Pick-A-Bundle: A Novel Bundling Strategy for Selling Multiple Items within Online Auctions

Ioannis A. Vetsikas, Alex Rogers and Nicholas R. Jennings
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK
{iv,acr,nrj}@ecs.soton.ac.uk

ABSTRACT

In this paper, we consider the design of an agent that is able to autonomously make optimal bundling decisions when selling multiple heterogeneous items within *existing* online auctions. We show that while bundling the items together into a single lot is effective at reducing listing costs, it also results in a loss in auction revenue. To address this loss we introduce a novel bundling strategy, that we call *pick-a-bundle*, that can be implemented within any existing auction format. We show, mainly using simulations, that this new bundling strategy generates greater expected revenue than the complete bundle of all items, and, by inducing additional competition between bidders, it usually generates greater expected revenue than using separate auctions for each item. In order for our agent to accurately and efficiently calculate its expected revenue when using our new strategy, we derive a novel polynomial time algorithm for calculating the probability distributions of the sum of the top order statistics of i.i.d. variables drawn from any arbitrary distribution. Furthermore, we include in our analysis the strategic behaviour, in terms of bid shading, that the buyers may consider in our new auction format.

1. INTRODUCTION

The bundling of a number of heterogeneous items into a single lot is a common strategy when sellers participate in auctions. For example, within an online auction, such as *eBay.com* or *taobao.com*, sellers may bundle together a small number of low cost items, such as DVDs or computer games, in order to avoid incurring separate listing fees. Likewise, in traditional auctions it is common to find furniture items bundled together into a single lot in order to reduce the time overhead (and cost) involved in selling each item separately.

While the cost saving of bundling items together is self-evident in the examples described above, these savings must be set against the effect that bundling has on the expected revenue of the auction. Specifically, when the items being sold exhibit complementary valuations (i.e. the valuation of the bundle is greater than the sum of the valuations of each individual item) then the rationale for bundling is clear. However, the formal economic literature has little to say regarding the seller's revenue when bundling items that exhibit non-complementary valuations within auctions (as in the example of a bundle of DVDs described above). In particular, research in this area has so far only addressed the problem faced by a multi-product monopolist offering single items or bundles of items at fixed prices [1, 12]. While this setting is somewhat different to the one that we consider, this work shows that the bundling of goods can yield an increase in revenue, even when the items offered exhibit non-complementary valuations. Similarly, more recent work on the bundling of information goods on the internet, again shows profitability even in the absence of network externalities or economies of scale [2, 3]. Jehiel *et al.* [2007] examine a setting with additive valuations, which is similar to ours, however as the authors point out in their conclusions, their results do not work within "standard auctions". This limitation is also present in the computer science literature, where the bundling of items is often studied within the context of combinatorial auctions (see [5] for a review). Furthermore, this work has largely addressed the issue of complementary valuations, and has proposed novel auction protocols that allow bidders to express their preferences for specific bundles of goods [13, 4, 6, 11]. While such results are useful to the designers of new online or real-world auctions, they do not help a seller who is attempting to use an existing auction format (such as the English auctions of eBay or the sealed bid second price auction) in which bidders may only submit bids on the lot offered (and not on subsets of items).

Against this background, in this paper we consider for the first time the effects of bundling non-complementary goods *within a standard auction format*[1]. Our goal is to develop an autonomous *auction agent* that can advise on, and ultimately automate, the process of selling multiple heterogeneous items within such online auctions. To this end, we present a novel bundling strategy, that we call *pick-a-bundle*, in which a seller lists a set of items and announces a bundle size, and the buyers bid for the right to select a number of items from this set, which is equal to the bundle size;[2] the remaining items, which were not selected by the winner in his bundle, are then sold in a second round of separate auctions. For example, five DVDs may be listed and a bundle size of three might be announced, thus the winner of the auction would select the three DVDs that he would like to receive and the remaining two would be sold in a second round of separate auctions afterwards.

The design of our pick-a-bundle auction is informed by the intuition that bundling items will in general reduce the transaction or

[1]We specifically consider a sealed bid second price auction in our analysis, but due to the revenue equivalence theorem, our results apply to any efficient auction protocol.

[2]The idea of choosing the top valued items has also been used in [9, 17]. However, in these settings, the prices are fixed. Furthermore, information goods can be copied, whereas in our setting the fact that the bidders are competing for limited unclonable resources is what creates additional revenue.

Cite as: Title, Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

listing fees of the seller, but by offering just a subset of the items to the buyer, it will also induce competition between bidders who may actually prefer different items. This second factor has been observed in real world auctions that are sometimes used to sell a number of individual apartments within an apartment block. Here, rather than bidding for individual apartments, the buyers bid for the opportunity to select one of the remaining remaining unsold apartments. This procedure generates competition between bidders who prefer different apartments, and results in increased revenue for the seller [Personal communication from Michael H. Rothkopf, 2007]. We show that the same factors influence our pick-a-bundle auction, and show that it generates greater expected revenue than an auction where all items are included in a single bundle, and moreover, in the vast majority of the cases examined, it also generates greater expected revenue than selling each item in separate single-item auctions. Thus, in more detail, this paper makes the following contributions:

- We prove that the standard bundling strategy, in which multiple items are sold together in a single bundle, generates, with few exceptions, less revenue than separate single-item auctions (proposition 1). We address the loss of revenue incurred in this complete bundle auction by describing our novel pick-a-bundle format. Contrary to pre-existing work on bundling, *this format can be implemented within any existing standard auction*, as it does not require a redesign of the auction.[3] We motivate our decision, by showing theoretically that there always exists a pick-a-bundle auction which generates greater revenue than selling the items separately (proposition 2).
- We empirically evaluate our new bundling strategy through simulation, and show that it generates greater expected revenue than separate single-item auctions and the complete bundle (regardless of whether listing costs are taken into account).
- We then address the problem that the aforementioned simulations scale poorly as the the size of the problem (i.e. the number of items and bidders) increases. In order to calculate, rather than simulate, the expected revenue of the pick-a-bundle auction, in the cases where the items are relatively similar,[4] we propose a novel polynomial time algorithm (section 5) for calculating the probability distributions of the sum of the top order statistics of i.i.d. variables drawn from arbitrary discrete distributions.
- Finally, we examine how the presence of the second round of auctions for selling the remaining items, which were not selected by the winner of the pick-a-bundle, affects the bids of the buyers. While in some cases, the bids will not be affected (claim 1), in some others the buyers are going to shade (reduce) their bids. We compute these bids and integrate this effect into our analysis of the pick-a-bundle revenue and our experiments.

