Foresighted Resource Reciprocation Strategies in P2P Networks

Hyunggon Park and Mihaela van der Schaar
Electrical Engineering Department, University of California, Los Angeles (UCLA)
Email: {hgpark, mihaela}@ee.ucla.edu

***Abstract*—We consider peer-to-peer (P2P) networks, where multiple peers are interested in sharing content. While sharing resources, autonomous and self-interested peers need to make decisions on the amount of their resource reciprocation (i.e. representing their actions) such that their individual rewards are maximized. We model the resource reciprocation among the peers as a stochastic game and show how the peers can determine their optimal strategies for the actions using a Markov Decision Process (MDP) framework. The optimal strategies determined based on MDP enable the peers to make foresighted decisions about resource reciprocation, such that they can explicitly consider both their immediate as well as future expected rewards. To successfully formulate the MDP framework, we propose a novel algorithm that efficiently identifies the state transition probabilities using representative resource reciprocation models of peers. Simulation results show that the proposed approach based on the reciprocation models can effectively cope with a dynamically changing environment of P2P networks. Moreover, we show that the foresighted decisions lead to the best performance in terms of the cumulative expected rewards.**

I. Introduction

P2P applications (e.g, [1], [2]) have become increasingly popular and P2P networks provide a cost effective and easily deployable framework for disseminating large files without relying on a centralized infrastructure [3]. Hence, it has been recently proposed to use P2P networks for general file sharing [2], [4] or multimedia streaming [3], [5], [6]. In this paper, we consider data-driven P2P systems such as CoolStreaming [5], Chainsaw [6], or BitTorrent systems [4], which adopt pull-based techniques [5], [6]. While this approach has been successfully deployed in real-time multimedia streaming and file distributions over P2P networks, key challenges such as resource reciprocation among autonomous and self-interested peers still remain open.

The resource reciprocation strategy deployed in BitTorrent (i.e., Tit-for-Tat (TFT) strategy) distributes the bandwidth equally: a peer equally divides its available upload bandwidth among multiple leechers [4]. The resource reciprocation in [5] is based on a heuristic scheduling algorithm, which enables the peers to determine the suppliers of required chunks and select the peer with the highest bandwidth. Alternatively, the resource reciprocation can be based on the random chunk selection [6]. Since the solutions in [5] and [6] are implemented assuming that the associated peers are altruistic, the resource reciprocation strategies based on these solutions do not consider the interactions of the autonomous, self-interested, and heterogeneous peers.

We model the resource reciprocation among the interested peers as a stochastic game, where peers decide their resource distribution by explicitly considering the probabilistic behaviors (reciprocation) of their associated peers. Unlike existing resource reciprocation strategies, which focus on myopic decisions, we formalize the resource reciprocation game as a Markov Decision Process (MDP) [7] to enable peers to make foresighted decisions on their resource distribution in a way that maximizes their cumulative rewards, i.e., their immediate as well as future rewards.

To successfully formulate the resource reciprocation game as an MDP problem, the peers need to identify the associated peers' probabilistic behaviors for resource reciprocation. The probabilistic behaviors of the associated peers can be estimated using the past history of resource reciprocation and are represented by *state transition probabilities* in the MDP framework. In this paper, the state of a peer is defined as the set of received resources from each of the associated peers. We propose a novel algorithm that can efficiently identify the state transition probabilities using peers' *reciprocation models*. The reciprocation models of the peers are motivated by [8], which classify the attitudes of players in a game towards their strategies as optimistic, pessimistic, and neutral archetypes. We show that this approach for resource reciprocation based on the reciprocation model provides a faster convergence. Hence, this approach enables peers to efficiently capture the changes of the state transition probabilities, resulting in an efficient solution for P2P networks.

This paper is organized as follows. In Section II, we model the resource reciprocation among peers as a resource reciprocation game. In Section III, the types of peers in the considered P2P networks are discussed. In Section IV, an algorithm that determines the state transition probabilities based on the reciprocation models is proposed. Simulation results are provided in Section V and conclusions are drawn in Section VI.

