Research Article

Evolutionary Game Theory-Based Evaluation of P2P File-Sharing Systems in Heterogeneous Environments

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*

Correspondence should be addressed to Masahiro Sasabe, sasabe@comm.eng.osaka-u.ac.jp

Received 28 July 2009; Accepted 19 November 2009

Academic Editor: Winston Seah

Copyright © 2010 Yusuke Matsuda et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-to-Peer (P2P) file sharing is one of key technologies for achieving attractive P2P multimedia social networking. In P2P file-sharing systems, file availability is improved by cooperative users who cache and share files. Note that file caching carries costs such as storage consumption and processing load. In addition, users have different degrees of cooperativity in file caching and they are in different surrounding environments arising from the topological structure of P2P networks. With evolutionary game theory, this paper evaluates the performance of P2P file sharing systems in such heterogeneous environments. Using micro-macro dynamics, we analyze the impact of the heterogeneity of user selfishness on the file availability and system stability. Further, through simulation experiments with agent-based dynamics, we reveal how other aspects, for example, synchronization among nodes and topological structure, affect the system performance. Both analytical and simulation results show that the environmental heterogeneity contributes to the file availability and system stability.

1. Introduction

File sharing is one of key technologies for achieving attractive multimedia social networking. Each user treasures multimedia contents, for example, video and music, in his/her own terminal and they are shared with other users. Since the number of users continues to increase and user terminals become more heterogeneous with the development of mobile technologies, it is desirable to achieve file sharing in a form of peer-to-peer (P2P) systems, which are self-organized and scalable without any specific infrastructure.

To realize effective P2P file-sharing, two key issues should be considered: User selfishness and heterogeneity. Since each node participating in the system is a user's terminal, it is controlled by the user. In general, each user is selfish and hesitates to actively share files due to costs for file caching, such as storage consumption, processing load, and bandwidth consumption [1]. On the other hand, the heterogeneity of user environments can be classified into self environments and surrounding environments. Each user potentially has a cooperative degree in file caching,

depending on its terminal performance, access link capacity, and sense of value to the file. On the other hand, the surrounding environment is represented as the connectivity relationship between nodes, which is determined by the topological structure of P2P networks. Each user has to determine its behavior so as to maximize its own benefit, taking account of the behavior of its neighboring nodes. These user environments are unique to each user and become more heterogeneous with the increase of nodes and the development of mobile technologies. In this paper, we reveal how these heterogeneity affects the performance of P2P file-sharing systems.

For this purpose, we utilize evolutionary game theory, which is a framework to examine what kinds of phenomena emerge from mutual interactions among many and heterogeneous individuals [2]. Evolutionary game theory was originally devised to reveal the mechanism that superior genes with high fitness for the environment are inherited from ancestors to offspring, through competition among individuals in the evolutionary process of organisms. In P2P file-sharing systems, the system behavior emerges from

interactions among potentially selfish nodes. In view of a similarity between biological societies and P2P file-sharing systems, we apply evolutionary game theory to revealing how the selfish nodes' behavior affects the performance of the whole system under heterogeneous environments.

Evolutionary game theory provides us with both theoretical and simulation-based approaches, that is, micro-macro dynamics [3] and agent-based dynamics [4]. Micro-macro dynamics mathematically describes the dynamic behavior of a micro-macro loop that defines interactions between micro behaviors and macro behaviors, where micro behavior and macro behavior are regarded as actions of heterogeneous individuals and the collective behavior resulting from mutual interactions among individuals, respectively. Thus micro-macro dynamics enables us to analytically reveal how interactions about file caching among nodes affect the caching condition of the whole system. In particular, it is a great advantage of micro-macro dynamics that the system performance and stability condition are expressed in terms of system parameters, including the degree of user selfishness.

Note that micro-macro dynamics assumes each node to be able to know the behavior of all other nodes, even though it may be difficult in large-scale networks. Agent-based dynamics can complement this shortcoming by introducing a network for communication and modeling the node behavior based only on local interactions between adjacent nodes in the network. It models a phenomenon that a superior strategy spreads over the network in a hop-by-hop manner. Consequently, we can find out the system characteristics in various network topologies, where nodes behave selfishly based on their local information. In addition, with agent-based dynamics, we can also examine the impact of synchronization/asynchronization among nodes, the distribution of cached files, and the impact of high-degree nodes' characteristics and behavior on the overall performance.

We model interactions about file caching among nodes as a caching game between two nodes, where payoff matrices are different among nodes. We then show that the cooperative degree of each node in file caching can be represented by a single parameter (called the threshold) deduced from its payoff matrix. We assume that threshold values of nodes follow Zipf's law, and represent the user heterogeneity in terms of this parameter. Applying evolutionary game theory to this model, we reveal the performance of P2P file-sharing systems in heterogeneous environments.

The rest of the paper is organized as follows. In Section 2, we introduce several research backgrounds and related works. We explain the system model in Section 3 and describe the overview of evolutionary game theory in Section 4. Section 5 provides theoretical analysis of micro-macro dynamics and Section 6 shows the result of simulation experiments with agent-based dynamics. Finally, we conclude this paper in Section 7.

