Strategy-Proof and Efficient Kidney Exchange Using a Credit Mechanism

Chen Hajaj
Bar-Ilan University
chen.hajaj@biu.ac.il

John P. Dickerson
Carnegie Mellon University
dickerson@cs.cmu.edu

Avinatan Hassidim
Bar-Ilan University
avinatan@cs.biu.ac.il

Tuomas Sandholm
Carnegie Mellon University
sandholm@cs.cmu.edu

David Sarne
Bar-Ilan University
sarned@cs.biu.ac.il

Abstract

We present a credit-based matching mechanism for dynamic barter markets—and kidney exchange in particular—that is both strategy proof and efficient, that is, it guarantees truthful disclosure of donor-patient pairs from the transplant centers and results in the maximum global matching. Furthermore, the mechanism is individually rational in the sense that, in the long run, it guarantees each transplant center more matches than the center could have achieved alone. The mechanism does not require assumptions about the underlying distribution of compatibility graphs—a nuance that has previously produced conflicting results in other aspects of theoretical kidney exchange. Our results apply not only to matching via 2-cycles: the matchings can also include cycles of any length and altruist-initiated chains, which is important at least in kidney exchanges. The mechanism can also be adjusted to guarantee immediate individual rationality at the expense of economic efficiency, while preserving strategy proofness via the credits. This circumvents a well-known impossibility result in static kidney exchange concerning the existence of an individually rational, strategy-proof, and maximal mechanism. We show empirically that the mechanism results in significant gains on data from a national kidney exchange that includes 59% of all US transplant centers.

Introduction

In the United States alone, over 3.8 million people—roughly 1.6% of the population—have kidney disease.[1] For many, kidney disease will progress to outright kidney failure—and with it the need for a kidney transplant. Transplant organs can be sourced from cadavers or willing living donors. However, there is a severe supply and demand mismatch with donor organs; in 2013, 36,395 people were added to the US national kidney waiting list, while only 16,462 left due to receiving a living or deceased donor kidney.[2] Furthermore, roughly half of the over 100,000 candidates on the US list have been waiting for a kidney for more than two years.

Kidney exchange aims to reduce the transplant organ supply-demand imbalance by making it easier to match willing donors to needy patients. A person in need of a kidney may have one or more healthy donors with two healthy kidneys who are *willing but unable* to donate a kidney to that person, typically due to a medical incompatibility like a blood or tissue type mismatch. These incompatible donor-patient pairs can exchange donors with other pairs, through cycles or chains of exchanges, in such a way that each involved pair's donor gives to a (medically compatible) patient of another pair. Recommending a "good" set of swaps is thus the basic kidney exchange problem.

In recent years, numerous multi-transplant center kidney exchanges have been fielded around the world, where donor-patient pairs from different transplant centers can switch donors. Multi-center kidney exchange programs are run through clearinghouses that recommend a matching from the full set of donor-patient pairs.

Transplant centers have different incentives than patients and donors. Performing an organ transplant surgery is typically very profitable to a center. Thus, centers may have incentive to only reveal some subset of their donor-patient pairs to the clearinghouse, and match other pairs internally at their center.[3] Such strategic behavior is rampant: today most transplant centers not only hide their easy-to-match pairs, but all their internally matchable pairs (Stewart et al. 2013). It has also been shown to reduce the overall efficiency of kidney exchange in theory (Ashlagi and Roth 2014).

While most prior work on mechanism design for kidney exchange has focused on static models, in this paper we consider a more realistic multi-period dynamic model. For this dynamic model, we study how to construct a mechanism that makes it a dominant strategy for each center to reveal its pairs truthfully, and hence maximizes the number of pairs in the central pool from which the clearinghouse will construct a global matching. Our contributions in this work are the following. First, we design a credit-based *strategy-proof* matching mechanism for the dynamic model that considers the incentives of the different transplant centers. The mechanism is *efficient* and guarantees long-term *individual rationality* (IR). Our results apply not only to matching via 2-cycles—a restriction present in some kidney exchange papers: our matchings can also include cycles of any length and altruist-initiated chains, which is important at least in

---

[1]The 2012 US National Health Interview Survey (Blackwell, Lucas, and Clarke 2014) counts patients with "weak or failing kidneys" (see their Tables 7 and 8 for methodology details).

[2]http://optn.transplant.hrsa.gov

[3]Internal matches are also logistically easier to handle.

kidney exchanges. Second, we show that the mechanism can be adjusted to guarantee *immediate individual rationality* at the expense of economic efficiency, while preserving strategy proofness via the credits. Our experiments show that the efficiency loss is very small in practice. This variant of the mechanism circumvents—via the use of the credits—a well-known impossibility result, proved for static kidney exchange, that no IR mechanism is both maximal and strategy proof. Our mechanism does not require assumptions about the underlying distribution of compatibility graphs—a nuance that has previously produced conflicting results in other aspects of theoretical kidney exchange. Finally, experiments with the mechanism, both on real data from a large fielded kidney exchange in the US as well as from a data generator, show that the number of resulting matches is substantially greater than without the mechanism.

