Simulation-Based Evaluation of P2P File-Sharing Systems under Heterogeneous Environments: Evolutionary Game Theoretic Approach

Yusuke Matsuda†, Masahiro Sasabe†, and Tetsuya Takine†

†Department of Information and Communications Technology, Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita 565-0871, Japan
Email: yusuke@post.comm.eng.osaka-u.ac.jp, {sasabe, takine}@comm.eng.osaka-u.ac.jp

Abstract—In Peer-to-Peer (P2P) file-sharing systems, cooperative file-caching improves the file availability. Users have different degree of cooperativity in caching files and they are in different surrounding environments arising from the topological structure of the P2P network. In this paper, with evolutionary game theory, we experimentally evaluate the performance of P2P file sharing systems in such heterogeneous environments. Simulation results show that the environmental heterogeneity contributes to conservation of the file availability and stability of the system.

1. Introduction

In Peer-to-Peer (P2P) file sharing systems, cooperative file-caching improves the file availability which is defined as the ratio of the number of nodes caching the file to the total number of nodes. Note that file caching carries costs such as storage consumption and processing load. As a result, the file availability deteriorates with the emergence of free-riders which only download files and never upload any files [1]. The cooperative degree in caching files is heterogeneous among users because the terminal performance, access link capacity, and sense of value to the file are different among users. Furthermore, each user has a different surrounding environment depending on the structure of the P2P network.

With evolutionary game theory [2], we experimentally try to reveal the performance of P2P file sharing systems in such heterogeneous environments. Evolutionary game theory is a framework to figure out the phenomena caused by mutual interactions among many heterogeneous individuals which compose the society. In evolutionary game theory, the mutual interaction is modeled by a game and the dynamics of the game is described in a mathematical form (*micro-macro dynamics* [3]) or a procedure (*agent-based dynamics* [4]). In recent years, several researchers have applied evolutionary game theory to revealing what mechanisms lie behind the emergence of the cooperative behavior in the human society and cooperative P2P networks [5, 6].

Our research group also applied micro-macro dynamics to P2P file sharing systems by modeling bargains about file caching among nodes as games [7]. Through theoretic analysis, we found that the number of cooperative nodes changes according to nonlinear dynamics. Note that in Ref. [7], bargains about file caching among nodes were assumed to be synchronized and nodes directly communicate with each other. In this paper, we evaluate the system performance through simulation experiments based on agent-based dynamics, and reveal how the synchronization property and the locality of information which nodes can obtain affect the system performance.

The rest of the paper is organized as follows. In Section 2, we explain the system model considered in this paper. We describe the overview of evolutionary game theory in Section 3 and present the result of simulation experiments based on agent-based dynamics in Section 4. Finally, we conclude this paper in Section 5.

2. System Model

2.1. Overview

We describe the P2P file sharing system considered in this paper. There are N nodes in the system. Each node first obtains a set of file holders (i.e., nodes that have the desired file), using one of existing search methods. It then retrieves the desired file from one or more nodes in the set. After retrieval, it can become a new holder of the file. Because costs accompany file caching, each node bargains with other file holders about whether or not to keep caching the file. In this paper, we model bargains among file holders as caching games under the presupposition that each file holder plays caching games with other file holders successively. Each node determines if it keeps caching the file, based on the results of a series of games. For simplicity, we focus on a single file and assume that any node in the system can be a file holder. Note that the result in this paper is applicable to the situation that there are many files and nodes make independent decisions of caching each file.

2.2. Caching Game

According to Ref. [7], we briefly introduce the caching game. To model the caching game, we first define a payoff matrix that determines the relationship between node's behavior (strategy) and payoff obtained by the strategy. In the P2P file sharing system, each node has two strategies: Caching ($\mathcal{S}_c$) and no caching ($\mathcal{S}_n$). The node with $\mathcal{S}_c$ (resp.

Table 1: Payoff matrix with user heterogeneity.

player i \ player j	cooperator	defector
cooperator	(R_i, R_j)	(S_i, T_j)
defector	(T_i, S_j)	(P_i, P_j)

$\mathcal{S}_\mathrm{n}$) corresponds to the cooperator (resp. defector). We define the payoff matrix between two noes as Table 1, where we assume $T_k > R_k \geq S_k > P_k$ for any player k. Table 1 indicates that the parameters of the payoff matrix can vary from node to node and selecting different strategy from the opponent's strategy yields more payoff.