2. Related Works

There are several studies on existing P2P file-sharing systems such as Napster [5], Gnutella [6], KaZaA [7], and BitTorrent

[8], based on real measurement [1, 9–12]. They revealed the system characteristics, for example, the system scale, the number and kinds of shared files, and the distributions of available bandwidth, search latency, and life time. Furthermore, some of them focused on user selfishness [1, 9, 12]. Those works discussed the characteristics about nodes' cooperative degree to file sharing in terms of the number of cached files per node and access link capacity declared to other nodes, and they insisted that there are free riders that only download files.

The increase of free-riders deteriorates the file availability, which is one of the most important performance metric for each user. For a specific file, the file availability is defined as the ratio of the number of nodes caching the file to the total number of nodes in the system. The file availability should be positive, otherwise nodes could not retrieve the file. Note that users may hesitate to cooperatively cache files due to the cost for file caching, such as storage consumption, processing load, and bandwidth consumption. This would be the major factor in the increase of free riders. The impact of the user selfishness on the system performance, however, have not been figured out completely.

In recent years, several researchers applied game theory to studying the relationship between the user selfishness and the system performance [13, 14]. They insisted that the free riders can be reduced by introducing incentive mechanisms, where in compensation for caching files, nodes can obtain payments and/or reputations from other nodes retrieving them. Such incentive mechanisms, however, require exchange of correct information about the payments and/or reputations, and thus they are not applicable to such systems as Freenet [15] that provides nodes with anonymity in communication.

The relationship between the user selfishness and the cooperative behavior among users has also been attracting attention in research fields other than information networking. Refs [16–18] employ evolutionary game theory for studying the human society consisting of heterogeneous individuals and explore what mechanisms lie behind the emergence of the cooperative behavior under temptation to be selfish. They found that the emergence of the cooperative behavior are closely related to the topological structure of the human society, users' sense of value to cooperativity, and so on.

With the help of these knowledge, several researchers have been applying evolutionary game theory to the control of information networks and the performance evaluation of information systems [19, 20]. Hales and Arteconi proposed SLACER algorithm that controls the topological structure of P2P networks, based on evolutionary game theory [19]. In SLACER, each node updates its neighboring nodes so as to maximize its own payoff, under the assumption that cooperative groups can yield more payoffs than uncooperative groups. Consequently, multiple groups of cooperative nodes are constructed, which look like a protein-structure topology [21].

Using evolutionary game theory, Sasabe et al. evaluated the performance of P2P file sharing systems [20]. They modeled interactions about file caching among nodes as

TABLE 1: Payoff matrix.

Player 1	Player 2	
	Cooperator	Defector
cooperator	(R, R)	(S, T)
defector	(T, S)	(P, P)

a caching games. Through theoretical analysis and simulation experiments, they showed the game structure for maintaining the positive file availability. Note that they considered only the situation that the benefit and the cost for file caching are identical among nodes. In practice, however, they can vary from node to node.

3. System Model

3.1. 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 communicates with other file holders about whether or not to keep caching the file. In this paper, we model interactions 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 that make independent decisions of caching each file.

3.2. Caching Game

3.2.1. Strategy and Payoff. Table 1 shows a general payoff matrix between two players used in game theory. A defector exploiting a cooperator obtains T and the exploited cooperator receives S . Both players receive R (P) when they cooperate (defect) each other. Prisoner's dilemma game ($T > R > P > S$, $2R > T + S$) and snowdrift game ($T > R > S > P$) are well-known examples.

In P2P file sharing systems, each node has two strategies: caching ($\mathcal{S}_c$) and no caching ($\mathcal{S}_n$). The node with $\mathcal{S}_c$ (resp., $\mathcal{S}_n$) corresponds to a cooperator (resp., defector). The work in [20] shows that cooperative caching can be accomplished if $T > R \geq S > P$.

The parameters of the payoff matrix can vary from node to node in practice, because a large variety of users can participate in P2P file sharing systems and the sense of value to a specific file is different among users. In order to represent such a user heterogeneity, we extend Tables 1 and 2, where we assume $T_k > R_k \geq S_k > P_k$ ($k = 1, 2, \dots, N$) for each node k . As a result, our game model can be categorized into an asymmetric complementarity game with complete

TABLE 2: Payoff matrix with user heterogeneity ($T_k > R_k \geq S_k > P_k$).

Player i	Player j	
	Cooperator	Defector
cooperator	(R_i, R_j)	(S_i, T_j)
defector	(T_i, S_j)	(P_i, P_j)

information for two players in which the payoff matrix is asymmetric, taking different strategy of the opponent yields better payoffs, and the player has complete information about the payoff matrix.

3.2.2. Strategy Selection. 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 selects one of strategies $\mathcal{S}_c$ and $\mathcal{S}_n$, which acquires more payoff than the other. In what follows, we describe the detail of the strategy selection.

