Sequential Auctions for the Allocation of Resources with Complementarities

Craig Boutilier
Dept. of Computer Science
University of British Columbia
Vancouver, BC V6T 1Z4
cebly@cs.ubc.ca

Moisés Goldszmidt and Bikash Sabata
SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025
{*moises,sabata*}*@erg.sri.com*

Abstract

Market-based mechanisms such as auctions are being studied as an appropriate means for resource allocation in distributed and multiagent decision problems. When agents value resources in combination rather than in isolation, one generally relies on *combinatorial auctions* where agents bid for resource bundles, or *simultaneous auctions* for all resources. We develop a different model, where agents bid for required resources *sequentially*. This model has the advantage that it can be applied in settings where combinatorial and simultaneous models are infeasible (e.g., when resources are made available at different points in time by different parties), as well as certain benefits in settings where combinatorial models are applicable. We develop a dynamic programming model for agents to compute *bidding policies* based on estimated distributions over prices. We also describe how these distributions are updated to provide a learning model for bidding behavior.

1 Introduction

A great deal of attention has been paid to the development of appropriate models and protocols for the interaction of agents in distributed and multiagent systems (MASs). Often agents need access to specific resources to pursue their objectives, but the needs of one agent may conflict with those of another. A number of market-based approaches have been proposed as a means to deal with the resource allocation and related problems in MASs [5, 21].

Of particular interest are *auction mechanisms*, where each agent bids for a resource according to some protocol, and the allocation and price for the resource are determined by specific rules [13]. Auctions have a number of desirable properties as a means for coordinating activities, including minimizing the communication between agents and, in some cases, guaranteeing Pareto efficient outcomes [13, 21].

An agent often requires several resources before pursuing a particular course of action. Obtaining one resource without another—for example, being allocated trucks without fuel or drivers, or processing time on a machine without skilled labor to operate it—makes that resource worthless. When resources exhibit such *complementarities*, it is unknown whether simple selling mechanisms can lead to efficient outcomes [21, 1]. Moreover, groups of resources are often *substitutable*: obtaining the bundle needed to pursue one course of action can lower the value of obtaining another, or render it worthless. For instance, once trucks and drivers are obtained for transporting material in an optimal fashion, helicopters and pilots lose any value they may have had.

Two methods for dealing with complementarities have been studied: simultaneous auctions for multiple goods [1, 17]; and *combinatorial auctions* in which agents submit bids for *resource bundles* [16, 18, 19, 9, 21]. Specific models sometimes deal (possibly implicitly) with substitution effects, and sometimes not. In this paper, we explore a model that combines features of both simultaneous and combinatorial auctions. Our *sequential auctions model* supposes that the set of resources of interest are auctioned in sequence. Agents bid for resources in a specific, known order, and can choose how much (and whether) to bid for a resource depending on past successes, failures, prices, and so on.

Our model has several advantages over standard combinatorial and simultaneous models. The chief benefit of such a model is that is can be applied in situations where combinatorial and simultaneous models cannot. Specifically, when multiple sellers offer various resources of interest, or when the resources are sold at different points in time, one does not have the luxury of setting up either combinatorial or simultaneous auctions. As such, our model is suitable for agents who are required to interact with multiple suppliers over time. Even in settings where combinatorial models can be applied, there may be some advantages to using a sequential model. Unlike combinatorial models, our model relieves the (computational) burden of determining a final allocation from the seller, effectively distributing computation among the buyers (as in the simultaneous case); note that determining an optimal allocation that maximizes the seller's revenue is NP-hard [18]. Our sequential model also has the advantage that buyers are not required to reveal information about their valuations for specific resource bundles that they do not obtain. Furthermore, it has greater flexibility in that agents can enter and leave the market without forcing recomputation of entire allocations. In contrast to simultaneous models, agents in the sequential model may lessen their exposure. If an agent does not obtain a certain resource early in the sequence, it need not expose itself by bidding on complementary resources occurring later in the sequence. Agents are typically bidding in a state of greater knowledge in the sequential model, at least in later stages; however, in earlier stages agents may have lesser information than they would in a simultaneous model.

One difficulty that arises in the sequential model is how an agent computes bids for individual resources (the same difficulty arises in simultaneous models). An agent has a valuation for a particular resource bundle $b = \{r_1, \cdots, r_k\}$, but has no independent assignment of value to the individual resources.[1] While auction theory can tell us how an agent should bid as a function of its valuation of resource r_i for specific auction mechanisms, in our setting no such valuation exists. If b is worth $v(b)$, how is an agent to "distribute the value" among the resources r_i in order to compute bids?

In this paper, we develop a dynamic programming algorithm for doing just this. We assume that each agent has a probabilistic estimate of the size of the maximum bids for each resource (excluding its own). It can then compute a *bidding policy* that maximizes its expected utility, and apply this policy as dictated by its initial endowment. Bids for resources early in the sequence are computed as a function of the odds of being able to obtain their complements and substitutes, while bids for later resources are conditioned on the resources obtained early in the sequence.

We also interested in adaptive bidding behavior, and to this end investigate a *repeated sequential auction model* in which agents repeatedly bid for the same resources over time. We consider the problem of estimating the probability distributions over maximal bids in this repeated scenario. If agents persistently find themselves requiring resources to pursue their aims, we want them to learn which resources they will be able to obtain and which they will not. This is related to recently proposed learning models for auctions [11, 12], though our focus is on learning prices and its effect on the valuation of individual resources in bundles.