As mentioned above, every node decides whether to cache the file after a certain number of caching games with other nodes. At each game, the node rationally behaves: It compares the expected payoff obtained by strategy $\mathcal{S}_\mathrm{c}$ and that obtained by strategy $\mathcal{S}_\mathrm{n}$, and selects the strategy which acquires more payoffs. In what follows, we describe the detail of the strategy selection.

We denote the ratio of the number of cooperators, i.e., nodes selecting $\mathcal{S}_\mathrm{c}$, to the total number of nodes at time t by $p(t)$. The expected payoff $U_i(\mathcal{S}_\mathrm{c})$ (resp. $U_i(\mathcal{S}_\mathrm{c})$) that node i obtains when selecting $\mathcal{S}_\mathrm{c}$ (resp. $\mathcal{S}_\mathrm{n}$) is given by

$$U_i(\mathcal{S}_\mathrm{c}) = p(t)R_i + (1 - p(t))S_i, \qquad (1)$$

$$U_i(\mathcal{S}_\mathrm{n}) = p(t)T_i + (1 - p(t))P_i. \qquad (2)$$

Node i selects the strategy with higher expected payoff as the next strategy after comparing $U_i(\mathcal{S}_\mathrm{c})$ and $U_i(\mathcal{S}_\mathrm{n})$:

$$X_i(t+1) = \begin{cases} \mathcal{S}_\mathrm{c}, & U_i(\mathcal{S}_\mathrm{c}) > U_i(\mathcal{S}_\mathrm{n}), \\ \mathcal{S}_\mathrm{n}, & U_i(\mathcal{S}_\mathrm{c}) \leq U_i(\mathcal{S}_\mathrm{n}), \end{cases} \qquad (3)$$

where $X_i(t+1)$ denotes the strategy of node i at time $t+1$. With Eqs. (1) and (2), Eq. (3) is rewritten to be

$$X_i(t+1) = \begin{cases} \mathcal{S}_\mathrm{c}, & p(t) < \theta_i, \\ \mathcal{S}_\mathrm{n}, & p(t) \geq \theta_i, \end{cases} \qquad (4)$$

where $\theta_i = (S_i - P_i)/(T_i - R_i + S_i - P_i)$. Note that the threshold θ_i ($0 < \theta_i < 1$) represents the degree of cooperativity of node i, and a larger θ_i implies node i being more cooperative.

2.3. Heterogeneous Environments

We consider two kinds of heterogeneity: User heterogeneity and network heterogeneity. The user heterogeneity stands for the difference in the cooperative degree in file caching among users. On the other hand, the network heterogeneity arises from the topological structure of P2P file-sharing networks.

2.3.1. Cooperative Degree to Caching

In practical P2P file sharing systems, it is reasonable to assume that uncooperative nodes constitute majority in the system because costs accompany file caching. Thus we assume that the distribution of θ_i follows Zipf's law. More specifically, we assume that θ_i's ($i = 1, 2, \ldots, N$) are independent and identically distributed according to a truncated Pareto distribution with shape parameter -1 and support $[1/K, 1]$. Let $f(\theta)$ denote the probability density function of θ_i. We then have

$$f(\theta) = \frac{1}{\log K} \cdot \theta^{-1}. \qquad (5)$$

Note here that K is closely related to varieties of the threshold and the users' sense of value becomes more heterogeneous with the increase of K.

2.3.2. Topological Structure

The topological structure determines the locality of information that each node can obtain through interactions with other nodes. The larger the hop count between two nodes is, the more they require communication overheads in terms of the number of messages transferred. In general, a topological structure is determined by rules for updating links between nodes and the number of links per node, i.e., node degree. In this paper, we use two kinds of topologies: Full-mesh and scale-free networks. The full-mesh network allows all nodes to directly communicate with each other. This is the same situation considered in [7].

The full-mesh network may not be realistic in large-scale systems. To gain a deep insight about the impact of the topological structure, we also evaluate the system performance in scale-free networks based on Barabási-Albert (BA) model [8]. The scale-free network is a network with a degree distribution following a power law, i.e., $p(k) \propto k^{-\gamma}$, where $p(k)$ denotes the probability that a node has degree k and γ is a constant. In the scale-free network, there are a small number of high-degree nodes and a large number of low degree nodes.

3. Evolutionary Game Theory

Evolutionary game theory originally aims to figure out a mechanism in which optimum behaviors are inherited to offspring in the evolutionary process of organisms. It provides us with two approaches for this purpose: Micro-macro dynamics and agent-based dynamics. With micro-macro dynamics, Ref. [7] analytically revealed the system characteristics of the P2P file-sharing system under heterogeneous environments: The equilibrium of $p(t)$ exists for $K > 1$ and it is stable if $K > e^e$ (≈ 15.2).