The remainder of this paper proceeds as follows: In section 2 we formalize the problem setting. In section 3 we describe and motivate the pick-a-bundle strategy, before discussing how to optimize its performance (by selecting the optimal bundle size) in section 4. In section 5 we present the algorithm for computing the expected revenue without simulations, for the case of similar items. In section 6 we examine the effect of the second round of auctions on the buyers' bids. In section 7 we generate empirical results that, unlike those of section 4, take the effect of bid shading into account; the algorithm of figure 2 is also used to generate the results without the need for simulations. Finally, we conclude.

[3]While we have specifically addressed the needs of an autonomous *auction agent* here, we note that all the results of this paper are also of immediate use to current sellers using existing traditional or online auctions.

[4]We assume that they have the same prior distributions $f_i(u)$ (defined in section 2).

2. PROBLEM STATEMENT

In this section we formally describe the auction setting to be analyzed and the auction format choices that a seller has in order to maximize her revenue. First, however, we introduce the basic notation. We assume that a seller has $m > 1$ items to sell. There are n buyers interested in purchasing some of these items. The valuation u_{ij} of item i to each bidder j is drawn from a known distribution with probability density function (pdf), $f_i(u)$, and associated cumulative density function (cdf), $F_i(u)$. The exact valuations $u_{ij}, \forall i$ are known only to the bidder j himself. These distributions represent the prior knowledge available to all participants about the valuations of the other agents in the auction and are common knowledge to all.[5] We assume that $f_i(u)$ take non-zero values in $[0, L]$, and thus for $\forall u < 0$ and $\forall u > L$, it is $f_i(u) = 0, \forall i$. Although there is nothing to limit the applicability of our experimental results to any distribution, in the experiments we assume that the item valuations and corresponding bids take discrete values. This represents no loss in accuracy since in all real auctions bidders can only place bids that are whole denominations of the currency used (i.e. a bid can't be 95.3 pence, but rather either 95 or 96 pence). Given this, the distributions $f_i(u)$ used, are discrete, and can take values $u = 0, \ldots, L$. In our analysis, we assume that the value of items is additive, meaning that a set of items is worth the sum of the individual item valuations; this is a result of the limited number of results in the literature about how bidders bid in cases of complementary and substitutable items (they don't bid truthfully), which makes it impossible to analyze the expected revenue in such cases.[6]

Sellers sell the items in second price auctions (i.e. auctions in which the top bidder is awarded the item for sale, and the payment is equal to the second highest bid). These auctions are incentive compatible, and thus we assume that bidders place bids equal to their true valuations [10]. Because of the revenue equivalence theorem, our results are also valid in cases that a first price auction would be used (or indeed any other efficient auction mechanism).

Finally, we assume that the seller has a certain degree of freedom when she sets up the auction; even though the auction format must follow that of the online or traditional auction house being used (i.e. in our case it must be a sealed bid second price auction). In particular, the good for sale is determined by the seller, and thus, instead of selling a single item, bundles of items may also be listed.

In real world auctions, there is often a *listing cost* associated with this auction which is levied on the buyer, the seller or both. In our experiments we assume a fixed listing cost C extracted from the

[5]These valuations can model a wide variety of scenarios, simply by selecting the correct distribution. For example, the distribution could incorporate the value of a potential resale by selecting the item valuations (given by the distribution) to be equal to the resale value in the cases when the actual bidder value from keeping the item is lower than the resale value.

[6]On eBay, for example, sometimes a seller will offer to sell substitutable items, like "choose a black or a white iPod". In such cases, the analysis is much easier than in our model, because here bidders are interested in exactly one item; in our model they could be interested in buying both.

seller's revenue for each auction that is conducted. This is inspired from the actual listing costs levied by online auction houses such as eBay, which charge a listing cost consisting of a fixed fee C and a percentage cost (which can be ignored as it reduces the revenue in all cases by the same percentage).[7]

3. THE BUNDLING STRATEGY

A seller with multiple items for sale may choose to do so in *separate auctions each selling a single item*. Another option, which is the main bundling technique used by sellers on sites such as eBay, is to offer a *complete bundle* of items for sale; the winning bidder gets all of the items listed. This is usually done in order to reduce the listing costs. However, as we show below (for the case of identical distributions $f_i(u)$), the complete bundle generates less revenue, except in the case of few bidders:

Proposition 1 *For a symmetric distribution $f(u)$, when the item distributions $f_i(u) = f(u), \forall i$, an auction for a complete bundle of m items, and m separate single-item auctions generate the same expected revenue for the seller when there are $n = 3$ bidders, more for $n = 2$ bidders, and less for every other case ($n > 3$ bidders).*

PROOF (SKETCH). Due to space constraints, we only present here the proof for the case when $n = 3$. The proofs for the other cases (when $n = 2$ and $n > 3$) follow some of the same steps as the part of the proof presented here and also use the observation that taking the convolution of a pdf with itself (any number of times) will concentrate the probability mass towards the mean of the distribution and that the effect is more pronounced as the number of convolutions performed increases. Also note that since f is symmetric, it is $f(u) = f(L - u), \forall u \in [0, L]$, and the cdf is thus $F(u) = 1 - F(L - u), \forall u \in [0, L]$.