We define $p(t)$ as the ratio of the number of cooperators (i.e., nodes selecting $\mathcal{S}_c$) to the total number N of nodes in the systems at time t . The expected payoff $U_i(\mathcal{S}_c)$ that node i ($i = 1, 2, \dots, N$) obtains when selecting $\mathcal{S}_c$ is given by

$$U_i(\mathcal{S}_c) = p(t)R_i + (1 - p(t))S_i. \quad (1)$$

Similarly, we have the expected payoff $U_i(\mathcal{S}_n)$ that node i obtains when selecting $\mathcal{S}_n$:

$$U_i(\mathcal{S}_n) = p(t)T_i + (1 - p(t))P_i. \quad (2)$$

Node i then selects the strategy with higher expected payoff. When $U_i(\mathcal{S}_c) = U_i(\mathcal{S}_n)$, node i is assumed to select $\mathcal{S}_n$ so that we evaluate the system performance in a less cooperative situation. Thus the strategy $X_i(t+1)$ of node i at time $t+1$ is given by

$$X_i(t+1) = \begin{cases} \mathcal{S}_c, & U_i(\mathcal{S}_c) > U_i(\mathcal{S}_n), \\ \mathcal{S}_n, & U_i(\mathcal{S}_c) \leq U_i(\mathcal{S}_n). \end{cases} \quad (3)$$

It then follows from (1)–(3) that

$$X_i(t+1) = \begin{cases} \mathcal{S}_c, & p(t) < \theta_i, \\ \mathcal{S}_n, & p(t) \geq \theta_i, \end{cases} \quad (4)$$

where θ_i denotes the threshold parameter of node i , which represents the degree of cooperativity of node i :

$$\theta_i = \frac{S_i - P_i}{T_i - R_i + S_i - P_i}. \quad (5)$$

Under the condition $T_i > R_i \geq S_i > P_i$, θ_i ranges from 0 to 1. From (4), we find that node i prefers taking the cooperative strategy if θ_i approaches to one. Thus the larger θ_i indicates that node i is potentially cooperative. We should also note here that each node i can determine its behavior (strategy) based only on its own θ_i and surrounding information $p(t)$.

3.3. Heterogeneous Environments. We consider two kinds of heterogeneity: user heterogeneity and network heterogeneity. The user heterogeneity represents the diversity of cooperative degrees of nodes in caching a specific file. On the other hand, the network heterogeneity arises from the topological structure of P2P file-sharing networks.

3.3.1. Cooperative Degree in File Caching. In P2P file sharing systems, it is natural that each node should have different senses of value to a specific file. We model this user heterogeneity by assuming that θ_i 's are different among nodes. It is also reasonable to assume that most of users are uncooperative because costs accompany file caching. Thus we assume that the distribution of θ_i follows Zipf's law; namely, θ_i 's ($i = 1, 2, \dots, N$) are assumed to be independent and identically distributed (i.i.d.) according to a truncated Pareto distribution with support $[1/K, 1]$ and shape parameter $\alpha > 0$. Let $f(\theta)$ denote the probability density function of θ_i . We then have

$$f(\theta) = \begin{cases} \frac{1}{\log K} \cdot \theta^{-1}, & \alpha = 1, \\ \frac{\alpha - 1}{K^{\alpha-1} - 1} \cdot \theta^{-\alpha}, & \alpha \neq 1. \end{cases} \quad (6)$$

Note here that K is closely related to varieties of the threshold and α determines the degree of the bias of threshold values. Thus the user heterogeneity is mainly dominated by K , and the users' sense of value becomes more heterogeneous with the increase of K .

3.3.2. Topological Structure. The topological structure determines the locality of information that each node can obtain through interactions with other nodes. Note that communication overhead in terms of the total number of transferred messages increases with the hop count between two nodes. In most P2P networks, a topological structure is determined by rules for updating links between nodes and the average number of links per node, that is, node degree. In this paper, we consider three kinds of topologies: full-mesh, random, and scale-free networks.

In the full-mesh network, all nodes are directly connected each other. However, the assumption of the full-mesh network may not be realistic in large-scale systems. Therefore, to gain a deep insight about the impact of the topological structure, we also evaluate the system performance under Waxman random networks [22] and scale-free networks based on Barabási-Albert (BA) model [23].

The Waxman random network is constructed by connecting each node pair (u, v) by a link with probability $P(u, v)$:

$$P(u, v) = ae^{-D(u, v)/(bL)}, \quad (7)$$

where $D(u, v)$ denotes the distance between nodes u and v , L denotes the maximum distance between arbitrary nodes, and $a > 0$ and $b \leq 1$ are control parameters. Equation (7) indicates that the probability of establishing a link between two nodes decreases exponentially with the increase of their distance.

The scale-free network is a network whose degree distribution follows a power law: $p(k) \propto k^{-\gamma}$, where $p(k)$ denotes the probability that a randomly chosen node has degree k and γ denotes a parameter. In the scale-free network, there are a small number of high-degree nodes and a large number of low-degree nodes.

4. Evolutionary Game Theory

In societies of organisms, various individuals influence each other. Evolutionary game theory [2] originally aims to reveal the mechanism in which optimum behaviors are inherited by offspring through the evolutionary process of organisms. In the frame work of evolutionary game theory, an individual behavior defined by genes is regarded as a strategy in game theory and the number of offspring selecting the behavior is proportional to payoff acquired by the strategy. In such a case, various individuals are in strategically mutual dependence relation of game theory. Thus we can apply game theory to explaining the phenomenon that superior behaviors spread over a society of organisms through inheritance from ancestors to offspring. Moreover, in sociology and economics, there are several studies using evolutionary game theory, which aim to reveal such a phenomena that valuable information and behavior spread over human societies [16–18].