The problem we study is part of a more general research program designed to study the impact of specific resource allocation schemes on the solution of sequential multiagent decision problems. We motivate the problem studied here as follows. We suppose that a number of agents have certain tasks and objectives to pursue, and for any objective there may exist a number of potential courses of action that are more or less suitable. For instance, an agent may construct a policy for a Markov decision process [15, 2], from which it can determine the value of various courses of action, their likelihood of success, and so on. Any specific course of action will require certain resources, say, bundle b^k, whose value can be determined as a function of the expected value of that course of action (and the expected value of alternative courses of action). As such, we suppose each bundle b^k has an associated value $v(b^k)$ and that the agent will use only one bundle (the one associated with the highest-valued course of action among those bundles it possesses). It is from these valuations that the agent must determine its bidding policy for individual resources. This is the problem considered here.

Ultimately, the decision problem we hope to study is far more complex. Determining appropriate courses of action will depend on perceived probability of obtaining requisite resources, uncertainty in that course of action, alternatives available and so on. We envision very sophisticated reasoning emerging regarding the interaction bidding behavior and "base-level" action choice (in the MDP), such as taking a few critical steps along a specific course of action before deciding to enter the market for the corresponding resources (e.g., perhaps because this policy is fraught with uncertainty). We also foresee interesting interactions with other coordination and communication protocols.

In Section 2 we describe the basic sequential bidding model. We note a number of dimensions along which our basic model can vary, though we will focus only on specific instantiations of the model for expository reasons. We describe our dynamic programming model for constructing bidding policies in Section 3. We also describe the motivation for using the specific model proposed here instead of using explicit equilibrium computation. We discuss *repeated* sequential auctions in Section 4, focusing on the problem of highest-bid estimation. In Section 5 we describe some preliminary experimental results, and conclude in Section 6 with discussion of future research directions.

While bidding strategies for sequential auctions would seem to be an issue worthy of study, there appears to have been little research focussed on this issue. What work exists (see, e.g., [8, 10]) tends to focus the seller's point of view—for example, will simultaneous or sequential sales maximize revenue—and does not address the types of complementarities we consider here. Generally, existing work assumes that single items are of interest to the buyer.

2 Basic Model

We assume we have a finite collection of agents, all of whom require resources from a pool of n resources $R = \{r_1, \cdots, r_n\}$. We denote by R^t the subset $\{r_1, \cdots, r_t\}$, $t \leq n$, with $R^0 = \emptyset$ by convention. We describe the quantities relevant to a specific agent a below, assuming that these quantities are defined for each agent. Agent a can use exactly one bundle $b^i = \{r^i_1, \cdots, r^i_{|b^i|}\}$ of resources from a set of k *possible bundles*: $B = \{b^1, \cdots, b^k\}$. We denote by $U(a) = \cup B$ the set of *useful resources* for our agent.

Agent a has a positive valuation $v(b^i)$ for each resource bundle $b^i \in B$. Suppose the *holdings* of a, $H(a) \subseteq U(a)$, are those resources it is able to obtain. The value of these holdings is given by $v(H(a)) = \max\{v(b^i) : b^i \subseteq H(a)\}$; that is, the agent will be able to use the resource bundle with maximal value from among those it holds in entirety, with the others going unused. This is consistent with our interpretation given in Section 1 where resource bundles correspond to alternative plans for achieving some objective (though other value combinators can be accommodated).

The resources will be auctioned sequentially in a commonly known order: without loss of generality, we assume that this ordering is $r_1, r_2, \cdots, r_n$ We use A_i to denote the auction for r_i. We refer to the sequence of auctions $A_1, A_2, \cdots, A_n$ as a *round* of auctions. There may be a single round, some (definite or indefinite) finite numbers of rounds, or an infinite number of rounds.

Supposing for the moment only one round, we assume that agent a is given an initial endowment e which it can use to obtain resources. At the end of the round, a has holdings $H(a)$

[1] In fact, we will assume that several bundles can be valued, with possible overlap. This accounts for possible substitution effects.

and d dollars remaining from its endowment.[2] We assume that the utility of being in such a state at the end of the round is given by $v(H(a)) + f(d)$, where f is some function attaching utility to the unused portion of the endowment. Other utility functions could be considered within this framework.

There are a wide range of options one could consider when instantiating this framework. We define a specific model here, but list the options that could be explored. We develop the algorithms in this paper for the specific model, but where appropriate, indicate how they should be modified for other design choices. The main design choices are:

- What auction mechanism is used for the auctions A_i?
- What rules are instituted for reselling or speculation?
- What information is revealed to the agents? When?
- What information do agents have when a round begins?

We assume that the individual auctions will be first-price, sealed-bid—each agent will provide a single bid and the highest bidder will be awarded the resource for the price bid. We adopt this model because of the ease with which it fits with our approach to bid computation; however, we believe our model could be adapted for other auction protocols. We also assume that bids are discrete (integer-valued); but we do describe the appropriate amendments to deal with continuous bids. Agents, once obtaining a resource, cannot resell that resource to another agent. This, of course, means that an agent may obtain one resource r_i, but later be unable to obtain a complementary resource r_{i+k}, essentially being "stuck" with a useless resource r_i. We do this primarily for simplicity, though in certain settings this assumption may be realistic. We are currently exploring more sophisticated models where agents can "put back" resources for re-auctioning, or possibly resell resources directly to other agents.

