请求者提供服务的概率和请求者为别人提供服务的概率相同,具体公式如公式(1)所示.

$$P_{Mirror}=\frac{\text{请求者提供服务的次数}}{\text{请求者接到别人请求的次数}} \tag{1}$$

比例互惠机制(Prop)

这个策略与上一种策略的描述类似,也需要了解请求者的交易历史.当一个比例互惠策略的个体收到另外一个个体的请求时,它为请求者提供服务的概率等同于请求者提供服务与消耗服务数量的比值,具体计算见公式(2).

$$P_{Prop}=\min\left(\frac{\text{请求者为别人提供的服务数}}{\text{请求者得到的服务数}},1\right) \tag{2}$$

3.2　奖惩机制

对于自组织的 P2P 网络环境而言,由于无法控制个体在系统中的参与行为,P2P 网络时常面临严重的可用性问题.因此提出一种机制来抑制不良的行为,鼓励合作行为是非常必要的[12].同时,在基于演化博弈的激励机制中,仅仅参考信誉(对应本文的交易历史)会使得系统在策略模仿学习过程中的公平性和合作率降低[13].因此,为了使得系统更好的促进合作的提升,抑制不合作行为,本文在已有的激励机制中引入奖惩机制.即对历史交易中,表现较好的个体进行奖励,对表现不好的个体进行惩罚.在文中,就是以历史交易概率为 0.5 作为基准,历史合作概率高于 0.5 的,对其得到服务的概率进行提升(奖励其合作行为);而低于 0.5 的,对其得到服务的概率进行降低(惩罚其不合作行为).而以前的激励机制的方法均是按照个体历史交易合作情况,进行线性的回馈,不能很好地激励个体向利于合作的方向发展.公式(2)正好满足这个特性,因此本文引入作为奖惩函数.其函数曲线如图 1 所示.使用该奖惩公式,本将上文中提到的两种激励机制修改为带有奖惩的激励策略,如公式(4)和(5)所示.

$$y=(9/4)\hat{}x-1 \tag{3}$$

$$P_{power\_mirror}=\min((9/4)\hat{}P_{Mirror}-1,1) \tag{4}$$

$$P_{power\_prop}=\min((9/4)\hat{}P_{Prop}-1,1) \tag{5}$$

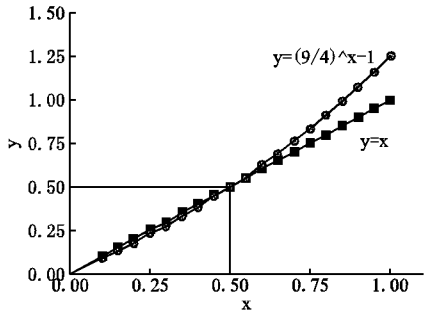

图 1　线性与指数型函数对比

Fig. 1　Linear and power function

3.3　学习模型

本文中,将使用经典的学习模型,以当前最大收益学习模型(CBLM)和机会学习模型(OLM)两种为例.

当前最大收益学习模型(CBLM)

这是一种最简单的学习模型,每轮博弈结束之后,所有的个体以一定概率将自己的策略改变为当前系统中平均收益最高的策略.具体的方法可描述为每轮描述后,个体以适应率 γa 选择是否要改变自己的策略,决定改变自己策略的个体以收益敏感度 γs 和最高平均收益与自己策略的平均收益差值的乘积作为概率(如公式 6).

$$P=\gamma_S*(payoff_{strat_i}-payoff_{best}) \tag{6}$$

将自己的策略改变为当前系统中收益最高的策略.这种学习模型需要有全局变量记录每种策略的平均收益.该学习模型的伪代码如算法 1 所示.

算法 1. CBLM

步骤 1. 以 γa 选择改变自己的策略,若改变,则步骤 2,否则结束

步骤 2. 计算学习概率 p

步骤 3. 以概率 p 学习当前收益最高的策略

机会学习模型(OLM)

每个时间步末尾,每个个体以概适应率 γa 随机选择系统中的另外一个个体作为自己的"老师".如果选择的"老师"和自己采取了不同的策略并且收益比自己高,则会以敏感度 γs

学习它的策略. 在这里,我们仍然采用每种策略的平均收益作为计算学习概率的一个重要标准. OLM 学习模型伪代码如算法 2 所示.