4.1. Micro-Macro Dynamics. Micro-macro dynamics [3] is a mathematical tool of modeling the dynamic behavior of a micro-macro loop, that is, interactions between micro behavior and macro behavior. Note here that the microbehavior is regarded as actions of heterogeneous individuals and the macrobehavior as a collective behavior arising from mutual interactions among individuals. Thus, using micro-macro dynamics, we can reveal how the interactions about file caching among nodes affect the caching condition in the whole system.

Let $p(t)$ denote the ratio of the number of cooperators to the total number N of nodes in the system at time t . Micro-macro dynamics describes the transition of $p(t)$ as a result of interactions among nodes, which can be obtained as follows. Recall that each node determines the next strategy according to the current $p(t)$ and its threshold θ_i (see (4)). We define $F(\theta)$ as the cumulative distribution function of the threshold:

$$F(\theta) = \int_{1/K}^{\theta} f(x)dx = \begin{cases} 1 + \frac{\log \theta}{\log K}, & \alpha = 1, \\ \frac{\theta^{-\alpha+1} - K^{\alpha-1}}{1 - K^{\alpha-1}}, & \alpha \neq 1. \end{cases} \quad (8)$$

Note that $p(t+1)$ is given by the ratio of the number of nodes with $p(t) < \theta_i$ to the total number of nodes in the system. Because $F(p(t))$ gives the ratio $1 - p(t+1)$ of the number of nodes with $\theta_i \leq p(t)$ to N , we obtain

$$p(t+1) = 1 - F(p(t)). \quad (9)$$

Equation (9) provides us with the relationship between the user selfishness and the system performance/stability

condition. The details will be given in Section 5. Note that (9) represents the system dynamics but each node does not solve this equation. As shown in (4), each node i determines its strategy based only on θ_i and $p(t)$ which is calculated by itself.

4.2. Agent-Based Dynamics. Micro-macro dynamics is a powerful mathematical tool because it can reveal the impact of system parameters on the system performance. Note that in micro-macro dynamics, all nodes are assumed to be able to obtain the global parameter $p(t)$. In actual systems, however, it would be difficult for each node to obtain $p(t)$ in a timely manner due to communication overheads.

Agent-based dynamics complements this shortcoming by modeling the decision making of nodes based on local interactions only with neighboring nodes in the network. As a result, agent-based dynamics models a phenomenon that a superior strategy spreads over the network in a hop-by-hop manner. Note that micro-macro dynamics is equivalent to agent-based dynamics in a full-mesh network where all nodes connect to each other.

In applying evolutionary game theory to P2P file-sharing systems, we assume that each node determines its behavior through local interactions with nodes located within d ($d \geq 1$) hops. $d = 1$ and $d = \infty$ result in agent-based dynamics and micro-macro dynamics, respectively. By controlling d , we can reveal how the locality of information exchanged among nodes affect the system performance. Furthermore we can observe the strategy distribution in the network, which can be regarded as the distribution of cached files. In what follows, we give the detail of agent-based dynamics for P2P file-sharing systems.

We first define a *generation* as the time 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 in the asynchronous system, only one node chosen randomly from the population updates its strategy at the beginning of each generation.

In the g th ($g = 1, 2, \dots$) generation, node i (which is eligible to update its strategy at the beginning of the $g + 1$ st generation) plays a game once with every node located within d hops and calculates the ratio $p_i(g)$ of cooperators including itself. Node i then determines the next strategy $X_i(g + 1)$ in the $g + 1$ st generation as follows:

$$X_i(g + 1) = \begin{cases} \delta_c, & p_i(g) < \theta_i, \\ \delta_n, & p_i(g) \geq \theta_i. \end{cases} \quad (10)$$

Thus, each node i does not require the global information $p(t)$ to determine its strategy.

Note that the initial strategy $X_i(0)$ ($i = 1, 2, \dots, N$) is arbitrary. The impact of the initial strategy distribution will be discussed in Section 6.

5. Analysis of Micro-Macro Dynamics

In this section, we analyze micro-macro dynamics given in (9). In particular, we are concerned with the file availability,

FIGURE 1: Complementary distribution function $\bar{F}(x) = 1 - F(x)$.

which is defined as the ratio $p(t)$ of the number of cooperators to the total number N of nodes for a sufficiently large t .

5.1. File Availability. Suppose that the system is in steady state. The file availability in steady state should be the fixed point p^* of (9):

$$p^* = 1 - F(p^*), \quad (11)$$

which is called the equilibrium. We can numerically obtain the equilibrium p^* with Newton's method [24]. Because $\bar{F}(x) = 1 - F(x)$ ($1/K \leq x \leq 1$) is monotonically decreasing function with $\bar{F}(1/K) = 1$ and $\bar{F}(1) = 0$, p^* is unique when $K > 1$ and it is given by the intersection of $y = x$ and $y = \bar{F}(x)$ (see Figure 1).