II. A New Framework for Resource Reciprocation

In this section, we model the resource reciprocation among the peers as a resource reciprocation game. Resource reciprocation games in the P2P networks are played by the peers that are interested in each other. A resource reciprocation game is played in a *group*, where a group consists of a peer and its associated peers, which is similar to concepts such as the swarms in [4], partnerships in [5], or neighbors in [6]. We

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

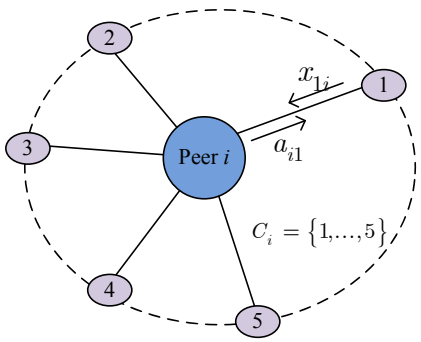

Fig. 1. An illustrative example for C_i with 5 matched peers.

denote the associated peers in a group of a peer i by C_i, and an illustrative example of C_i is depicted in Fig. 1. As shown in Fig. 1, C_i does *not* include peer i. The peers in C_i are indexed by $1, \ldots, N_{C_i}$, i.e., $C_i = \{1, \ldots, N_{C_i}\}$. Note that peer $k \in C_i$ also has its own group C_k which includes peer i.

The resource reciprocation game in C_i is a stochastic game [7]. To play the resource reciprocation game, a peer can deploy an MDP. An MDP of a peer i is a tuple $< \mathbf{S}_i, \mathbf{A}_i, P_i, R_i >$, where $\mathbf{S}_i$ is the state space, $\mathbf{A}_i$ is the action space, $P_i : \mathbf{S}_i \times \mathbf{A}_i \times \mathbf{S}_i \rightarrow [0, 1]$ is a state transition probability function, and $R_i : \mathbf{S}_i \rightarrow \mathbb{R}$ is a reward function. The details are explained as follows.

1) State Space $\mathbf{S}_i$: A state s_i of peer i represents the set of received resources from the peers in C_i, and the state space of peer i can be expressed as

$$\mathbf{S}_i = \left\{ s_i | \, s_{ik} = \psi_i(x_{ki}) \in \mathbf{\Delta}_i = \left\{\Delta_i^1, \ldots, \Delta_i^{n_i}\right\}, k \in C_i \right\},$$

where $s_i = (s_{i1}, \ldots, s_{iN_{C_i}})$ and x_{ki} denotes the provided resources (i.e., rate) by peer $k \in C_i$, which is limited by its maximum upload bandwidth L_k. s_{ik} is determined by a peer dependent function $\psi_i : \mathbb{R} \rightarrow \mathbf{\Delta}_i = \{\Delta_i^1, \ldots, \Delta_i^{n_i}\}$ that maps x_{ki} into one of n_i discrete values[1], which are called *state descriptions* in this paper.

2) Action Space $\mathbf{A}_i$: An action of peer i is its resource allocation to the peers in C_i. Hence, the action space of peer i in C_i can be expressed as

$$\mathbf{A}_i = \left\{ \mathbf{a}_i | 0 \leq a_{ik} \leq L_i, 1 \leq k \leq N_{C_i}, \sum\nolimits_{k \in C_i} a_{ik} \leq L_i \right\},$$

where $\mathbf{a}_i = (a_{i1}, \ldots, a_{iN_{C_i}}) \in \mathbf{A}_i = A_i \times \cdots \times A_i$ and $a_{ik} \in A_i$ denotes the allocated resources to peer k by peer i in C_i. Hence, peer i's action a_{ik} to peer k becomes peer k's received resources from peer i, i.e., $a_{ik} = x_{ik}$. We assume that the actions are the number of allocated "units" of bandwidth [9] to the associated peers in their groups. We define the *resource reciprocation* $(\hat{\mathbf{a}}_i, \hat{s}_i) = \left((\hat{a}_{i1}, \ldots, \hat{a}_{iN_{C_i}}), (\hat{s}_{i1}, \ldots, \hat{s}_{iN_{C_i}})\right)$ as a pair comprising peer i's action, $\hat{a}_{ik}$, and the corresponding modeled resource reciprocation $\hat{s}_{ik}$, which is determine as $\hat{s}_{ik} = \psi_i(\hat{x}_{ki})$ for all $k \in C_i$.

