Evolution of Resource Reciprocation Strategies in P2P Networks

Hyunggon Park, *Member, IEEE*, and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—In this paper, we consider the resource reciprocation among self-interested peers in peer-to-peer (P2P) networks, which is modeled as a stochastic game. Peers play the game by determining their optimal strategies for resource distributions using a Markov decision process (MDP) framework. The optimal strategies enable the peers to maximize their long-term utility. Unlike in conventional MDP frameworks, we consider heterogeneous peers that have different and limited ability to characterize their resource reciprocation with other peers. This is due to the large complexity requirements associated with their decision making processes. We analytically investigate these tradeoffs and show how to determine the optimal number of state descriptions, which maximizes each peer's average cumulative download rates given a limited time for computing the optimal strategies. We also investigate how the resource reciprocation evolves over time as peers adapt their reciprocation strategies by changing the number of state descriptions. Then, we study how resulting download rates affect their performance as well as that of the other peers with which they interact. Our simulation results quantify the tradeoffs between the number of state descriptions and the resulting utility. We also show that evolving resource reciprocation can improve the performance of peers which are simultaneously refining their state descriptions.**

***Index Terms*—Evolution of resource reciprocation, Markov decision process (MDP), peer-to-peer (P2P) network, resource reciprocation, stochastic game.**

I. Introduction

In recent times, peer-to-peer (P2P) architectures for content distribution have been emerging as a key technology in the Internet and various research has been devoted to making P2P applications more efficient and robust (e.g., [1]–[5]).

In this paper, data-driven approaches adopting pull-based techniques [3], [4], [6], [7] are considered, where different types of data are divided into chunks of uniform length and are then disseminated over the P2P network. Each peer possesses several chunks, which are shared among interested peers, and information about the availability of the chunks is periodically exchanged among the associated peers. Based on this information, peers can form groups, collections of peers, with which they can continuously exchange their chunks. While this approach has been successfully deployed in various applications over P2P networks, key challenges such as determining optimal resource reciprocation strategies among self-interested peers still remain largely unaddressed. For example, several multimedia streaming protocols over P2P networks have been designed based on the assumption that peers cooperate with each other on disseminating multimedia chunks or peers are altruistic [4], [8], [9]. However, such protocols do not provide efficient solutions for a self-interested peer, which would like to maximize its own utility (e.g., [10]) by reciprocating its available resources.

The peer selection strategy deployed in BitTorrent systems is based on a tit-for-tat (TFT) strategy, where a peer selects some of its associated peers (i.e., leechers), which are *currently* uploading at the highest rates, and provides them its content for downloading [3]. While this strategy is currently deployed in BitTorrent systems, a key disadvantage of the peer selection strategies in BitTorrent systems is that peers decide how to determine their resource reciprocation by evaluating only the current upload rates which it receives from its associated peers, and does not consider how this reciprocation will impact their upload rates in the future. Moreover, these strategies do not consider the P2P system's dynamics, which will impact the resource reciprocation, because a peer should optimally determine its uploading policy based on the experienced dynamics and its expectation of how these dynamics will evolve in the future. In other words, each peer decides its peer selection and resource reciprocation to the other peers *myopically*. Since peers in P2P networks are generally involved in repeated and long-term interactions, such myopic decisions on peer selection and bandwidth allocation can result in a suboptimal performance for the involved peers. To take into account the repeated nature of the resource reciprocation among self-interested peers, an evolutionary instantiation of the prisoner's dilemma and the generalized prisoner's dilemma is proposed in [11] and [12] as a resource reciprocation model. However, these works only consider the case where peers have a limited set of simple actions, i.e., full cooperation (allowing download) or defection (ignoring download requests), but does not address how to divide each peer's available resources (i.e., the level of cooperation). Thus, they do not provide solutions for self-interested peers to allocate their resources such that they can maximize their cumulative download rates through long-term interactions.

Manuscript received April 05, 2009; accepted September 12, 2009. First published October 13, 2009; current version published February 10, 2010. This work was supported by the National Science Foundation (NSF) under CCF 0830556 and CCF 0541867. The material in this paper was presented in part at the 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, Taiwan, April 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Z. Jane Wang.

H. Park was with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA. He is now with the Department of Electronics Engineering, Ewha Womans University, Seoul 120-750, Korea (e-mail: hyunggon.park@gmail.com).

M. van der Schaar is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: mihaela@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2033731

To address these challenges, in our previous work [13], we model the resource reciprocation among self-interested peers as a resource reciprocation game, where each peer decides its actions (i.e., peer selection and resource allocations) by considering the probabilistic resource reciprocation behavior of its associated peers. The resource reciprocations among the peers are formulated as a Markov decision process (MDP) [14], which enables peers to take foresighted actions that maximize their expected cumulative utilities (e.g., download rates and multimedia quality). Note that this game theoretic modeling is only used here to formulate the dynamic and repeated interactions emerging among peers. However, the focus of this paper is on determining optimal strategies based on which the peers can “play” the reciprocation game in order to maximize their long-term performance. As shown in [13], the MDP-based foresighted strategies improve the performance of existing P2P applications, and thus, they can be incorporated into existing solutions as peer selection and resource allocation strategies. However, it is not investigated in [13] how heterogeneous peers decide the state granularity (i.e., how accurately describe the download rates) and how they interact with each other based on their different abilities.

Unlike in conventional MDP frameworks, where peers can completely recognize their states from their associated peers, we consider heterogeneous peers that have different and limited abilities to characterize their resource reciprocation due to their computational complexity and delay constraints. Hence, the resource reciprocation of each peer is described based on a finite number of *state descriptions*. While using more state descriptions enables the peers to more accurately model their resource reciprocation, it enlarges the state space, leading the peers to require more time to compute their optimal policies in an MDP framework. To address this tradeoff, we analytically show that an optimal number of state descriptions exists, which leads peers to maximize their time averaged cumulative download rates and multimedia quality given a limited time for computing the optimal strategies.

The optimal strategy computed based on a peer’s MDP model only identifies one optimal action for each state. However, peers cannot differentiate among all possible download rates from their associated peers because they have a limited number of state descriptions to characterize them. Consequently, a peer may have multiple actions that are optimal because these actions do not alter its associated peers’ states, and thus, they do not alter the resource reciprocation of these peers. This observation motivates the peers to *mutually* improve their resource reciprocation strategies over time. We analytically show that multiple optimal actions exist for heterogeneous peers who have different abilities to describe their resource reciprocation. Moreover, we show that peers can mutually improve their download rates (in addition to the download rates obtained by the optimal strategy determined using the MDP) only if they simultaneously refine their state descriptions. We also investigate that the heterogeneity of peers will drive their group formation, and conclude that peers prefer to form groups with other peers, which not only have similar or higher upload rates, but also similar abilities to refine their state descriptions.

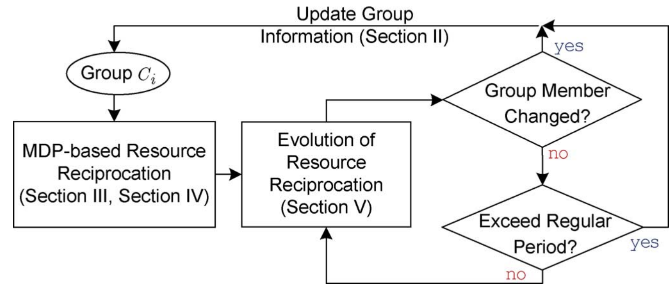


Fig. 1. Group update process and related processes.

This paper is organized as follows. In Section II, the proposed resource reciprocation model for P2P networks is presented. In Section III, an MDP-based resource reciprocation strategy is discussed. In Section IV, we analytically show that an optimal number of state descriptions exists for MDP-based resource reciprocation, which enables the peers to achieve their maximum average download rates. In Section V, we analytically study how the peers with different capabilities to refine their state descriptions can evolve their resource reciprocation strategies and discuss its impact on their download rates. Simulation results are shown in Section VI and conclusions are drawn in Section VII.

II. Resource Reciprocation Models

In general P2P networks, peers continuously interact with each other by repeatedly reciprocating their resources (i.e., upload bandwidth). In this paper, we focus on studying the resource reciprocation among directly connected peers based on the assumption that each peer can access its previous actions and observed actions from its associated peers.

A. Resource Reciprocation in Groups of P2P Networks

In data-driven P2P networks (e.g., [3], [4], [6], and [7]), peers would like to associate themselves with other peers which possess content in which they are interested. When peers possess content which another peer desires and *vice versa*, they negotiate how to share their content with each other (i.e., resource reciprocation). The set of peers which share their content and resources (i.e., associated peers) is referred to as a *group* in this paper.[1] The set of peers indexed by $1, \ldots, N_{C_i}$ in the group of peer i is denoted by C_i, $C_i = \{1, \ldots, N_{C_i}\}$. Note that if peer k is in group C_i, then it implies that peer i is also included in the group of peer C_k, i.e., $i \in C_k$. Due to the dynamics introduced by peers joining, leaving, or switching P2P networks, information about groups needs to be regularly (periodically) updated or it needs to be updated when group members change. The group update process and related processes are depicted in Fig. 1. We assume that each peer decides its resource reciprocation in its group using an MDP, which enables each peer to maximize its cumulative expected download rates.

Next, we briefly discuss the MDP-based resource reciprocation model in this paper. A similar MDP-based resource reciprocation model can be found in [13].

[1] A set of associated peers and a group are interchangeably used. Some examples of groups in P2P networks are swarms [7], partnerships [4], etc.

B. Resource Reciprocation Model

For a peer i, an MDP is a tuple $\langle \mathbf{S}_i, \mathbf{A}_i, P_i, R_i \rangle$, 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 that maps the state $s_i^{(t)} \in \mathbf{S}_i$ at time t, corresponding to an action $\mathbf{a}_i \in \mathbf{A}_i$ and the next state $s_i^{(t+\Delta t)} \in \mathbf{S}_i$ at time $t + \Delta t$ to a real number between 0 and 1, and $R_i : \mathbf{S}_i \rightarrow \mathbb{R}$ is the reward derived in state $s_i \in \mathbf{S}_i$.[2] The details are explained as follows.

1) State Space $\mathbf{S}_i$*:* A state of peer i represents the received resources from the peers in C_i. Let $(x_{1i}, \ldots, x_{N_{C_i}i})$ be peer i's resources received from its associated peers in C_i, where x_{ki} $(0 \leq x_{ki} \leq L_k, \forall k \in C_i)$ denotes the resources provided to peer i by peer k having its available maximum upload bandwidth L_k. Hence, x_{ki} is a peer i's download rate from peer $k \in C_i$ (i.e., upload rates of peer k to peer i). For peer i, download rates are recognized and represented by n_i discrete values[3] based on a function ψ_i, i.e., a download rate from peer x_{ki} is mapped into $\psi_i(x_{ki}) = s_{ik} \in \{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$. Hence, the state space of peer i can be expressed as

$$\mathbf{S}_i = \left\{ s_i = \left(s_{i1}, \ldots, s_{iN_{C_i}}\right) | s_{ik} = \psi_i(x_{ki}), k \in C_i \right\}.$$

Since a state s_i is represented by a set of N_{C_i} values of s_{ik} $(k \in C_i)$ and each of s_{ik} is selected among n_i discrete values of $\{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$, $\{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$ is referred to as a set of n_i *state descriptions* in this paper. Note that it can be easily observed that using more state descriptions results in higher accuracy for describing download rates. The impact of the number of state descriptions on resource reciprocation, evolution of resource reciprocation, and group formation is discussed in Sections IV–VI, respectively.