Each agent is told the winning price at the end of the each auction (and whether it was the winner). We could suppose that no information (other than winning or losing) is provided, that the distribution over bids is announced, or that the bids of specific individuals are made public; our assumption seems compatible with the first-price, sealed-bid model.

Finally, agent a believes that the highest bid that will be made for resource r_i, excluding any bid a might make, is drawn from some unknown distribution $\overline{Pr}^i$. Because bids are integer-valued, this unknown distribution is a multinomial over a non-negative, bounded range of integers.[3] To represent a's uncertainty over the parameters of this distribution, we assume a has a prior probability distribution $\Pr^i$ over the space of bid distributions. Agent a models $\Pr^i$ as a Dirichlet distribution with parameters $\beta_0^i, \cdots, \beta_{m_i}^i$ [6], where m_i is the (estimated) maximum possible bid for r_i. We elaborate on this probability model in Sections 3 and 4.

We make two remarks on this model. First, if the space of possible bids is continuous, a suitable continuous PDF (e.g., Gaussian) could be used to model bid distributions and the uncertainty about the parameters of this PDF. More questionable is the implicit assumption that bids for different resources are uncorrelated. By having distributions $\Pr^i$ rather than a joint distribution over *all* bids, agent a is reasoning as if the bids for different resources are independent. When resources exhibit complementarities, this is unlikely to be the case. For instance, if someone bids up the price of some resource r_i (e.g., trucks), they may subsequently bid up the price of complementary resource r_j (e.g., fuel or drivers). If agent a does not admit a model that can capture such correlations, it may make poor bids for certain resources. Again, we make this assumption primarily for ease of exposition. Admitting correlations does not fundamentally change the nature of the algorithms to follow, though it does raise interesting modeling and computational issues (see Section 4).

3 Computing Bids by Dynamic Programming

In this section we focus on the decisions facing an agent in a single round of auctions. A key decision facing an agent at the start of a round is how much to bid for each resource that makes up part of a useful bundle b_i. In standard single item auctions (e.g., first/second-price, sealed bid) rational agents with an assessment of the valuations of other agents can compute bids with maximum expected utility [13]. For example, in first-price, sealed bid auctions, an agent should bid a some amount below its true valuation, where this amount is given by its beliefs about the valuations of others.

Unfortunately, the same reasoning cannot be applied to our sequential setting, since individual resources cannot be assessed a well-defined valuation. For instance, if bundle $b^i = \{r_1^i, r_2^i\}$ has valuation $v(b^i)$, how should agent a apportion this value over the two resources? Intuitively, if there is a greater demand for r_1^i, a larger "portion" of the value should be allotted for bidding in the first auction rather than the second. If the agent fails to obtain r_1^i, the value of r_2^i goes to zero (ignoring other bundles). In contrast, should a obtain r_1^i, it is likely that the agent should offer a substantial bid for r_2^i, approaching the valuation $v(b^i)$, since the price paid for r_1^i is essentially a "sunk cost." Of course, if the agent expects this high price to be required, it should probably not have bid for r_1^i in the first place. Finally, the interaction with other bundles requires the agent to reason about the relative likelihood of obtaining any specific bundle for an acceptable price, and to focus attention on the most promising bundles.

3.1 The Dynamic Programming Model

These considerations suggest that the process by which an agent computes bids should not be one of assigning value to individual resources, but rather one of constructing a *bidding policy* by which its bid for any resource is conditioned on the outcome of events earlier in the round. The sequential nature of the bidding process means that it can be viewed as a standard sequential decision problem under uncertainty. Specifically, the problem faced by agent a can be modeled as a fully observable Markov decision process (MDP) [15, 2]. The computation of an optimal *bidding policy* can be implemented using a standard stochastic dynamic programming al-

[2] If speculation or reselling is allowed, there is the possibility that $d > e$, depending on the interaction protocols we allow. We will mention this possibility below, but we will examine only protocols that disallow it.

[3] We assume that a bound can be placed on the highest bid.

gorithm such as value iteration.

We emphasize that agents are computing optimal *bids*, not true valuations for individual resources. Thus issues involving revelation of truthful values for resources are not directly relevant (but see Section 4 on multiple rounds).

We assume the decision problem is broken into $n + 1$ stages, n stages at which bidding decisions must be made, and a terminal stage at the end of the round. We use a time index $0 \leq t \leq n$ to refer to stages—time t refers to the point at which auction A_{t+1} for r_{t+1} is about to begin. The *state* of the decision problem for a specific agent a at time t is given by two variables: $H^t(a) \subseteq R^t$, the subset of resources R^t held by agent a; and d^t, the dollar amount (unspent endowment) available for future bidding. We write $\langle h, d\rangle^t$ to denote the state of a's decision problem at time t. Note that although we could distinguish the state further according to which agents obtained which resources, these distinctions are not relevant to the decision facing a.[4]

The dynamics of the decision process can be characterized by a's estimated transition distributions. Specifically, assuming that prices are drawn independently from the stationary distributions $\overline{Pr}^i$, agent a can predict the effect of any action (bid) z available to it. If agent a is in state $\langle h, d\rangle^t$ at stage t, it can bid for r_{t+1} with any amount $0 \leq z \leq d^t$ (for convenience we use a bid of 0 to denote nonparticipation). Letting w denote the highest bid of other agents, if a bids z at time t, it will transition to state $\langle h \cup \{r_{t+1}\}, d - z\rangle^{t+1}$ with probability $\overline{\Pr}^{t+1}(w < z)$ and to $\langle h, d\rangle^{t+1}$ with $\overline{\Pr}^{t+1}(w \geq z)$.[5]

