Analysis of user-driven peer selection in peer-to-peer backup and storage systems

Laszlo Toka and Pietro Michiardi

ABSTRACT

In this paper we present a realistic model of peer-to-peer backup and storage systems in which users have the ability to selfishly select remote peers they want to exchange data with. In our work, peer characteristics (e.g., on-line availability, dedicated bandwidth) play an important role and are reflected in the model through a single parameter, termed profile. We show that selecting remote peers selfishly, based on their profiles, creates incentives for users to improve their contribution to the system. Our work is based on an extension to the Matching Theory that allows us to formulate a novel game, termed the stable exchange game, in which we shift the algorithmic nature of matching problems to a game theoretic framework. We propose a polynomial-time algorithm to compute (optimal) stable exchanges between peers and show, using an evolutionary game theoretic framework, that even semi-random peer selection strategies, that are easily implementable in practice, can be effective in providing incentives to users in order to improve their profiles.

Keywords

peer-to-peer system, backup, storage, peer selection, game theory, stable matching, user model, incentives

1. INTRODUCTION

Nowadays, the need for safe on-line storage and backup services with availability and reliability guarantees is a relevant issue. Since centralized client-server solutions are not scalable nor robust, alternatives based on distributed data structures have recently appeared, offering on-line storage as a web service (e.g., Amazon S3 [18]). Despite being often built on commodity hardware, on-line storage systems do not come for free because of the large amount of resources service providers need to dedicate and maintain. For example, for relatively small amount of storage space users pay roughly \$1/year/GB, but an excessive storage demand is "punished" by a yearly fee of roughly \$40 for 20 GB data at [18]; moreover bandwidth issues, as well as user requests for accessing data come also into the pricing picture. Alternative storage providers offer unlimited storage capacity for 60 \$ per year (e.g., AllMyData [1] and Mozy [16]).

Solutions based on distributed data structure generally do not exploit resources (storage and bandwidth) available at users, although very recently Amazon S3 adopted the BitTorrent protocol [5] to amortize on bandwidth costs for popular data stored in their system. Similarly to efficient content distribution, a peer-to-peer (p2p) approach seems to be a suitable scheme for on-line backup and storage services. In such a system, users are expected to cooperate, that is they are compelled to share their private resources (storage and bandwidth) with other participants to make the concept work. A notable example of a p2p backup and storage solution is Wuala [21]. In contrast to p2p file sharing systems where tit-for-tat based cooperation is temporary and lasts only for the data transfer time, in a p2p backup and storage system user cooperation must be *long-term* and dedicated to well-defined partners. Hence, more elaborate incentives schemes are required.

Although several works have defined subtle economic frameworks to design and analyze incentive schemes to enforce user cooperation (e.g., [2] and references therein), none of them have addressed the question of whether it would be possible to design a p2p backup and storage system with built-in incentives without requiring additional mechanisms. One of the main reasons is that existing p2p backup and storage systems do not constrain the interaction among peers: in systems such as AllMyData and Wuala peers exchange data with a randomly chosen neighbor set, composed of remote peers participating in the p2p system. An additional mechanism to degrade or eventually deny service to misbehaving peers is required.

In this work we make the case for a p2p system in which users can selfishly create their neighborhood. We build a model of a p2p backup and storage system in which users are described by a *profile*, that aggregates information such as on-line availability, bandwidth capacity (accessibility), behavior, etc. Our model is used to formulate a selfish optimization framework (in game theoretic terms, a *game*) in which peers can select the amount of data they wish to store in the system, and the remote peers they wish to exchange data with (termed peer selection). The novelty in our approach is that it allows users to selfishly determine their profile: e.g., availability and accessibility become optimization variables, all compacted in a user profile. Profiles are coupled with peer selection and we show that this is sufficient for providing incentives to users to improve their profile.

Due to the complexity of the joint optimization problem we originally formulate, in this paper we focus on the impact of peer selection on user profiles by extending the theory of Stable Matching. We define a novel game, termed the Stable Exchange Game, and propose a framework based on evolutionary game theory to analyze optimal and heuristic peer selection strategies and show that even semi-random choices, which are simple to implement in a real system, compel users to improve their profiles if they wish to obtain a better quality of service.

The remainder of the paper is organized as follows. In Section 2 we briefly discuss related works, while in Section 3 we formally introduce our system model. Section 4 focuses on the analysis of peer selection strategies; in Section 5 we show how prior results on matching theory can be extended to account for the requirements of our model. Section 6 introduces a framework based on evolutionary game theory through which we analyze the impact of several heuristic peer selection strategies on user profiles. Finally, we conclude the paper in Section 7 and discuss on our future work.

2. RELATED WORK

A number of related works focus on economic modeling of backup and storage systems and focus in particular on incentive mechanisms: we point the reader to [20] and its references for an overview of such works. In [20] selfish user behavior is described using non-cooperative game theory: users are modeled by their strategies on their demand (in terms of amount of data to backup) and their offer (offered resources such as storage space, available bandwidth and up-time). The user payoff function defined in [20] is linear and split in two parts: the first term represents users' willingness to participate to the backup service as a function of the amount of data they need to backup, the second term accounts for the costs for a peer to offer local resources to remote peers.

An example of a commercial p2p storage application is Wuala [21]: the system relies on symmetric exchanges of data between users. Each user has the right to backup the amount of data in the system that she offers locally, discounted by her availability, which must be kept above a certain threshold. Data persistence and integrity is periodically checked by the data owner, and important peer parameters (offered/used storage space, availability, bandwidth and malicious behavior) are maintained using a distributed hash table.

In [9] the authors present the performance evaluation of different peer selection strategies in presence of churn: they present a stochastic model of a p2p system and argue on the positive effects of randomization. Peer selection strategies in [9] are designed to mitigate the impact of churn while in our work peer selection itself provides incentives to users to increase their on-line time.

Papers on network formation (e.g., [7, 14]) discuss strategic peer selection to find equilibrium networks in which users selfishly minimize a cost function that accounts for end-to-end connectivity. Our model is related to these works in that users build their neighborhoods based on their local preferences. However, in our system end-to-end connectivity is not necessary. Moreover, we assume the creation of a link between peers to be bilateral, as discussed in [6]. Matching theory, a field of combinatorial optimization, provides useful tools to analyze peer selection in our setting: we discuss in details related works in Section 5.

Data management is a crucial issue in backup and storage systems. A vast literature exists tackling data redundancy [17], resilience to peer churn, [4] and reputation systems [10]. These problems are not addressed in our work. We also gloss over the problem of deciding which is the optimal amount of data to be stored in the system [20].

3. SYSTEM MODEL

In this section we define a general model of a p2p backup and storage system in which we assume *symmetric* exchange of data between peers. The model presented in this paper relies on the existence of a double-overlay structure. The first overlay is a distributed hash table that maintains information on the characteristics and behavior of all peers taking part to the system, as done in the Wuala application [21]. We combine users' features and behavior into a single parameter, that we termed *profile*. The second overlay is built by the users themselves through *peer selection*: every peer decides which remote peer to select and exchange data with.