2) Action Space $\mathbf{A}_i$*:* An action of peer i is its resource allocation to the associated peers in C_i. We assume that the upload bandwidth of peer i is decomposed into "units" of bandwidth [16], denoted by Δx_i. Thus, the actions are the number of units of bandwidth that are allocated to the associated peers, i.e.,

$$\mathbf{A}_i = \left\{ \mathbf{a}_i \middle| 0 \leq a_{ik} \leq \left\lfloor \frac{L_i}{\Delta x_i} \right\rfloor, \sum_{k \in C_i} a_{ik} \leq \left\lfloor \frac{L_i}{\Delta x_i} \right\rfloor, k \in C_i \right\}$$

where $\mathbf{a}_i \in \mathbb{Z}_+^{N_{C_i}}$. Hence, action a_{ik} determines peer k's download rate from peer i, i.e., $x_{ik} = a_{ik} \Delta x_i$. Note that the actions of a peer determine the states of its associated peers.

3) State Transition Probability $P_{\mathbf{a}_i}(s_i, s_i')$*:* A state transition probability represents the probability that an action of a peer in a state will lead to another state. Thus, given a state $s_i \in \mathbf{S}_i$ at time $t - \Delta t$, 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 with probability $P_{\mathbf{a}_i}(s_i, s_i')$, expressed as

$$P_{\mathbf{a}_i}\left(s_i, s_i'\right) = \Pr\left(s_i^{(t)} = s_i' | s_i^{(t-\Delta t)} = s_i, \mathbf{a}_i^{(t)} = \mathbf{a}_i\right).$$

[2]We assume that a resource reciprocation requires time duration Δt, which can be introduced due to estimating or measuring each peer's download rates [15].

[3]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., [16].

In this paper, we assume that the state transition probabilities of peers are known; that is, they have been efficiently identified based on, e.g., [13], and available.

4) Reward R_i*:* The reward of a peer in a state represents its total download rate in that state. Since the state s_i of peer i represents the set of individual download rates from its associated peers, the total download rates in a state can be explicitly determined. In this paper, we assume that the reward in a state $s_i = (s_{i1}, \ldots, s_{iN_{C_i}})$ represents a guaranteed download rate, i.e., $R(s_i) = \sum_{k \in C_i} U_i(s_{ik})$, where $U_i(s_{ik})$ denotes the guaranteed download rates from peer k. Note that the reward in a state can alternatively represent the average or maximum download rates by defining the U_i function accordingly. A more detailed discussion of the function U_i will be given in Section IV-A.

5) Resource Reciprocation Policy π_i^**:* The solution to the MDP is represented by peer i's optimal resource reciprocation policy π_i^*, which is a mapping from the states to optimal actions, i.e., $\pi_i^*(s_i) = \mathbf{a}_i^*$ for all $s_i \in \mathbf{S}_i$. Hence, peer i can decide its optimal actions in each state based on the optimal policy π_i^*. The optimal policy π_i^* can be obtained using well-known algorithms such as value iteration and policy iteration [14]. Note that it is known that these algorithms always converge to π_i^* given state space, action space, reward function, and state transition probability function. We assume that the optimal policy is periodically updated to capture and cope with changes in the resource reciprocation behavior of associated peers. Moreover, the policy can be updated when a new peer joins or leaves the group.

III. MDP-based Resource Reciprocation

A. MDP-Based Resource Reciprocation Process

As discussed in Section II, a set of download rates from associated peers within a group is modeled as a peer's state. Given a state, a peer can decide its corresponding optimal action (i.e., upload bandwidth allocations) based on its optimal policy determined by the MDP. In addition to the optimal action, as will also be shown in Section V, a peer may have multiple actions that are optimal. This is because these actions do not alter its associated peers' states, and therefore, they do not alter the resource reciprocation from these peers. By selecting such actions, resource reciprocation can evolve for the peers that can successively refine their states, which leads them to mutually improve their download rates besides the rates obtained by the optimal strategy determined using the MDP. These general resource reciprocation processes, including MDP-based resource reciprocation and the evolution of resource reciprocation, are depicted in Fig. 2.

To cope with the dynamics introduced by peers joining or leaving P2P networks, each peer needs to update its optimal policy. Some examples that require policy updates include: 1) when a new peer joins the group, 2) an associated group member leaves, or 3) the resource reciprocation behavior of an associated peer changes. Note that for 3), even if the group members of a peer do not change, the resource reciprocation behavior of its associated peers can change as a result of changes in their corresponding groups. Hence, to adapt to dynamics that can

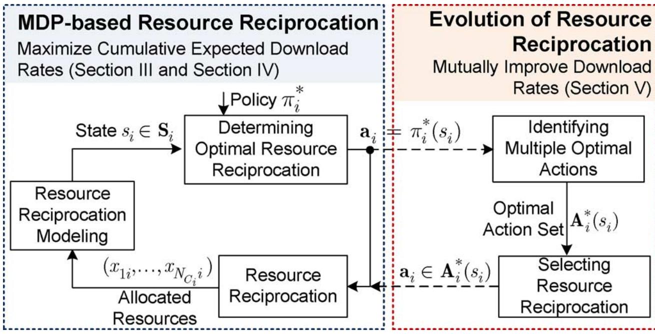


Fig. 2. Resource reciprocation process in group C_i.

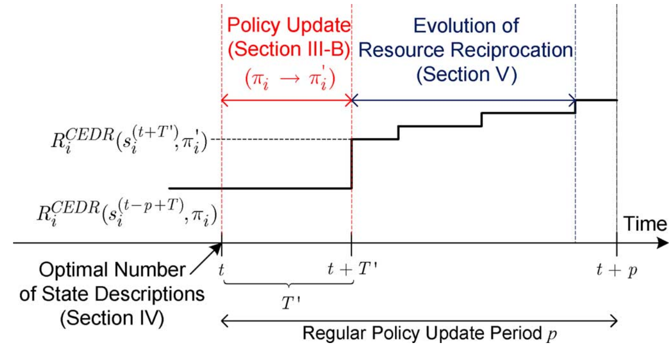


Fig. 3. Regular policy update and CEDRs. In previous period, time T was required for policy update.

propagate across the entire network, each peer needs to regularly update its own policy. The policy update processes play a similar role to the *periodic rechoke* in BitTorrent [3], [7], where a peer periodically (typically every 10 s) updates current download rates from its associated peers and shares its resources with the peers selected based on the TFT strategy. The rechoke process needs to be additionally performed whenever swarms are changed [7]. Hence, in the proposed approach, we assume that the optimal policy is updated regularly (with period p), or is additionally updated within the period (i.e., $p' < p$) when group member is changed. Note that it takes time for peers to update their policies as they need to find new policies, and the complexity of this computation depends on how the peers model their resource reciprocation (i.e., the number of states).[4] Thus, the peers should decide how accurately they model the resource reciprocation, as they may take actions that are not optimal while updating their policies, which can lead to worse cumulative download rates. An illustrative example for regular policy update and the corresponding cumulative (discounted) expected download rates (CEDR) is shown in Fig. 3.

Note that the time required to update policy π_i', denoted by T', and CEDR given policy π_i', denoted by $R_i^{\text{CEDR}}(s_i^{(t+T')}, \pi_i')$, depends on the accuracy for modeling the resource reciprocation, as it determines the size of the state space, which will be discussed in Section IV. In the next section, we study the MDP-based resource reciprocation strategy, where the policies are determined by the MDP framework.

B. MDP-Based Resource Reciprocation Strategy

As discussed in Section II-B, each peer's action represents the upload bandwidth allocation to its associated peers, and can be determined based on each peer's policy π. Peers that use the MDP framework for their resource reciprocation aim to maximize their CEDRs by strategically allocating their upload bandwidth. The CEDRs at time t can be expressed as

$$R_i^{\text{CEDR}}\left(s_i^{(t)}\right) = \sum_{h=0}^{\infty} \gamma_i^{h\Delta t} \cdot E\left[R\left(s_i^{(t+h\Delta t)}\right)\right] \qquad (1)$$

where γ_i $(0 \le \gamma_i < 1)$ is a constant referred to as a *discount factor*,[5] which determines a tradeoff between immediate reward and future reward, and $R(s_i^{(t)}) = \sum_{l=1}^{N_{C_i}} U_i(s_{il})$ for $s_i^{(t)} = (s_{i1}, \ldots, s_{iN_{C_i}})$. More precisely, the expression in (1) can be rewritten as (2), shown at the bottom of the page. Hence, peer i can determine a set of actions that maximizes $R_i^{\text{CEDR}}(s_i^{(t)})$ in (2) 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., $\pi_i^*(s_i) = \mathbf{a}_i^*$.

As shown in (2), an optimal policy and the corresponding optimal actions are primarily affected by the state transition probability. Moreover, the number of state descriptions determines the possible state transitions which in turn have an impact on the state transition probabilities. Therefore, the granularity of states (i.e., the number of state descriptions) of the peers has an impact on their policy as well as their corresponding actions, which ultimately affects the resulting download rates of all peers. This will be discussed next.

IV. Optimal Number of State Descriptions

A. Different State Granularity for Heterogeneous Peers

As discussed in Section II-B, for a peer i with n_i state descriptions, a function ψ_i can map a download rate x_{ki} from peer k into a state description s_{ik}, i.e.,

[4]For the value iteration algorithm, the complexity is quadratic in the number of states [17].