[1]A continuous value of x_{ki} can be discretized by peer i based on its quantization policy, as the bandwidth of each peer can be decomposed into several "units" of bandwidth by the client software, e.g., [9].

3) State Transition Probability $P_{\mathbf{a}_i}(s_i, s_i')$: A state transition probability represents the probability that by taking an action, a peer will transit into a new state. Given a state $s_i \in \mathbf{S}_i$ at time t, an action $\mathbf{a}_i \in \mathbf{A}_i$ of peer i can lead to another state $s_i' \in \mathbf{S}_i$ at $t+1$ with probability $P_{\mathbf{a}_i}(s_i, s_i') = \Pr(s_i' | s_i, \mathbf{a}_i)$. Hence, for a state $s_i = (s_{i1}, \ldots, s_{iN_{C_i}})$ of peer i in C_i, the probability that an action $\mathbf{a}_i$ leads the state transition from s_i to s_i' can be expressed as

$$P_{\mathbf{a}_i}(s_i, s_i') = \prod\nolimits_{l=1}^{N_{C_i}} P_{a_{il}}(s_{il}, s_{il}'). \quad (1)$$

4) Reward R_i: Since the peers prefer higher download rates, we consider that reward $R_i(s_i)$ for a peer i in state s_i is the total received resources in C_i, expressed as

$$R_i(s_i) = R_i(s_{i1}, \ldots, s_{iN_{C_i}}) = \sum\nolimits_{k \in C_i} r_i(s_{ik}), \quad (2)$$

where $r_i(s_{ik})$ represents the received resource in s_{ik}.

5) Reciprocation Policy π_i^*: The solution to the MDP is represented by peer i's optimal policy π_i^*, which is a mapping from the states to optimal actions. The optimal policy can be obtained using well-known methods such as value iteration and policy iteration [7]. Hence, peer i can decide its actions based on the optimal policy π_i^*, i.e., $\pi_i^*(s_i) = \mathbf{a}_i$ for all $s_i \in \mathbf{S}_i$.

Note that our focus is on the resource reciprocation game in a group, as the resource reciprocation games in a P2P network can be considered as multiple resource reciprocation games in groups.

We will discuss how the optimal policies can be determined for different types of peers in the next section.

III. Peer Types in P2P Networks

In this paper, we categorize the peers based on their adopted utilities and their resource reciprocation attitudes.

A. Peer types depending on their adopted utilities

Peers in the P2P networks can be considered as myopic or foresighted based on how they compute their rewards.

A peer i in state $s_i^{(t)}$ is myopic if it focuses on maximizing its immediate expected rewards $R_i^{myo}(s_i^{(t)})$, defined as

$$R_i^{myo}(s_i^{(t)}) \triangleq \sum\nolimits_{s_i^{(t+1)} \in \mathbf{S}_i} P_{\mathbf{a}_i}(s_i^{(t)}, s_i^{(t+1)}) R_i(s_i^{(t+1)}). \quad (3)$$

Hence, the myopic peer i determines its action $\mathbf{a}_i^*$ such that $\mathbf{a}_i^*$ maximizes its adopted expected reward $R_i^{myo}(s_i^{(t)})$, i.e.,

$$\mathbf{a}_i^* = \begin{array}{l} \arg\max_{\mathbf{a}_i \in \mathbf{A}_i} R_i^{myo}(s_i^{(t)}) \\ \text{subject to} \sum_{k \in C_i} a_{ik} \leq L_i \end{array}.$$

Unlike the myopic peers, the foresighted peers take their actions such that the actions maximize a cumulative discounted expected reward [7]. Hence, the objective of a foresighted peer i in state $s_i^{(t)}$ at time $t = t_c$ given a discount factor γ_i can be expressed as

$$R_i^{fore}(s_i^{(t)}) \triangleq \sum\nolimits_{t=t_c+1}^{\infty} \gamma_i^{(t-(t_c+1))} \cdot E\left[R_i(s_i^{(t)})\right], \quad (4)$$

where $E\left[R_i(s_i^{(t)})\right] = \sum\limits_{s_i^{(t+1)} \in \mathbf{S}_i} P_{\mathbf{a}_i^{(t+1)}}(s_i^{(t)}, s_i^{(t+1)}) R_i(s_i^{(t+1)})$.