Now, when $n = 3$ bidders participate, the expected revenue from each single-item auction, R_{single}, is the expected value of the 2^{nd} order statistic (the second highest valuation) of the valuation that each of the 3 bidders have for that item. This can be computed to be equal to $R_{single} = L - \int_0^L (F^3(u) + 3F^2(u)(1 - F(u)))du = \frac{L}{2}$, and can be seen to be equal to the mean of the original symmetric distribution with pdf $f(u)$.

Furthermore, if $g(u)$ and $G(u)$ are the pdf and cdf respectively of the distribution of the sum of m i.i.d. random variables which are drawn from distribution $f(u)$, then, since $f(u)$ is symmetric, we know that $g(u)$ must also be symmetric. Using the same reasoning as in the case of distribution f above, we deduce that the expected revenue from selling the complete bundle R_{bundle} is equal to the mean of distribution $g(u)$, which is equal to $R_{bundle} = \frac{m \cdot L}{2} = m \cdot R_{single}$. Therefore the expected revenue in both cases is the same. □

The complete bundle suffers from a loss in revenue, because the second highest valuation of the bundle is generally less than the sum of the second highest valuations for each separate item, since a bidder's high valuation for one item is likely to be balanced by a lower valuation for another item. We can generalize this observation based on our experimental evidence for any combination of distributions $f_i(u)$ (see sections 4 and 7) and conclude that, while for $n = 2$ the complete bundle is the best option, for $n \geq 4$ bidders participating, one would be better off using separate single-item auctions.

In order to address this loss in revenue, we propose our *pick-a-bundle* strategy. Under this scheme, the seller advertises all m items that are for sale, but each buyer bids for the right to buy k of these items (where $k \leq m$). The winner of the auction informs the seller which are the k items that it actually wants to receive (i.e. the k items that have the highest valuations for that agent). The remaining items, which are not selected, still need to be sold by the seller, and this is done by having another round of separate single-item auctions.[8] Note that the case when $k = m$ is exactly the complete bundle auction, which is why the complete bundle is a special case of pick-a-bundle. However, when $k = 1$, this is not the same as selling in separate auctions, because there is one auction in which the highest valuation among all agents for all items wins, and then the leftover items are sold in separate auctions.

The intuition behind our bundling strategy is clear. By reducing the size of the bundle compared to the complete bundle, the magnitude of the revenue loss due to the fact that a seller might be primarily interested in only some of the items is reduced. At the same time, the preference of different subsets of items by different bidders also induces additional competition into the auction, as buyers interested in different subsets of items will now compete against each other. This leads our pick-a-bundle auction to outperform both the complete and the separate single-item auctions in the vast majority of settings we examined.

In fact we can further motivate this approach even before we discuss how to evaluate the different auction settings:

Proposition 2 *For any distributions $f_i(u)$, using the pick-a-bundle auction, when $k = 1$ item is chosen, will always generate the same or more revenue than separate auctions each selling one of these items.*

PROOF. Let $u_{i,j}$ be the valuation that bidder i has for the j^{th} item. In the pick-a-bundle auction with bundle size $k = 1$, the revenue R_b is equal to the second highest among the highest values of each bidder. Let i_1 be the bidder with the highest valuation $u_{i,j}, \forall i, j$, for item j_1, and $i_2 \neq i_1$ the bidder with the second highest among the bidders' high valuations; this valuation is for item j_2. Thus $R_b = u_{i_2,j_2}$ and this revenue is obtained for selling object j_1.

The revenue R_{j_1} obtained in selling item j_1 on its own is the second highest value among $u_{i,j_1}, \forall i$. We examine the following two cases:

1. When $j_1 = j_2$. Then $R_{j_1} = u_{i_2,j_1} = u_{i_2,j_2} = R_b$. The revenue in this case is exactly the same.

2. When $j_1 \neq j_2$. We know that $u_{i_2,j_2} = \arg\max_{i \neq i_1, j} u_{i,j}$. Since R_{j_1} is equal to one of the elements of set $\{u_{i,j}\}, \forall i \neq i_1, j$, therefore $R_{j_1} \leq u_{i_2,j_2} = R_b$.

Thus, the pick-a-bundle auction, when one item is chosen (i.e. $k = 1$), will always generate the same or more revenue than any of the individual auctions selling these items. Thus, the revenue from our new auction setting (for $k = 1$) is always at least as high as that from selling the items in separate auctions. □

This shows that, in the case of scenarios where buyer strategic bidding is not an issue (see section 6), the pick-a-bundle auction for bundle size $k = 1$ generates more revenue than separate auctions. This does not mean that selecting $k = 1$ is indeed the best bundle size; in fact quite often it is not. Rather it means that in this setting, it is always preferable to use a pick-a-bundle auction in preference to separate auctions.

[7]To be more precise, eBay charges an "insertion fee", which is a fixed cost depending on the starting price of the auction and a "final value fee" which is, for the most part, a percentage of the final closing price of the auction.

[8]Note that given sufficient remaining items, another pick-a-bundle auction could be used on the remaining items. We do not explicitly consider this case, but the analysis that we present could easily be extended to cover it.

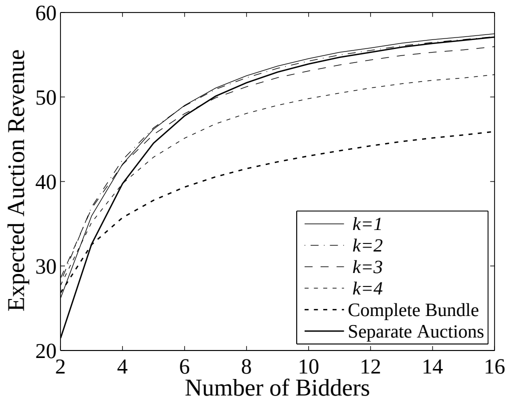

Figure 1: Seller's revenue for a bundle of $m = 5$ items, as the number of participating buyers n increases, when using (i) separate single-items auctions to sell each item individually, (ii) a pick-a-bundle auction initially where $k = 1, \ldots, 5$. The item valuations are drawn from five different, non-independent distributions.