[5]Note that the discount factor γ_i in the considered P2P network can represent the belief of peer i about the validity of the expected future rewards, which is discussed in the Appendix and quantitatively discussed in Section VI-A.

$$\underbrace{\sum_{s_i^{(t+\Delta t)} \in \mathbf{S}_i} P_{\mathbf{a}_i^{(t+\Delta t)}}\left(s_i^{(t)}, s_i^{(t+\Delta t)}\right) R\left(s_i^{(t+\Delta t)}\right)}_{\text{immediate expected download rates}} + \underbrace{\sum_{h=2}^{\infty} \gamma_i^{(h-1)\Delta t} \sum_{s_i^{(t+h\Delta t)} \in \mathbf{S}_i} P_{\mathbf{a}_i^{(t+h\Delta t)}}\left(s_i^{(t+(h-1)\Delta t)}, s_i^{(t+h\Delta t)}\right) R\left(s_i^{(t+h\Delta t)}\right)}_{\text{future(discounted)expected download rates}} \qquad (2)$$

$\psi_i(x_{ki}) = s_{ik} \in \{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$. Specifically, we use a uniform quantization function ψ_i defined as

$$\psi_i(x_{ki}) = \begin{cases} s_{ik}^1, & \text{if } 0 \le x_{ki} < L_k^m + \overline{L_k^M L_k^m}/n_i \\ s_{ik}^l, & \text{if } L_k^m + \frac{l-1}{n_i}\overline{L_k^M L_k^m} \le x_{ki} < L_k^m + \frac{l}{n_i}\overline{L_k^M L_k^m} \\ s_{ik}^{n_i}, & \text{if } L_k^m + \overline{L_k^M L_k^m}/n_i \le x_{ki} \le L_k \end{cases} \quad (3)$$

where L_k^M and L_k^m denote peer k's maximum and minimum upload bandwidths, respectively, and $\overline{L_k^M L_k^m} \triangleq L_k^M - L_k^m$. Moreover, guaranteed download rates of peer i from peer k in s_{ik} can be correspondingly defined as

$$U_i(s_{ik}) = \begin{cases} 0, & \text{if } s_{ik} = s_{ik}^1 \\ L_k^m + \frac{l-1}{n_i}\overline{L_k^M L_k^m}, & \text{if } s_{ik} = s_{ik}^l,\ l = 2, \ldots, n_i \end{cases}$$

which is the lower bound of the quantization interval in (3). As shown in (3), if a smaller number of state descriptions is used given L_k^M and L_k^m, each state description represents a larger interval, which leads to peer i becoming more insensitive to download rate variations from its associated peers. For example, suppose that a download rate x_{ki}, which is determined as an optimal action for peer k, is currently mapped into $s_{ik}^l = \psi_i(x_{ki})$ based on (3) for peer i with n_i state descriptions. Then, peer i *cannot* recognize a download rate variation Δx_k if

$$\psi_i(x_{ki} - \Delta x_k) = \psi_i(x_{ki}).$$

Hence, although peer k may reduce its resource reciprocation by Δx_k from x_{ki}, peer i cannot recognize this resource variation. Thus, it is possible for peer k to reduce its upload rate to peer i without affecting peer i's state transition model. Hence, peer k can reduce the resources to peer i, while achieving the same CEDRs from peer i, since peer i will continue to determine the same optimal policy as long as the state transitions are not affected. On the other hand, if peer i increases the number of state descriptions, such that it can recognize the resource reciprocation variation Δx_k, the resource variation will change peer i's state. Hence, peer i may take a different optimal action based on a new policy, resulting in a different CEDRs for peer k.

In this paper, we assume that the peers increase or decrease the number of state descriptions by a factor of 2, i.e., for integer b, one state description can be refined using 2^b (if $b \ge 0$) state descriptions or 2^{-b} (if $b < 0$) state descriptions can be merged, which is equivalent to using b more or less "bits" in order to describe the resource reciprocation.

B. *State Granularity and CEDRs*

Recall that the number of state descriptions of a peer determines the size of state space, and thus, it affects the peers' policies and the corresponding CEDRs. A basic result for the relationship between the number of state descriptions and the CEDRs is stated in Lemma 1.

Lemma 1: Given a policy π_i for n_i state descriptions, a peer i can improve its CEDRs by using more state descriptions.

Proof: Let x_{ki} be a resource allocation from peer $k \in C_i$ to peer i, and it is represented by $\psi_i(x_{ki}) \in \{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$ for peer i. As discussed in (1), CEDRs in state $s_i^{(t)} = (s_{i1}, \ldots, s_{iN_{C_i}})$, $s_{ik} \in \{s_{ik}^1, \ldots, s_{ik}^{n_i}\}$ for $k \in C_i$, are expressed as

$$R_i^{\text{CEDR}}\left(s_i^{(t)}\right) = \sum_{h=0}^{\infty} \gamma_i^{h\Delta t} \cdot E\left[R\left(s_i^{(t+h\Delta t)}\right)\right] \quad (4)$$

where $R(s_i^{(t)}) = \sum_{l=1}^{N_{C_i}} U_i(s_{il})$. If a peer i uses b more bits to describe the resource reciprocation of peer k, then s_{ik}^l is refined using 2^b state descriptions for all $l = 1, \ldots, n_i$. We denote the refined state descriptions of s_{ik}^l by $s_{ik}^{l_1}, \ldots, s_{ik}^{l_{2^b}}$. Thus, the state $s_i^{(t)} = (s_{i1}, \ldots, s_{iN_{C_i}})$ can be represented by $(s_i^{(t)})' = (s'_{i1}, \ldots, s'_{iN_{C_i}})$, where $s'_{ik} \in \{s_{ik}^{l_m} | 1 \le l \le n_i, 1 \le m \le 2^b\}$ for $k \in C_i$. Since $U_i(s_{ik}^l)$ is defined as download rates that are guaranteed in s_{ik}^l, we have $U_i(s_{ik}^{l_m}) \ge U_i(s_{ik}^l)$, for all $m = 1, \ldots, 2^b$ and $l = 1, \ldots, n_i$. This leads to

$$R\left(\left(s_i^{(t)}\right)'\right) \ge R\left(s_i^{(t)}\right).$$

Since policy π_i is determined for n_i state descriptions, actions that correspond to the refined state descriptions remain unchanged, i.e., $\pi_i(s_{ik}^{l_m}) = \pi_i(s_{ik}^l)$ for $1 \le m \le 2^b$. Therefore

$$R_i^{\text{CEDR}}\left(\left(s_i^{(t)}\right)'\right) \ge R_i^{\text{CEDR}}\left(s_i^{(t)}\right). \qquad \blacksquare$$

From Lemma 1, we can conclude that a peer does not worsen its benefit from (unilaterally) refining its state descriptions for the download rates from associated peers, as it guarantees an improvement to their CEDRs. In Lemma 2, we quantitatively show the improvement of the maximum CEDRs if the number of state descriptions increases.

Lemma 2: The maximum CEDR improvement for each additional bit used for state descriptions is exponentially decreasing.

Proof: Suppose that peer i uses 2^B state descriptions for peer k, and assume that $L_k^m = 0$ and $L_k^M = L_k$ without loss of generality. Assume that a resource allocation x_{ki} from peer k is mapped into $\psi_i(x_{ki}) = s_{ik}^l$, i.e., $(L_k/2^B)(l-1) \le x_{ki} < (L_k/2^B)l$. Hence, $U_i(s_{ik}^l) = (L_k/2^B)(l-1)$. If peer i increases the number of state descriptions by 2^{B+b} (i.e., 2^b more state descriptions are used in each state description), the expected $U'_i(s_{ik}^l)$ based on the refined state descriptions can be expressed as

$$U'_i\left(s_{ik}^l\right) = \frac{L_k}{2^B}(l-1) + \frac{L_k}{2^B}\sum_{d=1}^{2^b} p_{l_d}\frac{d-1}{2^b}$$

where $p_{l_d} = \Pr(\psi_i(x_{ki}) = s_{ik}^{l_d})$ is the probability that the download rate x_{ki} is mapped into the dth refined state description $s_{ik}^{l_d}$ of state s_{ik}^l. Since a set of possible probability distributions can be expressed as $\mathbf{p}_l^d = \{(p_{l_1}, \ldots, p_{l_{2^b}}) | 0 \le p_{l_d} \le$

$1, \sum_{d=1}^{2^b} p_{l_d} = 1\}$ for all d, the maximum improvement can be expressed as

$$
\begin{aligned}
\max_{\mathbf{p}_l^d} &\left\{U_i'\left(s_{ik}^l\right)\right\} - U_i\left(s_{ik}^l\right) \\
&= \max_{\mathbf{p}_l^d} \left\{ \frac{L_k}{2^B}(l-1) + \frac{L_k}{2^B} \sum_{d=1}^{2^b} p_{l_d} \frac{d-1}{2^b} \right\} - U_i\left(s_{ik}^l\right) \\
&= \left[\frac{L_k}{2^B}(l-1) + \frac{L_k}{2^B}\frac{2^b-1}{2^b} \right] - \frac{L_k}{2^B}(l-1) = \frac{L_k}{2^B}\frac{2^b-1}{2^b}
\end{aligned}
$$

where the maximum can be achieved if $d = 2^b$ and

$$
p_{l_d} = \begin{cases} 1, & \text{if } d = 2^b \\ 0, & \text{otherwise} \end{cases}
$$

as $U_i(s_{ik}^{l_{r_1}}) \geq U_i(s_{ik}^{l_{r_2}})$ for $l_{r_1} \geq l_{r_2}$, $r_1, r_2 \in \{1, \ldots, 2^b\}$. Thus, the maximum improvement of $U_i(s_{ik}^l)$ by using two additional state descriptions (i.e., 1-bit increase) from 2^a $(a < b)$ state descriptions can be represented by

$$
\frac{L_k}{2^B}\frac{2^{a+1}-1}{2^{a+1}} - \frac{L_k}{2^B}\frac{2^a-1}{2^a} = \frac{L_k}{2^B}\frac{1}{2^{a+1}}. \tag{5}
$$

Since (5) holds for all $l = 1, \ldots, 2^B$, the maximum CEDR improvement for each additional bit used for the state descriptions is exponentially decreasing. ■

Lemma 2 shows that the maximum CEDR improvement can be bounded as the number of bits used for the state descriptions increases. Thus, for example, while the first two more state descriptions (i.e., $a = 0$) can lead to 50% improvement, adding two more state descriptions (i.e., $a = 1$) can lead to only 25% additional improvement. However, the complexity required to compute the optimal policy based on an MDP increases as the number of state descriptions increases [17]. For example, as shown in [17], the value iteration algorithm requires $O(|\mathbf{A}||\mathbf{S}|^2)$ complexity for each iteration, where $|\mathbf{A}|$ and $|\mathbf{S}|$ denote the number of actions and the number of states. Moreover, the maximum number of iterations $N_{\text{iter}}^M(B, \epsilon, \gamma)$ needed to find an ϵ-optimal policy also increases, as it is given by

$$
N_{\text{iter}}^M(B, \epsilon, \gamma) = \frac{B + \log(1/\epsilon) + \log(1/(1-\gamma)) + 1}{1-\gamma} \tag{6}
$$

where B denotes bits used to represent state descriptions and γ is the discount factor [17]. Thus, increasing the number of state descriptions obviously leads to higher computational complexity, which implies that it takes a longer time to compute the optimal policy. Hence, peers cannot always increase the number of state descriptions in order to improve their CEDRs. These tradeoffs should be explicitly considered when peers determine their numbers of state descriptions. Considering the CEDR improvement and the corresponding time required to compute an optimal policy, a peer can determine an optimal number of state descriptions that can maximize its time averaged CEDRs in a period. As shown in Fig. 4, the average CEDRs in a period p can be expressed as

$$
\begin{aligned}
&R_i^{\text{CEDR}}\left(s_i^{(t-p+T(B))}, \pi_i\right) \frac{T'(B+b)}{p} \\
&\quad + R_i^{\text{CEDR}}\left(s_i^{(t+T'(B+b))}, \pi_i'\right) \frac{p - T'(B+b)}{p}
\end{aligned} \tag{7}
$$

where the CEDRs that can be achieved by policies π_i and π_i' from state s_i at time $t-p+T(B)$ and $t+T'(B+b)$ are denoted

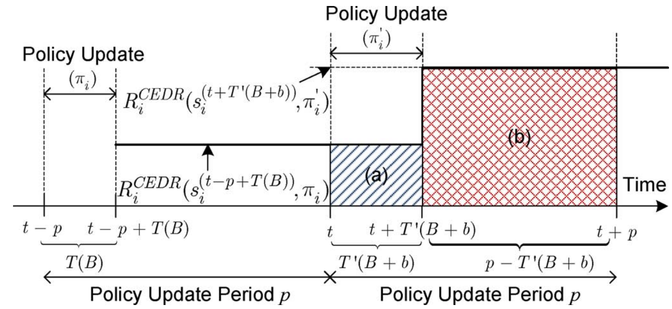


Fig. 4. Regular policy update and CEDRs. Time averaged CEDR is given by (a) + (b).

by $R_i^{\text{CEDR}}(s_i^{(t-p+T(B))}, \pi_i)$ and $R_i^{\text{CEDR}}(s_i^{(t+T'(B+b))}, \pi_i')$, respectively. $T(B)$ and $T'(B+b)$ represent the time required to compute the optimal policies π_i and π_i' with B and $B+b$ bits for state descriptions, respectively. Proposition 3 shows the existence of an optimal number of state descriptions that leads to the maximum average CEDR in a period.