Hence, the foresighted peer i can determine a set of actions

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

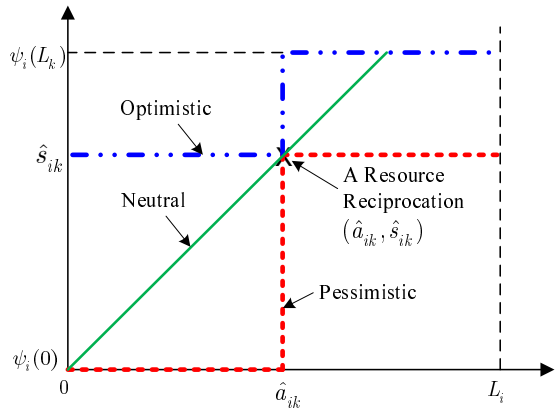

Fig. 2. Illustration of resource reciprocation based on peers' attitudes.

that maximizes $R_i^{fore}(s_i^{(t)})$ for every state in $\mathbf{S}_i$, which leads to an optimal policy π_i^*. The optimal policy π_i^* thus maps each state $s_i \in \mathbf{S}_i$ into a corresponding optimal action $\mathbf{a}_i^*$, i.e., $\pi_i^*(s_i) = \mathbf{a}_i^*$ for all $s_i \in \mathbf{S}_i$.

From (3) and (4), we can observe that the myopic decisions are a special case of the foresighted decisions when $\gamma_i = 0$. Note that the discount factor γ_i in the considered P2P network can alternatively represent the belief of the peer i about the validity of the expected future rewards, since the state transition probability can be affected by system dynamics such as the other peers' joining, switching, or leaving groups. Hence, for example, if the P2P network is in a transient regime, a small discount factor is desirable. However, a large discount factor can be used if the P2P network is in stationary regime [10]. Thus, we assume that the value of the discount factor can be determined by the peers using information based on their past experiences, reputation of their associated peers [11], [12], etc.

B. Peer types based on their attitudes

Peers in the considered P2P networks can also be characterized based on their attitudes towards the resource reciprocation, which are pessimistic, neutral, or optimistic [8].

Let $(\hat{a}_{ik}, \hat{s}_{ik})$ be a resource reciprocation between peer i and peer k. A peer i is *neutral* if it presumes that the resource reciprocation changes linearly depending on its actions. A peer i is *pessimistic* if it presumes that the resource reciprocation decreases fast for $a_{ik} \leq \hat{a}_{ik}$ and increases slow for $a_{ik} \geq \hat{a}_{ik}$. On the other hand, an optimistic peer i presumes that the resource reciprocation decreases slow for $a_{ik} \leq \hat{a}_{ik}$ and increases fast for $a_{ik} \geq \hat{a}_{ik}$. Illustrative examples of resource reciprocation shapes that correspond to peers' different attitudes are shown in Fig. 2. In this paper, we consider these resource reciprocation profiles, which will be referred to as *reciprocation models*.

These types of peers discussed above obviously affect their resource reciprocation strategies. In the following sections, we discuss how the peers' attitudes can impact the way in which peers model the other peers' resource reciprocation behavior.

IV. Determining The State Transition Probabilities

A. Empirical Frequency based State Transition Probabilities

A peer i can identify its state transition probabilities based on the frequency of the reciprocation. For this, we consider a table T_i^k that stores the history of resource reciprocation for peer k given actions of peer i. An element $T_i^k(\Delta_i^{l_1}, \Delta_i^{l_2}, a_{ik})$ of the table T_i^k denotes the number of state transitions from $s_{ik} = \Delta_i^{l_1}$ to $s'_{ik} = \Delta_i^{l_2}$, given an action a_{ik}. Hence, the state transition probability $P_{a_{ik}}(s_{ik} = \Delta_i^{l_1}, s'_{ik} = \Delta_i^{l_2})$ based on the empirical frequency can be expressed as:

$$P_{a_{ik}}(s_{ik} = \Delta_i^{l_1}, s'_{ik} = \Delta_i^{l_2}) = \frac{T_i^k(\Delta_i^{l_1}, \Delta_i^{l_2}, a_{ik})}{\sum_{h=1}^{n_i} T_i^k(\Delta_i^{l_1}, \Delta_i^{h}, a_{ik})}.$$