算法 2. OLM

步骤 1. 个体 a 以概率 γa 随机选择老师节点 teacher

步骤 2. 如果 a. strategy ！ = teacher. strategy 并且 a. payoff < teacher. payoff 步骤 3,否则结束

步骤 3. 如果 randf( ) < γs,步骤 4,否则结束

步骤 4. a 改变策略为 teacher 的策略

4 实验及结果分析

4.1 实验说明

4.1.1 基本流程

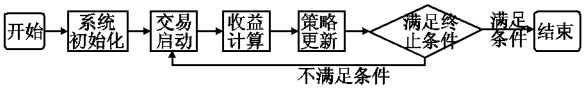

图 2 实验基本流程

Fig. 2 Logic flow of experiments

1) 系统初始化阶段

在系统初始化阶段,系统根据与预先设定的各种策略个体的比例,对网络中每个个体赋予初始策略,并将相应的个体属性进行初始化,网络中其他属性(每种策略的平均收益,各策略个体的数量,以及其他参数)进行初始化.

2) 交易启动阶段

在交易启动阶段,系统中的每个个体均随机向系统中的其他个体(不包括自己)发起资源请求,收到请求的个体根据自身的策略和对方的交易历史,决定是否为对方提供服务. 这些信息均有系统提供. 每个个体需要做 100 次. 在这一阶段,每次交易结束后,系统会记录每个个体发起请求的次数,得到响应的次数,收到请求的次数,提供服务的次数等相关信息.

3) 收益计算阶段

在这一阶段,要计算每种策略的平均收益和系统的总平均收益. 平均收益的计算与每种策略得到服务的概率和提供服务的概率有关.

4) 策略更新阶段

策略更新部分中,我们的实验采用同步更新的方式,即在每个时间步的末尾,所有的个体均根据同样的方式选择更新自己的策略. 本文不考虑个体策略的突变情况.

4.1.2 策略设定

网络中每个个体是有策略的,即当个体收到请求者的请求时,被请求者根据自己的策略进行响应,策略有 C、D 和 R.

1) C(Cooperator)策略:即无条件合作策略. C 策略个体为所有请求者提供服务.

2) D(Defector)策略:即无条件背叛策略. D 策略个体拒绝为任何请求者提供服务. 其中根据前文中对 Free-rider(“搭便车”者)的描述,本文的模型中将策略 D 的个体与现实系统中的 Free-rider 对应.

3) R(Reciprocator)策略:互惠策略,即有条件合作策略,这里指上文中提及的激励策略(后文中统称为 R 策略),采用有条件策略的个体他们对系统中所有的请求予以有条件提供服务. 这些 R 策略包括本文 3.1 和 3.2 中提到的改进前和改进后的镜像与比例激励策略. 在每组实验中采用在实验中我们分别使用上述其中一种激励策略,即 R 策略.

4.1.3 收益计算

在本文中我们计算平均收益,对于不同的策略更新方式我们分别计算每个个体的平均收益,和每种策略的平均收益. 本文严格按照系统中交易的情况计算收益,如公式(7).

$$payoff = alpha * Num\_Recieve/total\_send\_req - beta * Num\_Serve/total\_get\_req \quad (7)$$

该公式的含义是,获得服务的概率和提供服务的概率分别乘以获得的收益 alpha,和相应的消耗 beta,从而求得平均收益. 这里我们将提供服务的消耗 beta 设为 1.

4.1.4 实验参数设定

我们将网络抽象成混合均匀的网络,意味着任意两个个体均可以实现交互,这样 500 个个体的规模交互数量已经很庞大;而且 Nowak 在 Nature 上发表的文章[17],也是以 20,50,100 个节点个体来研究合作问题,因此本文给出 500 个节点的规模网络,进行仿真实验,其结论同样适用于更大规模的 P2P 网络,对试验结论没有定性影响.