Proposition 3: Given a limited time for computing the optimal policy (period p), there exists an optimal number of state descriptions such that peer i can maximize its average CEDRs in the current period. Moreover, the optimal number of state descriptions is upper-bounded by $2^{b_{\text{MAX}}}$, where $b_{\text{MAX}} = \lceil \ln(p/\overline{T})/(\ln 4 + 1/\alpha) \rceil$ for $\alpha = B + \log_2(1/\epsilon) + \log_2(1/(1-\gamma)) + 1$ and $\overline{T} = T\alpha/((1-\gamma)N_{\text{iter}}(B, \epsilon, \gamma))$.

Proof: The CEDRs are bounded and the number of state descriptions is determined by the number of used bits, which are discrete. Thus, it is enough to show the existence of bounds for the number of bits.

Let p be a period for a regular policy update. We assume that peer i is currently using B bits for its state descriptions for peer k, and it can refine each state description using 2^b more state descriptions (i.e., b more bits). Since maximum time T' required to compute an optimal policy cannot exceed the policy updating period, the maximum number of bits that can be increased can be obtained using (6) as follows:

$$
\begin{aligned}
T' &= T \cdot 2^{2b} \cdot \frac{N_{\text{iter}}^M(B+b, \epsilon, \gamma)}{N_{\text{iter}}(B, \epsilon, \gamma)} \\
&= T \cdot \frac{2^{2b}}{N_{\text{iter}}(B, \epsilon, \gamma)} \cdot \frac{b+\alpha}{1-\gamma} \leq p
\end{aligned}
$$

where $\alpha \triangleq B + \log_2(1/\epsilon) + \log_2(1/(1-\gamma)) + 1 \geq 1$, and T and $N_{\text{iter}}(B, \epsilon, \gamma)$ denote the time and the number of iterations required to compute the last optimal policy, respectively. To find the upper bound of b, let b_{MAX} be a minimum value of b such that it exceeds p, i.e.,

$$
b_{\text{MAX}} = \min\left\{ b | p/\overline{T} \leq 4^b(1 + b/\alpha), b \in \mathbb{Z} \right\} \tag{8}
$$

where $\overline{T} = T\alpha/((1-\gamma)N_{\text{iter}}(B, \epsilon, \gamma))$. Since $(1 + b/\alpha) \leq e^{b/\alpha}$, $p/\overline{T} \leq 4^b(1 + b/\alpha)$ in (8) can be rewritten as

$$
\frac{p}{\overline{T}} \leq 4^b \left(1 + \frac{b}{\alpha}\right) \leq 4^b e^{b/\alpha} = (4e^{1/\alpha})^b
$$

which leads to

$$b \geq \frac{\ln\left(p/\overline{T}\right)}{\ln 4 + 1/\alpha}. \tag{9}$$

Therefore, $b_{\mathrm{MAX}} = \lceil \ln(p/\overline{T})/(\ln 4 + 1/\alpha) \rceil$. Hence, the possible number of bits that peer i can use has a bound, expressed as

$$-B \leq b < b_{\mathrm{MAX}} = \left\lceil \frac{\ln(p/\overline{T})}{\ln 4 + 1/\alpha} \right\rceil$$

which completes the proof. ∎

Algorithm 1 Determining Optimal Number of State Descriptions

Given: $\{T(B), N_{iter}(B,\epsilon,\gamma), 2^B\}$: information from previous period, $R_i^{CEDR}\left(s_i^{(t-p+T(B))}, \pi_i\right)$, $2^{b'}$: initial number of state descriptions for a new joining peer

1: $b_{MAX} \leftarrow \lceil \ln(p/\overline{T})/(\ln 4 + 1/\alpha) \rceil$.
CASE: Regular Policy Update
// Find optimal b^* that maximizes the average CEDRs
2: $b \leftarrow -B$
3: **for** b such that $-B \leq b \leq b_{MAX}$,
4: $T'(B+b) \leftarrow T(B) \cdot 2^{2b} \cdot \frac{N_{iter}^M(B+b,\epsilon,\gamma)}{N_{iter}(B,\epsilon,\gamma)}$
5: $V_1 \leftarrow R_i^{CEDR}\left(s_i^{(t-p+T(B))}, \pi_i\right)$
6: $V_2 \leftarrow R_i^{CEDR}\left(s_i^{(t+T'(B+b))}, \pi_i'\right)$
7: $V \leftarrow V_1 \cdot T'(B+b) + V_2 \cdot (p - T'(B+b))$
8: **if** V is maximum,
9: $b^* \leftarrow b$
CASE: Policy Update due to Peer k's Leaving
10: $V_1 \leftarrow R_i^{CEDR}\left(s_i^{(t-p+T(B))}, \pi_i\right)$
11: $V_2 \leftarrow R_i^{CEDR}\left(s_i^{(t-p+T(B))} \backslash s_{ik}, \pi_i\right)$
12: $R_i^{CEDR}\left(s_i^{(t-p+T(B))}, \pi_i\right) \leftarrow V_1 - V_2$
13: **do** Step 2–9
CASE: Policy Update due to Peer k's Joining
14: **if** $T'(B+b') = T(B) \cdot 2^{2b'} \cdot \frac{N_{iter}^M(B+b',\epsilon,\gamma)}{N_{iter}(B,\epsilon,\gamma)} < p$
15: $s_i \leftarrow s_i \cup \{s_{ik}\}$
16: **do** Step 2–9
17: **else**
18: Reject joining

Negative values of b mean that peer i can decrease the number of its state descriptions. Note that b_{MAX} is determined by $\ln(p/\overline{T})$, which enables peers to consider only a small number of possible values of b in practice. Hence, an algorithm that simply computes and compares the average CEDRs obtained using different number of bits for state descriptions can be deployed.

Recall that we designed our system such that the number of state descriptions can be increased or decreased by a factor of 2. Such state description refinements are simple to implement, and more importantly, enable each peer to estimate its new state transition probability functions efficiently as shown in [18] when the number of state descriptions changes. However, even if an optimal number of bits b^* that maximizes the average CEDRs is determined, there are still 2^{b^*} possible state descriptions in the interval $[2^{b^*}, 2^{b^*+1})$. In order to find an optimal number of state description in the interval $[2^{b^*}, 2^{b^*+1})$, interpolation techniques can be used, such that average CEDRs for $K \in [2^{b^*}, 2^{b^*+1})$ state descriptions are computed. For this extension, however, peers need to estimate their new state transition probability functions by observing new resource reciprocations, whenever they change the number of their state descriptions. This may result in significant delays in addition to $T'(B+b)$, which is undesirable. Hence, we assume that only 2^b state descriptions are available in this paper.

Note that it may not be practical to explicitly compute the average CEDRs when peers decide the number of state descriptions. This is because computing the average CEDRs requires finding optimal policies. However, as shown in Lemma 2, since the CEDR improvement is exponentially decreasing in the number of additional bits used for state descriptions, peers can approximate the average CEDRs when they determine an optimal number of state descriptions. The process for determining the number of state descriptions of peer i is summarized in Algorithm 1.

After peers determine the number of their state descriptions that maximizes their CEDRs, they can additionally improve their mutual download rates by evolving their resource reciprocation strategies if multiple optimal actions are available to them. This will be discussed in the next section.

V. Evolution of Resource Reciprocation

As discussed in Section IV, each peer can determine its optimal number of state descriptions given a limited time for computing its policy. If peers have other optimal actions besides an optimal action determined by the policy, the peers can improve their download rates by selecting their actions cooperatively. In this section, we show the existence of the multiple optimal actions and study the impact of cooperative resource reciprocation. Note that we assume that each peer's optimal policy is determined given a number of state descriptions using the approach in Section IV. Hence, the peers do *not* need to compute their new policies, while evolving their resource reciprocation in this section.

To highlight the impact of the different number of state descriptions on the peers' interactions, we first study the case where a peer having finer state descriptions interacts with several associated peers having coarser state descriptions. We assume that peer i uses the finest state descriptions among its associated peers in C_i. Moreover, we assume that $\Delta x_i \ll (L_i^M - L_i^m)/n_k$ and $\Delta x_k \ll (L_k^M - L_k^m)/n_i$ for all $k \in C_i$, that is, the unit of bandwidth for each peer's upload bandwidth allocation is small enough compared to the interval that corresponds to each state description. We consider a peer i in a state $s_i \in \mathbf{S}_i$ interacting with the peers in C_i and determining its optimal actions based on its optimal policy π_i^*. A result for the interactions among the peers in C_i is summarized in Lemma 4.

Lemma 4: If the unit of bandwidth for each peer's upload bandwidth allocation is small enough, then peers can have multiple optimal actions in each of their states.

Proof: Let $\mathbf{a}_i^* = (a_{i1}^*, \ldots, a_{iN_{C_i}}^*)$ be an optimal action of a peer i in state s_i determined by π_i^*, i.e., $\pi_i^*(s_i) = \mathbf{a}_i^*$. There exists an action $\mathbf{a}_i = (a_{i1}, \ldots, a_{iN_{C_i}}) \neq \mathbf{a}_i^*$, where $a_{il} = a_{il}^* + v_l$ for all $l \in C_i$ and for some $v_l \in \mathbb{Z}$, such that

$$\psi_l(a_{il}) = \psi_l\left(a_{il}^*\right)$$

for all $l \in C_i$. Then, $\mathbf{a}_i$ is also optimal, since it does not change the states of peer $l \in C_i$. Hence, peer i can have multiple optimal actions given its state $s_i \in \mathbf{S}_i$ and optimal policy π_i^*. ■

Lemma 4 shows that peers can have multiple optimal actions. A set of multiple optimal actions, referred to as *optimal action set*, for peer i in state s_i is denoted by $\mathbf{A}_i^*(s_i)$, which is defined as

$$\mathbf{A}_i^*(s_i) = \{\mathbf{a}_i | a_{il} = a_{il}^* + v_l \Delta x_i, \psi_l(a_{il}) = \psi_l\left(a_{il}^*\right), \quad v_l \in \mathbb{Z} \text{ for all } l \in C_i\} \tag{10}$$

where $\pi_i^*(s_i) = \mathbf{a}_i^*$ for $s_i \in \mathbf{S}_i$. Hence, given a policy π_i^*, using different optimal actions in $\mathbf{A}_i^*(s_i)$ results in the same CEDRs for peer i, unless the associated peers refine their state descriptions such that they can differentiate the actions. As will be shown in Proposition 5, peer i and peer k, which are interacting with each other and have multiple optimal actions, can improve their download rates, only if both of them increase the number of their state descriptions *simultaneously*.