This does not form an MDP *per se*, since a may be uncertain about the true distribution $\overline{Pr}^{t+1}$, having only a Dirichlet distribution $\langle \beta_1^{t+1}, \cdots, \beta_{m_{t+1}}^{t+1}\rangle$ over the possible parameters of $\overline{Pr}^{t+1}$. However, the expectation that the highest bid is w is given by the relative weight of parameter β_w^{t+1}; thus,

$$\Pr^{t+1}(w < z) = \frac{\sum_{i=0}^{z-1} \beta_i^{t+1}}{\sum_{i=0}^{m_{t+1}} \beta_{i+1}^t}$$

While the observation of the true winning bid can cause this estimated probability to change (properly making this a partially observable MDP), the change cannot impact *future* transition probability estimates or decisions: we have assumed that the high bid probabilities are independent. Thus, treating this as a fully observable MDP with transition probabilities given by *expected* transition probabilities is sound.

The final piece of the MDP is a reward function q. We simply associate a reward of zero with all states at stages 0 through $n-1$, and assign reward $v(h) + f(d)$ to every terminal state $\langle h, d\rangle^n$. A *bidding policy* π is a mapping from states into actions: for each legal state $\langle h, d\rangle^t$, $\pi(\langle h, d\rangle^t) = z$ means that a will bid z for resource r_{t+1}. The *value* $V^\pi(\langle h, d\rangle^t)$ of policy π at any state $\langle h, d\rangle^t$ is the expected reward $E_\pi(q(\langle H(a), d\rangle^n) | \langle h, d\rangle^t)$ obtained by executing π. The expected value of π given the agent's initial state $\langle \emptyset, e\rangle^t$ is simply $V^\pi(\langle \emptyset, e\rangle^t)$. An *optimal bidding policy* is any π that has maximal expected reward at every state.

We compute the optimal policy using value iteration [15], defining the value of states at stage t using the value of states at stage $t+1$. Specifically, we set

$$V(\langle h, d\rangle^n) = v(h) + f(d)$$

and define, for each $t < n$:

$$\begin{aligned} Q(\langle h, d\rangle^t, z) &= \Pr^{t+1}(w < z) \cdot V(\langle h \cup \{r_t\}, d - z\rangle^{t+1}) \\ &\quad + \Pr^{t+1}(w \geq z) \cdot V(\langle h, d\rangle^{t+1}) \\ V(\langle h, d\rangle^t) &= \max_{z \leq d} Q(\langle h, d\rangle^t, z) \\ \pi(\langle h, d\rangle^t) &= \arg\max_{z \leq d} Q(\langle h, d\rangle^t, z) \end{aligned}$$

Given that V is defined for all stage $t+1$ states, $Q(\langle h, d\rangle^t, z)$ denotes the value of bidding z at state $\langle h, d\rangle^t$ and acting optimally thereafter. $V(\langle h, d\rangle^t)$ denotes the optimal value at state $\langle h, d\rangle^t$, while $\pi(\langle h, d\rangle^t)$ is the optimal bid.

Implementing value iteration requires that we enumerate, for each t, all possible stage t states and compute the consequences of every feasible action at that state. This can require substantial computational effort. While linear in the state and action spaces (and in the number of stages n), the state and action spaces themselves are potentially quite large. The number of possible states at stage t could potentially consist of any subset of resources R^t together with any monetary component. The action set at a state with monetary component d has size $d+1$. Fortunately, we can manage some of this complexity using the following observations: first, a never needs to bid for any resource outside the useful set $U(a)$, so its state space (at stage t) is restricted to subsets of $U^t(a)$; and second, if a resource r_t requires a complementary resource $r_{t'}, t' < t$, (that is, all bundles containing r_t also contain $r_{t'}$), then we need never consider a state where a has r_t but not $r_{t'}$.[6] Reducing the impact of the number of possible bids is more difficult. We can certainly restrict the state and action space to dollar values no greater than a's initial endowment e. If the PDF is well-behaved (e.g., concave), pruning is possible: e.g., once the expected value of a larger bids starts to decrease, search for a maximizing bid can be halted.[7]

This dynamic programming model deals with the complementarities and substitutability inherent in our resource model; no special devices are required. Furthermore, it automatically deals with issues such as uncertainty, dynamic valuation, "sunk costs," and so on. Given stationary, uncorrelated bid distributions, the computed policy is optimal.

3.2 Extensions of the Model

While the assumptions underlying our (single-round) model are often reasonable, there are two assumptions that must be relaxed in certain settings: the requirement for discrete bids

[4] This is true under the current assumptions, but may not be under different models; see below.

[5] For expository purposes, the model assumes ties are won. Several rules can be used for ties; none complicate the analysis.

[6] This reasoning extends to arbitrary subset complementarities.

[7] If we move to a continuous action space, the value function representation and maximization problems may become easier to manage for certain well-behaved classes of probability distributions and utility functions (see Section 3.2 and [3]).

and the prohibition of reselling or returning resources for resale. We are currently exploring these relaxations.