在每个时间步内,个体之间进行交互,该过程成为 Warming_Up. 在实验中,我们统计每个时间步结束时系统中各个策略个体所占比例,互惠者之间提供服务的概率以及系统中的平均收益. 由于系统中的 D 策略者过多,会使得 R 策略个体为别人提供服务的总概率降低[14]. 因此过多的背叛者会使得系统中互惠者之间自伤的现象加剧. 同时,在真实的系统中,大部分个体都不会始终保持某种确定的策略不变,一定会参照之前的交互以及自己的感受,作出决定,及大多数人都是有条件合作的. 因此在实验初始条件下,D 策略个体的比例应该较小,互惠者的比例较高.

表 1 基本参数设定

Table 1 Parameters

实验参数	参数值
节点个数	500
Warming_Up 请求数	100 次/节点
演化步数	3000
适应率 γa	0.1
敏感度参数 γs	0.04
初始 C,D,R 比例	(0.2,0.25,0.55)

4.2 实验结果

本文的实验分别采用 CBLM 和 OLM 两种不同的策略学习模型. 因为[15]中将系统的鲁棒性与演化博弈的演化稳定策略结合,定义为我们期望系统达到的最佳策略在演化稳定状态时,该策略个体占群体中的比例. 我们借鉴引文[16]中对鲁棒性评估的方法,将演化稳定状态时合作者与互惠者比例之和作为考量鲁棒性的重要标准. 同时,我们将统的平均收益作为另外一个考量系统鲁棒性的重要标准. 其中,为了能够统一比较在不同 alpha 下系统平均收益的变化,不同将该收益做归一化处理,即

$$payoff_{归一} = payoff/alpha \quad (8)$$

通过这样的方式,尽管随着 alpha 的增大,系统中最大平均收

益与 alpha 的比值在增大,但我们仍然可以从大体上看出在不同的 alpha 下,收益的变化情况.

如图 3 所示,当系统中获得服务的收益 alpha 与提供服务的消耗 beta(=1)相同时,由于困境的存在,系统最终会全部演化为背叛者,而当 alpha 与 beta 的比值较大的时候,困境变弱,此时系统会朝着促进合作的方向演化. 但是不同的机制表

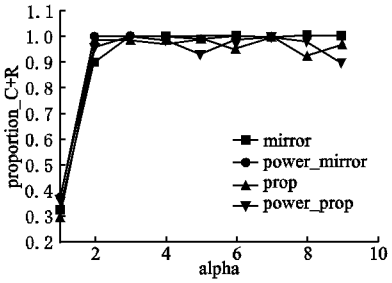

图 3　不同的 alpha 采用 CBLM 演化稳定时 C + R 的比例

Fig. 3　Proportion of C + R under CBLM with different alpha

现的鲁棒性不同. 从图 3 可以看到,改进的 power_mirror 策略会比原有的 mirror 策略在更小的 alpha 下达到最佳的稳定状态,即网络中全部都是合作者与互惠者,这是因为加入了奖惩机制后,有效的抑制了互惠者之间的自伤行为,使得互惠者可以获得更多的收益,避免向其他策略演化,而 mirror 的个体对于背叛者可以起到很好的抑制作用. 因此,我们在图 4 中可以看到,系统在达到稳态时使用引入奖惩的 power_mirror 策略会更多. 从图 5 中也可以看到,改进后的策略可以更快的达到在该 alpha 下的最优收益.