Proposition 5: Suppose that two peers, peer i and peer k, have multiple optimal actions and interact with each other. If they simultaneously refine their state descriptions, then actions exist which can improve both peers' download rates.

Proof: Let $\mathbf{A}_i^*(s_i^{(t)})$ and $\mathbf{A}_k^*(s_k^{(t)})$ be optimal sets of peer i and peer k for all $s_i \in \mathbf{S}_i$ and for all $s_k \in \mathbf{S}_k$ at time t, respectively. Consider the following interactions.

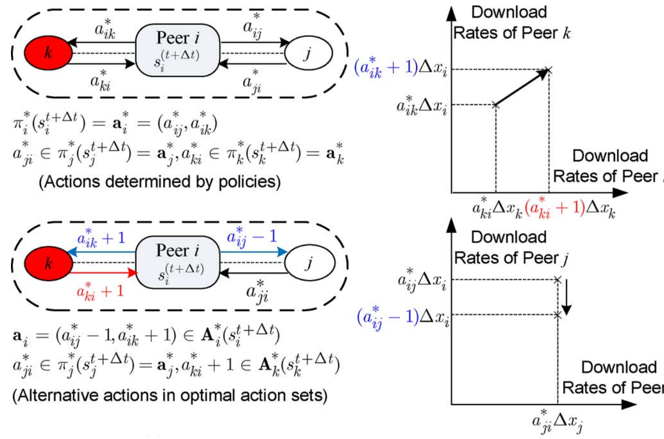
Case 1: No Associated Peer Is Refining Its States. By definition, different actions $\mathbf{a}_i = (a_{i1}, \ldots, a_{iN_{C_i}}) \in \mathbf{A}_i^*(s_i^{(t+w_i\Delta t)})$ of peer i at time $t + w_i\Delta t$ ($w_i \in \mathbb{N}$) do not change the states of its associated peers $l \in C_i$ because

$$\psi_l(a_{il}) = \psi_l\left(a_{il}^*\right)$$

for all $l \in C_i$. This implies that peer i may not improve its average CEDRs by taking different actions in $\mathbf{A}_i^*(s_i^{(t+w_i\Delta t)})$, i.e., peer i is indifferent to actions $\mathbf{a}_i \in \mathbf{A}_i^*(s_i^{(t+w_i\Delta t)})$.

Case 2: Only One Peer Is Refining Its States. Suppose that peer $k \in C_i$ can refine its states for peer i and let $\pi_k^*(s_k^{(t+w_k\Delta t)}) = \mathbf{a}_k^* = (a_{k1}^*, \ldots, a_{ki}^*, \ldots, a_{kN_{C_k}}^*) \in \mathbf{A}_k^*(s_k^{(t+w_k\Delta t)})$ be an optimal action determined by π_k^* at time $t + w_k\Delta t$. While peer k can take $\mathbf{a}_k = (a_{k1}, \ldots, a_{ki}, \ldots, a_{kN_{C_k}}) \in \mathbf{A}_k^*(s_k^{(t+w_k\Delta t)})$ such that $a_{ki} > a_{ki}^*$ instead of $\mathbf{a}_k^*$ at time $t + w_k\Delta t$, peer i cannot differentiate between a_{ki}^* and a_{ki} unless peer i refines its state. Thus, this interaction becomes Case 1.

Case 3: Both Peer i and Peer k Are Refining Their States Simultaneously. Suppose that peer i and peer $k \in C_i$ can both refine their states and let $\pi_i^*(s_i^{(t+w_i\Delta t)}) = \mathbf{a}_i^* = (a_{i1}^*, \ldots, a_{ik}^*, \ldots, a_{iN_{C_i}}^*) \in \mathbf{A}_i^*(s_i^{(t+w_k\Delta t)})$ and $\pi_k^*(s_k^{(t+w_k\Delta t)}) = \mathbf{a}_k^* = (a_{k1}^*, \ldots, a_{ki}^*, \ldots, a_{kN_{C_k}}^*) \in \mathbf{A}_k^*(s_k^{(t+w_k\Delta t)})$ be optimal actions determined by their policies, respectively. If peer i takes action $\mathbf{a}_i (\neq \mathbf{a}_i^*) \in \mathbf{A}_i^*(s_i^{(t+w_i\Delta t)})$ and peer k takes action $\mathbf{a}_k (\neq \mathbf{a}_k^*) \in \mathbf{A}_k^*(s_k^{(t+w_k\Delta t)})$, where



(a) (b)

Fig. 5. Illustrative example for evolution of resource reciprocation. (a) Peers can take actions determined by their policies or take alternative optimal actions. (b) Corresponding changes of the download rates. Peer i and peer k mutually improve their download rates.

$w_i \neq w_k$, then this interaction becomes Case 2. However, if $w_i = w_k = w$, then they can agree on taking actions $\mathbf{a}_i = (a_{i1}, \ldots, a_{ik}, \ldots, a_{iN_{C_i}}) \in \mathbf{A}_i^*(s_i^{(t+w_i\Delta t)})$ and $\mathbf{a}_k = (a_{k1}, \ldots, a_{ki}, \ldots, a_{kN_{C_k}}) \in \mathbf{A}_k^*(s_k^{(t+w_k\Delta t)})$ such that $a_{ik} > a_{ik}^*$ and $a_{ki} > a_{ki}^*$, as both peer i and peer k *mutually* improve their download rates.

Therefore, both peers can improve their download rates only if they simultaneously increase the number of their state descriptions. ■

Since multiple optimal actions are available for peers, successively increasing the number of state descriptions can provide the peers successive improvement of their download rates, i.e., the resource reciprocation evolves. The evolution of resource reciprocation is quantitatively evaluated through simulation results in Section VI.

Fig. 5 shows an illustrative example for Proposition 5. In this example, we assume that peer i has an optimal action $\mathbf{a}_i^* = (a_{ij}^*, a_{ik}^*)$ determined by $\pi_i^*(s_i^{(t+\Delta t)})$. Based on $\mathbf{a}_i^*$, peer i finds an alternative optimal action $\mathbf{a}_i$, leading to the optimal action set $\mathbf{A}_i^*(s_i^{(t+\Delta t)}) = \{\mathbf{a}_i^*, \mathbf{a}_i\}$ at time $t + \Delta t$. Similarly, peer k has multiple optimal actions $\mathbf{A}_k^*(s_k^{(t+\Delta t)}) = \{\mathbf{a}_k^*, \mathbf{a}_k\}$, where $\pi_k^*(s_k^{(t+\Delta t)}) = \mathbf{a}_k^*$ and $a_{ki} \in \mathbf{a}_k$ such that $a_{ki} > a_{ki}^*$. We also assume that both peer i and peer k can refine their states but peer j cannot. As discussed in Proposition 5, if peer i and peer k refine their states, their resource reciprocations can evolve by enabling peers to select cooperative actions which improve their download rates. In this example, peer i selects $\mathbf{a}_i$ instead of $\mathbf{a}_i^*$, and peer k can also select its action $\mathbf{a}_k$ such that it benefits peer i, e.g., $a_{ki} = a_{ki}^* + 1$. Therefore, both peer i and peer k can mutually improve their download rates. However, peer j would not select its alternative action against peer i's alternative action a_{ij}, because peer j cannot differentiate between a_{ij}^* and a_{ij}. The resulting download rate changes are shown in Fig. 5(b). We note that the maximum improvement of the download rates through this evolution of resource reciprocation is bounded by the number of state descriptions when the optimal policy is computed. This is because the multiple optimal actions can be found

Procedure 2 Evolution of Resource Reciprocation Process for Peer i and Peer k in C_i

Given: b^*, $\hat{b}_i$: number of bits for current and maximum state descriptions, $s_i = \left(s_{i1}, \ldots, s_{iN_{C_i}}\right)$, π_i^*, $\mathbf{a}_i^* = \pi_i^*(s_i)$, $\mathbf{A}_i^*(s_i) \leftarrow \mathbf{a}_i^*$
1: **while** $b^* < \hat{b}_i$
2: Refine state descriptions of peer i; $b^* \leftarrow b^* + 1$
3: **for** $\mathbf{a}_i (\neq \mathbf{a}_i^*) \in \mathbf{A}_i^*(s_i)$ // randomly selected action in $\mathbf{A}_i^*(s_i)$
4: **if** $\psi_i(x_{ki}) > s_{ik}$ // if cooperative resource reciprocation
5: $\mathbf{A}_i^*(s_i) \leftarrow \{\mathbf{a}_i | \, a_{ik} > a_{ik}^*, \text{ for all } \mathbf{a}_i \in \mathbf{A}_i^*(s_i)\}$

only after optimal actions are determined based on the optimal policy, as shown in (10).

The result from Proposition 5 can be extended to the case where multiple peers increase their number of state descriptions simultaneously, which is summarized in Corollary 6.

Corollary 6: If a peer i and its associated peers $k_1, k_2, \ldots, k_l$ ($l \geq 2$) have multiple optimal actions and they simultaneously increase their number of state descriptions, actions exist which can improve each peer's download rates.

Proof: Let $\mathbf{A}_i^*(s_i)$ be an optimal action set for peer i in state s_i. If peers $k_1, k_2, \ldots, k_l$ ($l \geq 2$) increase the number of their state descriptions, then there exists a set of actions $\mathbf{A}_i(s_i) \subseteq \mathbf{A}_i^*(s_i)$ for peer i such that

$$\mathbf{A}_i(s_i) = \left\{\mathbf{a}_i | a_{ik_h} \geq a_{ik_h}^*, \text{ for all } 1 \leq h \leq l\right\}. \quad (11)$$

With the same arguments in Proposition 5, if peer i and peers $k_1, k_2, \ldots, k_l$ refine their states, they can agree that peer i takes an action $\mathbf{a}_i \in \mathbf{A}_i(s_i)$ in (11) and each peer k_h also takes its action such that $a_{k_h i} > a_{k_h i}^*$. This lead them to mutually improve their download rates. ■

From Proposition 5 and Corollary 6, we can conclude that the resource reciprocation among peers which simultaneously increase their number of state descriptions can constructively evolve, leading to higher download rates for the involved peers. Moreover, these results also highlight that peers that use a fixed number of state descriptions in their resource reciprocation games will be penalized by their associated peers that can keep refining their state descriptions. Therefore, peers tend to form groups with peers having similar abilities to refine their state descriptions. These observations are quantitatively verified through several illustrative simulation results in Sections VI-B and VI-C. The evolution of resource reciprocation process between peer i and peer k is summarized in Procedure 2.