We begin by defining the degrees of freedom of the system: these are the variables a given peer is allowed to locally optimize. We then present the utility function that characterizes a peer: this allows to define a non-cooperative game that we will dissert throughout the paper[1].

3.1 User profile

DEFINITION 1. *Let $\mathcal{I}$ denote the set of participants in the system. We introduce the parameter vector $\alpha = (\alpha_i)$ for $\forall i \in \mathcal{I}$, which we will refer to as peers' profiles.*

Hence, peers' characteristics are combined in one scalar, $\alpha_i \in [0, 1]$. α_i accounts for peer i's: data possession behavior (i.e., liability in storing data), availability (probability to be found on-line), and accessibility (available bandwidth). We assume α_i to be computed, maintained and advertised through a dedicated DHT overlay. We note that the definition of a method to compute users' profiles calls for realistic measurements on peers' behavior: we will focus on this issue in our future work, while in this paper we gloss over the details of how profiles are computed.

In this work we make the case for users to *control their profiles*: users' behavior, as well as their (economic) efforts directed to improve their availability and accessibility are considered optimization variables that can be adjusted by a peer when participating to the p2p backup and storage system.

In the next section we discuss on data exchange strategies between peers: these rules are necessary to ensure data availability despite peer churn. We suggest to use peers' profiles α as an important ingredient to drive peer exchanges.

3.2 Backup data exchange

DEFINITION 2. *We denote by c the set of c_i for $\forall i \in \mathcal{I}$, c_i being the amount of data user i needs to backup or store in the system.*

Most of the existing works on p2p backup and storage systems consider a specific exchange rule in order to address

[1] In this paper the terms *user*, *peer*, *player* and *participant* are synonyms.

the data availability issue and to enforce symmetric collaboration between peers. For example, [21] imposes on peer i, with c_i data units stored in the system, the duty of storing $\frac{c_i}{p_i}$ data units for others, where p_i is peer i's availability, $p_i \leq 1$. This means that a peer can store only discount$(o) < o$ in the system when she offers o storage capacity to others. The redundancy factor introduced through the discount function is achieved using techniques such as erasure coding and replication, and guarantees a permanent backup service despite peer churn.

In our model, discounting is achieved using peers' profiles α: if peer i offers $\frac{c_i}{\alpha_i}$ storage space she can backup c_i data units in the system (note that $0 \leq \alpha_i \leq 1$). We further note that a fair barter-based p2p system should lie on *symmetric* exchanges of backup space between peers. In our model, the *profile-dependent* redundancy factor is used to reach symmetric exchanges.

Our general model considers peers being able to locally decide (and optimize) the amount of data c_i they will store in the p2p system.

3.3 Peer selection

Along with the two optimization variables α, c discussed in the previous sections, our model explicitly accounts for the strategic selection of remote connections a peer establishes. This process, termed *peer selection*, defines neighborhood relations among peers: the union of all peers' neighborhoods defines an overlay network through which peers exchange data.

DEFINITION 3. *We define the set $n = \{n_i\}$ $\forall i \in \mathcal{I}$, where n_i is a $|n_i|$-tuple describing user i's neighborhood, and $|n_i|$ is the number of user i's neighbors. Each element of n_i has the form (c_{ij}, α_j), where c_{ij} is the discounted amount of data exchanged between user i and user j. Obviously, $c_{ij} > 0$ holds for $\forall j \in n_i$; moreover $c_{ij} = c_{ji} \in n_j$ for $\forall i, j$; moreover, $c_i \geq \sum_{j \neq i} c_{ij}$ must hold.*

Note that although the definition of peer selection we give in this section resembles to what has been previously studied in network creation games [7], the utility function we next define implies that peers are not interested in end-to-end connectivity.

3.4 User utility function

The utility function is the key component of our model. Every peer in the system is assumed to selfishly optimize the utility function by appropriately selecting their strategy which is a combination of three elements: their profiles, the amount of data they want to store in the system and the remote connections they establish. The utility function we define in this section accounts for peers' availability, accessibility and behavior through the peers' profiles.

DEFINITION 4. *For $\forall i \in \mathcal{I}$, player i's payoff $P_i(\alpha_i, c_i, n_i)$ can be described by the following form:*

$$P_i(\alpha_i, c_i, n_i) = U_i(c_i) - D_i(c_i, n_i) \\ - O_i(\alpha_i, c_i) - T_i(\alpha_i, c_i, n_i) - E_i(\alpha_i, \hat{\alpha_i}),$$

where

- $U_i(c_i)$ *stands for user i's* benefit *which is assumed to be positive, continuously differentiable, increasing and quasi-concave in its argument;*
- $D_i(c_i, n_i)$ *indicates the* service degradation *due to non-optimal neighbors. It takes peer i's neighbors' α_js, weighted by the amount of data c_{ij} stored at each remote peer as inputs. It is decreasing in the remote peers' profiles: connecting to a remote peer with a higher α_j value implies less degradation. If the cardinality $|n_i|$ of a peer's neighborhood drops drastically, the service degradation increases;*
- $O_i(\alpha_i, c_i)$ *is the* opportunity cost *of offering private resources (i.e., storage). This is a user specific function of user i's c_i and α_i, since it is assumed to be positive, continuously differentiable, increasing and convex in the offered storage space, which is given by α_i and c_i as discussed in Subsection 3.2;*
- $T_i(\alpha_i, c_i, n_i)$ *represents the* transfer cost *related to the service, and it is a user specific function of user i's c_i, α_i and weighted profile set of her neighborhood. T_i is decreasing in α_i, increasing in c_i and shows similar characteristics to D_i on the neighbor set n_i;*
- $E_i(\alpha_i, \hat{\alpha_i})$ *describes the* effort cost *that peer i has to bear when improving her initial effortless profile $\hat{\alpha_i}$ to α_i. E_i is assumed to be a positive increasing convex function of $\alpha_i > \hat{\alpha_i}$.*

Authors in [20] model the utility as a function of the amount of data stored at remote peers, while the uploading/retrieving process is assumed to be ideal. In this work we argue that the "quality of service" of a backup system should appear in the user's payoff. As a concrete example, a peer may offer a large amount of storage space to remote peers, but the value to other peers should be weighted by her up-link capacity: 1 TeraByte of data is worth little if the up-link capacity is only a few bytes per second.

3.5 Formal game description

We now formally define the dynamic, non-cooperative game that can be built around the system model we discussed in the previous sections:

- $\mathcal{I}$ denotes the player set ($|\mathcal{I}|$ is the number of players);
- $\mathcal{S}$ depicts the collection of strategy sets ($\mathcal{S} = (\mathcal{S}_i)$ for $\forall i \in \mathcal{I}$), $\mathcal{S}_i$ being the combination of the three different strategy sets: $\alpha_i \in \mathbb{R}_{[0,1]}$, $c_i \in \mathbb{R}^+$, $n_i \subseteq \mathcal{N}_i = \{\{i, j, c_{ij}\} : j \in \{\mathcal{I} \setminus i\}, c_{ij} \in \mathbb{R}_{[0, \min(c_i, c_j)]}\}$;
- $\mathcal{P}$ function gives the player payoffs ($P = (P_i)$ for $\forall i \in \mathcal{I}$) on the combination of strategy sets ($\mathcal{P} : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \rightarrow \mathbb{R}^{|\mathcal{I}|}$).