Continuous bidding models are important for computational reasons. Though money is not truly continuous, the increments that need to be considered generally render explicit value calculations for all discrete bids infeasible. Continuous function maximization and manipulation techniques are often considerably more efficient that discrete enumeration, and approximately optimal "integer" bids can usually be extracted. We are currently exploring specific continuous models, specifically using parameterized bid distributions (such as Gaussian and uniform distributions) and linear utility functions (as described above). The key difficulty in extending value iteration is determining an appropriate value function representation. While the maximization problem (over bids) for a specific state is not difficult, we must represent V^t as a function of the continuous state space. This function is linear (in d) at all states where the remaining endowment d is greater than the maximal worthwhile bid. But a different function representation is needed for states with endowment less than the best bid. We are currently exploring a value function representation with piecewise, continuous representations of V for each (discrete) set of holdings $H(a)$ [3].

Reselling may be appropriate in many settings and can allow agents to bid more aggressively with less risk. We are currently developing a simple model in which agents are allowed, at the end of a round, to "put back" resources for re-auction that are not needed (e.g., are not part of the agent's max-valued complete bundle).[8] Several difficulties arise in this setting, including the fact that agents may need to estimate the probability that an unobtained resource may be returned for re-auction.

3.3 Equilibrium Computation

The model described above does not allow for strategic reasoning on the part of the bidding agent. The agent takes the expected prices as given and does not attempt to compute the impact of its bids on the behavior of other agents, how they might estimate its behavior and respond, and so on; that is, no form of equilibrium is computed. Standard models in auction theory generally prescribe bidding strategies that are in Bayes-Nash equilibrium: when each agent has beliefs about the *types* of other agents (i.e., how each agent values the good for sale), and these beliefs are common knowledge, then the agents' bidding policies can be prescribed so that no agent has incentive to change its policy.[9] This, for instance, is the basis for prescribing the well-known strategies for bidding in first- and second-price auctions [20].

Our approach is much more "myopic." There are several reasons for adopting such a model rather that a full Bayes-Nash equilibrium model. First, equilibrium computation is often infeasible, especially in a nontrivial sequential, multi-resource setting like ours. Second, the information required on the part of each agent, namely a distribution over the possible types of other agents, is incredibly complex—an agent type in this setting is its set of valuations for *all* resource bundles, making the space of types unmanagable. Finally, the common knowledge assumptions usually required for equilibrium analysis are unlikely to hold in this setting.

We expect that the MDP model described here could be extended to allow for equilibrium computation. Rather than do this, we consider an alternative, adaptive model for bidding in which agents will adjust their estimates of prices—hence their bidding policies—over time. Implicitly, agents learn how others value different resources, and hopefully some type of "equilibrium" will emerge. We turn our attention to this process of adaptation.

4 Repeated Auctions and Value Estimation

In certain domains, agents will repeatedly need resources drawn from some pool to pursue ongoing objectives. We model this by assuming that the same resource collection is auctioned repeatedly in rounds. While agents could compute a single bidding policy and use it at every round, we would like agents to use the behavior they've observed at earlier rounds to update their policies. Specifically, observed winning prices for resource auctions A_i in the past can be used by an agent to update its estimate of the true distribution $\Pr^i$ of high bids for r_i. Its bidding strategy at the next round can be based on the updated distributions.

If each agent updates its bidding policy based on past price observations, the prices observed at earlier rounds may not be reflective of the prices that will obtain at the next round. This means that the agents are learning based on observations drawn from a nonstationary distribution. This setting is common in game theory, where agents react to each others past behavior. Myopic learning models such as *fictitious play* [4] (designed to learn strategy profiles) can be shown to converge to a stationary distribution despite the initial nonstationarity. This type of learning model has been applied to repeated (single-item) auctions and shown to converge [11]. Our model is based on similar intuitions—namely, that learning about prices will eventually converge to a steady state. Hu and Wellman [12] also develop a related model for price learning in a somewhat different context.

The advantage of a learning model is that agents can come to learn which resources they can realistically obtain and focus their bidding on those. If agents A and B have similar endowments and both equally value having either r_1 or r_2, they may learn over time not to compete for r_1 and r_2; instead they may learn to anticipate (implicitly, through pricing) each other's strategy and (implicitly) coordinate their activities, with one pursuing r_1 and the other r_2. If one agent has a greater endowment than another (e.g., it may have higher priority objectives in a distributed planning environment), the poorer agent should learn that it can't compete and focus on less contentious (and perhaps less valued) resources. Another important feature of learning models is that they can be used to overcome biased or weak prior assessments.

Given the form of the probabilistic model described in Section 3, an agent can update its estimate of a bid distribution rather easily. Suppose agent a has parameters $\langle \beta_1^t, \cdots, \beta_{m_t}^t \rangle$

[8] More complicated models that allow agents to put back resources during the round or resell directly are also possible.

[9] We use *type* here in the sense used in game theory for games with incomplete information [14].

that characterize its distribution $\Pr^t$ over the true distribution $\overline{\Pr}^t$ of high bids for resource r_t. After auction A_t the winning bid w is announced to each agent.[10] If a fails to win the resource, it should update these Dirichlet parameters by setting β_w^t to $\beta_w^t + 1$; at the next round, its estimate that the highest bid will be w is thus increased. If a wins resource r_t for price z, the only information it gets about the highest bid (excluding its own) is that it is less than z. The Dirichlet parameters can then be updated with an algorithm such as EM [7]. Roughly, the expectation step computes an update of the parameters of the Dirichlet using current estimates to distribute the observation over the parameters $\beta_1^t, \cdots, \beta_{z-1}^t$: each β_j^t ($j < z$) is increased by $\beta_j^t / \sum_{i=0}^{z-1} \beta_i^t$. The maximization step corresponds to the actual update followed by the substitution of these parameters in $\Pr$. Whereas the EM algorithm requires an iteration of these two steps until convergence, we performed this iteration about 10 times.[11]

In the specific probability model developed here, agents cannot profitably use this updated estimate during the current round. Because prices are assumed independent, learning about one price cannot influence an agent's bidding strategy for other resources.[12] Thus the agent continues to implement the bidding policy computed at the start of the round. The updated bid distributions are used prior to the start of the next round of auctions to compute an new bidding policy.