Figure 2 illustrates the equilibrium p^* as a function of K ($K \geq 2$). In all cases of $\alpha = 0.8, 1.0$, and 1.2 , p^* steeply drops in the range of $2 \leq K \leq 10$ and converges to a value depending on α . We also find that high user heterogeneity, that is, large K and α , leads to a small p^* . The reason of this phenomenon is that the number of uncooperative nodes increases with K and/or α (see (6)). The most important result for P2P file sharing systems is that p^* is always positive; there is at least one cached file in the network.

5.2. System Stability. In this section, we discuss the stability of the equilibrium p^* . When the equilibrium is stable, $p(t)$ converges to p^* once it reaches the vicinity of the equilibrium. Otherwise, $p(t)$ continues to oscillate forever even if it starts from the neighborhood of the equilibrium p^* . Therefore the stability of the equilibrium is significant for P2P file sharing systems to achieve effective and stable file caching. In what follows, we first consider the asymptotic stability of the equilibrium, which ensure the convergence of (9) to p^* from any point in the open neighborhood of p^* .

FIGURE 2: The equilibrium p^* as a function of K . FIGURE 3: Relationship between K and $|\bar{F}^{(1)}(p^*)|$.

We then extend the discussion to the region where $p(t)$ is certain to converge to p^* .

It follows from the standard theory of difference equations [25] that the equilibrium is asymptotically stable if the absolute value of the first derivative of $\bar{F}(x) = 1 - F(x)$ at the equilibrium $x = p^*$ is less than one. Because $\bar{F}(x)$ is a decreasing function of x , we have

$$\bar{F}^{(1)}(p^*) = \lim_{x \rightarrow p^*} \frac{d}{dx} \bar{F}(x) > -1. \quad (12)$$

The asymptotic stability of the equilibrium depends on parameters K and α , which determine the degree of the user

heterogeneity and its bias, respectively. When $\alpha = 1$, we obtain

$$\bar{F}^{(1)}(p^*) = -\frac{1}{p^* \log K} = \frac{1}{\log(p^*)}, \quad (13)$$

where we use $p^* = \bar{F}(p^*)$ in the second equality. It then follows from (12) that a sufficient condition of the asymptotic stability for $\alpha = 1$ is given by

$$p^* < e^{-1}. \quad (14)$$

Recall that $\bar{F}(x)$ is in $[0, 1]$ and monotonically decreases with x ($1/K \leq x \leq 1$). Therefore (14) is equivalent to

$$1 - F(e^{-1}) < e^{-1}. \quad (15)$$

Manipulating (8) and (15), we finally obtain a sufficient condition of the asymptotic stability of the equilibrium for $\alpha = 1$:

$$K > e^e \approx 15.2. \quad (16)$$

Thus the system is asymptotically stable for a sufficiently large K when $\alpha = 1$.

On the other hand, when $\alpha \neq 1$, we obtain

$$\lim_{x \rightarrow p^*} \frac{d}{dx} \bar{F}(x) = \frac{\alpha - 1}{1 - K^{\alpha-1}} \left(\frac{1 - (p^*)^{-\alpha+1}}{1 - K^{\alpha-1}} \right)^{-\alpha}. \quad (17)$$

Since this equation is too complicated to analyze, we proceed the discussion with numerical examples. Figure 3 depicts $|\bar{F}^{(1)}(p^*)|$ for $K \geq 2$. We find that the minimum integer K that satisfies (12) is equal to 15, 16, and 18 for $\alpha = 0.8, 1.0$, and 1.2 , respectively. Note that when $\alpha = 1$, the minimum integer $K = 16$ satisfies (16).

We now consider the region where $p(t)$ converges to the asymptotically stable p^* . A simple sufficient condition can be obtained as follows. Suppose that there exists an open region R such that $p^* \in R$ and for $p(t) \in R$,

$$\bar{F}(p(t)) \in R, \quad |p(t) - p^*| > |\bar{F}(p(t)) - p^*|. \quad (18)$$

If $p(0) \in R$, it is easy to see that $p(t)$ steadily approaches to p^* because $\bar{F}(p(t)) = p(t+1)$.

It can be shown that if

$$|p(0) - p^*| > |\bar{F}(p(0)) - p^*|, \quad (19)$$

holds for all t ($t = 1, 2, \dots$) because $\bar{F}(x)$ is a decreasing convex function (see (8)). We assume that p^* is asymptotically stable and let $p_{\min}$ ($p_{\min} < p^*$) denote the unique solution of $\bar{F}(x) - p^* = p^* - x$. We then have $R = (r_{\min}, r_{\max})$, where $r_{\min}$ is given by

$$r_{\min} = \max\left(\frac{1}{K}, p_{\min}\right), \quad (20)$$

and $r_{\max}$ is given by the unique solution of $r_{\min} = \bar{F}(x)$. Since R is determined by K and α , the system stability depends on K and α .

FIGURE 4: Transient behavior ($\alpha = 1.0$).