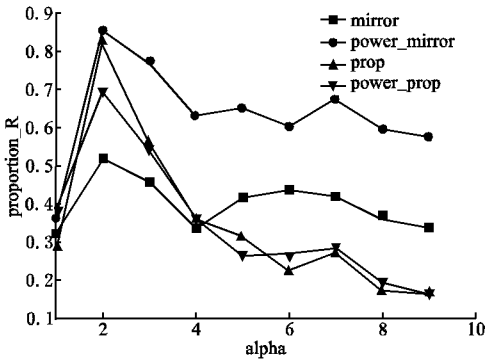

图 4　不同的 alpha 采用 CBLM 演化稳定时 R 的比例

Fig. 4　Proportion of R under CBLM with different alpha

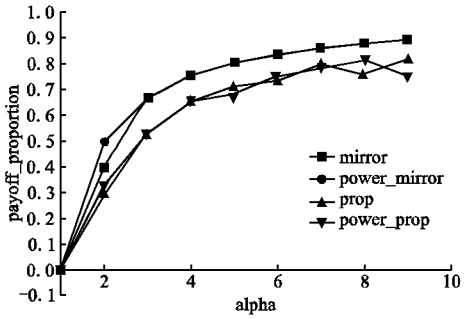

图 5　不同的 alpha 采用 CBLM 演化稳定时的系统平均收益

Fig. 5　Average payoff under CBLM with different alpha

对于比例互惠策略,改变前后的现象相差不大. 这与该机制本身有关. 这种策略考虑请求者的提供服务与得到服务的比值. 对于互惠者来说,无论系统中的背叛者有多少,自身为其他人提供服务和得到服务的比值相对稳定,接近于 1,即互惠者之间总是相互提供服务. 因此,引入了奖惩机制对其影响不大,从图 3 到图 5 的现象中也可以看到,prop 与 power_prop 两条曲线接近,且互有高低.

在 OLM 学习模型下重复同样的实验,我们得到了与 CBLM 下相似的结论. 尽管由于学习方式的变化,使得不同状态下系统在稳态时达到的各项指标接近(图 6,图 8),但是仍然可以看出引入了奖惩机制的激励机制,对于保证系统快速到达最优稳定状态,即系统中全部为C和R个体以及最优平

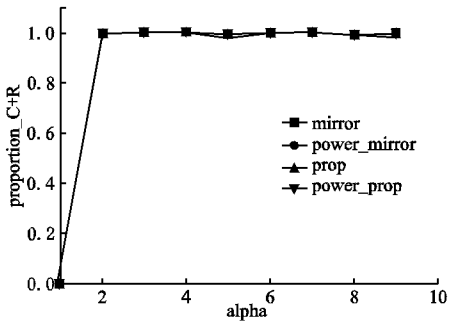

图 6　不同的 alpha 采用 CBLM 演化稳定时 C + R 的比例

Fig. 6　Proportion of C + R under OLM with different alpha

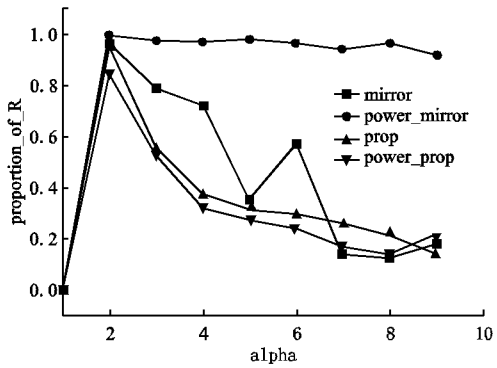

图 7　不同的 alpha 采用 CBLM 演化稳定时 R 的比例

Fig. 7　Proportion of R under OLM with different alpha

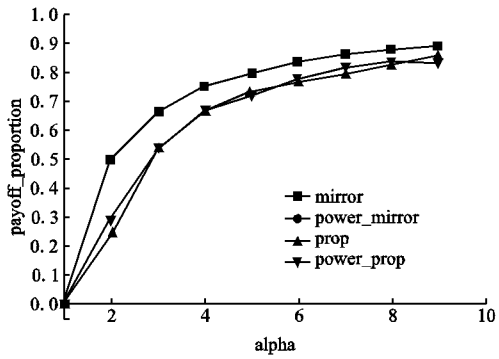

图 8　不同的 alpha 采用 CBLM 演化稳定时的系统平均收益

Fig. 8　Average payoff under OLM with different alpha

均收益稍有优势. 同时,采用这样的方式可以较好的避免互惠者之间的自伤,提高互惠者之间相互提供服务的概率(图 7).

5 结 论

本文针对 P2P 网络中的“搭便车”问题,改进了现有的激励机制,将奖惩机制与基于演化博弈的激励机制相结合,构建了一个更加完整的 P2P 网络激励模型. 该模型继承了传统的激励机制的优点,并进行了完善和改进. 通过仿真实验,我们证明了该模型能够有效的提高系统中的合作频率,同时在系统演化过程中,使得系统能更快的达到最优稳定状态,和最优平均收益.