Related Work

The idea of kidney exchange was introduced by Rapaport (1986), and the first organized kidney exchange started in 2003 (Roth, Sönmez, and Ünver 2004). The topic has attracted researchers from non-medical fields such as economics (Roth, Sönmez, and Ünver 2005; 2007b; Akbarpour, Li, and Gharan 2014) and computer science (Abraham, Blum, and Sandholm 2007; Biró, Manlove, and Rizzi 2009; Dickerson, Procaccia, and Sandholm 2012a; 2013; 2014; Dickerson and Sandholm 2014; Anshelevich et al. 2013; Liu, Tang, and Fang 2014; Li et al. 2014). The National Organ Transplant Act of 1984 makes it illegal to buy or sell a kidney in the US, thus making donation the only viable option for kidney transplantation (Roth 2007; Leider and Roth 2010). Similar legislation exists throughout most of the world. The initial proposal for a large-scale kidney exchange that was made by Roth, Sönmez, and Ünver (2004) included the ability to use both cycles (where each donor donates a kidney to the next patient, with the final donor donating her kidney to the first patient) and chains (that are like cycles except that the cycle does not close).[4]

In the bulk of the work on mechanism design for kidney exchange (including our paper), the agents that need to be incentivized are the transplant centers (not the donor-patient pairs or altruistic donors) (Ashlagi and Roth 2014; Toulis and Parkes 2011; Sönmez and Ünver 2013; Ashlagi et al. 2013). A center can decide to reveal none, some, or all of its pairs and altruists to the clearinghouse. Roth, Sönmez, and Ünver (2007a) proved that no individually rational (IR) and maximal mechanism can also be strategy proof (in a static model, where patients and donors do not arrive and leave over time). According to their definition, a matching is maximal if there exists no larger matching that fully encompasses the former. *One of the main contributions of our paper is the circumvention of this impossibility result via a credit mechanism in a dynamic model, and real kidney exchanges are dynamic.* Ashlagi and Roth (2014) contin-

ued their work in the static model, showing that the efficiency loss due to lack of strategy proofness can be very high but that there exists an ϵ-Bayesian incentive compatible IR mechanism that tends to become nearly efficient in the large under a dense random graph model.

Ashlagi et al. (2013) presented a randomized strategy-proof IR mechanism for a static setting that delivers at least 50% of efficiency. Their model, though, is for 2-cycles only, while ours supports cycles and chains of any length.

Model

We start from the standard kidney exchange model with a set of n transplant centers, $T = \{\tau_1, \ldots, \tau_n\}$, and a central clearinghouse. The process is modeled as a game that is divided into time periods: at each period, each center τ_i receives a set $V_i = \{v_i^1, \ldots, v_i^{\hat{N}}\}$ of donor-patient pairs.[5]

We assume that a pair must be matched in the current period or not at all. This could be motivated in a setting where the patients are in critical condition and will not survive to the next period. Whether the assumption is a good approximation of reality depends on the patient pool and the cadence of matching in the particular exchange; there is enormous variability in practice. For example, the National Kidney Registry (a private exchange in New York) conducts match runs multiple times per day, while Canada's and the UK's national exchanges match four times per year. Extending our mechanism to survival spans beyond one period does not seem straightforward; we leave that for future research.

We treat altruistic donors as pairs. Because pairs only last for one period, any chain will end during the period, and the last pair donates to the deceased-donor waiting list. Assuming that surgery is allocated to the center that submitted the altruist that started that chain (Woodle et al. 2010), the number of surgeries that a center gets from an altruist is the same as that from a pair, so we can treat altruists as pairs.[6]

The arrival rate distribution of pairs to each transplant center τ_i is of a general form with a mean of k_i, which is known by the center and the clearinghouse. (For example in the case of the UNOS nationwide exchange in the US, this could be justified by the fact that UNOS has visibility to the full history of both the deceased-donor waiting list and the transplants that each center in the country has conducted.) The distributions are defined over finite intervals, which can be justified by each transplant center's capacity for patients. For exposition purposes, we assume each transplant center's upper bound on capacity is $2k_i$, though the mechanism can be adjusted to any bound. We also make the innocuous assumption that for every transplant center, in every period, there is a non-zero probability of getting a pair that the center cannot match internally and the pair participates in one or more (but not all) global maximum matchings; this motivates the centers to consider participation in the mechanism.

---

[4]Their proposal was to have chains start with a pair that has received a kidney from the deceased-donor waiting list. In practice, chains start with an altruistic donor that does not expect anything in return.

---

[5]A donor can donate one organ and a patient needs one organ; each belongs to a single pair. (A pair can have multiple donors.)

[6]The number of surgeries a center gets also depends on whether donors fly to recipients' centers or have their kidneys shipped. Both modes are used in practice; our mechanism supports each.

Centers are assumed to be self-interested and risk-neutral, so their goal is to maximize the expected number of pairs per period out of those they "own" that will be matched (internally or through the clearinghouse) over an infinite horizon.

Our mechanism manages a credit balance for the different transplant centers as explained in the following sections. The matching decisions made by the mechanism at each time period are solely based on the centers' states, captured for each center by the tuple (V_i, c_i), where V_i is the transplant center's set of pairs and c_i is the transplant center's credit balance. The strategy of any transplant center τ_i is thus a mapping $S_i : (V_i, c_i) \rightarrow V_i^{disclosed} \subseteq V_i$, where $V_i^{disclosed}$ is the set of pairs the center discloses to the clearinghouse.

The clearinghouse can use combinations of two structures for matching: *cycles* and *chains*.[7] Chains enable the inclusion of altruistic donors, who enter the pool without a paired patient, enabling more flexibility in the matching. Chains significantly increase efficiency in theory (Dickerson, Procaccia, and Sandholm 2012b; Ashlagi et al. 2012) and practice (Rees et al. 2009). The proposed mechanism and its properties do not change even if only cycles or only chains were used, rather than both—which is the status quo in most modern kidney exchanges in practice.