A disadvantage of this approach is that it may require a considerable amount of observations of the resource reciprocation over time to accurately identify the state transition probabilities. To reduce the number of required observations, we propose an alternative algorithm that is based on the resource reciprocation models.

B. Reciprocation Model based State Transition Probabilities

A set of the state transition probability that correspond to the resource reciprocation models is called reciprocation matrix. The set of m available reciprocation matrices of peer i in s_{ik} for peer k is denoted by $\mathbf{M}_i^k(s_{ik}) = \{M_{i1}^k(s_{ik}), \ldots, M_{im}^k(s_{ik})\}$, where $M_{il}^k(s_{ik})$ is a $|A_i| \times n_i$ matrix with its element $M_{il}^k(s_{ik})[a_{ik}, s'_{ik}] = P_{a_{ik}}(s_{ik}, s'_{ik})$. A reciprocation matrix $M_{il}^k(s_{ik})$ for a pessimistic peer i taking action a_{ik} in s_{ik} shown in Fig. 2 can be expressed by

$$M_{il}^k(s_{ik})[a_{ik}, s'_{ik}] \quad (5)$$
$$= \begin{cases} 1, \text{ if } a_{ik} < \hat{a}_{ik}, s'_{ik} = \psi_i(0) \text{ or } a_{ik} > \hat{a}_{ik}, s'_{ik} = s_{ik} \\ 1/W_P, \text{ if } a_{ik} = \hat{a}_{ik}, \psi_i(0) \leq s'_{ik} \leq s_{ik} \\ 0, \text{ otherwise,} \end{cases},$$

where $W_P = |\{\, l |\, \psi_i(0) \leq s_{il} \leq s_{ik}\}|$ is the number of state descriptions between s_{ik} and $\psi_i(0)$, and a_{ik} denotes the available actions that can be taken.

Similarly, a reciprocation matrix of an optimistic peer i in s_{ik} for peer k shown in Fig. 2 can be represented by

$$M_{il}^k(s_{ik})[a_{ik}, s'_{ik}] \quad (6)$$
$$= \begin{cases} 1, \text{if } a_{ik} < \hat{a}_{ik}, s'_{ik} = s_{ik} \text{ or } a_{ik} > \hat{a}_{ik}, s'_{ik} = \psi_i(L_k) \\ 1/W_O, \text{if } a_{ik} = \hat{a}_{ik}, s_{ik} \leq s'_{ik} \leq \psi_i(L_k) \\ 0, \text{otherwise} \end{cases},$$

where $W_O = |\{\, l |\, s_{ik} \leq s_{il} \leq \psi_i(L_k)\}|$ is the number of state descriptions between s_{ik} and $\psi_i(L_k)$.

A neutral peer i presumes that an action $a_{ik} \neq \hat{a}_{ik}$ will lead to linear changes in resource reciprocation. Hence, the reciprocation matrix of a neutral peer i can be expressed as

$$M_{il}^k(s_{ik})[a_{ik}, s'_{ik}] = \begin{cases} 1, \text{if } s'_{ik} = \psi_i(x_{ki} = \alpha \cdot a_{ik}) \\ 0, \text{otherwise} \end{cases}, \quad (7)$$

where $\alpha = s_{ik}/\hat{a}_{ik}$ denotes the slope determined based on the current resource reciprocation. In the following subsection,

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

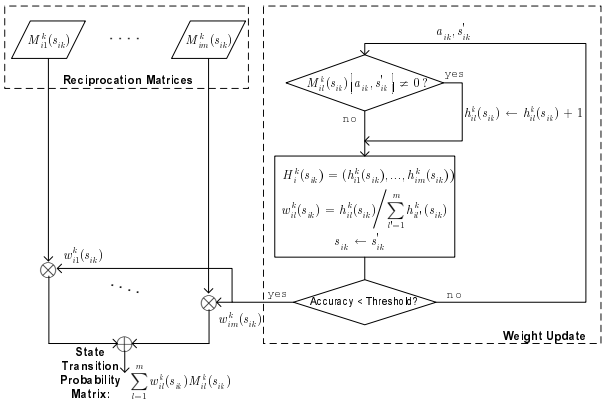

Fig. 3. Block diagram for updating the weights of the reciprocation matrices.