VI. Simulation Results

In this section, we quantify the impact of the number of state descriptions on each peer's performance, and study the evolution of resource reciprocation among peers that have multiple optimal actions. Then, we investigate how heterogeneous peers form their groups. Note that in our previous work [13], we compared the proposed MDP-based resource reciprocation strategies to existing solutions such as the TFT strategy in BitTorrent systems, and showed that the MDP-based approach improves the performance.

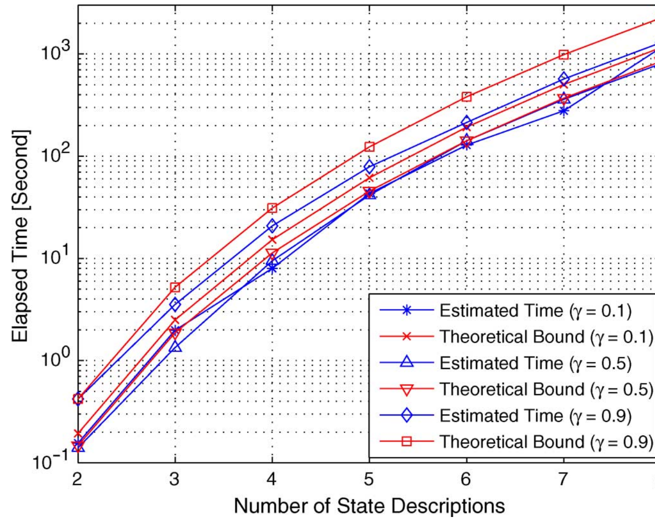


Fig. 6. Time required to compute optimal policies for different number of state descriptions.

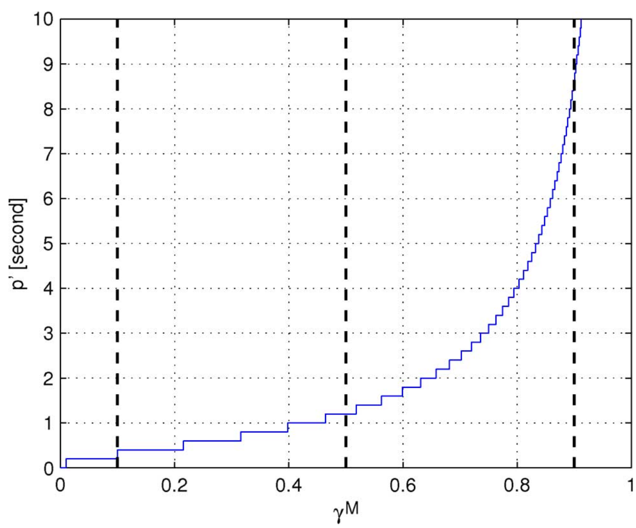


Fig. 7. Relationship between γ^M and p' ($\Delta t = 0.2, \theta = 0.01$).

A. Interactions Among Peers With Different Number of State Descriptions

In this section, we evaluate the impact of the number of state descriptions and the corresponding average CEDRs, which is studied in Sections III and IV. In our simulations, we consider an illustrative interaction scenario, where a peer forms a group with three other associated peers. To study the impact of the discount factor γ on the CEDRs, we compare among three different discount factors $\gamma = 0.1, 0.5, 0.9$. The solution to the MDP is implemented based on a well-known value iteration method [14]. The stopping criteria of this method is assumed to be 0.1 (i.e., $\epsilon = 0.1$), which affects the computation time. The simulation results are shown in Fig. 6. In Fig. 6, the elapsed time is actually measured using a MATLAB simulation and the theoretical upper bounds are computed considering the required complexity and the maximum number of required iterations, which are discussed in Section IV. In these results, we assume that the state transition probability functions are given, i.e., these results

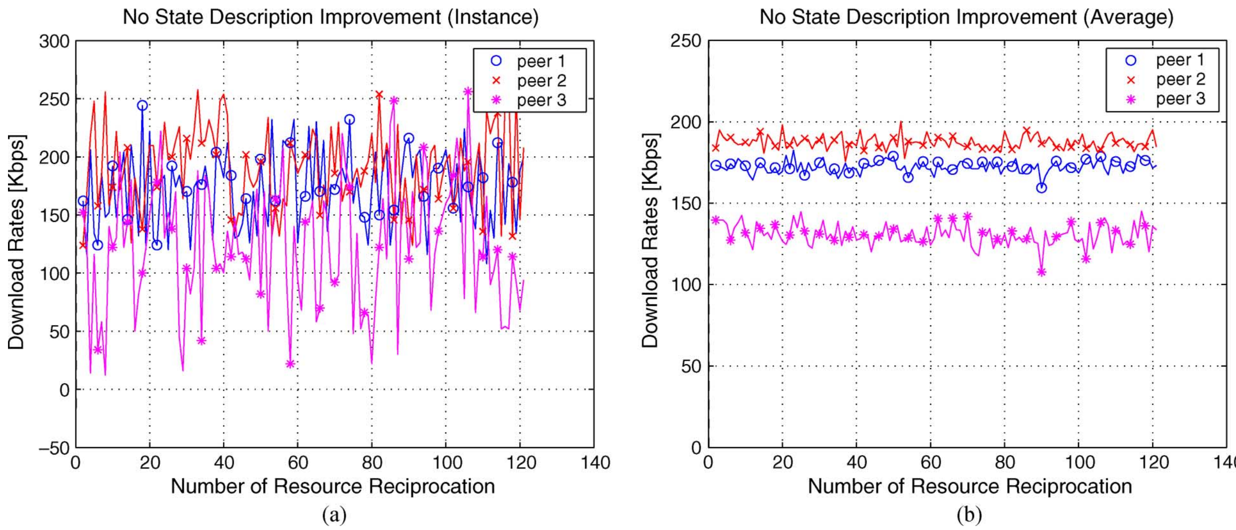


Fig. 8. Download rates of peers when no state description is refined. (a) Instance. (b) Average.

TABLE I
AVERAGE CEDRS FOR DIFFERENT VALUES OF γ AND DIFFERENT NUMBER OF STATE DESCRIPTIONS

γ		Number of State Descriptions (2^{B+b})		
		2	4	8
0.1	CEDRs [Kbps]	212.7	361.6	427.9
	Average CEDRs [Kbps]	211.9	272.6	83.6
0.5	CEDRs [Kbps]	459.7	612.0	733.9
	Average CEDRs [Kbps]	457.6	411.5	83.6
0.9	CEDRs [Kbps]	2461.0	3363.8	3660.5
	Average CEDRs [Kbps]	2420.9	633.0	83.6

are generated with no consideration of the time required to estimate the state transition probability functions for each of state descriptions. Because increasing the number of state descriptions enlarges the state space, and deploying a larger discount factor requires more consideration of the future impact of current actions on the rewards, it is obvious that both of them require more time to compute the optimal policy.

The average CEDRs corresponding to different values of discount factors and different numbers of state descriptions in a regular policy update period are shown in Table I. As discussed in Section IV-A, we assume that the peer can increase or decrease its state descriptions by the power of 2 in these results. We assume that the regular policy update period $p = 25$ s. The CEDRs of the previous period is 83.6 kb/s and the average CEDRs are computed based on (7) given period p.

Note that the CEDRs for different discount factors should not be compared against each other, since different discount factors reflect different P2P networks' dynamics, as discussed in Section III-B and the Appendix. However, as shown in Table I, it is clearly observed that a higher value of discount factor can lead to higher CEDRs given a number of state descriptions, as it can consider a longer impact of current actions on the CEDRs. This can illustrate the case of a stationary regime of P2P networks, where each peer's state transition probability does not change for a longer time. For example, the relationship between the maximum discount factor γ^M and a limited time for computing the optimal policy p', which is discussed in the Appendix, is shown in Fig. 7 given parameters of $\Delta t = 0.2$ s (required time for a resource reciprocation) and a threshold $\theta = 0.01$. Hence, the values of $\gamma = 0.1, 0.5, 0.9$ in Table I can be used for P2P systems if the state transition probability holds for 0.2, 1.2, and 8.6 s (i.e., $p' = 0.2, 1.2, 8.6$). Note that p' is determined based on how fast the system dynamics change.

The optimal number of state descriptions can also be determined for this illustrative example. Although CEDRs generally improve as the number of state descriptions increases, computing optimal strategy based on the MDP framework can exceed the time limit (e.g., period p or p'). Based on the average CEDRs, the peer can determine the number of its state descriptions in this period as 4, 2, and 2.

In summary, we can conclude that if the networks are stationary, more refined state descriptions or a higher value of discount factor can be used, which leads to higher average CEDRs. Inversely, a coarsely refined state descriptions or a lower value of the discount factor needs to be used if the networks are highly dynamic.

B. Evolution of Resource Reciprocation in a Period

In this section, we investigate and study the evolution of resource reciprocation among peers that have multiple optimal actions. To highlight the impact of the evolution of resource reciprocation, we assume that each peer currently adopts the optimal policy and the initial resource reciprocation is determined by the policy. In the following simulations, we consider a total of 1000 peers that are reciprocating resources. While reciprocating resources, each peer counts the downloaded data measured in bytes from its associated peers and estimates the download rates every 10 s. The groups of peers can be formed based on the chunk availability in the network. Direct and indirect impacts on resource reciprocation among the group members are captured in the process of estimating the state transition probabilities. To simplify and highlight the presentation of the main observations—the evolution of resource reciprocation—we focused only on showcasing the results of one group of peers (i.e., peer 1, peer 2, and peer 3) out of the total 1000 peers.

1) No State Description Improvement: If peers cannot improve their state descriptions during a period, then they simply select one available optimal action out of their set of optimal

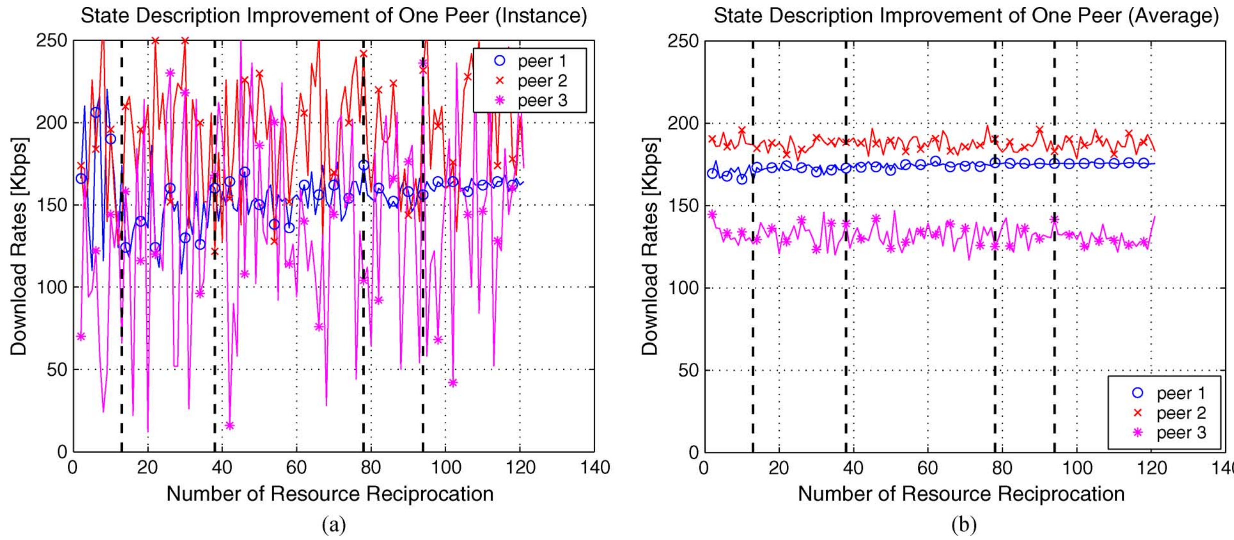


Fig. 9. Download rates of peers when a peer (peer 1) is refining its states. (a) Instance. (b) Average.

actions. In the simulations, an action is randomly chosen. The simulation results are shown in Fig. 8.