The user strategies and their effects on the payoff function lead to the maximization problem a selfish user faces in the system: user i always maximizes her payoff P_i on her three strategy variables (i.e., her on-line characteristics described by α_i, her backup c_i and her strategy regarding peer selection n_i), all of them having effects on her payoff function as the previous section presented, but not independently, e.g., a user may decide to make costly efforts to increase her α_i in order to have a better neighborhood (in terms of neighbors' profiles).

The optimal user strategy tuple $s_i^* = (\alpha_i^*, c_i^*, n_i^*) \in \mathcal{S}_i$ is defined by solving the equation $\arg_i(\max(P_i))$ with the constraint that $n = (n_i)$ for $\forall i \in I$ must ensure pairwise and symmetric exchanges. In (Nash) equilibrium $P_i(s_i^*, s_{-i}^*) \geq P_i(s_i', s_{-i}^*)$ for any player i and for any alternative strategy tuple $s_i' \neq s_i^*$, where $s_{-i}^* = (\alpha_{-i}, c_{-i}, n_{-i})^*$ depicts the composition of equilibrium strategy tuples of players other than i.

In summary, the game defined in this section is a joint, distributed optimization problem that turns out to be very difficult to analyze. In Section 4 we restrict our attention to *peer selection* and derive some simplifying assumptions to improve the tractability of the problem.

4. MODELING PEER SELECTION

To the best of our knowledge, little work has been done in studying the strategic selection of remote peers to exchange data with in a p2p backup and storage system. In this section we simplify the generic model defined in Section 3 and focus on strategic peer selection. We leverage on the literature of p2p content sharing (see for example [15]) and cast the peer selection as a *stable matching problem*. However, we improve on previous models by allowing matchings to be the results of a dynamic game in which peers can *both* select remote connections based on some preference ordering *and* can operate on their profiles α_i to modify their rankings.

In the following section we primarily focus on determining the existence of a dominant strategy for peer selection and compare it against what is typically implemented in current systems, *i.e.,* a random peer selection. Next, we define a simplified utility function derived from Definition 4 that induces the peer selection game.

4.1 Dominant strategies

Before delving into the analysis of peer selection strategies (n), we simplify the game previously discussed by constraining the degrees of liberty of the system: we simply "downgrade" two strategic variables to play the role of parameters in our simplified system:

ASSUMPTION 1. *We assume that for $\forall i \in \mathcal{I}$ user i's strategies c_i, α_i are fixed; moreover, peer i equally splits the c_i data units among the neighbor set, i.e., $c_{ij} = c_{ik}$ for $\forall j, k \in n_i$.*

DEFINITION 5.
- *Player i adopts a **selective strategy** if neighbors with high profile are preferred over neighbors with low profile;*
- *Player i adopts a **non-selective strategy** if neighbors are chosen at random.*

DEFINITION 6. *A dominant strategy of a game $\langle \mathcal{I}, \mathcal{S}, \mathcal{P} \rangle$, where $\mathcal{I}$ is the set of players, $\mathcal{N} = (\mathcal{N}_i)$ for $\forall i \in \mathcal{I}$ is the strategy set, and $\mathcal{P} = (\mathcal{P}_i)$ gives player preferences over the strategy set, is a strategy $n_i^* = s_i^* \in \mathcal{S}_i$ with the property that for player $i \in \mathcal{I}$ we have $\mathcal{P}(n_i^*, n_{-i}^*) \geq \mathcal{P}(n_i, n_{-i}^*)$ for $\forall n_i \in \mathcal{S}_i$ and for any counter strategy set $n_{-i}^* = s_{-i}^* \in \mathcal{S}_{-i}$.*

PROPOSITION 1. *The selective strategy is dominant for* every *player.*

PROOF. The proof's key is that cooperation is bilateral, i.e., needs consent from both parties. Under Assumption 1 a player's payoff depends only on her selected partners' profiles; moreover based on the payoff function, collaborating with player j is less beneficial to a given user i than cooperating with player k instead, such that $\alpha_k > \alpha_j$. Let us separate player i's partners based on their strategies: selective and non-selective partners. Then i's payoff is the function of the average profiles of her non-selective and selective partners, depicted in the form of $P_i(\alpha^{non-selective}, \alpha^{selective})$. Since cooperation is pairwise, a selective player will not collaborate with a worse player, if she can pick a better one. Thus, assuming large number of heterogeneous profile players, a selective (resp. non-selective) player's P_i is a function of $(\widetilde{\overline{\alpha_i}}, \alpha_i)$ (resp.$(\widetilde{\alpha}, \underline{\widetilde{\alpha_i}})$), where $\widetilde{\overline{\alpha_i}}$ is the average profile of non-selective players having at least α_i (resp. $\underline{\widetilde{\alpha_i}}$ stands for the average profile of selective partners worse than α_i). Since $P(\widetilde{\overline{\alpha_i}}, \alpha_i) > P(\widetilde{\alpha}, \alpha_i)$, and similarly $P(\widetilde{\alpha}, \alpha_i) > P(\widetilde{\alpha}, \underline{\widetilde{\alpha_i}})$, being selective is always the best strategy as it assures the highest payoff regardless the counter-strategy set. $\square$

4.2 Peer selection game: a simplified utility function

Proposition 1 indicates that selfishly selecting remote peers to connect to dominates a random strategy. We now define a formal setting to study the problem of the existence of stable matchings between peers that selfishly select *both* remote connections based on some preference ordering *and* that operate on their profiles α_i to modify their ranking. To improve the tractability of the problem we suppose peer homogeneity in the amount of data that needs to be stored and an initial (effortless) profile parameter α_i. The combinatorial problem that arises in our game is thoroughly discussed in Section 5.

ASSUMPTION 2. *We assume $c_i = C \in \mathbb{N}^+$ $\forall i$ (also exchanges are discrete) and that $\hat{\alpha}_i = 0$ for $\forall i \in \mathcal{I}$, where $\hat{\alpha}_i$ is the initial, effortless profile.*

Leveraging on Assumption 2 we can define the following simplified utility function, which is selfishly optimized by all peers participating to the system:

DEFINITION 7. *We assume that user i's payoff $\forall i \in \mathcal{I}$ is defined as follows: $P(\alpha_i, C, n_i) = U(C) - D(C, n_i) - O(\alpha_i, C) - T(\alpha_i, C, n_i) - E_i(\alpha_i, 0)$, where:*

- *the utility of service $U(C)$ and the opportunity cost $O(\alpha_i, C)$ are such that exchanging backup data brings positive gain with any α_j peer, for simplicity $U(C) = 2$ and $O = 0$ in the subsequent analysis;*
- *the degradation cost, which increases (convex) in the backup fraction exchanged with a particular peer but decreases in the latter's profile, has the following form: $D = \sum_{j \in n_i} (\frac{c_{ij}}{C})^{(1+\alpha_j)}(1 - \alpha_j)$;*
- *the transfer cost, which depends linearly on the backup fractions, is $T = (1 - \alpha_i) \sum_{j \in n_i} \frac{c_{ij}}{C}(1 - \alpha_j)$;*
- *and the effort cost is equal to $E_i = \alpha_i^2$, assuming $E = (\alpha_i - \hat{\alpha}_i)^2$.*

5. THE PEER SELECTION GAME

In this section we show that under the assumptions made in Section 4.2 we can anticipate optimal peer selection strategies and stable overlay graphs for any given α vector when users selfishly optimize their utility from participating to the system. Since we cast the problem as a stable matching problem, we first provide a brief introduction to matching theory and then focus on how to shift from the algorithmic domain that characterizes simple matching problems to a game theoretic framework.