As mentioned above, the price-independence assumption may be unrealistic. If prices are correlated, the observed price of a resource can impact the estimated price of another resource that will be available later in the round. Agents in this case should revise their bidding policies to reflect this information. Two approaches can be used to deal with correlations. First, agents can simply recompute their bidding policies during a round based on earlier outcomes. An alternative is to model this directly within the MDP itself: this entails making the MDP partially observable, which can cause computational difficulties.

One thing we do not consider is agents acting strategically within a round to influence prices at subsequent rounds. Agents are reasoning "myopically" within a specific round. By formulating multi-round behavior as a sequential problem, we could have agents attempting to manipulate prices for future gain. Our current model does not allow this.

5 Results

We now describe the results of applying this model to some simple resource allocation problems. These illustrate interesting qualitative behavior such as adaptation and coordination. We also explain why such behavior arises. In all runs, multiple rounds are considered and remaining endowment d is valued at $0.5d$ ($\alpha = 0.5$). Agent priors have slightly increasing weights on higher bids.[13]

The first series of examples illustrates bidding behavior in allocation problems with specific parameter settings.

Example There are two agents whose optimal bundles are disjoint: a_1 requires $b_1^1 = \{r_1, r_3\}$ (value 20) or $b_2^1 = \{r_4, r_5, r_6\}$ (value 30), while a_2 requires $b_1^2 = \{r_2, r_3\}$ (value 20) or $b_2^2 = \{r_7, r_8, r_9\}$ (value 30). Initially, both agents focus on the smaller (and lower-valued) bundles. At the first round, a_1 obtains b_1^1, while a_2 gets "stuck" with r_2 (a_1 outbid it for r_3). The next round sees a_2 bid less for r_2, and more for r_3 (outbidding a_1). Since it obtains b_1^2, it does not attempt to bid for b_2^2. But without b_1^1, and its estimated prices for resources in b_2^1 lowered, a_1 now bids for and gets b_2^1 (its optimal bundle). Up to the 14th round, one of the agents gets its best bundle and the other its worst. At the 14th round, each gets its best bundle, and after the 16th round, the *socially optimal* allocation (the one with maximal total bundle value) is reached each time: the agents (more or less) "realize" that they need not compete. The agents do "hedge their bets" and still keep bidding for resources r_1, r_2 and r_3. They also offer fairly high bids for the nonconflicting resources, though these bids are reduced over time.

This first example shows that optimal allocations will emerge when agents are not in direct competition. It also illustrates general behavioral phenomenon that occur in almost all examples. (1) Agents tend to bid more aggressively (initially) for resources in bundles with smaller size, since the odds of getting all resources in a larger bundle are lower. (2) Agents tend to bid more aggressively for resources that occur later in the sequence. Once an agent obtains all resources in a bundle but one, the last resource is very valuable (for example, in round 16 above, a_1 obtains b_2^1 by paying 1 for r_4 and r_5, and 27 for r_6. (3) Agents tend to initially offer high bids for certain resources, and gradually lower their bids over time (realizing slowly that there is no competition). For example, a_1 reduces its bid for r_6 to 26 only at round 36. This is a consequence of the simple priors and belief update rules we use, and the lack of information it obtains when it wins the resource consistently: it is not told what the next highest bid is (it is zero), and can only conclude that it was less than 27, making belief update slow. The equivalent sample size of our priors also makes adjustment somewhat slow. Domain-specific (more accurate) priors, and the use of exponential decay (or finite histories) in price-estimation would alleviate much of this slowness of response.

Example There are 25 resources and five agents with four bundles each (with an average of four resources per bundle). There exists an allocation of five disjoint bundles, one to each agent. For each agent three of the resources occur only in its bundles, so the agents are competing for only 10 of the 25 resources. The socially optimal allocation has value 100. Over fifty rounds, the agents generally find very good (but not optimal) allocations. Figure 1 shows the value of the allocations obtained at each round,

[10] Our model can accommodate both more (e.g., the bids of all agents) and less (e.g., only whether an agent won or lost) revealed information about the auction outcome rather easily.

[11] Preliminary experiments showed this sufficient.

[12] With correlated prices, an agent could attempt to provide misleading information about its valuation of one resource in order to secure a later resource at a cheaper price. This type of deception, studied for identical item auctions in [10], cannot arise within a single round in our current model, even if strategic reasoning is used.

[13] More realistic priors could reflect perceived demand.

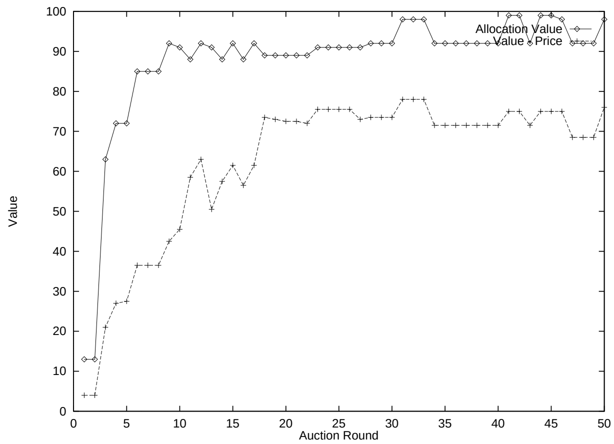

Figure 1: Behavior over 50 Auction Rounds: 5 agents with disjoint bundles (optimal allocation has value 100).