Fig. 8(a) and (b) shows each peer's instant or averaged download rates (over 100 independent simulations) over time determined by the resource reciprocations, respectively. Due to each peer's random action selections at each resource reciprocation, fluctuations in download rates are observed in Fig. 8(a). However, Fig. 8(b) shows nearly constant average download rates over time, as the optimal multiple actions of each peer are determined such that the other peers do not change their current states. Hence, we can conclude that if there is no state description improvement, there is no evolution of resource reciprocation among peers.

2) One Peer's State Description Improvement: Unlike the above experiment, we investigate the impact of one peer's state description improvement. In the simulations, we assume that only one peer improves its state descriptions over time, while interacting with its associated peers that keep the same number of their current state descriptions. The simulation results are shown in Fig. 9.

Fig. 9(a) and (b) shows each peer's instant or average download rates over time determined by the resource reciprocations, respectively. Without loss of generality, we assume that peer 1 refines its state description at its 13th, 38th, 78th, and 94th resource reciprocation. As shown in Fig. 9(a), we can observe that the variation of download rates for peer 1 decreases as it increases the number of its state descriptions. Since increasing the number of state descriptions of peer 1 also increases the number of states, peer 1 can differentiate smaller variations of resource reciprocation. As the associated peers select their actions to peer 1 such that they do not change peer 1's state, the variation of download rates of peer 1 can decrease. However, as shown in Fig. 9(b), note that increasing the number of state descriptions of one peer does not guarantee its download rate improvement, as it cannot enforce the associated peers to select actions that are beneficial to peer 1. Rather, the associated peers still randomly select their actions among the multiple optimal actions. By comparing the results from Fig. 8(b) and Fig. 9(b), the average download rates of peer 1 are the same when it does not increase the number of state descriptions.

3) Evolution of Resource Reciprocation Among Evolving Peers: Unlike the above two cases, where no peer or only one peer improves the number of state descriptions, we now consider the case where multiple peers consecutively refine their state descriptions. For illustration, in the considered group, peer 1 and peer 2 can refine their state descriptions, while peer 3 cannot. Without loss of generality, we assume that the peers simultaneously refine their state description by doubling the number of state descriptions for each state (i.e., using 1 bit more for state descriptions) at their 13th, 38th, 78th, and 94th resource reciprocation. Fig. 10 shows the resulting download rates among the peers in the group.

Fig. 10 clearly shows that two peers (i.e., peer 1 and peer 2) can improve their download rates by simultaneously increasing the number of state descriptions, while peer 3 cannot. In Fig. 10(a), we can verify that increasing the number of both peers' state descriptions can improve their download rates as well as decrease the variation of download rates, as already discussed in Proposition 5 and Section VI-B2. Fig. 10(b) shows that the average download rates of peer 1 and peer 2 improve as refined state descriptions are used. With a similar argument to Lemma 2, the additional download rate improvement is exponentially decreasing. For example, in Fig. 10(b), the average download rate improvement for each additional state-description bit of peer 1 is 48.6%, 27.1%, and 11.4% respectively, which coincides with the conclusion of Lemma 2. We can also verify that the improvement by increasing the number of state descriptions in this resource reciprocation process is bounded, as this resource reciprocations are initiated by a fixed number of state descriptions that is determined by the MDP. Therefore, we can conclude that if peers simultaneously increase their number of state descriptions, they can improve their download rates. If only two peers can evolve their resource reciprocation strategies in a group as shown in this example, then they can directly use the maximum available state descriptions to maximize their download rates. However, the same conclusion does not apply when more than two peers evolve their resource reciprocation

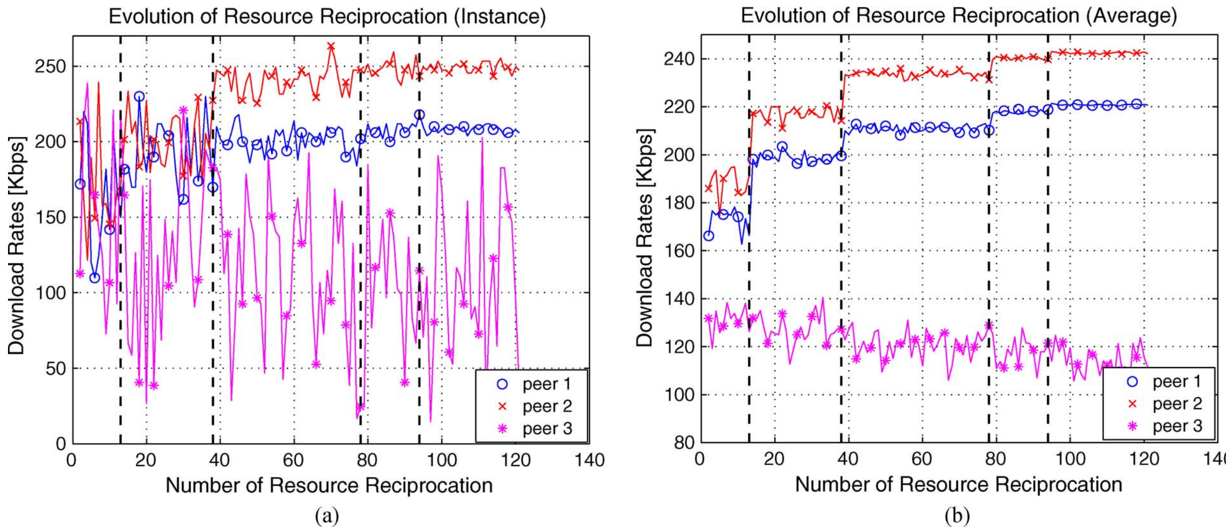


Fig. 10. Evolution of resource reciprocation. Peer 1 and peer 2 are refining their states. (a) Instance. (b) Average.

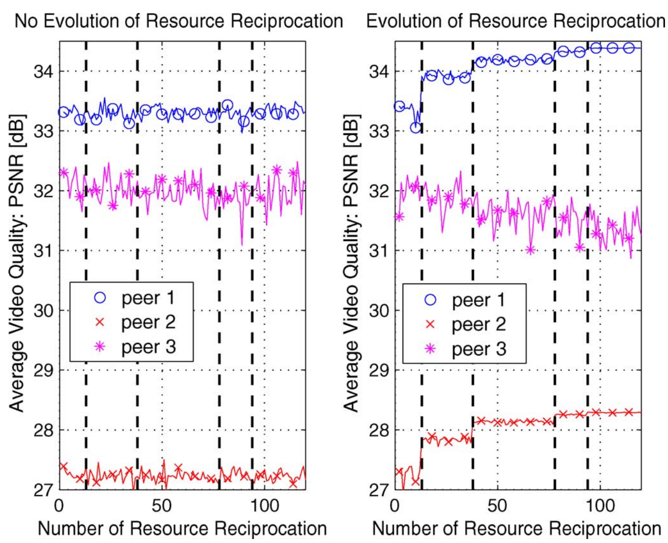


Fig. 11. Resulting average video qualities.

strategies, because they do not know how much their associated peers can refine their states. In this case, each peer needs to successively refine its state descriptions if its associated peers refine their state descriptions. Hence, this result can also be extended and applied to the case where more than two peers evolve.

The resulting average video qualities, measured in peak-signal-to-noise ratio (PSNR) for Y-component, for the above cases are shown in Fig. 11. In these experiments, the video files are at CIF (352×288) resolution, 30 frames/s, and encoded in a prioritized manner using the H.264/AVC encoder [19]. Specifically, the encoding structure for all sequences is "IPP $\cdots$ P," where each group of pictures (GOP) consists of 30 frames. We assume that packets for the I-frame have higher priority than packets for P-frames. For P-frame packets, higher priority is given to packets with shorter delay deadlines. The quantization parameters (QPs) for sequences are *Foreman* (I:28, P:32), *Coastguard* (I:29. P:38), and *Silent* (I:28, P:37) targeting bitrates at 250, 250, and 150 kb/s, respectively. The encoded video files are partitioned into uniform 20-kb chunks, and peers first download the chunks that have higher priority. Moreover, a single video file has 100-s duration, which was obtained by concatenating ten identical MPEG test sequences. Peers 1, 2, and 3 are downloading video sequences *Foreman*, *Coastguard*, and *Silent*, respectively. If there is no state description improvement among peers, or only one peer refines its state descriptions, then their download rates do not improve, as discussed in Sections VI-B1 and VI-B2 [see Figs. 8(b) and 9(b)]. Therefore, the resulting average video qualities do not improve. However, if several peers refine their state descriptions, and hence, they can improve their download rates through the evolution of resource reciprocation, then their download rates can improve [see Fig. 10(b)], thereby enabling the peers to achieve higher resulting average video qualities. If further quality improvement is required, several modules such as scheduling schemes, coding techniques, error resilient techniques, etc. (e.g., [16], [19]–[23]) can be incorporate into the proposed framework. These strategies for efficient multimedia transmission can be incorporated into the proposed framework.

C. Clustering Among Heterogeneous Peers

In this section, we study how heterogeneous peers are clustered, i.e., how they form their groups in the proposed MDP framework. For this, peers consider the expected download rates achieved by the CEDR as well as the evolution of resource reciprocation. Note that several clustering behaviors of peers in BitTorrent systems are discussed in [7].

Fig. 12 shows the distribution of probabilities that each peer forms a group with the other peers. To highlight the impact of peers' heterogeneity on clustering, in these simulations, we assume that peers are classified into five categories based on their ability to refine their state descriptions or based on their available upload bandwidth.