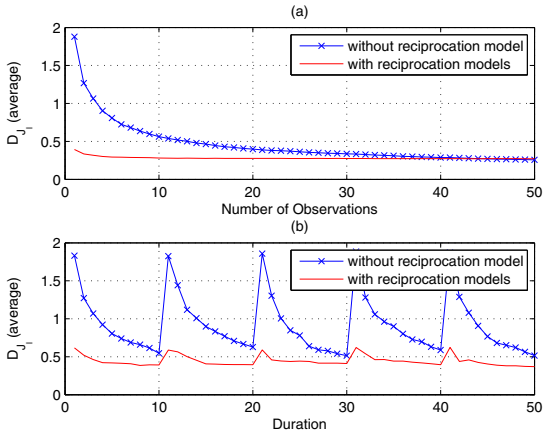

Fig. 4. (Averaged) D_{J_l} for estimated state transition probability matrices with/without reciprocation models.

we propose an algorithm that uses the discussed reciprocation matrices to efficiently identify the state transition probability.

C. Building State Transition Probabilities based on Reciprocation Matrices

We assume that each peer has a pre-determined initial action $\mathbf{a}_i^I = (a_{i1}^I, \ldots, a_{iN_{C_i}}^I) \in \mathbf{A}_i$ that is used for initializing the reciprocation matrices, i.e., a peer i has a pre-determined action $a_{ik}^I \in A_i$ for peer k and the resulting s_{ik}. Based on the initial resource reciprocation (a_{ik}^I, s_{ik}) between peer i and peer k, the reciprocation matrices of peer i can be initialized based on (5), (6), and (7).

Let $\mathbf{M}_i^k(s_{ik})$ be the set of m reciprocation matrices that are initialized based on (a_{ik}^I, s_{ik}). The weights of peer i for the reciprocation matrices are denoted by $w_i^k(s_{ik}) = (w_{i1}^k(s_{ik}), \ldots, w_{im}^k(s_{ik}))$ for peer $k \in C_i$. We define $H_i^k(s_{ik}) = (h_{i1}^k(s_{ik}), \ldots, h_{im}^k(s_{ik}))$ as the set of number of hits, where the resource reciprocation between peer i and peer k are matched to non-zero elements in $\mathbf{M}_i^k(s_{ik})$. Specifically, if (a_{ik}, s'_{ik}) is matched up to a non-zero element in $M_{il}^k(s_{ik})\,[a_{ik}, s'_{ik}]$, then $h_{il}^k(s_{ik}) \leftarrow h_{il}^k(s_{ik}) + 1$. Based on $H_i^k(s_{ik})$, the weights of reciprocation matrices for peer k can be computed as

$$w_{il}^k(s_{ik}) = \frac{h_{il}^k(s_{ik})}{\sum_{l'=1}^m h_{il'}^k(s_{ik})}.$$

This update process is depicted in Fig. 3. Finally, based on the identified weights, the probability $P_{a_{ik}}(s_{ik}, s'_{ik})$ can be obtained by

$$P_{a_{ik}}(s_{ik}, s'_{ik}) = \left\{ \sum_{l=1}^m w_{il}^k(s_{ik}) M_{il}^k(s_{ik}) \right\} [a_{ik}, s'_{ik}]\,. \quad (8)$$

V. Simulation Results

A. Comparison of Various Approaches for Identifying the State Transition Probabilities

To illustrate the tradeoffs between the efficiency and the accuracy, we deploy the two approaches discussed in Section IV to identify the true state transition probability.