References:

[1] Saroiu S, Gummadi P K, Gribble S D. Measurement study of peer-to-peer file sharing systems[C]. Electronic Imaging, International Society for Optics and Photonics, 2001: 156-170.

[2] Lin Hai, Wu Chen-xu. Repeated prisoner's dilemma game strategy based on genetic algorithm evolution in complex networks [J]. Physics, 2007, 56(8): 4313-4318.

[3] Ouyang Jing-cheng, Lin Ya-ping, Zhou Si-wang, et al. Global trust value based incentive mechanism in P2P networks [J]. Journal of System Simulation, 2013, 25(5): 1046-1052.

[4] Guo Dong, Lu Shan, Yin Bao-qun. A novel incentive model of market mechanisms P2P file sharing system [J]. Journal of Chinese Computer Systems, 2012, 33(1): 1-6.

[5] Liao Xin-kao, Wang Li-sheng. Research incentives based on social norms and guidelines boycott nodes[J]. Computer Science, 2014, 41(4): 28-30, 35.

[6] Golle P, Leyton Brown K, Mironov I, et al. Incentives for sharing in peer-to-peer networks [M]. Electronic Commerce, Springer Berlin Heidelberg, 2001: 75-87.

[7] Irwin D, Chase J, Grit L, et al. Self-recharging virtual currency[C]. Proceedings of the 2005 ACM SIGCOMM workshop on Economics of Peer-to-Peer Systems, ACM, 2005: 93-98.

[8] Antoniadis P, Le Grand B. Incentives for resource sharing in self-organized communities: From economics to social psychology [C]. Digital Information Management 2nd International Conference on (ICDIM'07), IEEE, 2007, 2: 756-761.

[9] Kamvar S D, Schlosser M T, Garcia Molina H. The eigentrust algorithm for reputation management in p2p networks[C]. Proceedings of the 12th International Conference on World Wide Web, ACM, 2003: 640-651.

[10] Palomar E, Alcaide A, Ribagorda A, et al. The peer's dilemma: a general framework to examine cooperation in pure peer-to-peer systems[J]. Computer Networks, 2012, 56(17): 3756-3766.

[11] Feldman M, Chuang J. Overcoming free-riding behavior in peer-to-peer systems[J]. ACM SIGecom Exchanges, 2005, 5(4): 41-50.

[12] Gui Chun-mei, Jian Qiang, Wang Huai-min, et al. Repeated game theory based penalty-incentive mechanism in Internet-based virtual computing environment[J]. Journal of Software, 2010, 21(12): 3042-3055.

[13] Sigmund K, Hauert C, Nowak M A. Reward and punishment[J]. Proceedings of the National Academy of Sciences, 2001, 98(19): 10757-10762.

[14] Zhao B Q, Lui J C S, Chiu D M. A mathematical framework for analyzing adaptive incentive protocols in P2P networks [J]. IEEE/ACM Transactions on Networking (TON), 2012, 20(2): 367-380.

[15] An Y L J Z B, Sen S. A practical robustness measure of incentive mechanisms[J]. 2014.

[16] Liu Y, Zhang J. Robustness evaluation of incentive mechanisms [C]. Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2013: 1293-1294.

[17] Nowak M A, Sigmund K. Evolution of indirect reciprocity by image scoring[J]. Nature, 1998, 393(6685): 573-577.

附中文参考文献:

[2] 林 海,吴晨旭. 基于遗传算法的重复囚徒困境博弈策略在复杂网络中的演化[J]. 物理学报,2007,56(8):4313-4318.

[3] 欧阳竟成,林亚平,周四望,等. P2P 网络中基于全局信任值的激励机制[J]. 系统仿真学报,2013,25(5):1046-1052.

[4] 郭 东,芦 珊,殷保群. 一种新颖的市场机制的 P2P 文件共享系统的激励模型[J]. 小型微型计算机系统,2012,33(001):1-6.

[5] 廖新考,王力生. 基于社会规范准则和联合抵制的节点激励机制研究[J]. 计算机科学,2014,41(4):28-30,35.

[12] 桂春梅,蹇 强,王怀民,等. 虚拟计算环境中基于重复博弈的惩罚激励机制[J]. 软件学报,2010,21(12):3042-3055.