Next we consider the speed of convergence. The high-speed convergence is beneficial because nodes can determine whether to cache a file shortly. In Figure 4(a) ($K = 30$) and Figure 4(b) ($K = 100$), we show the transient behavior of $p(t)$ governed by (9), where $\alpha = 1.0$. Comparing these figures, we find that the larger K leads to the shorter convergence time, independent of $p(0)$. The reason of this phenomenon is that the slope at the vicinity of the equilibrium becomes gentle with the increase of K (cf. Figure 1). This result indicates that high user heterogeneity contributes to fast decision-making for file caching. We obtained similar results for $\alpha \neq 1$, too.

Figure 4(c) shows the case that the equilibrium p^* is not asymptotically stable, where $K = 5$, and $\alpha = 1.0$. In such a case, $p(t)$ does not converge to p^* . Rather, $p(t)$ oscillates permanently between 0 and 1 after a certain period of time, depending on $p(0)$. Thus, we cannot achieve effective file caching due to the unpredictable file availability. We confirmed that this property also emerged when $\alpha \neq 1$.

From the above results, we conclude that high user heterogeneity leads to a stable file caching, while keeping the file availability positive.

6. Simulation Experiments with Agent-Based Dynamics

With micro-macro dynamics and full-mesh networks, the preceding section analytically derived the relationship between the user heterogeneity and the system performance. In this section, we proceed to evaluate the impact of the network heterogeneity on the system performance, using agent-based dynamics. In particular, we reveal how the system synchronization, the topological structure, and the characteristics and behavior of high-degree nodes affect the system performance. Note that those issues cannot be investigated through micro-macro dynamics. Because the distribution of the cooperators can be obtained through simulation experiments, we evaluate search latency, that is, the number of hops to reach the nearest cooperative node, in addition to the performance metrics used in the preceding section.

6.1. Simulation Model. We use NetLogo [26] in our simulation experiments. We prepare full-mesh, random and scale-free networks of 1,000 nodes. Random networks are generated by Waxman algorithm [22] with $a = 0.15$ and $b = 0.2$. On the other hand, scale-free networks are generated based on Barabási-Albert (BA) model [23], where the number m of links that a newly participating node will establish is set to be 1, 2, or 4. For the sake of comparison, we adjust the average degree in random networks to $2m$. We use the topology generator BRITE [27] to generate those random and scale-free networks.

Recall that the threshold θ_i ($i = 1, 2, \dots, 1,000$) represents the cooperative degree to file caching of node i , which are assumed to be i.i.d. according to (8) with $\alpha = 1$. We also set the range d of local interactions to be one, so that each node can communicate only with adjacent nodes. Note that

FIGURE 5: Transient behavior (full-mesh network).

increasing d brings results similar to those in micro-macro dynamics, because micro-macro dynamics is considered to be agent-based dynamics with $d = \infty$. The initial ratio $p(0)$ of the number of cooperators to the total number of nodes is set to be 0.5, unless otherwise stated. We prepare ten independent samples of networks for each topological structure with a set of specific parameter values, and we will

show only the average of those ten independent simulation results for each set of specific parameter values.

6.2. Influence of Synchronization. Figures 5(a) and 5(b) illustrate the transient behavior of the synchronous and asynchronous systems, respectively, in the full-mesh network. Comparing Figure 5(a) with Figure 4, we observe that the synchronous system shows transient behaviors similar to the micro-macro dynamics, both in the unstable case ($K = 5$) and in the stable case ($K = 100$). The reason of this phenomenon is that the synchronous system is configured according to the same assumption as the micro-macro dynamics and the stability condition follows (16). From Figures 5(a) and 5(b), we observe that when the synchronous system converges to the equilibrium (i.e., $K = 100$), the asynchronous system also converges to the same equilibrium. Furthermore the asynchronous system converges even in the unstable case ($K = 5$), independent of the initial ratio of cooperators. Since in the asynchronous system, only one node updates its strategy in each generation, global synchronization among nodes is avoided, and thus the ratio of cooperators does not oscillate. Note that the file availability after convergence is higher than that in the stable case, which would arise from the fact that nodes in the unstable system with $K = 5$ are more cooperative by nature than those in the stable system with $K = 100$.

We turn our attention to the speed of convergence. From the lower panels of Figures 5(a) and 5(b), we observe that the asynchronous system takes much generations before convergence than the synchronous system; the former takes about 3,000 generations, while the latter takes about 15 generations. Note here that in the synchronous system, all 1000 nodes update their strategies simultaneously at the beginning of generations, whereas only one node updates its strategy at the beginning of each generation in the asynchronous system. Therefore, in terms of the total number of occasions to update the strategy, the speed of convergence in the asynchronous system is faster than that in the synchronous system, because the number of cooperators does not oscillate in the asynchronous system.

From the above results, we conclude that when there exists the stable equilibrium, the synchronous and asynchronous systems exhibit a similar behavior. In addition, the asynchronous system can be stable even if the equilibrium in micro-macro dynamics is not stable. Note also that the asynchronous system is more realistic than the synchronous system. Therefore, in what follows, we consider only the asynchronous system.