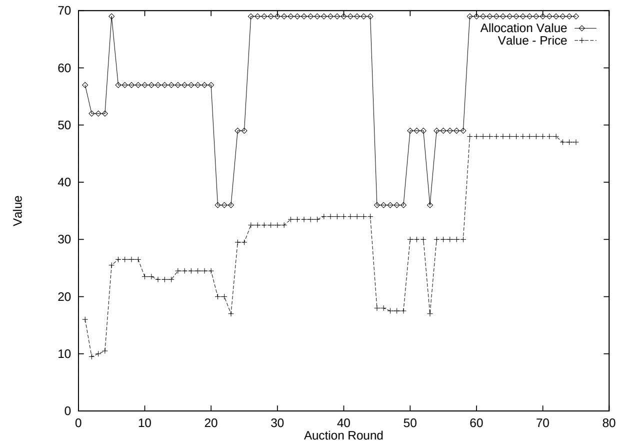

Figure 2: Sample Behavior over 75 Auction Rounds: 5 agents (optimal allocation has value 69)

as well as the collective "surplus" (total value minus α-adjusted prices paid). The agents quickly find good allocations (by the ninth round, no allocation has value less than 88), and also learn to pay less for the resources.

Example An interesting phenomenon emerges in a two-agent example of [21] that has no price equilibrium: assume resources r_1, r_2, with a_1 valuing bundle $\{r_1, r_2\}$ at 6, and a_2 valuing either of r_1, r_2 at 4. The agents have equal endowments. Though there is no price equilibrium, in our adaptive protocol a_2 wins one of its bundles much more frequently than a_1. It bids for r_1, and if it wins it need not bid for r_2; if it loses it can outbid a_1 for r_2 (since a_1 has paid for r_1). a_2 experiments with r_1 and wins it occasionally. a_2 gradually lowers its bid for r_2 and, since it does not model correlations in prices, occasionally loses r_2, allowing a_1 to get both r_1 and r_2. When this occurs, a_2 will quickly raise its bids and win one of the resources again. By modeling price correlations, or estimating the requirements of a_1, agent a_2 *could* guarantee that it obtains one of its resources (see Section 6).

Example We have 3 resources and 2 agents, each valuing $\{r_1, r_2\}$ at 10 and r_3 at 5, but differing in initial endowment: a_1 begins with 6, a_2 with 8. Initially, a_1 gets the first (higher-valued) bundle (at prices 2 and 5) and a_2 the second (at price 3). By the fourth round, a_2 realizes that it can win r_1 with bids of 3 and 5. It spends 8 on $\{r_1, r_2\}$, leaving a_1 to bid 4 for r_3. These prices persist, with a_2 not bidding on r_3 and a_2 eventually not bidding on r_1 or r_2. This illustrates that agents with larger endowments (or less relative value for money compared to bundles) have greater odds of obtaining their most important bundles, leaving "poorer" agents to get what is left.[14]

[14] This last property is useful for teams if agents with higher priority objectives are given larger endowments.

We also studied the bidding behavior on randomly generated allocation problems. Here we describe two sets of experiments. In problem set PS1, five allocation problems were randomly generated with the following characteristics: four agents are competing for 12 resources with an initial endowment of 30 each; each agent has a random number of needed bundles (normally distributed with mean 4 and s.d. 1); each bundle contains a random number of resources (normally distributed with mean 3 and s.d. 1, where the resources are themselves drawn uniformly from the set of 12); and the value of each bundle is random (normally distributed with mean 16 and s.d. 3). Problem set PS2 is identical except there are five agents and the mean number of resources per bundles is 4: hence problems in PS2 are more constrained, with more competition among the agents.

Typical behavior for one trial from PS2 (the more constrained problem set) is shown in Figure 2, which plots the the value of the allocations obtained at each round, as well as the collective surplus. The agents find good allocations in this problem, reaching the (socially) optimal allocation (with value 69) at many of the rounds. On average, over the 75 rounds, the allocation obtained has value 59 (85% of optimal). Note that once the agents "find" a good allocation, they may not stick with it—generally such allocations are not in equilibrium in the sequential game induced by a round of auctions. At the very least, agents have a tendency to attempt to lower the prices they bid after consistently winning a good, due to the lack of information about what other agents bid and how they update their beliefs (as mentioned above). This itself can cause some instability. The greater cause of instability however is the fact that a socially optimal allocation does not generally make self-interested agents happy.

Other trials illustrate similar qualitative behavior. When comparing PS1 (the less constrained problem set) to PS2 (the more constrained), we find that the allocations in PS1 have value that is, on average, within 87% of the optimal, while with PS2, allocations are within 80% of optimal. This suggests that for less constrained problems, sequential auctions among self-interested agents can lead to allocations with higher social welfare value. Given that agents "discover" many different allocations, one might view sequential auctions as a heuristic search mechanism for combinatorial auctions.[15] However, we emphasize that the main goal of our model is to compute bidding policies when combinatorial and simultaneous auctions are not possible.