We design our mechanism so that the central clearinghouse can use a myopic matching rule that performs a maximum cardinality matching in the reported pool at every period, possibly subject to external constraints imposed by the mechanism. This is a desirable property of the model, as myopic matching is the status quo in fielded kidney exchanges.

Mechanism Desiderata

Before presenting our mechanism, we define the concepts of individual rationality, efficiency, and strategy proofness.
Individual Rationality (IR). We address two common notions of individual rationality. *Long-term individual rationality* requires that the participation in the mechanism results in at least the same expected utility as not participating overall (Parkes 2001). In our setting this maps to having at least the same long-run number of matches if participating in the mechanism. The second, stricter IR notion, *immediate individual rationality*, requires that at each time period, the allocation guarantees the transplant center at least the number of matches that it would have received by matching internally (Ashlagi and Roth 2014). While the two notions coincide in static settings, they differ in dynamic settings.
Efficiency. A mechanism is *efficient* if it results in an allocation that maximizes global utility, i.e., social welfare. In our setting, this is measured as the number of matches produced. Thus an efficient algorithm produces a maximum matching given *all* pairs that the transplant centers have (regardless of

which pairs they reveal). We also consider *IR efficiency*, defined as the maximum matching that guarantees IR in each period (i.e., immediate IR). We further consider the notion of *maximal matching* (Ashlagi and Roth 2014), which is a matching that cannot be extended to a larger matching without changing the original matching.
Strategy proofness. A mechanism is strategy proof (aka. dominant-strategy incentive compatible) if under that mechanism, in every state of the world (i.e., credits and sets of pairs of each transplant center) each center is no worse off reporting all of its pairs than reporting a subset thereof.

We use a strong notion of dominance between a center's strategies; it does not depend on what other centers reveal.

Definition 1. *Under a given mechanism, strategy S_i for a center τ_i dominates strategy S_i' if by being applied starting from any state of the world, the expected number of matches for τ_i is strictly greater under S_i than under S_i'.*

Transplant centers do not always have incentive to reveal all their pairs (Toulis and Parkes 2011; Sönmez and Ünver 2011; Ashlagi et al. 2013; Ashlagi and Roth 2014), and in practice, hiding of pairs is rampant (Stewart et al. 2013). This causes the number of overall matches (counting both clearinghouse matches and centers' internal matches) to be suboptimal—that is, efficiency is compromised.

Our Credit-Based Mechanism

In this section, we present our mechanism. It incentivizes truthful revelation by lowering the probability that in the future, a transplant center's disclosed pairs will be included in a global matching, if that center reveals in the present a smaller number of pairs than it is expected to have.

The mechanism is based on managing an account of credits for each transplant center. Decisions of whether to prefer one transplant center over another in the global matching are then influenced by the credit balances of the different centers. In general, having more credits will increase the likelihood of a center having more of its disclosed pairs matched by the mechanism. By giving and taking credits from the different centers, the clearinghouse incentivizes each center to adopt a truthful strategy, that is, to reveal all of its pairs. Thus, the number of pairs in the pool from which the clearinghouse will construct the global matching is maximized.

At each period, the mechanism performs 1) an initial credit balance update, 2) a global matching, and 3) a final credit balance update. Below we describe each of these steps in detail. We use M to denote a matching, $u_i(M)$ to denote the number of matched pairs belonging to transplant center τ_i in the matching M, and $u(M)$ to denote the total number of matched pairs (i.e., $u(M) = \sum_{\tau_i \in T} u_i(M)$).

Upon receiving the sets of pairs disclosed by the different transplant centers, the mechanism constructs a directed, labeled *compatibility graph* $G = (V, E)$ with a vertex $v \in V$ for each disclosed donor-patient pair (labeled by its transplant center) and directed edge $e \in E$ from v_i to v_j for any two pairs (v_i, v_j) for which the donor of v_i is compatible with the patient of v_j. The graph is used as a basis for determining the matches for the different transplant centers.

---

[7]There are two kinds of chains in practice: *domino chains* where the last pair in the chain donates to the deceased-donor waiting list (Montgomery et al. 2006), and *nonsimultaneous, extended, altruistic-donor (NEAD) chains* (Rees et al. 2009), where the donor in the final pair is used as a virtual altruist in later matching periods to potentially continue the chain, thus enabling the chain to potentially go on forever. In our model, the life span of a pair is assumed to be one period, so our chains are domino chains.

Algorithm 1 Credit-based matching mechanism.

1: **function** RUNMECHANISM(G, T, c)
2:     $c \leftarrow$ INITIALCREDITUPDATE(T, c)   // Update centers' credit balances based on reported vs. expected type
3:     $\mathcal{O}^= \leftarrow \max_{M \in \text{ALLLEGALMATCHINGS}(G)} u(M)$ // Compute a maximum global matching
4:     **return** COMPUTEMATCHING($\mathcal{O}^=, G, T, c$)   // Run the iterative matching process and return optimal matching

---

1: **function** INITIALCREDITUPDATE(T, c)
2:     **for each** transplant center $i \in T$ **do**
3:         $c_i \leftarrow c_i + 4k_i \left(\left| V_i^{disclosed} \right| - k_i \right)$
4:     **return** c

1: **function** MATCHINGS($G, T, \mathcal{C}^=, \mathcal{O}^=$)
2:     $\mathcal{M} \leftarrow$ ALLLEGALMATCHINGS(G)
3:     $\mathcal{M} \leftarrow \{M \mid M \in \mathcal{M} \land u(M) = \mathcal{O}^=\}$
4:     **for each** transplant center $i \in T$ **do**
5:         **if exists** $\mathcal{C}^=(i)$ **then**
6:             $\mathcal{M} \leftarrow \{M \mid M \in \mathcal{M} \land u_i(M) = \mathcal{C}^=(i)\}$
7:     **return** $\mathcal{M}$