4. OPTIMIZING THE BUNDLE SIZE

Now that we have presented the seller's choices (i.e. separate auctions, complete bundle and pick-a-bundle), she needs to find out which of these choices gives the highest revenue. This involves computing the expected revenue for all these choices, and in the case of the pick-a-bundle auction, also selecting the best value of k (the size of the bundle offered). For the general scenario described in section 2, we need to run a simulation in order to compute the expected revenue of the various cases; we take a large number of random samples of item valuations for all bidders, and average the revenue of the auction across all samples.

To illustrate the wide applicability of our new auction format we carried out a number of simulations with different distributions. Here we show the results for a particular representative case. More specifically, we assume that item valuations u_{ij} are drawn from the following 5 distributions: (i) u_{1j} are drawn uniformly from $\{3, \ldots, 12\}$ (prob. $\frac{1}{10}$ for each value), (ii) u_{2j} are correlated with u_{1j} and are $u_{2j} = u_{1j} + z$, where z is drawn uniformly at random from $\{-2, \ldots, 2\}$, so u_{2j} take values from 1 to 14, (iii) u_{3j} are drawn uniformly from $\{0, \ldots, 14\}$, (iv) u_{4j} are equal to 0 with prob. $\frac{4}{10}$ and take values $\{5, \ldots, 10\}$, each with prob. $\frac{1}{10}$, and (v) u_{5j} are equal to 0 with prob. $\frac{1}{2}$ and equal to 12 with prob. $\frac{1}{2}$.[9]

We present the mean value of the revenue generated in the simulation for all possible cases in figure 1. We repeat the simulation up to 100000 times such that the standard error in the mean values shown is less than the thickness of the line, and hence, we do not show error bars in the figure. As we expected (based in part on the theoretical results of propositions 1 and 2), the complete bundle is not a good option no matter how many bidders participate in the auction. The same is also true for the revenue from separate auctions. The pick-a-bundle auction generates more revenue, no matter whether the listing costs are considered or not.[10] Now, we consider how the choice of the bundle size (i.e. the value of k) affects the revenue of the pick-a-bundle auction, and hence, show how to determine which bundle size is optimal in any given case. For only $n = 2$ bidders, $k = 3$ is the best choice, for a small number of bidders ($n = 3, \ldots, 5$), $k = 2$ is the best choice, whereas for larger numbers ($n \geq 6$), $k = 1$ is the best choice. Depending on the prior belief about possible numbers of bidders participating, n, the auction used should be the pick-a-bundle auction with $k = 1$ or $k = 2$. If the listing cost is included, then $k = 2$ is the best choice for almost all cases ($n \geq 4$), and $k = 3$ is the best choice only when $n \leq 3$.

While these simulations allows us to compute the auction parameters that maximize the seller's expected revenue for any specific setting, they are impractical in cases where the number of bidders, n, or items, m, increases significantly. This is due to the fact that, in this case, each individual simulation run takes longer to compute, and also that each simulation must be repeated multiple times in order to ensure the statistical significance of the final results (e.g. in the experiments of figure 1 we repeat the simulation 100000 times). Thus, in the next section we address this practical issue by presenting an alternative algorithmic approach.

5. ALGORITHMIC COMPUTATION OF THE EXPECTED REVENUE

Given the comments above, we now present an algorithm that allows us to compute the revenue of our new bundle auction, in the case that the items are similar. This algorithm is exact, is much faster than the simulation approach presented in the previous section, and scales extremely well (i.e. as the number of bidders participating in the auction increases, the corresponding increase in computational cost is minimal). However, it is restricted to cases where the valuations of each item are i.i.d. and are drawn from the same distribution, $f(u)$. (for other cases the simulations are still the best method) This constraint implies that all of the items have similar values. While not always applicable, it certainly applies to our motivating example of selling a number of DVD movies when these movies are of the same class (e.g. new blockbusters).

To compute the expected revenue of the pick-a-bundle auction, we first need to calculate the probability density function that describes the bidders' valuation for the subset of k items that they value most highly (and hence will bid for). Since the top k order statistics are not independent, this is not at all trivial, and in fact, algorithms that can compute this pdf accurately for general valuation distributions and arbitrary numbers of items do not exist.[11] Here, we propose an algorithm that is polynomial with respect to variables k, m and L; it is shown in pseudo-code in figure 2. Once this pdf is computed, we compute the expected revenue of the pick-a-bundle auctions in time polynomial to n.

The main idea behind our algorithm is that we use three "for" loops to execute the computation. The main loop changes the value of variable x, which represents the value of the k^{th} order statistic. Once this is known, we use the other two loops to enumerate all the possible values of variables ω_b and ω_a, which represent the number of order statistics, which are respectively below (i.e. $(k+1)^{th}$ to m^{th} order statistics) and above[12] (i.e. 1^{st} to k^{th} order statistics)

[9]We selected to present here an example with many different distributions and valuations that are not always independent of each other (here u_{1j} and u_{2j} are not), in order to show how generally applicable our work is. In all the cases that we considered, the same broad pattern of results were observed.

[10]Based on the closing price of these auctions, the proper listing cost would be $C = 0.35$, if these were eBay auctions. The listing fees saved through bundling are equal to $(\lambda - 1) \cdot C$, where $\lambda = k$ for the pick-a-bundle auction, $\lambda = 1$ for separate auctions, and $\lambda = m$ for the complete bundle.

[11]There are asymptotic methods that give approximations of this pdf as the size of the problem increases to infinity [15]. However, we consider relatively small numbers of items, and these methods do not provide sufficiently accurate approximations in these cases.