At the end of the section we illustrate a heuristic algorithm to study the game described in Section 4.2.

5.1 Matching problems

The game presented in Section 3 and its simplified version discussed in Section 4.2 incorporates a matching problem on the strategy vector n: we are interested in stable outcomes of these games. Here we emphasize the complexity that the pairwise symmetric backup exchanges introduce to the system model. We define our problem starting from traditional matching problems. In each case, we assume complete preference lists and that if player i prefers one of her strategies to an other, it is because the *strict* preference order over the payoff P_i for the *given* best response strategy set yields so.

The first works on matching theory focused on bipartite, stable marriage problems [8]. However, in our setting there is no such distinction of genders (womanhood and manhood), hence the bipartite approach is not suitable. Single linking between players belonging to the same set was first introduced in the stable roommates problem.

5.1.1 Stable roommates problem

In a stable roommates (SR) problem player i's strategy is $n_i \in \mathcal{N}_i$, where $\mathcal{N}_i$ is the set $\{\{i,j\} : j \in \mathcal{I} \setminus \{i\}\}$, and $\mathcal{P}$ is assumed to give strict order on i's possible pairs, termed *preference list*. The formal definition of the SR problem is to find a matching $\mathcal{M}$ on the setting presented above, $\mathcal{M}$ being a set of $\frac{|\mathcal{I}|}{2}$ disjoint pairs of players, which is stable if there are no two players, each of whom prefers the other to his partner in $\mathcal{M}$. Such a pair is said to block $\mathcal{M}$. Following the statement of the SR problem by Gale and Shapley in [8], Irving's [12] presents a polynomial-time algorithm to determine whether a stable matching exists for a given SR instance, and if so to find one such matching.

For the case where a given player may be part of multiple pairs, the stable fixtures problem was introduced.

5.1.2 Stable fixtures problem

Irving and Scott present in [13] the stable fixtures (SF) problem, which is a generalization of the SR problem. Formally, the notion of *capacity* is introduced such that for each $i \in \mathcal{I}$ a positive integer c_i, which is player i's capacity, denotes the maximum number of matches, i.e., pairs (i, j) in which player i can appear. i's strategy is $n_i \subseteq \mathcal{N}_i = \{\{i,j\} : j \in \{\mathcal{I} \setminus i\}\}$ and $\mathcal{P}$ gives again the strictly ordered *preference list* on i's matches. It is straightforward to see that the SR problem is a special case of the SF problem when $c_i = 1\ \forall i \in \mathcal{I}$, i.e., each player may have 1 match at most. A matching $\mathcal{M}$ here is a set of acceptable pairs $\{i, j\}$ such that for $\forall i \in \mathcal{I}\ |\{j : \{i,j\} \in \mathcal{M}\}| \leq c_i$, where a pair $\{i, j\}$ is acceptable if i appears in n_j and j appears in n_i. $\mathcal{M}$ is stable if there is no blocking pair, i.e., an acceptable pair $\{i,j\} \notin \mathcal{M}$ such that

- either i has fewer matches than c_i or prefers j to at least one of her matches in $\mathcal{M}$; *and*
- either j has fewer matches than c_j or prefers i to at least one of her matches in $\mathcal{M}$.

[13] describes a linear-time algorithm that determines whether a stable matching exists, and if so, returns one such matching.

In this work, we define a more general problem by further extending the SF problem with the possibility of multiple matches between two given players. We call this problem the *stable exchange problem*. In [3] the authors arrive at a very similar extension of the SF problem through the definition of SR problem generalizations under the names of stable activities problem, where parallel edges in the underlying graph are allowed, and stable multiple activities problem, where multiple partners are allowed.

5.1.3 Stable exchange problem

In the stable exchange (SE) problem player i's strategy is $n_i \subseteq \mathcal{N}_i = \{\{i, j, c_{ij}\} : j \in \{\mathcal{I} \setminus i\}, 0 \leq c_{ij} \leq \min(c_i, c_j)\}$, where c_{ij} (resp. c_i) denotes player i's number of matches towards player j (resp. towards all the players). A matching $\mathcal{M}$ is a set of matches $\{i, j, c_{ij}\}$ such that $\{i, j, c_{ij}\} \in n_i$, $\{j, i, c_{ij}\} \in n_j$ for $\forall i, j \in \mathcal{I}$, and $\sum_{j:\{i,j,c_{ij}\}\in\mathcal{M}} c_{ij} \leq c_i$ holds for $\forall i \in \mathcal{I}$. To avoid inconsistency in the consequence order of consecutive matches between given players, we make the following assumption regarding the preference list:

ASSUMPTION 3. *$P_i(i, j, c') > P_i(i, j, c'')$ holds for any pair of matches between players i and j if $c' < c''$ for $\forall i, j$, where, by an abuse of notation, we denote players i and j's c'th* pairwise *match's payoff for i by $P_i(i, j, c')$.*

$\mathcal{M}$ is stable if, similarly to the SF problem, there is no blocking match, i.e., no match $\{i, j, c'\} \notin \mathcal{M}$, thus $c' > c_{ij}$ for $\forall i, j : (i, j, c_{ij}) \in \mathcal{M}$, such that

- either i has fewer matches than c_i or $P_i(i, j, c') > P_i(i, k, c_{ik}) \in \mathcal{M}$, such that $j \neq k$, i.e., $\{i, j, c'\}$ is more beneficial for i than at least one of her matches in $\mathcal{M}$; *and*
- either j has fewer matches than c_j or $P_j(j, i, c') > P_j(j, l, c_{jl}) \in \mathcal{M}$, such that $i \neq l$, i.e., $\{j, i, c'\}$ is more beneficial for j than at least one of her matches in $\mathcal{M}$;

In other words, in a stable matching no two players could have a new match between themselves which is preferred by both of them to any of their existing matches.

5.2 The capacity-uniform stable exchange problem

In this section we analyze a simplified instance of the stable exchange problem: we assume an homogeneous case in which all users store the same amount of data in the system while selfishly optimizing the simplified utility model (see Section 4). Let us suppose that the payoff function $\mathcal{P}$ (and thus the preference order on $\mathcal{N} = (\mathcal{N}_i)$ for $\forall i \in \mathcal{I}$) is defined based on the player parameter set α. The implications of α on $\mathcal{P}$ are compacted in the following proposition.