Note here that the micro-macro dynamics is based on the global information $p(t)$. In actual systems, however, it may be difficult for each node to obtain up-to-date $p(t)$ due to communication overheads. Agent-based dynamics complements this shortcoming by introducing a network for interactions among nodes and modeling the decision making of nodes based only on local interactions with neighboring nodes. As a result, agent-based dynamics models a phenomenon that a superior strategy spreads over the network in a hop-by-hop manner. In what follows, we give the detail of agent-based dynamics for P2P file-sharing systems.

We define a *generation* as the interval of strategy updates. We consider two kinds of systems, *synchronous* and *asynchronous* systems. In the synchronous system, all nodes update their strategies simultaneously at the beginning of each generation, while only one node, chosen randomly from the population, updates its strategy in the asynchronous system.

In the gth ($g = 1, 2, \ldots$) generation, node i (which is eligible for updating its strategy) plays a game once with every neighboring node and obtain the ratio $p_i(g)$ of cooperators including itself. I then determines the strategy $X(g+1)$ at the $g+1$st generation as follows:

$$X_i(g+1) = \begin{cases} \mathcal{S}_{\rm c}, & p_i(g) < \theta_i, \\ \mathcal{S}_{\rm n}, & p_i(g) \geq \theta_i. \end{cases} \quad (6)$$

Note that the initial distribution of strategies is arbitrary and its influence on the system performance will be discussed in Section 4.

4. Simulation Results

In this section, we reveal how the synchronization among nodes and the topological structure affect the system performance.

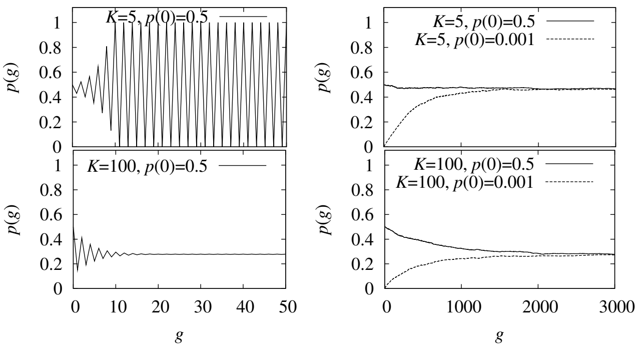
4.1. Simulation Model

We use NetLogo [9] in our simulation experiments. We prepare full-mesh and scale-free networks of 1,000 nodes. Scale-free networks are generated based on Barabási-Albert (BA) model, where the number m of links that a newly participating node will establish is set to be 1, 2, or 4. We use the topology generator BRITE [10] to generate scale-free networks.

We assign the threshold θ_i that represents the cooperative degree in file caching of node i as follows: We first generate 1,000 samples of thresholds by Eq. (5) and then we randomly assign them to nodes. We configure that the initial ratio of the number of cooperators to the total number of nodes is 0.5, unless otherwise stated. We prepare ten independent samples of networks for each K, and we will show only the average of those ten independent simulation results for each K.

4.2. Impact of Synchronization

Figs 1(a) and 1(b) illustrate the transient behavior of the synchronous and asynchronous systems, respectively, in the full-mesh network. Both results for $K = 5$ and $K = 100$ in the synchronous system are the same as those in the micro-macro dynamics in Ref. [7]. The reason is that the synchronous system is configured according to the same assumption as in the micro-macro dynamics. On the other hand, the asynchronous system converges even in the unstable case ($K = 5$) of the micro-macro dynamics, independent of the initial ratio of cooperators. Since only one node updates per generation in the asynchronous system, global synchronization among nodes is avoided, and thus the ratio of cooperators does not oscillate.



(a) Synchronous system. (b) Asynchronous system.

Figure 1: Transient behavior (full-mesh network).

Next, we turn our attention to the speed of convergence. Note that all 1,000 nodes update their strategies at the beginning of each generation in the synchronous system, whereas only one node updates its strategy at each generation in the asynchronous system. Comparing the lower panels of Fig. 1(a) with Fig. 1(b), we find that the number of occasions to update strategies required to reach the steady state in the synchronous system is approximate five times as large as that in the asynchronous system: The synchronous and asynchronous systems require about 15,000 and 3000 generations, respectively. The fast convergence in the asynchronous system arises from avoiding the oscillation of the ratio of cooperators.

From the above results, we conclude that the file availability is independent of the synchronization and the asynchronous system has more stable and faster transient behavior than the synchronous system. Since the asynchronous system is more realistic than the synchronous system, we further examine the characteristics of the asynchronous system in the following.