[12]Actually here we also count the k^{th} order statistic as well, so it is $\omega_a \geq 1$. When we say that some order statistic is above the k^{th}, this means it is one of the 1^{st} to $(k-1)^{th}$ order statistics and

```
p_a := 1, p_b := 0, p_eq := 0, g_m^k(u) = 0, ∀u
for x := 0 to L do
    p_b := p_b + p_eq   [Note: p_b = Σ_{i=0}^{x−1} f(i)]
    p_eq := f(x)
    p_a := p_a − p_eq   [Note: p_a = Σ_{i=x+1}^{L} f(i)]
    h⃗ := [f(x+1) ... f(L)]
    h⃗′ := [1]
    for ω_a := k to 1 step −1 do
        p := 0
        for ω_b := 0 to (m − k) do
            p := p + m! / ((m−k−ω_b)!(ω_a+ω_b)!(k−ω_a)!) · p_eq^(ω_a+ω_b) · p_b^(m−k−ω_b)
        end for loop (variable ω_b)
        if (ω_a = k) then increase g_m^k(k·x) by p
        else
            h⃗′ := h⃗′ ⊗ h⃗
            Increase g_m^k(i + k·x + k − ω_a) by p · g⃗′(i), ∀i
        end else
    end for loop (variable ω_a)
end for loop (variable x)
```

Figure 2: Algorithm for computing the probability distribution of the sum of the k top order statistics of m i.i.d. variables drawn from distribution $f(u)$

the k^{th} order statistic x, and that have a value equal to x. The probability that $(\omega_a+\omega_b)$ variables have value equal to x, $(k-\omega_a)$ have values greater than x and $(m-k-\omega_b)$ have values smaller than x, is:

$$p(x,\omega_a,\omega_b)=\frac{m!(\sum_{i=0}^{x-1}f(i))^{m-k-\omega_b}f(x)^{\omega_a+\omega_b}(\sum_{i=x+1}^{L}f(i))^{k-\omega_a}}{(m-k-\omega_b)!(\omega_a+\omega_b)!(k-\omega_a)!} \tag{1}$$

The precise values that the bottom $(m-k-\omega_b)$ order statistics have are not important, because they correspond to the least valued items; only the values of the top $(k-\omega_a)$ order statistics must be taken into account. What we know is that in this case, all these values are greater than x. Therefore, we can take the pdf $h(u)$ to denote the conditional probability when we know that $u > x$. This is given by:

$$h(u) = \frac{f(u)}{\sum_{i=x+1}^{L} f(i)}, \text{ if } u \in \{x+1,\ldots,L\} \tag{2}$$

and $h(u) = 0$, otherwise. Given that the valuations are i.i.d. we can compute the distribution of their sum by taking the convolution:

$$h_{k-\omega_a}(u) = \underbrace{h(u) \otimes \ldots \otimes h(u)}_{(k-\omega_a)\text{ times}} \tag{3}$$

Now, we also know that another ω_a order statistics are equal to x, meaning that the pdf of the k top order statistics for this case is:

$$\widetilde{h}_{k-\omega_a}(u+\omega_a x) = h_{k-\omega_a}(u), \forall u \in \{0,\ldots,(k-\omega_a)L\} \tag{4}$$

and $\widetilde{h}_{k-\omega_a}(u) = 0$, otherwise. As this case happens with probability $p(x,\omega_a,\omega_b)$, and we account for all possible values of x, ω_a and ω_b with the three loops, the pdf of the sum of the top k order statistics is finally given by:

$$g_m^k(u) = \sum_{x=0}^{L}\sum_{\omega_a=1}^{k}\sum_{\omega_b=0}^{m-k} p(x,\omega_a,\omega_b)\widetilde{h}_{k-\omega_a}(u) \tag{5}$$

It should be noted that the algorithm in figure 2 has some additional optimizations (which are not described here due to limited space). Its complexity depends on k, m and L and is $O(k^2mL^3)$, because we use a convolution algorithm with quadratic cost $O(kL^2)$. An even more efficient algorithm could make use of a Fast Fourier Transform to compute the convolution, which would reduce the complexity to $O(k^2mL^2\ln(kL))$.

Given that the probability distribution of the sum has been computed, the expected revenue, ER, of each auction is given by:

$$ER_G = n(n-1)\int_0^{mL}[1-G(u)]^{n-2}g(u)uG(u)du \tag{6}$$

where $G()$ and $g()$ are, respectively, the cdf and pdf of the distribution of the valuation of the commodity sold in the auction; hence $g = g_m^k$ for the pick-a-bundle auction, and $g = f$ for the separate auctions. This can be done in $O(mL\ln(n))$ time, since the computation of a power x^n is done in $O(\ln(n))$ time. Thus, the complexity of computing the revenue of one pick-a-bundle auction is $O(k^2mL^2\ln(kL) + mL\ln(n))$, which is low polynomial complexity with the size of the problem. Consequently, this allows us to compute the best auction very fast and efficiently.

6. ACCOUNTING FOR THE SECOND ROUND OF AUCTIONS

In this section, we examine the effect of the second round of auctions on the buyers' bids, which has not been included in the work presented until this point. Initially, we examine dynamic settings, where the effect of the second round can be ignored (thus for these scenarios the analysis of the previous sections can be applied as presented), and then we examine settings in which it cannot. For the second case, we first compute accurately the bid shading that occurs due to the second round of auctions in the case where similar items are being sold (as discussed in the previous section) and we then proceed to examine more general settings.