In the simulations, we assume that peer i and peer k are in a group and reciprocate their resources. To study the impact of the two approaches on the discounted expected rewards, we consider discounted expected rewards $\mathbf{J}^* = \left[J^*(s_{ik}^1), \ldots, J^*(s_{ik}^{n_i})\right]^T$ and $\mathbf{J'}^* = \left[J'^*(s_{ik}^1), \ldots, J'^*(s_{ik}^{n_i})\right]^T$ of peer i from peer k obtained based on true and estimated state transition probabilities P ad P' and the corresponding optimal policy π_i^* and $\pi_i'^*$, respectively. We assume that P is stationary. Note that $\mathbf{J}^*$ can be computed by $\mathbf{J}^* = P\mathbf{r} + \gamma_i P\mathbf{J}^*$ [7], or equivalently,

$$\mathbf{J}^* = [I - \gamma_i P]^{-1} P\mathbf{r} = [P + \gamma_i P^2 + \gamma_i^2 P^3 + \cdots]\mathbf{r},$$

where $\mathbf{r} = [r(s_{ik}^1), \cdots, r(s_{ik}^{n_i})]^T$. Similarly, $\mathbf{J'}^*$ can also be computed by $\mathbf{J'}^* = [I - \gamma_i P']^{-1} P'\mathbf{r}$. Without loss of generality, we consider a discounted expected reward from s_{ik}^l, i.e., $J^*(s_{ik}^l)$ and $J'^*(s_{ik}^l)$. To quantify them, we use a metric D_{J_l}, defined by $D_{J_l} = \left|J^*(s_{ik}^l) - J'^*(s_{ik}^l)\right|$. The results are shown in Fig. 4.

In Fig. 4 (a), since the true state transition probability matrix is stationary, if there are enough observations of resource reciprocation, the true state transition probability matrix can be well-identified by the empirical frequency based approach. However, it may require the observation of resource reciprocation among peers to obtain accurate state transition probabilities. In contrast, the approach based on the reciprocation models can identify the state transition probabilities based on fewer observations. However, unlike the empirical frequency based approach, the improvement gained for more observations diminishes rapidly (before reaching the best performance of the empirical frequency based approach), as the estimation relies on pre-determined reciprocation models.

As shown in Fig. 4 (a), the approach based on the reciprocation models provides a faster convergence, which becomes important when the state transition probabilities vary over time. Illustrative simulation results are shown in Fig. 4 (b). To emulate a dynamic environment, randomly generated different state transition probability matrices of a peer are deployed ev-

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

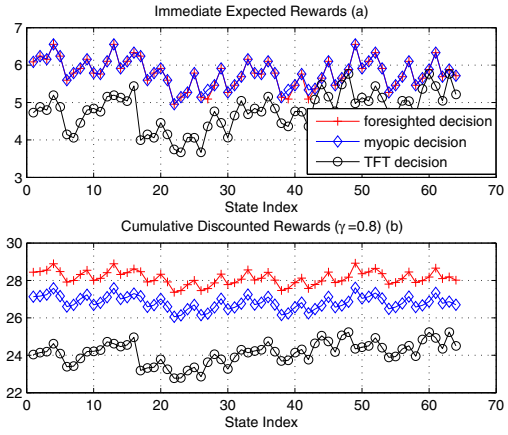

Fig. 5. Immediate (a) and cumulative expected discounted rewards (b) achieved by different policies.

ery 10 resource reciprocation. As expected, the approach based on the reciprocation models provides a faster convergence, thereby enabling peers to efficiently capture the changes of the state transition probability. Therefore, the proposed approach can cope with a dynamic environment, thereby making it more suitable than the empirical frequency based approach for P2P networks.

B. Impact of Myopic and Foresighted Policies on Rewards

The impact of the myopic and foresighted policies for their actions on achieved cumulative expected rewards are quantified. To focus on the impact of myopic and foresighted policies, we assume that the associated peers' behaviors are randomly generated and stationary. The solution to the MDP is implemented based on a well-known policy iteration method [7]. In addition, the TFT strategy implemented in BitTorrent-like system is compared, which supports two leechers simultaneously as an illustration. The simulation results are shown in Fig. 5.