1: **function** ALLLEGALMATCHINGS(G)
2:     Computes the set of all feasible matchings in G

1: **function** COMPUTEMATCHING($\mathcal{O}^=, c$)
2:     $\mathcal{C}^= \leftarrow \varnothing$; $M^* \leftarrow \varnothing$
3:     **for each** transplant center $i \in$ RANDPERM(T) **do**
4:         $M_i^+ \leftarrow \arg\max_{M \in \text{MATCHINGS}(G,T,\mathcal{C}^=,\mathcal{O}^=)} u_i(M)$
5:         $M_i^- \leftarrow \arg\min_{M \in \text{MATCHINGS}(G,T,\mathcal{C}^=,\mathcal{O}^=)} u_i(M)$
6:         **if** $c_i \geq 0$ **then**
7:             $M^* \leftarrow M_i^+$
8:             $c_i \leftarrow c_i - (u_i(M_i^+) - u_i(M_i^-))$
9:         **else**
10:            $M^* \leftarrow M_i^-$
11:            $c_i \leftarrow c_i + (u_i(M_i^+) - u_i(M_i^-))$
12:        $\mathcal{C}^=(i) \leftarrow u_i(M^*)$
13:    **return** (c, M^*)

The pseudocode for the matching mechanism is given as Algorithm 1. The first step, function INITIALCREDITUPDATE, is updating the credit balance of each center based on its expected arrival rate. This is done by comparing the number of pairs disclosed by each transplant center ($|V_i^{disclosed}|$) in this period to the number of pairs the transplant center is expected to get in this period, k_i, and updating the credit balance of each transplant center τ_i, denoted c_i, as follows:

$$c_i = c_i + 4k_i(|V_i^{disclosed}| - k_i) \tag{1}$$

The idea is that the more pairs reported by the transplant center, the greater the number of credits awarded (or, depending on k_i, the fewer the number of credits taken). Thus, if we can show that receiving more credits yields greater expected utility than hiding a pair from the clearinghouse, a transplant center has incentive to disclose all of its pairs. Since the arrival of pairs to the transplant center is probabilistic, in periods where the transplant center receives fewer than k_i pairs, it will lose some credits despite being truthful.

Next, the mechanism conducts the global matching. This is done by iteratively switching between different maximum matchings where, on each iteration, a center is rewarded or punished base on its credits (Step 4 in the main RUNMECHANISM, which calls COMPUTEMATCHING). The function COMPUTEMATCHING iterates over the different centers according to a random permutation (RANDPERM), gradually building the set of constraints under which the next maximum matching to be evaluated will be selected. The constraints are stored in the structure $\mathcal{C}^=$ (initially set to $\varnothing$) and they specify the exact number of matches for each center. For each center in the permutation, the mechanism extracts a matching that maximizes the center's number of matches and one that minimizes it, denoted M_i^+ and M_i^- respectively, subject to $\mathcal{C}^=$ and requiring that the total matching size equals that of the fully uncon-

strained one (via function MATCHINGS). This latter value, $\mathcal{O}^=$, is obtained in Step 3 of RUNMECHANISM. It is used to guarantee that the global matching produced is also a maximum matching. The mechanism then takes transplant center τ_i's current credit balance into account by choosing to proceed with either M_i^+ or M_i^- as a constraint for the following iterations. A positive balance results in choosing M_i^+ and a negative balance in M_i^- (Steps 6–11). Therefore, the function COMPUTEMATCHING returns a maximum matching that corresponds to all of the reward and punishment decisions made throughout the algorithm in this period.

Theorem 1. *Under this mechanism, with probability 1, any untruthful strategy S is dominated by every strategy S' such that $S'(\cdot, \cdot) = S(\cdot, \cdot)$ except for one (arbitrary) time step in which τ_i's (arbitrary) state is (V_i, c_i), and S hides j pairs while S' hides $j - 1$ pairs.*

Proof. Let c_i^S and $c_i^{S'}$ denote the credit balance of transplant center τ_i when using strategy S and S', respectively. Let t^* denote the time period when S' and S differ. (Since pairs exist for only one time period each, the only part of the state that carries across periods is the credits, and therefore there is exactly one such period t^*.) Then, to prove that S' dominates S, it suffices to show that, at some period $t' > t^*$, (i) the two strategies enter that period with $c_i^{S'} = c_i^S$, thus resulting in identical treatment by the mechanism from that point forward, and (ii) strategy S' has received at least one more match than S in aggregate.

We track two variables at each time period t: the difference in credit balances $D_c^t = c_i^{S'} - c_i^S$, and the aggregate difference in total matches (in favor of S') D_m^t. For each $t^- < t^*$, S and S' are identical: $D_c^{t^-} = D_m^{t^-} = 0$. At the deviation time period t^*, we have $D_c^{t^*} = 4k_i$ (because S' revealed one additional pair, resulting in $4k_i$ additional credits (Equation 1) and $D_m^{t^*} \geq -2k_i + 1$. This latter lower bound

follows from the maximum additional number of matches a transplant center could receive by withholding a single additional pair, $2k_i - 1$, which follows from the center's maximum number of pairs at any period being $2k_i$. We will now focus on the time periods after the single deviation between S' and S has occurred and show that, with high probability, there is a future time period $t' > t^*$ where $D_c^{t'} = 0$ and $D_m^{t'} \geq 1$.