In more detail, online auctions constitute *dynamic scenarios*, in which there are many external opportunities for buyers to acquire equivalent items (e.g. from other sellers in an online auction house such as eBay), and thus, the second round of the pick-a-bundle auction does not represent the last chance to acquire any particular item. In this case, there is no bid shading since its effect is already incorporated within the buyers' valuations for the items:

Claim 1 *In a dynamic auctions setting, we can ignore the effect of bid shading due to the second round of auctions.*

DISCUSSION. We can assume that the true internal valuations of bidder i in this setting are $\overline{u}_{i,j}$, which are drawn from a distribution with pdf $\overline{f}_j(u)$ for each item j. Since online auctions are often dynamic scenarios where bidders and buyers come and go and the same item is being sold in many different auctions, the bidders will indeed shade their bids in any auction in which they participate, because of the possibility of getting the item that they are interested in from another seller (in another auction). To prove this fact more rigourously, we will use the sequential auction model and the subsequent analysis presented in [14]. This analysis provides the mapping $\overline{g}_j : \overline{u}_{i,j} \rightarrow u_{i,j}$ from the real valuations $\overline{u}_{i,j}$ to the "shaded" ones $u_{i,j}$. Using the same mapping, we can generate the new "shaded" prior distributions $f_j(u)$ from which the shaded valuations $u_{i,j}$ are drawn. Because the number of potential sellers is sufficiently large in an online auction setting, following the analysis in [14], we can conclude that the bidder's bids will not change (more than a tiny amount), because of the presence of one more seller, or, equivalently, one additional chance to buy the item. Therefore, we can use the shaded valuations $u_{i,j}$, in our analysis of how bidders will bid in the pick-a-bundle format, to represent the valuations of bidders in both rounds of bids; the effect of additional bid shading is minuscule, if not non-existent. □

therefore its value is the same or higher than that of the k^{th} order statistic.

However, there are also *static scenarios*, in which a fixed number of buyers are participating and the items being sold are not available from any other sources (i.e. because they are rare or difficult to find). Within our pick-a-bundle auction, the buyers will thus have two opportunities to acquire each item: in the pick-a-bundle auction as part of a bundle of k items, and in the second round of separate single-item auctions, if that item was not sold in the bundle auction. Because of this second opportunity to purchase items if they don't win, buyers shade (i.e. strategically reduce) the bids placed in the bundle auction. We can initially consider the case of similar items:

Proposition 3 *Assume w.l.o.g. that $u_{i1} > u_{i2} > \ldots > u_{im}$ are the valuations of buyer i for the items offered for sale. If he participates in a pick-a-bundle auction where k items are sold in the first round, then, because of the chance to get the remaining $(m-k)$ of these items in the second round of auctions, he will bid:*

$$b_i = \sum_{j=1}^{k} \left(u_{ij} - \frac{m-k}{m} \sum_{x=0}^{u_{ij}-1} \Omega(x) \right) \quad (7)$$

where $\Omega(x)$ is the distribution of the top bid of the opponents it will face in the second round.

PROOF. Since the seller sells k out of m items in the pick-a-bundle auction initially (first round), then there is a $\frac{m-k}{m}$ chance that any item that a buyer wishes to purchase might be available for sale in the second round (i.e. there is an equal probability that each item is not among the top k most desired items of the winner of the first auction and therefore left for sale in the second round). Let us assume that item j for which a certain buyer i has valuation u_{ij}, is relisted in the second round. From his point of view, buyer i competes against a number of other bidders, whose maximum bid is B (the top order statistic of these bidder's private valuations). Since we assumed that B is drawn from distribution with pdf $\Omega()$, it is $Prob[B \leq x] = \Omega(x)$. Then bidder i with valuation u_{ij}, will bid truthfully and win this auction with probability $\Omega(u_{ij})$. The payment he will make is equal to B. Thus, his expected profit in this auction is given by:

$$EProfit_{ij} = \sum_{x=0}^{u_{ij}-1} (u_{ij} - x) \cdot Prob[B = x] = \sum_{x=0}^{u_{ij}-1} \Omega(x) \quad (8)$$

Therefore, if we assume w.l.o.g. that $u_{i1} > u_{i2} > \ldots > u_{im}$ are the valuations of buyer i, then he has a chance to win each of his k most valued items in the second round, if he does not win in the first. His expected total profit in this case is:

$$C_i = \frac{m-k}{m} \cdot \sum_{j=1}^{k} EProfit_{ij} = \frac{m-k}{m} \cdot \sum_{j=1}^{k} \sum_{x=0}^{u_{ij}-1} \Omega(x) \quad (9)$$

Now this buyer i will gain a profit of $(\sum_{j=1}^{k} u_{ij} - t)$, where t is the payment and is equal to the second highest bid, if he wins the pick-a-bundle auction, and an expected profit C_i if he loses it (since he then has a chance to buy some of the items he desires in the second round auctions). It is then trivial to show that the bid that maximizes his overall expected profit is $(\sum_{j=1}^{k} u_{ij} - C_i)$, and thus, he should shade his true valuation for the bundle (given by $\sum_{j=1}^{k} u_{ij}$), by an amount equal to his expected profit C_i. □

This effectively means that each bidder shades his true valuation u_{ij} for each item by $\frac{m-k}{m} \sum_{x=0}^{u_{ij}-1} \Omega(x)$.[13]

Now, we describe how to modify the algorithm of figure 2 in order to incorporate the bid shading. The easiest way to do so is to modify the prior distribution of the valuations $f(u)$, and to map every value $u \in \{0, \ldots, L\}$ to $u - \frac{m-k}{m} \cdot \sum_{x=0}^{u-1} \Omega(x)$. Note, however, that this equation does not always give integer values. For example, value $u = 3$ might need to be mapped to 2.4. If the new pdf $\widehat{f}(u)$, that we input in the algorithm, were generated from these mappings by rounding each value to the closest integer (i.e. 2.4 becomes 2), then the sum of bids could have a significant error (as much as $\lfloor \frac{k}{2} \rfloor$). To alleviate this, we can simply create k virtual valuations. For example, if f is defined at $\{0, 1, \ldots\}$ we can decide to define it at values $\{0, \frac{1}{k}, \frac{2}{k}, \frac{3}{k}, \ldots\}$ and then round the shaded bids to these values rather than the integer values. By doing this, we guarantee that the total error in the computation is less than $\frac{1}{2}$. As the bidders must bid integer values, this means that we would have to round the final bid b_i anyway, and therefore with this trick we can guarantee that our algorithm computes as a value either $\lfloor b_i \rfloor$ or $\lceil b_i \rceil$. The algorithm now should have complexity $O(k^5 m L^3)$, but by taking advantage of the sparsity of the pdf in this case, we derive an algorithm with lower complexity $O(k^3 \cdot m \cdot L^3)$.