PROPOSITION 2. *In a capacity-uniform stable exchange problem determined by Assumption 2 and Definition 7,*

$$P_i(i, j, c') > P_i(i, k, c')$$

holds for a given $0 < c' \leq C$ for any given pair $j, k \in \{\mathcal{I} \setminus i\}$ if, and only if $\alpha_j > \alpha_k$ for $\forall i \in \mathcal{I}$. In the case $\alpha_j = \alpha_k$, $P_i(i, j, c') = P_i(i, k, c')$ for any $0 < c' \leq C$. Also, Assumption 3 holds as direct consequence of Definition 7.

PROOF. Statement comes directly from Assumption 2 and Definition 7. Player i's payoff of the cth match with player j is $P_i(i,j,c') = U_i - D_i - O_i - T_i - E_i$, where the different terms are given as follows:

- $U_i = \frac{2}{C}$;
- $D_i = \left((\frac{c'}{C})^{(1+\alpha_j)} - (\frac{c'-1}{C})^{(1+\alpha_j)} \right)(1-\alpha_j)$;
- $O_i = 0$;
- $T_i = (1-\alpha_i)(1-\alpha_j)\frac{1}{C}$;
- $E_i = \frac{\alpha_i^2}{C}$.

This gives straightforwardly the proposition. □

For the given capacity-uniform stable exchange problem, constructed by the assumptions and holding properties given in Proposition 2, we now prove that it is always possible to find the *optimal* stable matching $\mathcal{M}$.

PROPOSITION 3. *At least one stable matching exists for a given uniform backup exchange problem instance, and a slightly extended version of Irving's algorithm (presented in [13]) finds the optimal one in polynomial time.*

PROOF. The proposition comes directly from our extension of Irving's algorithm for SF problems to SE problems and from Proposition 2. By supposing deterministic behavior of our algorithm, the statement becomes straightforward. Note that non-determinism has no effect on the outcome, therefore let us suppose that players place their bids in the profile order. The best profile player i bids the first C matches on her preference list, and based on Proposition 2, all of them will be accepted and reciprocated, since if $P_{i,j,c_{ij}} > P_{i,k,c_{ik}}$ then $P_{j,i,c_{ji}} > P_{j,k,c_{jk}}$ with $c_{ji} \geq c_{ij}$ for $\forall i,j,k$ such that $\alpha_i > \alpha_j$ and $0 < c_{ij}, c_{ik}, c_{ji} \leq C$. In other words, this means that if a higher profile peer is interested in a match with a lower profile one, then the latter is interested also at least to the same extent. After the best profile peer has found her stable matches, all the bids of the other players targeting her are dropped, and the same reasoning stands for the best profiled player of the rest, and so forth. This deterministic sequence of the algorithm also assures that the optimal matching will be found, since there is no possible further pairwise match which yields higher payoff than the ones in $\mathcal{M}$. □

5.3 The capacity-uniform stable exchange game

We now shift from the basic algorithmic setting of matching problems to a game theoretic setting. Formally, let α be a *strategy* variable vector the players can decide on, which indirectly influences the payoff function $\mathcal{P}$: in this setting, supposing that Assumption 2 and Definition 7 hold, the uniform stable backup exchange problem becomes a game.

5.3.1 Game definition

In the capacity-uniform stable exchange game, using the Section 3's notations, the *joint* strategy s_i for player i consists of $\alpha_i \in [0,1]$ and an instance $n_i \subseteq \mathcal{N}_i = \{(i,j,c_{ij}) : j \in \{\mathcal{I} \setminus i\}, 0 \leq c_{ij} \leq C\}$. Player $i \in \mathcal{I}$ selfishly maximizes her payoff P_i, given by $\mathcal{P}$ on the α strategy vector *and* the peer-selection strategy vector n, i.e., $\mathcal{P} : \alpha \times \mathcal{N} \to \mathbb{R}^{|\mathcal{I}|}$.

5.3.2 Equilibrium

In Nash equilibrium, which must be a stable matching, the $P_i(\{\alpha_i^*, \alpha_{-i}^*\}, \{n_i^*, n_{-i}^*\}) \geq P_i(\{\alpha_i, \alpha_{-i}^*\}, \{n_i, n_{-i}^*\})$ holds for any α_i, n_i and for $\forall i \in \mathcal{I}$, where α_{-i}^* and s_{-i}^* depict the best response counter strategy sets.

The optimal player strategy tuple is

$$(\alpha_i^*, n_i^*) = \arg_i(\max(\mathcal{P}_i(\alpha, \mathcal{N})))$$

for $\forall i \in \mathcal{I}$ with the constraint that stable matching is symmetric in $n^* = (n_i^*)$ for $\forall i \in \mathcal{I}$, since every match is pairwise. The social welfare is given by $\max(\sum_{i \in \mathcal{I}} \mathcal{P}_i(\alpha, \mathcal{N}))$ also with the stable matching constraint.

We suspect that showing the existence of the pairwise Nash equilibrium of the game, as well as the joint optimization problem defined above, are NP-hard problem, but defer to an extended version of this work a formal proof.

5.4 An illustrative example

Based on Proposition 3, we propose a heuristic distributed algorithm which approximates the optimal equilibrium strategies in polynomial time. Players' decisions regarding their two strategic variables (i.e., α and n) are *interleaved* and carried out *iteratively*. Players' decisions on α follow a heuristic based on evolutionary game theory [11]: players switch to their partners' average strategy α if they experience lower payoff then their neighborhood's average. In each iteration we find the optimal matching on the actual profile vector based on a variation of the Irving algorithm. The pseudocode of the algorithm is depicted in Algorithm 1.

Algorithm 1 Iterative distributed dynamic uniform exchange algorithm

$k = 0$, initial strategy set α^k, initial fitness set P^k
repeat
  compute stable matching $\mathcal{M}^k$ by Irving's algorithm's extended version based on $n_i^k = \arg\max \mathcal{P}_i(\mathcal{N})|_{\alpha^k}$ for $\forall i \in \mathcal{I}$, where $\mathcal{M}^k = \bigcup_{i \in \mathcal{I}} \mathcal{M}_i^k$, i.e., player i's matches in $\mathcal{M}^k$
  compute $\mathcal{P}_i^k$ given α^k and $\mathcal{M}^k$ for $\forall i \in \mathcal{I}$
  compute $\bar{\mathcal{P}}_{-i}^k = \sum_{j \in \mathcal{M}_i^k} \frac{c_{ij}^k}{C} \mathcal{P}_j^k$ for $\forall i \in \mathcal{I}$
  for all $i \in \mathcal{I}$ **do**
    if $\mathcal{P}_i^k < \bar{\mathcal{P}}_{-i}^k$ **then**
      $\alpha_i^{k+1} := \sum_{j \in \mathcal{M}_i^k} \frac{c_{ij}^k}{C} \alpha_j^k$
    else
      $\alpha_i^{k+1} := \alpha_i^k$
    end if
  end for
  $k := k + 1$
until $\alpha^k = \alpha^{k-1}$

We implemented Algorithm 1 in a custom simulator and studied its convergence properties. Our experiments involve $|\mathcal{I}| = 100$ players with strategies and payoffs as defined in Assumption 2 and Definition 7; we assume $C = 30$, and starting user profiles to be uniformly distributed on the $[0,1]$ interval.