For each $t^+ > t^*$, the difference in credits $D_c^{t^+}$ can only change in two ways: first via the initial credit update part of the mechanism (Equation 1), and second, in Algorithm 1 on steps 6–11 of COMPUTEMATCHING. Since S and S' (after their single deviation) reveal the same pairs, Equation 1 will not affect $D_c^{t^+}$. Thus, steps 6–11 fully define the future dynamics of $D_c^{t^+}$ and $D_m^{t^+}$. This yields the following bounds: $D_c^{t^+} \in [-4k_i + 4, 4k_i]$ and $D_m^{t^+} \in [-2k_i + 1, 2k_i - 1]$. Indeed, from period t^* onward, $D_c^{t^+}$ will keep changing, with the two strategies possibly alternating in taking the lead in number of overall matches. If $sign(c_i^S) \neq sign(c_i^{S'})$, the change in the difference $D_m^{t^+}$, denoted $\Delta(D_m^{t^+}) = D_m^{t^+} - D_m^{t^+-1}$, will be

$$\Delta(D_m^{t^+}) = \begin{cases} u_i(M_i^+) - u_i(M_i^-) & \text{if } c_i^{S'} \geq 0 \\ -(u_i(M_i^+) - u_i(M_i^-)) & \text{otherwise} \end{cases} \quad (2)$$

Furthermore, the change in difference $D_c^{t^+}$, denoted $\Delta(D_c^{t^+}) = D_c^{t^+} - D_c^{t^+-1}$, will be exactly

$$\Delta(D_c^{t^+}) = -2\Delta(D_m^{t^+}). \quad (3)$$

Because the mechanism's matching favors the strategy that is ahead in credits, and centers favored in the matching lose credits, the strategy that is ahead in credits loses more credits over time. (By the assumption made earlier in the paper that there is a non-zero probability of receiving a pair that *cannot* be matched internally but *can* be matched in one but not all maximum global matchings, at each time period there is a constant probability of spending a credit to receive a match, so the credit balance will change over time with probability 1.) Formally, the random walk of $D_c^{t^+} \in [-4k_i + 4, 4k_i]$ is such that it goes to zero by some time period $t' > t^*$; that is, it starts as $D_c^{t^*} = 4k_i$ and moves to 0 by way of

$$D_c^{t'} = D_c^{t^*} + \sum_{t^+ \in t^*+1,\dots,t'} \Delta(D_c^{t^+}). \quad (4)$$

Because $D_c^{t'} = 0$ and $D_c^{t^*} = 4k_i$, we have

$$-4k_i = \sum_{t^+ \in t^*+1,\dots,t'} \Delta(D_c^{t^+}). \quad (5)$$

During the same time periods, in lock step, $D_m^{t^*} \geq -2k_i + 1$ moves to

$$D_m^{t'} = D_m^{t^*} + \sum_{t^+ \in t^*+1,\dots,t'} \Delta(D_m^{t^+}). \quad (6)$$

From Equation 3, we have $\Delta(D_m^{t^+}) = -\frac{1}{2}\Delta(D_c^{t^+})$, so

$$D_m^{t'} = D_m^{t^*} + \sum_{t^+ \in t^*+1,\dots,t'} -\frac{1}{2}\Delta(D_c^{t^+}). \quad (7)$$

By Equation 5 and the lower bound $D_m^{t^*} \geq -2k_i + 1$, it follows that

$$D_m^{t'} \geq (-2k_i + 1) - \frac{1}{2}(-4k_i) = 1. \quad (8)$$

So, by the time t' when the credit balances of the two strategies have equalized, strategy S' is ahead of S by at least one match. $\qquad\square$

Corollary 1. *Under this mechanism, the truthful strategy of revealing all pairs dominates all untruthful ones (w.p.1).*

Proof. Using Theorem 1, with probability 1, any untruthful strategy S is dominated by a more truthful strategy S' that is identical to S except that it is truthful in a single period in which S is not truthful. Hence, each untruthful strategy is dominated by a strategy that drops the first untruthful step. Repeating this process recursively, we end up with the truthful strategy dominating any other strategy. $\qquad\square$

Corollary 2. *The mechanism is long-term IR (w.p.1).*

Proof. Not participating in the mechanism is equivalent to participating but hiding all pairs. From Corollary 1, w.p.1, this strategy is dominated by participating truthfully. $\qquad\square$

Using the multiplier $4k_i$ in the credit update step

The initial credit balance update (Equation 1, used in function INITIALCREDITUPDATE in the mechanism) uses a somewhat cryptic-seeming credit multiplier $4k_i$. Theorem 1 and its proof hold for any multiplier $\lambda > 4k_i - 4$. We now provide an example that shows that for any multiplier $\lambda \leq 4k_i - 4$, the mechanism can fail to be strategy proof.

Fix some constant $\lambda \leq 4k_i - 4$. Consider a setting with two transplant centers. We will focus on transplant center τ_1, which has an arrival distribution with mean $k_1 = 3$. On this particular time step, τ_1 received the maximum number of pairs it could get (i.e., $2k_1 = 6$ pairs). Transplant center τ_2 also receives some pairs in this period, and reveals them to the clearinghouse. Two possible cycles exist in the revealed pool; Figure 1 shows these cycles, with τ_1's pairs in white and τ_2's pairs in dark gray. Note that the cycles overlap (pair v_2^1 from transplant center τ_2 is present in both cycles), so the clearinghouse can only choose one cycle.
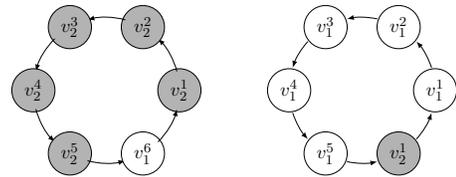

Figure 1: Two cycles in the global compatibility pool.