6.3. Influence of Topological Structure. Figure 6 illustrates the file availability p^* in the asynchronous system as a function of K , where $m = 1, 2$, and 4 in the scale-free and random networks. Recall that users becomes more heterogeneous with the increase of K . We observe that for a fixed m , the scale-free and random networks have almost the same file availability whose value is always greater than that in the full-mesh network. We also observe that a small m leads to a high value of the file availability. To understand these phenomena,

FIGURE 6: File availability p^* as a function of K . FIGURE 7: Search latency as a function of K .

we further examine the distribution of cooperators in the network in steady state.

We define the search latency as the average of the number of hops from each node to the nearest cooperator, where the number of hops is equal to zero when the node is a cooperator. Figure 7 shows the relationship between K and the search latency in the scale-free and random 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 lies in $(0.3, 0.4)$.

To see the distribution of cooperators over the network, we further define the reachability to cooperators as the average ratio of the number of cooperators within H hops from each node to the total number N of nodes in the network, where $H = 0, 1, \dots$. Thus, for $H = 0$ and $N - 1$, the reachabilities to cooperators are given by p^*/N and p^* ,

FIGURE 8: Reachability to cooperators ($K = 100$, $m = 2$).

respectively. Figure 8 shows the reachability to cooperators as a function of H , where $K = 100$ and $m = 2$. For the sake of comparison, we also plot the reachability to cooperators in the case that p^*N cooperators are distributed randomly over the network. There is almost no difference between these two cases, and therefore the agent-based dynamics uniformly distributes cooperators over the network. We confirmed that the same result was obtained for $m = 4$. Note that this is caused by the structure of the game: each node can acquire higher payoff by selecting the different strategy of the opponent. Therefore defectors are likely to gather around a cooperator, which leads to the uniform distribution of cooperators. For the same reason, a larger m yields more defectors because of the increase of neighboring nodes.

Figure 9 may help understand these phenomena, which illustrates an example of topological structure with transition of strategy distribution in a scale-free network with $N = 100$ and $m = 1$. For instance, the ratio of pairs composed of a cooperator and a defector increases from 0.38 ($g = 0$) to 0.66 ($g = 2500$). Although we set N to 100 so as to improve the visibility, we confirmed the similar results under the above mentioned simulation settings.

Next, we investigate the influence of the topological structure on the transient behavior. Figure 10 illustrates the transient behavior of the file availability in the scale-free network with $m = 2$. The transient behavior in the scale-free network is similar to that in the full-mesh network. Note that the speed of convergence in the scale-free network is slightly faster than that in the full-mesh network when $K = 100$.

6.4. Influence of High-Degree Nodes' Characteristics and Behavior. There are a small number of high-degree nodes in the scale-free network, which are expected to have a greater impact on the system performance than low-degree nodes. We examine this hypothesis in two ways: threshold assignment evaluation and strategy assignment evaluation.

FIGURE 9: An example of topological structure with transition of strategy distribution (a scale-free network with $N = 100$ and $m = 1$).

In the threshold assignment evaluation, we first generated 1,000 samples of the threshold according to (8) and sorted them in a descending order. We then assigned them to nodes in a descending (resp., ascending) order of degree, which makes high-degree nodes cooperative (resp., uncooperative). Figure 11 shows the results of the cooperative and uncooperative cases. For the sake of comparison, we also present the result of the random case, which has already been shown in Figure 6. We observe that the uncooperative case achieves the higher file availability than the cooperative case. This result arises from two factors: game structure and topological structure. Because the game structure encourages a node to take a strategy different from the opponent and high-degree nodes have a lot of neighbors, the uncooperative behavior of high-degree nodes yields many cooperators. The result of the random case is biased towards that of the uncooperative case because each node is subject to be uncooperative potentially due to (8).

Next, we present the results of the strategy assignment evaluation in Figure 12. We call the top 5% of nodes in a descending order of degree *hub nodes*. In the cooperative (resp., uncooperative) case, we fixed the strategies of hub nodes to cooperators (resp., defectors). The initial strategy of the rest of nodes is determined according to (8). Figure 12

FIGURE 10: Transient behavior (asynchronous system, scale-free network with $m = 2$).

FIGURE 11: File availability p^* as a function of K (scale-free network with $m = 2$, threshold assignment evaluation).

shows the file availability as a function of K . We observe similar results with those in Figure 11, which comes from the same reason as discussed in the above.

In actual P2P file sharing systems, both the cooperative and uncooperative cases can occur. The cooperative case corresponds to the situation where a node acquires a lot of

FIGURE 12: File availability p^* as a function of K (scale-free network with $m = 2$, strategy assignment evaluation).

connections from others because of its cooperative behavior. On the other hand, the uncooperative case is regarded as the situation where an uncooperative node exploits many cooperative nodes. Note that in either case, the system becomes robust to file disappearance.

7. Conclusions

With evolutionary game theory, we discussed how the 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 P2P networks. We first modeled the interactions about file caching among nodes as caching games, assuming that nodes are heterogeneous. With micro-macro dynamics, we analytically showed that high user heterogeneity leads to the stable file availability. Furthermore, we conducted simulation experiments based on agent-based dynamics and revealed how local interactions among nodes affect the system performance in the heterogeneous topological structure. The results of the simulation experiments are summarized as follows: (i) the asynchronous system is more stable than the synchronous system, (ii) 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, and (iii) the existence (resp., absence) of cooperative high-degree nodes decreases (resp., increases) the number of cooperative nodes in the system.