Fig. 12(a) shows the probability distributions for five categories of peers, where the peers in each category have the same ability to refine their state descriptions. We assume that peers

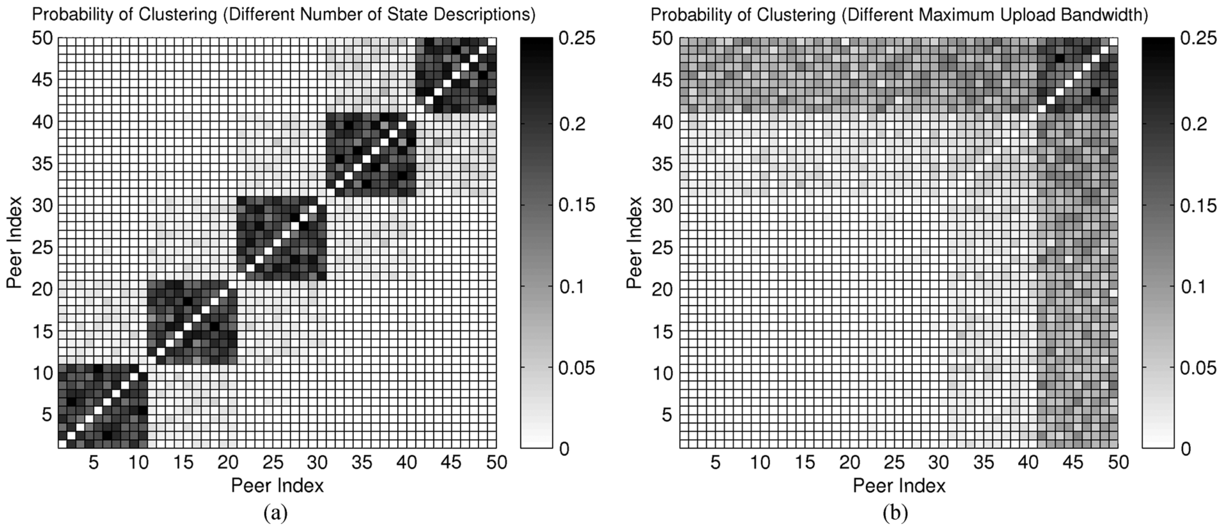


Fig. 12. Clustering among heterogeneous peers. (a) Peers with different ability to refine state descriptions. (b) Peers with different available upload bandwidth.

with peer index 1–10, 11–20, 21–30, 31–40, and 41–50 are in each category, categories 1–5, respectively. It can be observed that if peers are differentiated only by their state description refinement, peers with similar abilities form a group. This is because peers that can increase the number of state descriptions can improve their download rates only by reciprocating their resources with the other peers that can also increase the number of state descriptions, as discussed in Proposition 5. Note that, however, peers that cannot increase the number of state descriptions can be penalized if they are reciprocating resources with the peers refining their state descriptions. Fig. 12(b) shows the probability distribution that peers form their groups if their maximum available bandwidths are different, while their abilities to refine their state descriptions are the same. In this simulation, we assume that peers in each category have the same maximum available bandwidth, and the available bandwidth for peers in category i is higher than category j for $i > j$, $i, j \in \{1, 2, 3, 4, 5\}$. Unlike the results shown in Fig. 12(a), peers would like to associate with peers that have higher bandwidth. This is because the peers with higher available upload bandwidth may provide higher CEDRs.

From the above results, we can conclude that peers prefer to make groups with peers who have the same ability to refine their state descriptions, and who have a higher available bandwidth.

VII. Conclusion

In this paper, we study the interactions among multiple self-interested peers by sharing their content and resources in P2P networks. Since the resource reciprocation based on the MDP framework is largely affected by the number of each peer's state descriptions, we analytically investigate the impact of peers' heterogeneous ability to refine their state descriptions on their interactions of resource reciprocation and group formation. We show that more refined state descriptions enable peers to achieve higher download rates, while requiring more time to compute the optimal policy. Therefore, we conclude that both the number of state descriptions and the discount factor need to be determined by considering these tradeoffs. Moreover, we also investigate the impact of the number of state descriptions on the evolution of resource reciprocation, and show that peers can achieve higher download rates only if they simultaneously improve their state descriptions. Finally, group formation for heterogeneous peers is discussed, showing that peers prefer to form groups with other peers that have a similar ability to refine state descriptions, but also have higher upload bandwidths.

Appendix

The discount factor γ_i in the considered P2P network can represent the belief of peer i about the validity of the expected future rewards.

Let p' be a time constraint for computing optimal policy and Δt be a time required for one resource reciprocation. Then, the maximum number of resource reciprocation in p' can be computed as

$$h^M(p', \Delta t) = \left\lfloor \frac{p'}{\Delta t} \right\rfloor$$

since $h^M(p', \Delta t)\Delta t \le p'$. Therefore, given a threshold θ $(0 < \theta < 1)$ that indicates the effectiveness of the impact of current action on future rewards (i.e., future rewards discounted by $\gamma^n < \theta$ will be ignored in CEDRs), a set of possible values of discount factor γ can be expressed as $\Gamma = \{\gamma | \gamma^n \ge \theta, n \in \mathbb{N}\}$. Hence, the maximum value of discount factor γ^M can be parameterized by parameters of p', Δt, and θ, and can be found in set Γ, which is expressed as

$$\gamma^M(p', \Delta t, \theta) = \theta^{1/h^M(p', \Delta t)}.$$

Hence, given Δt and θ, it is observed that $\gamma^M(p', \Delta t, \theta)$ increases (or decreases) as p' increases (or decreases) because $h^M(p', \Delta t)$ is a nondecreasing function of p'. Note that time constraint p' can be determined based on dynamics introduced by peers joining, leaving, or switching groups in P2P systems [24]. For example, if a P2P system is so dynamic that the policy update process needs to be performed frequently, then $p' \ll$

p for a regular policy update period p (e.g., transient regime [25]). However, if the system is in, e.g., a stationary regime [25], then the policy update process can be performed at each regular policy update period, i.e., $p' \approx p$. Therefore, the discount factor can represent P2P networks' dynamics.

REFERENCES

[1] Napster [Online]. Available: http://www.napster.com

[2] KaZaA [Online]. Available: http://www.kazaa.com

[3] B. Cohen, "Incentives build robustness in BitTorrent," in *Proc. P2P Econom. Workshop*, Berkeley, CA, 2003.

[4] X. Zhang, J. Liu, B. Li, and T. S. P. Yum, "Cool streaming/DONet: A data-driven overlay network for efficient live media streaming," in *Proc. IEEE Conf. Comput. Commun. Soc.*, Mar. 2005, vol. 3, pp. 2102–2111.

[5] PPLive [Online]. Available: http://www.pplive.com

[6] S. Androutsellis-Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies," *ACM Comput. Surv.*, vol. 36, no. 4, pp. 335–371, Dec. 2004.

[7] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and sharing incentives in BitTorrent systems," *SIGMETRICS Perform. Eval. Rev.*, vol. 35, no. 1, pp. 301–312, 2007.

[8] V. Pai, K. Kumar, K. Tamilmani, V. Sambamurthy, and A. E. Mohr, "Chainsaw: Eliminating trees from overlay multicast," in *Proc. Int. Workshop Peer-to-Peer Syst.*, Feb. 2005, pp. 127–140.

[9] D. Kostic, A. Rodriguez, J. Albrecht, and A. Vahdat, "Bullet: High bandwidth data dissemination using an overlay mesh," in *Proc. ACM SIGOPS Operat. Syst. Rev.*, Dec. 2003, vol. 37, no. 5, pp. 282–297.

[10] C. E. Luna, L. P. Kondi, and A. K. Katsaggelos, "Maximizing user utility in video streaming applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 2, pp. 141–148, Feb. 2003.

[11] M. Feldman, K. Lai, I. Stoica, and J. Chuang, "Robust incentive techniques for peer-to-peer networks," in *Proc. ACM Conf. Electron. Commerce*, 2004, pp. 102–111.

[12] K. Lai, M. Feldman, I. Stoica, and J. Chuang, "Incentives for cooperation in peer-to-peer networks," in *Proc. Workshop Econom. Peer-to-Peer Syst.*, 2003.

[13] H. Park and M. van der Schaar, "A framework for foresighted resource reciprocation in P2P networks," *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 101–116, Jan. 2009.

[14] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. New York: Academic, 1976.

[15] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do incentives build robustness in BitTorrent?," in *Proc. USENIX Symp. Netw. Syst. Design Implement.*, Apr. 2007.

[16] K. Jain, L. Lovász, and P. A. Chou, "Building scalable and robust peer-to-peer overlay networks for broadcasting using network coding," *Distrib. Comput.*, vol. 19, no. 4, pp. 301–311, 2007.

[17] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the complexity of solving Markov decision problems," in *Proc. Conf. Uncertainty Artif. Intell.*, May 1995, pp. 394–402.

[18] H. Park and M. van der Schaar, "On the impact of bounded rationality in peer-to-peer networks," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 675–678, Aug. 2009.

[19] M. van der Schaar and P. A. Chou, Eds., *Multimedia Over IP and Wireless Networks*. New York: Academic, 2007.

[20] T. Stockhammer, L. Gunther, and M. Walter, "Optimized H.264/AVC-based bit stream switching for mobile video streaming," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–19, Jan. 2009.

[21] E. Setton and B. Girod, "Congestion-distortion optimized scheduling of video," in *Proc. Multimedia Signal Process. Workshop*, Oct. 2004, pp. 99–102.

[22] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[23] F. Boulos, W. Chen, B. Parrein, and P. L. Callet, "A new H.264/AVC error resilience model based on regions of interest," in *Proc. Int. Packet Video Workshop*, May 2009.

[24] Q. Zhang, H.-F. Xue, and X. D. Kou, "An evolutionary game model of resources-sharing mechanism in P2P networks," in *Proc. Workshop Intell.Inf. Technol. Appl.*, Dec. 2007, pp. 282–285.

[25] G. de Veciana and X. Yang, "Fairness, incentives and performance in peer-to-peer networks," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2003.

Hyunggon Park (S'08–M'09) received the B.S. degree (*magna cum laude*) in electronics and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles (UCLA), in 2006 and 2008, respectively.

Currently, he is an Assistant Professor at the Department of Electronics Engineering, Ewha Womans University, Seoul, Korea. His research interests are game theoretic approaches for distributed resource management (resource reciprocation and resource allocation) strategies for multiuser systems and multiuser transmission over wireless/wired/peer-to-peer (P2P) networks. In 2008, he was an intern at IBM T. J. Watson Research Center, Hawthorne, NY, and he was a Senior Researcher at the Signal Processing Laboratory (LTS4), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2009–2010.

Dr. Park was a recipient of the Graduate Study Abroad Scholarship from the Korea Science and Engineering Foundation during 2004–2006 and a recipient of the Electrical Engineering Department Fellowship at UCLA in 2008.

Mihaela van der Schaar (SM'00–F'10) received the M.S. and Ph.D. degrees in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1996 and 2001, respectively.

Currently, she is an Associate Professor at the Electrical Engineering Department, University of California, Los Angeles (UCLA). Her research interests are in multimedia communications, networking, processing, and systems.

Prof. van der Schaar received the National Science Foundation (NSF) Career Award in 2004; the Best Paper Award from the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2005; the Okawa Foundation Award in 2006; the IBM Faculty Award in 2005, 2007, and 2008; and the Most Cited Paper Award from *EURASIP: Image Communications* journal in 2006. She was an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the SIGNAL PROCESSING MAGAZINE, etc. She holds 30 granted U.S. patents and three ISO awards for her contributions to the MPEG video compression and streaming international standardization activities.