Fig. 5 shows the immediate and cumulative expected discounted rewards with $\gamma = 0.8$ obtained based on the myopic, foresighted, and TFT policies. State index represents available states in the state space. To focus on the comparison among the considered policies, we assume that the rewards that can be obtained in each state are normalized as the state index. We can observe that the obtained rewards based on the TFT policy are the worst, since the actions determined by the TFT policy do not consider the expected rewards. Moreover, the constraints of fixed concurrent allowable uploads to the leechers can prevent the decision process from selecting better actions. As discussed previously, the myopic decisions are made based on (3), which maximize the immediate expected rewards. Hence, we can verify that the immediate expected rewards obtained by the actions of myopic policy are always higher (or equal) than the other policies in Fig. 5 (a). However, the foresighted decisions are made based on (4), such that they maximize the cumulative discounted expected rewards, as shown in Fig. 5 (b). Therefore, the foresighted policy enables the peers to determine their decisions that lead to the highest cumulative expected discounted rewards.

VI. Conclusion

In this paper, the resource reciprocation among the peers is modeled as a reciprocation game, and the game is formulated based on the MDP framework. Hence, peers can determine their resource reciprocation such that they can maximize their cumulative expected rewards. To successfully formulate the MDP framework in P2P networks, we propose the reciprocation model based approach that enables peers to efficiently identify the state transition probability matrix. We study the tradeoffs between efficiency and accuracy when different numbers of reciprocation models are deployed. In the simulations, we show that the proposed reciprocation based approach is more suitable for P2P networks. We also show that the proposed foresighted decisions lead to the best performance in terms of the cumulative expected rewards. Therefore, if the proposed approach is deployed in existing resource reciprocation strategies such as TFT in BitTorrent, it enables peers to make foresighted decisions on their resource allocation, leading to performance improvement.

References

[1] "Napster." [Online]. Available: http://www.napster.com
[2] "Gnutella." [Online]. Available: http://www.gnutella.com
[3] J. Liu, S. G. Rao, B. Li, and H. Zhang, "Opportunities and challenges of peer-to-peer internet video broadcast," *Proc. of the IEEE, Special Issue on Recent Advances in Distributed Multimedia Commun.*, 2007.
[4] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and sharing incentives in BitTorrent systems," INRIA-00112066, Tech. Rep. 1-21, Nov. 2006.
[5] X. Zhang, J. Liu, B. Li, and T. S. P. Yum, "CoolStreaming/DONet: A data-driven overlay network for efficient live media streaming," in *Proc. INFOCOM '05*, 2005.
[6] V. Pai, K. Kumar, K. Tamilmani, V. Sambamurthy, and A. E. Mohr, "Chainsaw: Eliminating trees from overlay multicast," in *Proc. 4th Int. Workshop on Peer-to-Peer Systems (IPTPS)*, Feb. 2005.
[7] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. Academic P, 1976.
[8] E. Haruvy, D. O. Stahl, and P. W. Wilson, "Evidence for optimistic and pessimistic behavior in normal-form games," *Economics Lett.*, vol. 63, pp. 255–259, 1999.
[9] K. Jain, L. Lovsz, and P. A. Chou, "Building scalable and robust peer-to-peer overlay networks for broadcasting using network coding," *Distributed Computing*, vol. 19, no. 4, pp. 301–311, 2007.
[10] G. de Veciana and X. Yang, "Fairness, incentives and performance in peer-to-peer networks," in *41th Annual Allerton Conference on Communication, Control and Computing*, 2003.
[11] E. Damiani, D. C. di Vimercati, S. Paraboschi, P. Samarati, and F. Violante, "A reputation-based approach for choosing reliable resources in peer-to-peer networks," in *Proc. the 9th ACM Conf. on Comput. and Commun. Security (CCS '02)*. ACM Press, 2002, pp. 207–216.
[12] M. Gupta, P. Judge, and M. Ammar, "A reputation system for peer-to-peer networks," in *Proc. the 13th Int. Workshop on Netw. and Operating Systems Support for Digital Audio and Video (NOSSDAV '03)*. ACM Press, 2003, pp. 144–152.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.