In this case, τ_1 has the following two options (among others): disclosing all of its pairs and hoping that the right cycle will be chosen (otherwise, only one of its pairs will be

matched) or hiding only pair v_1^6 (resulting in all disclosed pairs being matched). Choosing the first option results in $\lambda(2k_i - k_i)$ credits, but the left cycle may be selected (resulting in only one of τ_1's pairs being matched, v_1^6) if τ_1 falls behind τ_2 in the random permutation over centers. But, by simply hiding v_1^6 (which breaks the left cycle, forcing the smaller cycle to be chosen), this transplant center will instead lose λ credits, but gain an additional four matches to τ_1 relative to the truthful case.

To ensure that transplant center τ_1 will be truthful, we must set λ such that $\lambda > (2k_i - 1 - 1)$. Still, in this case, by adopting a truthful strategy there exist two possible matchings (with the same overall utility), so if the right cycle will be selected, τ_1 will lose $(2k_i - 1) - 1$ credits in the final credit balance update (the difference between the maximum and minimum matchings). Thus, the inequality that we need to satisfy is $\lambda > [(2k_i - 2) + (2k_i - 2)]$, which simplifies to $\lambda > (4k_i - 4)$.

Trading Efficiency for Immediate IR

While the above mechanism is efficient and strategy proof, the individual rationality it guarantees is only in the long term. We now show how the mechanism can be adjusted to guarantee individual rationality at each period (immediate IR)—of course, taking the value of the credits for the future into account. The immediate individual rationality comes at the cost of a decrease in the amount of guaranteed efficiency. Practically, guaranteeing immediate IR does not contribute much; even with only long-term IR guaranteed, all centers are incentivized to participate and report truthfully. Still, the revised mechanism circumvents a well-known impossibility result in static kidney exchange (Roth, Sönmez, and Ünver 2007a), so it is important in its own right.

In a given time period, each transplant center τ_i has some (potentially empty) subset of pairs $V_i^{IM} \subseteq V_i$ that it can match internally. The main challenge in designing a truthful yet efficient and immediate IR mechanism is that, relative to just V_i^{IM}, the mechanism can only provide at most $|V_i^{IM}|$ matches in the worst case—which is the same number that τ_i would have obtained by matching each pair in V_i^{IM} internally. One possible workaround for this is to have the mechanism deliver a global solution where each transplant center receives at least as many matches as it could have matched internally (computed based on the pairs it revealed). Still, as discussed by Ashlagi and Roth (2014), this property is not enough to incentivize truth-telling for each transplant center.

By moving from a single-period model to a (more realistic) multi-period, dynamic model, we can overcome this problem—even though pairs do not carry over across periods. Specifically, in our model we can extend the state of each transplant center to $(V_i, \mathcal{C}^\geq(i), c_i)$, where $\mathcal{C}^\geq(i)$ is the number of pairs transplant center τ_i could have matched internally out of its disclosed set $V_i^{disclosed}$. Since the clearinghouse can compute $\mathcal{C}^\geq(i)$, all that remains is to change the mechanism so that instead of generating maximum matchings, it generates maximum constrained matchings where each transplant center is guaranteed at least $\mathcal{C}^\geq(i)$ matches. Note that there is at least one matching that guarantees at least $\mathcal{C}^\geq(i)$ matches for each transplant center τ_i:

in the worst case, each center receives exactly its internal matching (among disclosed vertices).

The fact that the revised mechanism is immediate IR follows from the fact that each transplant center receives at least the size of its internal matching. Similarly, the mechanism is IR-efficient, as it guarantees the maximum matching available subject to the internal-matching constraints. The proof for its strategy proofness is the same as the one given for the original mechanism.

Experiments

In this section, we evaluate the mechanism experimentally. We use data from a large, fielded kidney exchange in the US run by the United Network for Organ Sharing (UNOS). The exchange started in 2010 and now includes over 140 transplant centers. We include 3-cycles and 4-chains, in accordance with the current practice of the UNOS exchange.

We also include experiments on an older family of graphs in kidney exchange research, from a generator due to Saidman et al. (2006). While the UNOS family of graphs more closely mirrors reality (see, e.g., (Ashlagi et al. 2012; Ashlagi, Jaillet, and Manshadi 2013; Dickerson, Procaccia, and Sandholm 2013; 2014)), we include the Saidman et al.-style graphs for posterity. This distribution roughly mimics a US-based kidney exchange that draws patients uniformly from the national waiting list of people in need of kidneys (UNOS), and draws donors uniformly from the US adult population. Indeed, while this may someday mimic the composition of a US-wide kidney exchange in its steady state, it does not accurately reflect the composition of currently fielded exchanges.

Experimental Setup. Our simulation framework was built in Java 1.6 on top of a vetted open source kidney exchange software suite;[8] we incorporated our mechanisms into this suite. At a high level, a single run of simulation executes as follows. First, fix the number of transplant centers $|T|$, arrival distribution and time limit η. Then, for each time period $t \in \{1, \ldots, \eta\}$, we execute the mechanism as described in Algorithm 1. We track the number of vertices matched by the clearinghouse and by each center internally.