4.3. Impact of Topological Structure

We define the file availability as $p(g)$ for $g = 4,000$, at which the system is in steady state, and search latency as the smallest number of hops to reach a cooperator. Fig. 2 illustrates the relationship between the user heterogeneity and the file availability in the asynchronous system in scale-free networks with $m = 1, 2, 4$. We observe that for a fixed m, the scale-free network achieves higher file availability than the full-mesh network. We also observe that a small m leads to a high value of the file availability. To understand these phenomena, we further examine the distribution of cooperators in the network in steady state.

Fig. 3 shows the relationship between K and the search

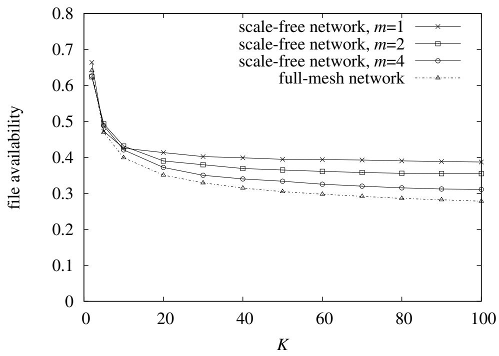


Figure 2: Relationship between K and file availability (full mesh and scale-free networks).

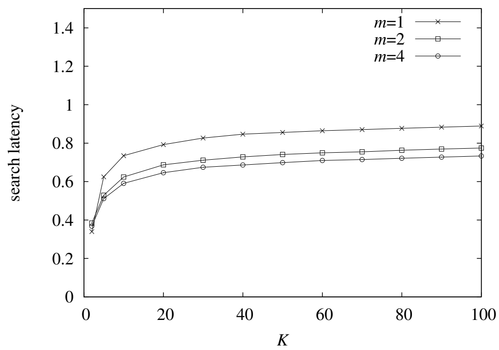


Figure 3: Relationship between K and search latency (scale-free networks).

latency in scale-free networks. We find that a randomly chosen node can reach a cooperator within one hop on average even when $K = 100$ where the file availability is about 0.3. The reason is that due to the game structure, each node tries to acquire higher payoff by selecting the different strategy of the opponent and prefers being a defector. As a result, defectors gather around each cooperator. Note also that a large m yields more defectors because of the increase of neighboring nodes.

We also observed that the topological structure hardly affects the convergence properties, which is omitted due to the shortage of space.

5. Conclusions

With evolutionary game theory, we discussed how selfish and autonomous users' behavior in P2P file sharing systems affects the system performance. In particular, we focused on the heterogeneous users' sense of value to file caching and the heterogeneous topological structure of the P2P network. From the simulation results, we obtained following two characteristics: The asynchronous system is more stable than the synchronous system, and compared with the topological structure, the game structure has a greater impact on both the file availability and the distribution of cooperators in the system.

Acknowledgements

This research was supported in part by a Grant-in-Aid for Young Scientists (B) 20700061, and "Global COE (Centers of Excellence) Program" of the Ministry of Education, Culture, Sports, Science and Technology in Japan.

References

[1] E. Adar and B. A. Huberman, "Free Riding on Gnutella," *Technical Report Xerox PARC*, 2000.

[2] J. W. Weibull, *Evolutionary Game Theory*. The MIT Press, 1997.

[3] M. Granovetter, "Threshold Models of Collective Behavior," *The American journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, May 1978.

[4] G. Szabó and G. Fáth, "Evolutionary Games on Graphs," *Physics Reports*, vol. 446, pp. 97–216, Jul. 2007.

[5] F. Fu, L. Liu, and L. Wang, "Evolutionary Prisoner's Dilemma on Heterogeneous Newman-Watts Small-World Network," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 56, no. 4, pp. 367–372, 2007.

[6] D. Hales and S. Arteconi, "SLACER: A Self-Organizing Protocol for Coordination in Peer-to-Peer Networks," *IEEE Intelligent Systems*, vol. 21, no. 2, pp. 29–35, 2006.

[7] M. Sasabe, Y. Matsuda, and T. Takine, "How Does User Heterogeneity Affect Performance of P2P Caching?: Evolutionary Game Theoretic Approach," in *Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems*, 2008.

[8] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, 1999.

[9] U. Wilensky, "NetLogo," 1999, available at http://ccl.northwestern.edu/netlogo/.

[10] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: An Approach to Universal Topology Generation," in *Proceedings of the Ninth International Symposium in Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'01)*. Washington, DC, USA: IEEE Computer Society, 2001, pp. 346–354.