The previous case described accounts for bid shading in the case that the items are similar. So it's now time to examine the case that they are not and therefore they have different priors $f_j(u)$. In this case, we assume that the winner of the first round (i.e. the pick-a-bundle auction) will always select his top valued items. Then we can extend the results of the previous subsection to get:

$$b_i = \sum_{j=1}^{k} \left(u_{ij} - \pi_j \sum_{x=0}^{u_{ij}-1} \Omega_j(x) \right) \quad (10)$$

where $\Omega_j(x)$ is now the distribution of the top opponent bid for the j^{th} item, which is computed in the same way as in the previous case, and π_j is the probability that this item will be available in the second round; this probability is computed from the distributions $f_j()$. Due to the assumption made that the k highest valued items are always selected, this is really a fair, albeit not absolutely accurate, estimate of the bid placed, unlike equation 7 which is accurate for the previous case. We would like to explain why, in a few cases, the winner will not select his k most valued items in the first round. Let us give an example of when this happens. Assume that the k^{th} top valued items of winning bidder i has value $u_{i,j_k} = 10$ and the prior for that item is $f_{j_k} = U[0, 10]$ (uniform). Also the next highest valued item (i.e. the $(k+1)^{th}$ order statistic among bidder i's valuations) has value $u_{i,j_{k+1}} = 9$ and the prior distribution for that item is $f_{j_k} = U[9, 12]$ (also a uniform distribution). In most cases, it would be in bidder i's interest to select item j_{k+1} rather than j_k to get from the pick-a-bundle auction, even though its value is lower. This is due to the fact that this bidder will have a much easier time winning item j_k compared to item j_{k+1} in the second round, and thus make even more profit.

Even though this case of a winner not selecting one of his k highest valued items in the first round is not very common, it does complicate the analysis quite a bit. This is the reason why, we leave this analysis for future work.

7. EMPIRICAL RESULTS

[13]In the case that this bidder faces $\mathcal{N} = n - 1$ other bidders, the valuation of each being drawn from distribution with cdf $F(u)$, then the top opponent bid is drawn from distribution with cdf $\Omega(x) = F(x)^{n-1}$. This is not exactly the distribution that we will use in our experiments in section 7, because we need to account for the fact that the valuations of one bidder (the winner of the pick-a-bundle auction) are drawn from a different distribution, which we denote $H(u)$. In this case $\Omega(x) = F(x)^{n-2} H(x)$. In the experiments presented in the next section, we used a simulation to compute $H(u)$. The expected revenue of the separate auctions selling the remaining items is now computed by taking the expected value of distribution $\Psi(x) = (F(x))^{n-1} + (N-1) \cdot H(x) \cdot (1 - F(x)) \cdot F(x)^{n-2}$, rather than equation 6.

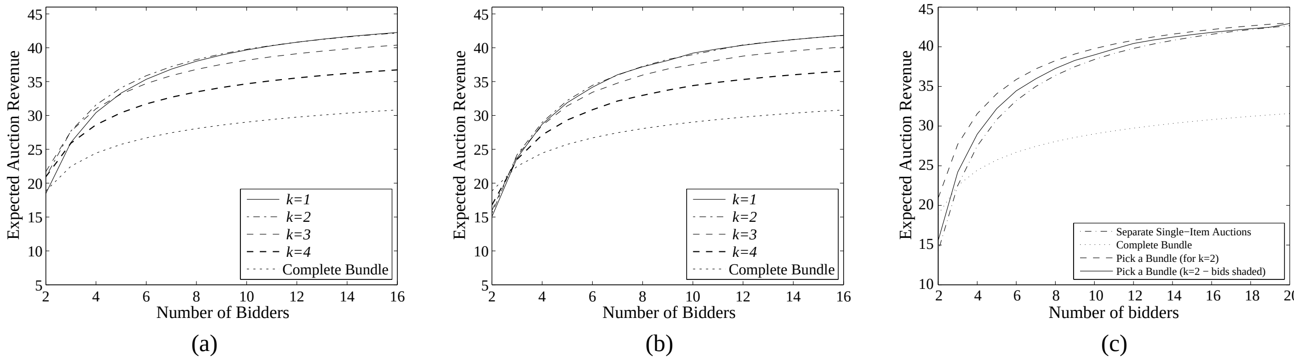

Figure 3: Seller's revenue for a bundle of $m = 5$ items, as the number of participating buyers n increases. In (b) the bids are shaded, whereas in (a) they are not. In (c) we include only the best pick-a-bundle auction ($k = 2$). The distribution of the buyer valuations $f(u)$ is uniform for values $\{0, \dots, 9\}$.

In section 4, we showed how to compute the best bundle size for the pick-a-bundle auction and determine whether this is the best choice available to the seller. In this section, we give additional empirical results that make use of both the algorithmic computation (from section 5), which is faster and more efficient, and our analysis of the strategic buyer behaviour (from section 6).

The first experiment uses the algorithms of figure 2 and its extension which accounts for the buyers shading their bids for several scenarios (i.e. distributions) when all the items are similar. More specifically, we examine the seller's revenue, who has $m = 5$ items for sale, to n participating bidders. The valuations of the items are i.i.d. variables drawn from the same distribution $f(u)$. In figure 3 we graph the expected revenue, when $f(u)$ is a uniform distribution for values $\{0, \dots, 9\}$, so $f(u) = 0.1, \forall u \in \{0, \dots, 9\}$, and $f(u) = 0, \forall u \notin \{0, \dots, 9\}$. We use uniform distributions, because they are the ones most commonly used in analyzing auction scenarios in the literature. We have also done the same analysis with several other distributions[14] and, though we don't present the results here, they lead to similar conclusions as those drawn here.