Solving the maximization/minimization problem (e.g., $M_i^+ \leftarrow \arg\max_{M \in \text{MATCHINGS}(G,T,\mathcal{C}=,\mathcal{O}=)} u_i(M)$) described in Algorithm 1 is NP-complete (Abraham, Blum, and Sandholm 2007). We use an integer programming (IP) formulation with constraints that enforce the space of feasible matchings described by the function MATCHINGS. The IP is built on top of the "cycle formulation" of the kidney exchange problem, where one binary decision variable x_c is associated with each cycle or chain c in the set of all legal cycles and chains in a graph (Abraham, Blum, and Sandholm 2007). Our code uses IBM CPLEX as the IP solver.

Simulations were performed over $\eta = 100$ time periods and were parameterized in two major ways: number of transplant centers ($|T|$) and arrival rate distribution. Arrival rates varied from very low (e.g., $\sim U[1, 5]$, where each transplant center receives at least 1 pair, at most 5 pairs, and on average 3 pairs per time period) to very high (e.g.,
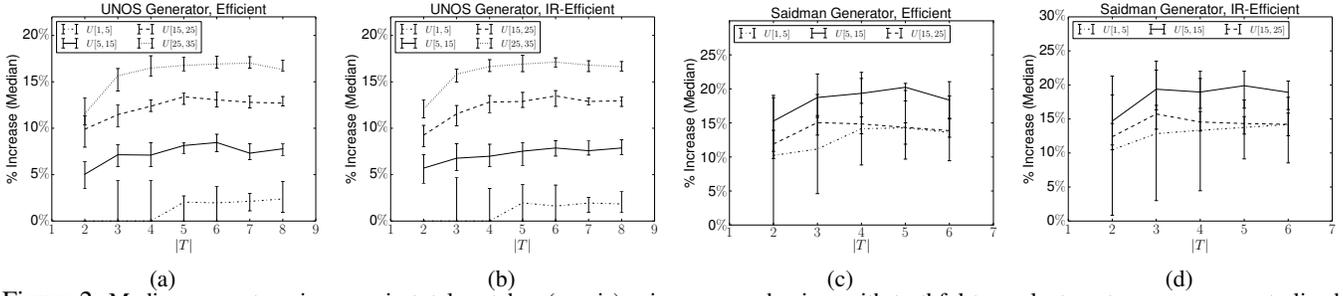
Figure 2: Median percentage increase in total matches (y-axis) using our mechanism with truthful transplant centers versus a centralized mechanism with non-truthful transplant centers, for increasing $|T|$ (x-axis) and arrival rates (different lines), for the efficient (2a,2c) and IR-efficient (2b,2d) mechanisms, for the real UNOS (2a,2b) and generated Saidman (2c,2d) data.

$\sim U[25, 35]$, where each transplant center receives on average 30 pairs per time period). The percentage of altruists received was determined endogenously by the UNOS data, or set to 5% of the number of pairs in the Saidman case. We simulated a homogeneous environment: the transplant centers used the same strategy and had the same arrival rate distribution. Transplant centers were either *truth-telling*, revealing all their pairs to the clearinghouse, or *strategic*, first performing a maximum matching internally and then revealing the unmatched pairs to the clearinghouse. Runs for each parameter setting were performed at least 25 times.

Experimental Results. Figure 2 shows the median percentage increase in total number of matches from using our strategy-proof mechanisms (Algorithm 1 on Figures 2a and 2c, and the adapted IR-Efficient mechanism on Figures 2b and 2d) with truth-telling transplant centers compared to strategic ones. Error bars denote upper and lower quartiles. In general, as the number of transplant centers increases (shown on the x-axis), the relative gain stays flat or increases. Similarly, as the arrival rates increase (shown as different lines), relative gains increase. Intuitively, as either of these variables increases, the thickness of the reported pool available to the clearinghouse increases relative to the thickness of the internal pools of the transplant centers. As has been shown theoretically and empirically (Roth, Sönmez, and Ünver 2004; Ashlagi and Roth 2014; Dickerson, Procaccia, and Sandholm 2012b), thicker pools enable a larger *percentage* of pairs in the pool to be matched. Also, the efficient mechanism and the IR-efficient mechanism perform similarly.

Conclusions and Future Research

We proposed the first kidney exchange mechanism for multi-period dynamic settings that incentivizes truthful revelation of pairs and altruistic donors from transplant centers while guaranteeing efficiency, and long-term individual rationality. It is based on centers receiving credits when they reveal more pairs than they are expected to have per period and losing credits when they reveal fewer pairs than that. The credits are used by the clearinghouse to favor transplant centers in the global matching process in a specific way. The mechanism can also be adjusted to guarantee immediate individual rationality, though in this case it guarantees IR-efficiency rather than efficiency. This latter result circumvents impossi-

bility results in static kidney exchange (Roth, Sönmez, and Ünver 2007a). We showed experimentally that both mechanisms result in significant gains on real data from a large national kidney exchange, as well as on generated data.

Our mechanisms do not rely on the structure of the compatibility graph. In contrast, for example, theoretical results on the length of chains in the Saidman et al. model (Dickerson, Procaccia, and Sandholm 2012b) are essentially contradictory to those in a sparser model (Ashlagi et al. 2012). Our results hold in either model—and in any other.

We believe that there is great potential for credit-based mechanisms of the kind presented in this paper for kidney exchange and other barter markets—and possibly applications beyond barter. There are many interesting directions for future research. Most important is extending the model to include pairs and altruistic donors that remain in the pool for more than a single time period. Indeed, this was not an issue in most prior theoretical kidney exchange literature, as the focus was on static settings.