Figure 1 illustrates the distribution of profiles at the stable state. It is possible to distinguish 5 major groups of players holding nearly the same profile level. These players are colluded into clusters, and in each cluster players end up

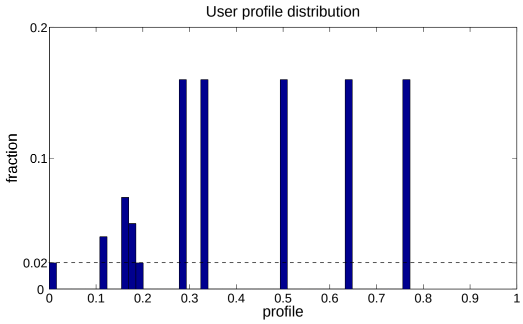

Figure 1: Distribution of player profiles in equilibrium

with the same profile strategy since they all have the same neighborhood to compare themselves to.

We point out that the initial profile distribution includes a fraction of players with a profile α_i close to zero: this is the case for peers having, e.g., poor connectivity or malicious tendency. Figure 1 shows that "low-profile" peers do not improve their profiles throughout subsequent generations. This phenomenon is due to the fact that no peer is willing to cooperate with such peers: user-driven peer selection induces robustness against low-profile peers, that are eventually isolated from the system.

Figure 2 depicts the evolution in time of players' profile strategies: we show the average system profile in each generation. It is possible to deduce that user-driven peer selection results in higher average peer contribution (profile) than a random overlay creation scheme. Here we outline the main reasons for this outcome, but we defer the details to Section 6. Initially, when profiles are uniformly distributed on the $[0,1]$ interval, the average profile is 0.5. Although it appears to decrease throughout the iterations of the algorithm, the average profile when stability is reached is above 0.4; without user-driven peer selection the equilibrium average profile is bounded by the value 0.33: selfish players would not increment their profiles beyond that level (note that $\alpha_i > 0$ implies $E_i = \alpha_i^2 > 0$ cost). While we formalize this result in Proposition 4, here we observe that a peer's best response to the uniform profile distribution is $\alpha^* = 0.25$ by maximizing the payoff's T and E ($\alpha_i^* = \max_{\alpha_i}(\alpha_i(0.5 - \alpha_i))$) while the neighborhood is not profile-dependent. As we show later, evolution makes $\alpha^* = \frac{1}{3}$ to be the dominant strategy.

Finally, Figure 3 depicts the evolution of player payoffs compacted in the the social welfare (i.e., sum of all players' payoffs) for each generation: the heuristic scheme we defined in this section (that is, copying winning strategy from within the neighborhood) drives the system to higher social welfare.

6. PEER SELECTION STRATEGIES: AN EVOLUTIONARY FRAMEWORK

In this section we build a framework based on evolutionary game theory to analyze the properties of a range of peer selection strategies. Our goal is to study the impact of peer selection on peers' profile α, that is we address the following question: *how much effort will a peer dedicate to improve her profile, given a specific peer selection strategy?*

We build upon the illustrative example presented in section 5.4, and define the following evolutionary game: players

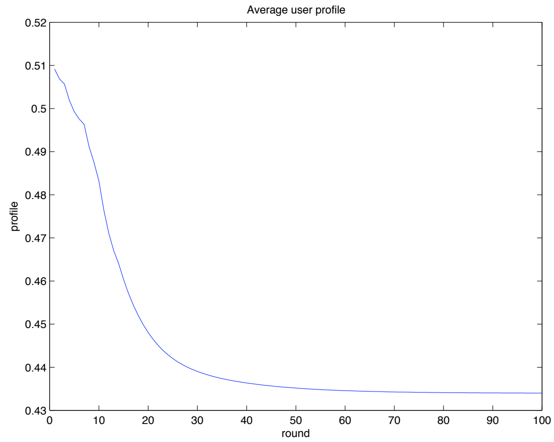

Figure 2: Evolution of player profiles

execute an *interleaved* sequence of peer selection and profile selection. In the first phase, peers adopt one of the peer selection strategies we define hereafter, in the second phase, they adapt their profiles based on the average profile computed over the first phase's strategy set. Only an increment in their fitness will motivate an additional effort in improving their profile. These two phases are repeated throughout generations until an equilibrium is reached. The *asymptotic* evaluation of the system we present aims at determining evolutionary stable strategies (ESS) over peers' profile α.

Assumption 4. *We assume the number of players to tend to infinity: $|\mathcal{I}| \to \infty$; the initial strategy profile α is assumed to be uniformly distributed on $[0,1]$ and continuous over the infinite population. Moreover, we assume that in each generation, every player attempts to establish $\frac{C}{2}$ matches (i.e., data exchanges).*

Definition 8. *Let $\Delta v(\alpha_i, \alpha_j)$ denote the variation in fitness for a player with profile strategy α_i when establishing a match to a player with profile strategy α_j.*

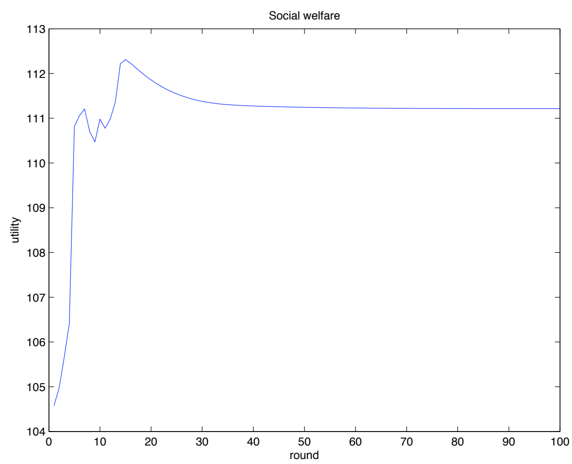

Figure 3: Evolution of player payoffs

Assumption 5. *We assume a cumulative fitness function that accounts for a player's payoff obtained in previous generations. Effort cost is assumed to be equally attributed to the maximum number of matches (C) a player can establish. Therefore*[2]

$$\Delta v(\alpha_i, \alpha_j) = \frac{1}{C}\left(\alpha_j(2-\alpha_i) + \alpha_i(1-\alpha_i)\right).$$

We now turn our attention to a range of heuristic peer selection techniques. Instead of selfishly maximizing a utility function as illustrated in Section 5.4, we make the case for simpler strategies. The following peer selection strategies can be easily implemented in a realistic setting and they do not require global knowledge. Informally, we first propose a completely random, un-biased and unilateral peer selection. We then constrain peer selection accounting for the profile of the two peers involved in a matching. First, we explore a strategy in which the remote peer accepts a matching with a probability that is proportional to the profile of the initiator of the matching. Then, we propose a strategy in which random peer selection is biased by the profile of remote peers: a peer with a high profile will be more likely selected than one with a low profile. Remote peers accept a connection with a probability proportional to the initiator.