6 Concluding Remarks

We have described a model for sequential auctioning of resources that exhibit complementarities for different agents and described a dynamic programming algorithm for the computation of optimal bidding policies. We have also illustrated how price learning can be used to allow agents to adapt there bidding policies to those of other agents. The sequential model can be applied in settings where combinatorial and simultaneous models are infeasible (e.g., when agents enter or leave markets over time, or when agents require resources from multiple sellers). Preliminary results are encouraging and suggest that desirable behavior often emerges.

We have suggested several possible extensions of the model, some of which we are currently exploring. These include developing continuous bidding models, models with reselling/return, incorporating correlated bid distributions and exploring the interactions between decision theoretic planning and bidding for the resources needed to implement plans and policies.

There are several more immediate directions we hope to pursue. One is the investigation of models where prices are estimated with greater weight placed on more recent prices. Along with correlated price distributions, the use of limited "opponent" models may be helpful: by identifying which agents tend to need which resources, a bidder can make more informed decisions. Additional revealed information about specific auctions (such as who bid what amount) could also lead to more informed decisions. This information may be appropriate in team situations, where distributed decision makers are not directly in competition.

Apart from such myopic mechanisms, we would also like to develop a Bayes-Nash equilibrium formulation of the sequential model, and study the extent to which myopic models like our simple learning scheme approximate it. The conditions under which our model converges to interesting allocations (socially optimal allocations, equilibria, etc.) is also worthy of exploration. Other avenues to be considered are the development of different auction ordering heuristics to maximize social welfare, seller's revenue or other objective criteria; and the development of generalization methods to speed up dynamic programming. We are also integrating the sequential auction model for resource allocation into the general planning context described in Section 1.

Acknowledgments Craig Boutilier and Moisés Goldszmidt acknowledge partial support from the DARPA Co-ABS program (through Stanford University contract F30602-98-C-0214). Bikash Sabata was funded by DARPA through the SPAWARSYSCEN under Contract Number N66001-97-C-8525. Craig Boutilier was partially supported by NSERC Research Grant OGP0121843. Thanks to Bill Walsh, Piero La Mura, Tuomas Sandholm, Yoav Shoham and Mike Wellman for their helpful comments, suggestions and pointers to relevant literature.

[15] This is reminiscent of the mechanism suggested in [9].

References

[1] S. Bikhchandani and J. W. Mamer. Competitive equilibria in and exchange economy with indivisibilities. *J. Econ. Theory*, 74:385–413, 1997.

[2] C. Boutilier, T. Dean, and S. Hanks. Decision theoretic planning: Structural assumptions and computational leverage. *J. Artif. Intel. Res.*, 1999. To appear.

[3] C. Boutilier, M. Goldszmidt, and B. Sabata. Continuous value function approximation for sequential bidding policies. manuscript, 1999.

[4] G. W. Brown. Iterative solution of games by fictitious play. T. C. Koopmans, ed., *Activity Analysis of Production and Allocation*. Wiley, New York, 1951.

[5] S. Clearwater, ed. *Market-based Control: A Paradigm for Distributed Resource Allocation*. World Scientific, San Mateo, 1995.

[6] M. H. Degroot. *Probability and Statistics*. Addison-Wesley, New York, 1986.

[7] A. P. Dempster, N. M. Laird and D. B. Rubin Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, B-39:1–39, 1977.

[8] R. Engelbrecht-Wiggans and R. J. Weber. A sequential auction involving assymetrically informed bidders. *Intl. J. Game Theory*, 12:123–127, 1983.

[9] Y. Fujisima, K. Leyton-Brown, and Y. Shoham. Taming the computational complexity of combinatorial auctions. *IJCAI-99*, Stockholm, 1999. To appear.

[10] D. B. Hausch. Multi-object auctions: Sequential vs. simultaneous sales. *Mgmt. Sci.*, 32(12):1599–1610, 1986.

[11] S. Hon-Snir, D. Monderer, and A. Sela. A learning approach to auctions. *J. Econ. Theory*, 82(1):65–88, 1998.

[12] J. Hu and M. P. Wellman. Online learning about other agents in a dynamic multiagent system. *2nd Intl. Conf. Auton. Agents*, Minneapolis, 1998.

[13] R. P. McAfee and J. McMillan. Auctions and bidding. *J. Econ. Lit.*, 25:699–738, 1987.

[14] R. B. Myerson. *Game Theory: Analysis of Conflict*. Harvard, 1991.

[15] M. L. Puterman. *Markov Decision Processes*. Wiley, 1994.

[16] S. J. Rassenti, V. L. Smith, and R. L. Bulfin. A combinatorial auction mechanism for airport time slot allocation. *Bell J. Econ.*, 13:402–417, 1982.

[17] M. H. Rothkopf. Bidding in simultaneous auctions with a constraint on exposure. *Operations Research*, 25:620–629, 1977.

[18] M. H. Rothkopf, A. Pekeč, and R. M. Harstad. Computationally manageable combinatorial auctions. *Mgmt. Sci.*, 1998. To appear.

[19] T. Sandholm. An algorithm for optimal winner determination in combinatorial auctions. *Proc. of IJCAI-99*, Stockholm, 1999. To appear.

[20] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance*, 16(1):8–37, 1961.

[21] M. P. Wellman, W. E. Walsh, P. R. Wurman, and J. K. MacKie-Mason. Auction protocols for decentralized scheduling. (manuscript), 1998.