Other extensions include experimental investigation of how the gains in efficiency vary in the presence of larger and smaller transplant centers or when some of the centers are not fully rational (with some probability they deviate from their dominant, truthful strategy). We also plan to explore how robust our mechanism is to the expected arrival rates; this may also beget the design of refinements of the mechanism. Finally, we also plan to study time-discounted settings, although, in a steady-state kidney exchange our objective of maximizing the average number of transplant per period may be more appropriate since the real value of a human life should not change over time.

References

Abraham, D.; Blum, A.; and Sandholm, T. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 295–304.

Akbarpour, M.; Li, S.; and Gharan, S. O. 2014. Dynamic matching market design. In *Proceedings of the ACM Conference on Economics and Computation (EC)*.

Anshelevich, E.; Chhabra, M.; Das, S.; and Gerrior, M. 2013. On the social welfare of mechanisms for repeated batch matching. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 60–66.

Ashlagi, I., and Roth, A. E. 2014. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*. To appear.

Ashlagi, I.; Gamarnik, D.; Rees, M.; and Roth, A. E. 2012. The need for (long) chains in kidney exchange. NBER Working Paper No. 18202.

Ashlagi, I.; Fischer, F.; Kash, I. A.; and Procaccia, A. D. 2013. Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*.

Ashlagi, I.; Jaillet, P.; and Manshadi, V. H. 2013. Kidney exchange in dynamic sparse heterogenous pools. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 25–26.

Biró, P.; Manlove, D. F.; and Rizzi, R. 2009. Maximum weight cycle packing in directed graphs, with application to kidney exchange programs. *Discrete Mathematics, Algorithms and Applications* 1(04):499–517.

Blackwell, D.; Lucas, J.; and Clarke, T. 2014. Summary health statistics for US adults: National Health Interview Survey, 2012. *Vital and Health Statistics* 10(260):1–161.

Dickerson, J. P., and Sandholm, T. 2014. Multi-organ exchange: The whole is greater than the sum of its parts. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1412–1418.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2012a. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1340–1346.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2012b. Optimizing kidney exchange with transplant chains: Theory and reality. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 711–718.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2013. Failure-aware kidney exchange. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 323–340.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2014. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Leider, S., and Roth, A. E. 2010. Kidneys for sale: Who disapproves, and why? *American Journal of Transplantation* 10(5):1221–1227.

Li, J.; Liu, Y.; Huang, L.; and Tang, P. 2014. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Liu, Y.; Tang, P.; and Fang, W. 2014. Internally stable matchings and exchanges. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1433–1439.

Montgomery, R.; Gentry, S.; Marks, W. H.; Warren, D. S.; Hiller, J.; Houp, J.; Zachary, A. A.; Melancon, J. K.; Maley, W. R.; Rabb, H.; Simpkins, C.; and Segev, D. L. 2006. Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet* 368(9533):419–421.

Parkes, D. C. 2001. *Iterative combinatorial auctions: Achieving economic and computational efficiency*. Ph.D. Dissertation, University of Pennsylvania.

Rapaport, F. T. 1986. The case for a living emotionally related international kidney donor exchange registry. *Transplantation Proceedings* 18:5–9.

Rees, M.; Kopke, J.; Pelletier, R.; Segev, D.; Rutter, M.; Fabrega, A.; Rogers, J.; Pankewycz, O.; Hiller, J.; Roth, A.; Sandholm, T.; Ünver, U.; and Montgomery, R. 2009. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine* 360(11):1096–1101.

Roth, A.; Sönmez, T.; and Ünver, U. 2004. Kidney exchange. *Quarterly Journal of Economics* 119(2):457–488.

Roth, A.; Sönmez, T.; and Ünver, U. 2005. A kidney exchange clearinghouse in New England. *American Economic Review* 95(2):376–380.

Roth, A.; Sönmez, T.; and Ünver, M. U. 2007a. Notes on forming large markets from small ones: Participation incentives in multi-center kidney exchange. Unpublished.

Roth, A.; Sönmez, T.; and Ünver, U. 2007b. Efficient kidney exchange: Coincidence of wants in a market with compatibility-based preferences. *American Economic Review* 97:828–851.

Roth, A. 2007. Repugnance as a constraint on markets. *Journal of Economic Perspectives* 21(3):37–58.

Saidman, S. L.; Roth, A.; Sönmez, T.; Ünver, U.; and Delmonico, F. 2006. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation* 81(5):773–782.

Sönmez, T., and Ünver, M. U. 2011. Matching, allocation, and exchange of discrete resources. *Handbook of Social Economics* 1:781–852.

Sönmez, T., and Ünver, M. U. 2013. Market design for kidney exchange. *The Handbook of Market Design* 93–137.

Stewart, D.; Leishman, R.; Sleeman, E.; Monstello, C.; Lunsford, G.; Maghirang, J.; Sandholm, T.; Gentry, S.; Formica, R.; Friedewald, J.; and Andreoni, K. 2013. Tuning the OPTN's KPD optimization algorithm to incentivize centers to enter their easy-to-match pairs. In *American Transplant Congress (ATC)*. Talk abstract.

Toulis, P., and Parkes, D. C. 2011. A random graph model of kidney exchanges: efficiency, individual-rationality and incentives. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 323–332.

United Network for Organ Sharing (UNOS). http://www.unos.org/.

Woodle, S.; Daller, J.; Aeder, M.; Shapiro, R.; Sandholm, T.; Casingal, V.; Goldfarb, D.; Lewis, R.; Goebel, J.; and Siegler, M. 2010. Ethical considerations for participation of nondirected living donors in kidney exchange programs. *American Journal of Transplantation* 10:1460–1467.