Definition 9. *Heuristic peer selection strategies:*

- ***one-sided random matches****: remote peers are randomly chosen and the match is not pairwise, i.e., when chosen, a player has to cooperate with the initiator one;*
- ***pairwise random matches****: a player with α_i randomly selects a remote player, and the match is accepted with probability α_i;*
- ***pairwise strategic matches****: a player with profile α_i selects a remote player with profile α_j with probability α_j and the match will be accepted with probability α_i.*

Before delving into the analysis of the impact of peer selection strategies on profile selection, we briefly review two important concepts in evolutionary game theory.

6.1 Evolutionarily stable strategy

The definition of an ESS that Maynard Smith [19] gives cases involving two possible pure player strategies is the following. In order for a strategy to be evolutionarily stable, it must have the property that if almost every member of the population follows it, no mutant (i.e., an individual who adopts a novel strategy) can successfully invade. Let $v(\alpha')$ denote the total fitness of an individual following strategy α'; furthermore, suppose that each individual in the population has an initial fitness of v_0. If α^* is an evolutionarily stable strategy and α' a mutant attempting to invade the population, then

$$v(\alpha^*) = v_0 + (1-p)\Delta v(\alpha^*, \alpha^*) + p\Delta v(\alpha^*, \alpha');$$

$$v(\alpha') = v_0 + (1-p)\Delta v(\alpha', \alpha^*) + p\Delta v(\alpha', \alpha');$$

where p is the proportion of the population following the mutant strategy α'.

Since α^* is evolutionarily stable, the fitness of an individual following α^* must be greater than the fitness of an individual following α' (otherwise the mutant following α' would be able to invade), and so $v(\alpha^*) > v(\alpha')$. Now, as p is very close to 0, this requires that either that

$$\Delta v(\alpha^*, \alpha^*) > \Delta v(\alpha', \alpha^*) \quad (1)$$

or that

$$\Delta v(\alpha^*, \alpha^*) = \Delta v(\alpha', \alpha^*) \text{ and } \Delta v(\alpha^*, \alpha') > \Delta v(\alpha', \alpha'). \quad (2)$$

In other words, what this means is that a strategy α^* is an ESS if one of two conditions holds: (1) α^* does better playing against α^* than any mutant does playing against α^*, or (2) some mutant does just as well playing against α^* as α^*, but α^* does better playing against the mutant than the mutant does.

6.2 Replicator dynamics

An ESS is a strategy with the property that, once all members of the population follow it, then no "rational" alternative exists. To determine the stable equilibrium state, at first we need to study the replicator dynamics of the system from the initial state. During each generation, players establish matches, and their fitness improves thereby.

As mentioned above, the system is assumed to show discrete dynamics characters, i.e., generations follow each other. The proportion of the population following a given strategy in the next generation is related to the proportion of the population following the same strategy in the current generation according to the rule:

$$x_{\alpha_i}^{t+1} = x_{\alpha_i}^{t}\frac{v_{\alpha_i}(x)}{\bar{v}(x)},$$

where $x_{\alpha_i}^t$ (resp. $v_{\alpha_i}(x)$) denotes the proportion (resp. the average fitness) of population holding strategy α_i during the t-th generation. $\bar{v}(x)$ depicts the average fitness of the whole player set.

6.3 One-sided random matches

When considering one-sided random matches, each player randomly selects $\frac{C}{2}$ players to connect to: the match is established even if the remote peer profile is low compared to the initiator profile.

We now establish the ESS profile selection strategy:

Proposition 4. *In a SE game where one-sided random matching is used $\alpha^* = \frac{1}{3}$ is the only ESS.*

Proof. Let α' be a mutant strategy such that $\alpha' \neq \frac{1}{3}$. We show that $\Delta v(\alpha^*, \alpha^*) > \Delta v(\alpha', \alpha^*)$ always holds. Since $\Delta v(\alpha^*, \alpha^*) = \frac{7}{9}$, after some algebra we arrive at $(\alpha' - \frac{1}{3})^2 > 0$ inequality for the condition to hold, which is always true given $\alpha' \neq \frac{1}{3}$. Similarly, we can show that $\alpha' = \frac{1}{3}$ is successful as mutant strategy against any other strategy, thus it can invade any other overall population α^* strategy. □

6.4 Pairwise random matches

When supposing pairwise random matches, a player with α_i gets rejected with a probability of (1-α_i). In case of rejection, the match is not successful, therefore it does not increase the player's fitness. With this extension we reduce the success possibility of low profile players, so their fitness is expected to increase slower than a player with higher profile. The *expected* payoff of a match initiated by a player holding

[2] Assumption 4 allows to approximate $(\frac{c_{ij}}{C})^{(1+\alpha_j)}$ by $\frac{1}{C}$.

α_i becomes:

$$\Delta v(\alpha_i, \alpha_j) = \frac{1}{C}\left(\alpha_j(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha_i.$$

When considering pairwise random matches, no player has C expected matches, unless all players in the generation hold the maximum profile, i.e., $\alpha = \mathbb{I}_1$. A low profile peer will be rejected with high probability when she initiates a match, on the other hand she will be selected randomly by others, thus despite her bad profile, she might be matched to some peers. To embrace this duality, which does not occur in the previous case where every initiated match is supposed to be successful, we need to distinguish between the payoffs due to "incoming" matches from those obtained from "outgoing" matches. A player with profile α_i will improve her fitness by $\Delta v^b(\alpha_i)$ due to "outgoing" matches and by $\Delta v^t(\alpha_i)$ due "outgoing" matches.

The player fitness improvements depend on the distribution of the population proportions holding given profile strategies. This distribution is time variant due to the inter-generation strategy changes, thus the probability of picking a specific profiled player randomly for a match attempt evolves through subsequent generations. This evolution reacts to the fitness improvement of the player with a given strategy. Assuming uniform initial strategy distribution, we make the following proposition, *limited to the second generation.*

PROPOSITION 5. *Under Assumption 4, the proportion of population with* $\alpha_i < 0.31$ *and with* $\alpha_i > 0.89$ *will decrease, and the number of players with strategy profiles in between is going to increase in the second generation.*

PROOF. At the initial state, profile strategy set is uniformly distributed, i.e., $x_{\alpha_i} = 1$ where x_{α_i} denotes the probability density function (i.e., distribution) of players holding α_i as strategy. The average fitness of the population's proportion holding strategy α_i is

$$\begin{aligned} v_{\alpha_i}(x) &= \Delta v(\alpha_i) = \Delta v^b(\alpha_i) + \Delta v^t(\alpha_i) = \\ &\frac{1}{2}\int_0^1 \left(\alpha(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha_i x_\alpha d\alpha \\ &\quad + \frac{1}{2}\int_0^1 \left(\alpha(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha x_\alpha d\alpha, \end{aligned}$$

since based on the law of large numbers each player initiates and receives $\frac{C}{2}$ match attempts during a generation lifetime. After some algebra we get $v_{\alpha_i}(x) = -\frac{1}{2}\alpha_i^3 + \frac{7}{12}\alpha_i + \frac{1}{3}$ under the assumption $x_\alpha = 1$. Since the average fitness improvement is given by $\bar{v}(x) = \int_0^1 v_\alpha(x) x_\alpha d\alpha$, in our case it is equal to 0.5. Thus based on Subsection 6.2 only the proportion of players holding strategies such that $-\frac{1}{2}\alpha_i^3 + \frac{7}{12}\alpha_i + \frac{1}{3} > \frac{1}{2}$ will increase, establishing the proposition. $\square$

Pairwise random peer selection provides incentives to player for improving their profiles: compared to the ESS of the random matching strategy, the average profile will be higher.