Acknowledgments

This research was supported in part by a Grant-in-Aid for Young Scientists (B) 20700061, International Communication Foundation, and "Global COE (Centers of Excellence)

Program" of the Ministry of Education, Culture, Sports, Science and Technology in Japan. This paper was presented in part at the 1st International Workshop on Technologies for Ambient Information Society (TAIS'08), Awaji, Japan, 2008, and 2009 International Symposium on Nonlinear Theory and its Applications (NOLTA'09), Sapporo, Japan, 2009.

References

- [1] E. Adar and B. A. Huberman, "Free riding on Gnutella," Tech. Rep., Xerox PARC, Palo Alto, Calif, USA, 2000.
- [2] J. W. Weibull, *Evolutionary Game Theory*, MIT Press, Cambridge, Mass, USA, 1997.
- [3] S. Iwanaga and A. Namatame, "The Complexity of Collective Decision," *Journal of Nonlinear Dynamics and Control*, vol. 6, no. 2, pp. 137–158, 2002.
- [4] G. Szabó and G. Fáth, "Evolutionary games on graphs," *Physics Reports*, vol. 446, no. 4–6, pp. 97–216, 2007.
- [5] Napster, <http://free.napster.com>.
- [6] Gnutella, http://www9.limewire.com/developer/gnutella_protocol.0.4.pdf.
- [7] KaZaA, <http://www.kazaa.com>.
- [8] BitTorrent, <http://www.bittorrent.com>.
- [9] S. Saroiu, P. K. Gummadi, and S. D. Gribble, "A measurement study of peer-to-peer file sharing systems," in *Proceedings of the Multimedia Computing and Networking (MMCN '02)*, pp. 156–170, San Jose, Calif, USA, January 2002.
- [10] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," in *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP '03)*, pp. 314–329, New York, NY, USA, 2003.
- [11] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips, "The Bittorrent P2P file-sharing system: measurements and analysis," in *Proceedings of the 4th International Workshop on Peer-to-Peer Systems (IPTPS '05)*, vol. 3640 of *Lecture Notes in Computer Science*, pp. 205–216, Springer, New York, NY, USA, February 2005.
- [12] D. Hughes, G. Coulson, and J. Walkerdine, "Free riding on Gnutella revisited: the bell tolls?" *IEEE Distributed Systems Online*, vol. 6, no. 6, 2005.
- [13] B.-G. Chun, K. Chaudhuri, H. Wee, M. Barreno, C. H. Papadimitriou, and J. Kubiawicz, "Selfish caching in distributed systems: a game-theoretic analysis," in *Proceedings of the 23rd Annual ACM Symposium on Principles of Distributed Computing (PODC '04)*, pp. 21–30, St. John's, Canada, July 2004.
- [14] K. Ranganathan, M. Ripeanu, A. Sarin, and I. Foster, "Incentive mechanisms for large collaborative resource sharing," in *Proceedings of IEEE International Symposium on Cluster Computing and the Grid (CCGRID '04)*, pp. 1–8, Washington, DC, USA, 2004.
- [15] Freenet, <http://freenetproject.org/index.html>.
- [16] F. Fu, L.-H. Liu, and L. Wang, "Evolutionary Prisoner's Dilemma on heterogeneous Newman-Watts small-world network," *European Physical Journal B*, vol. 56, no. 4, pp. 367–372, 2007.
- [17] H. Ohtsuki and M. A. Nowak, "Evolutionary stability on graphs," *Journal of Theoretical Biology*, vol. 251, no. 4, pp. 698–707, 2008.

- [18] F. C. Santos, J. M. Pacheco, and T. Lenaerts, "Evolutionary dynamics of social dilemmas in structured heterogeneous populations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 9, pp. 3490–3494, 2006.
- [19] 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.
- [20] M. Sasabe, N. Wakamiya, and M. Murata, "User selfishness vs. file availability in P2P file-sharing systems: evolutionary game theoretic approach," *Peer to Peer Networking and Applications*, 2009, <http://www.springerlink.com/content/n5gv715v18834lj2/>.
- [21] D. Hales and S. Arteconi, "Motifs in evolving cooperative networks look like protein structure networks," *Networks and Heterogeneous Media*, vol. 3, no. 2, pp. 239–249, 2008.
- [22] B. M. Waxman, "Routing of multipoint connections," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1617–1622, 1988.
- [23] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [24] C. T. Kelley, "Solving nonlinear equations with Newton's method," in *Fundamentals of Algorithms*, SIAM, Philadelphia, Pa, USA, 2003.
- [25] S. N. Elaydi, *An Introduction to Difference Equations*, Springer, New York, NY, USA, 2nd edition, 1991.
- [26] U. Wilensky, NetLogo, 1999, <http://ccl.northwestern.edu/netlogo>.
- [27] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: an approach to universal topology generation," in *Proceedings of the 9th IEEE International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS '01)*, pp. 346–353, Washington, DC, USA, 2001.