6.5 Pairwise strategic matches

We have seen in section 5.4 that the case of utility-based pairwise matching yields stricter exclusion effect on low profile players. Pairwise strategic matches brings the heuristic strategy closer to the idea behind pairwise utility-based matching, yet it is simpler to implement.

In a SE game of pairwise strategic matches, the *expected* payoff of a match initiated by a player holding α_i is:

$$\Delta v(\alpha_i, \alpha_j) = \frac{1}{C}\left(\alpha_j(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha_i\alpha_j. \quad (3)$$

Based on Equation 3, we establish the following proposition:

PROPOSITION 6. *Under Assumption 4, the lowest profile (under* 0.4*) players will increase their profiles in a system implementing pairwise strategic matches.*

PROOF. The proof is similar to the one given for Proposition 5. Here $v_{\alpha_i}(x) = \frac{1}{2}\int_0^1 \left(\alpha(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha_i\alpha x_\alpha d\alpha$ $+ \frac{1}{2}\int_0^1 \left(\alpha(2-\alpha_i) + \alpha_i(1-\alpha_i)\right)\alpha\alpha_i x_\alpha d\alpha = -\frac{1}{2}\alpha_i^3 + \frac{1}{6}\alpha_i^2 + \frac{2}{3}\alpha_i$, so at the initial state $\bar{v}(x) = \frac{19}{72}$. This result implies that the proportion of population holding higher profile than 0.4 will increase, thus players worse than this threshold will increase their profiles. $\square$

Proposition 6 indicates that if peer profiles are part of the peer selection strategy, the consequence is that peers will be compelled to improve their profiles in order to obtain better matching. In summary, even simple techniques that are not based on any local optimization of a utility function, provide incentives for peers to improve their profiles.

7. CONCLUSION AND FUTURE WORK

In this paper we presented a realistic model of a p2p backup and storage system that accounts for the characteristics (profiles) of peers participating to the system, including their availability, accessibility and (malicious) behavior. We used game theory to define a game in which peers can selfishly optimize the amount of data they wish to store in the system, the set of remote peers to exchange data with, and their profile.

Hindered by the complexity of the joint optimization problem, we focused on the important problem of peer selection with the aim of understanding if peer selection alone can be used to provide incentives to peers for improving their profiles. We cast the problem of peer selection and profile selection as a game, and showed how to extend Stable Matching Theory to fit our problem setting. We extended a known polynomial-time algorithm to compute multiple stable matching and showed through a numerical evaluation that matching alone can be used to compel peers to improve their profile. We also showed that the consequence of the proposed peer selection strategy for the whole system is to have an increased aggregate utility.

We then established a framework based on evolutionary game theory to study simplified peer selection strategies and showed that even semi-random peer selection can be sufficient to provide incentives to peer for improving their profile.

As part of our research agenda, we plan to perform measurements on existing backup and storage solutions (not necessarily p2p systems) in order to build realistic data-sets on peer availability, accessibility and behavior. This will allow us to focus on a clear formulation of the profile set α: which ingredient has an outstanding importance for incentive compatibility to arise? We will also design a real system implementing our heuristic peer selection strategies, study its performance in terms of aggregate utility (benefit for peers) and investigate on the benefit a service provider could derive in managing such a "self-improving" system.

Acknowledgements

This work has been partially funded by the Integrated Project CASCADAS (FET Proactive Initiative, IST-2004-2.3.4 Situated and Autonomic Communications) within the 6th IST Framework Program. This work has also been partially funded by the STREP Project NADA within the 7th IST Framework Program.

8. REFERENCES

[1] Allmydata. Unlimited online backup, storage, and sharing. http://www.allmydata.com.

[2] P. Antoniadis, C. Courcoubetis, and R. Mason. Comparing economic incentives in peer-to-peer networks. *Computer Networks*, 46(1):133–146, 2004.

[3] K. Cechlárová and T. Fleiner. On a generalization of the stable roommates problem. *ACM Trans. Algorithms*, 1(1):143–156, 2005.

[4] B. Chun, F. Dabek, A. Haeberlen, E. Sit, H. Weatherspoon, M. F. Kaashoek, J. Kubiatowicz, and R. Morris. Efficient replica maintenance for distributed storage systems. In *Symposum on Networked Systems Design and Implementation (NSDI)*, 2006.

[5] B. Cohen. Incentives build robustness in bittorrent. In *Proc. of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON'03)*, 2003.

[6] J. Corbo and D. Parkes. The price of selfish behavior in bilateral network formation. In *Proc. of 24th annual ACM Symposium on Principles of Distributed Computing (PODC'05)*, pages 99–107, 2005.

[7] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker. On a network creation game. In *Proc. of 22nd ACM Symposium on Principles of Distributed Computing (PODC)*, 2003.

[8] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[9] P. B. Godfrey, S. Shenker, and I. Stoica. Minimizing churn in distributed systems. *SIGCOMM Comput. Commun. Rev.*, 36(4):147–158, 2006.

[10] D. Grolimund, L. Meisser, S. Schmid, and R. Wattenhofer. Havelaar: A Robust and Efficient Reputation System for Active Peer-to-Peer Systems. In *1st Workshop on the Economics of Networked Systems (NetEcon)*, 2006.

[11] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[12] R. W. Irving. An efficient algorithm for the "stable roommates" problem. *Journal of Algorithms*, 6(4):577–595, 1985.

[13] R. W. Irving and S. Scott. The stable fixtures problem - a many-to-many extension of stable roommates. *Discrete Appl. Math.*, 155(16):2118–2129, 2007.

[14] N. Laoutaris, G. Smaragdakis, A. Bestavros, and J. W. Byers. Implications of selfish neighbor selection in overlay networks. In *INFOCOM*, pages 490–498, 2007.

[15] D. Lebedev, F. Mathieu, L. Viennot, A.-T. Gai, J. Reynier, and F. de Montgolfier. On using matching theory to understand p2p network design. *CoRR*, 2006.

[16] Mozy. Online backup service. http://mozy.com.

[17] R. Rodrigues and B. Liskov. High availability in dhts: Erasure coding vs. replication. In *International Workshop on Peer-to-Peer Systems (IPTPS)*, 2005.

[18] Amazon S3. Simple storage service. http://aws.amazon.com/s3.

[19] J. M. Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[20] L. Toka and P. Maillé. Managing a peer-to-peer backup system: Does imposed fairness socially outperform a revenue-driven monopoly? In *GECON*, pages 150–163, 2007.

[21] Wuala. Peer-to-peer storage system. http://wua.la/en/home.html.