In figure 3(a), the expected revenue of the pick-a-bundle format is examined for all the values of the bundle size k,[15] when no bid shading occurs. The expected revenue of all cases is computed using equation 6. This figure allows us to consider how the choice of the bundle size (i.e. the value of k) affects the revenue of the auction, and hence, show how to determine which bundle size is optimal in any given case. As seen in the plot, the pick-a-bundle auction with $k = 2$ generates more revenue over a wide range of the number of bidders. More specifically, it yields the highest revenue for $n \in \{3, \dots, 19\}$ and is thus the best choice for these values; when $n = 2$ (resp. $n \geq 20$), then the best choice is $k = 3$ (resp. $k = 1$). These results must also be set against the $(k - 1) \cdot C$ listing fees saved through bundling. In this case, $k = 2$ is the best choice when $n \geq 4$, while $k = 3$ is better for very small numbers of participating bidders, such as $n \leq 3$.

Figure 3(b) presents the equivalent results to figure 3(a), when bid shading does take place. The results in this case are very similar to those of figure 3(a); the pick-a-bundle auctions for $k = 1$ and $k = 2$ give almost the same expected revenue for most values of n (although, of course in practice, $k = 2$ would be preferred due to the listing fee savings). For very small values of n ($n \leq 3$), $k = 3$ can be the best choice, especially if listing costs are considered.

Based on the results of figures 3(a) and 3(b), the overall best choice for the bundle size is $k = 2$. We use this value in figure 3(c), where we compare the revenue obtained from the pick-a-bundle auction with those of the complete bundle and of using separate single-item auctions. When bid shading does not occur, the new bundling strategy is the clear winner, and when bid shading is taken into account it still outperforms the other two in almost all cases (the exception being when $n = 2$, when one should select the complete bundle). If we consider listing costs these observations do not change; for small values of n the complete bundle might be better (and only in the case where the bids are shaded) and for every other case the pick-a-bundle format is the best choice.

The second experiment is the same as that presented in section 4. Here, however, we also include the effect of bid shading and, thus, the buyers' bids are computed using equation 10. The results are presented in figure 4 and they reinforce the observations made in the previous experiments. Furthermore, comparing with figure 1, we notice that, while accounting for the bid shading reduces the benefit of using pick-a-bundle compared to the complete bundle and the separate auctions, it is still the best choice. More specifically, $n = 2$ bidders is the only case in which the complete bundle would be preferable (especially when including listing costs), while for small number of bidders ($n = 3, \dots, 6$) the pick-a-bundle with size $k = 2$ is best, and for $n > 6$ bidders the pick-a-bundle of size $k = 1$ yields that highest revenue (even accounting for the listing costs). In addition, when the number of bidders becomes significant ($n \geq 14$ in this experiments) the difference between the pick-a-bundle and the separate auctions becomes small (less than the standard error for the number of samples that we used in our simulations). From all these, we can conclude that pick-a-bundle is the best option, with the optimal bundle size depending on the number of participating bidders n; however when there is sufficient competition (a large number of bidders n) the best choice is to use pick-a-bundle with bundle size $k = 1$ closely followed by separate auctions.

General Observations In general, from all the experiments we conducted, both the representative ones we presented in this paper and others we also performed, we observe that the complete bundle

[14] For example, as there is evidence to suggest that in many cases, there is a substantial probability that a buyer is either not interested at all or is interested very little in a specific item, e.g. the experimental distribution used for the TAC game analysis in [16], we used (among others) distributions that place substantial probability at value 0. This is a situation which is very likely to occur within the motivating example concerning bundles of DVDs; for example, if the buyer already has a copy of a DVD, he is likely to have little interest in acquiring a second one.

[15] The reader is reminded that selecting $k = m$ is the complete bundle case.

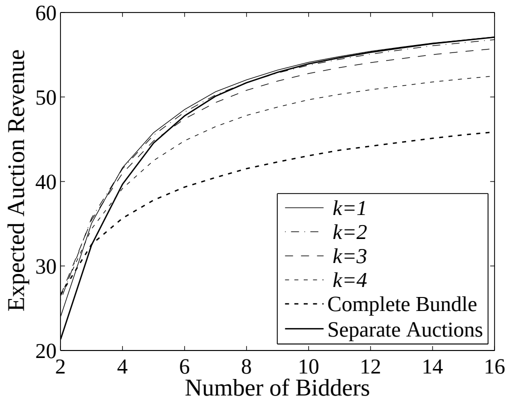

Figure 4: The same experiment as that of figure 1, when bid shading is also considered.

only makes sense in the case of few bidders ($n = 2$) and only in cases that the effect of listing costs and bid shading on the auction revenue is considered. The cases when relatively few (i.e. $n \leq 10$) bidders participate is when the pick-a-bundle gives the highest benefit compared, in particular, to the case of separate single-item auctions. While bid shading causes a drop of the pick-a-bundle revenue, this mainly happens when the number of bidders is small, because the chance of winning an item in the second round is highest for a small number of opponent bidders; and even with this effect, the pick-a-bundle is better (provided that the correct bundle size is choosen) than separate auctions. Finally, when we know that the number of bidders n is large (the exact number depending on the scenario), then, while the pick-a-bundle format with bundle size $k = 1$ is best in most cases, the revenue improvement over that of selling in separate auctions is small. The observations should provide a rule of thumb in the case that the seller would not wish to perform the complete analysis and yet still make a good choice as to the auction format that would yield maximal revenue.

8. CONCLUSIONS

In this paper, we examined the revenue of the auctioneer, when she sells items in bundles, rather than individually. Because we want these results to be applicable within existing online auctions, we rule out the design of new auction mechanisms; this is the main reason why the preexisting literature on bundling is not applicable here. On eBay, for example, it not uncommon to see a seller listing bundles of several similar items (i.e. DVDs), in order to save on the listing costs. However, our analysis has shown that this does not yield a higher profit for the seller due to the loss in expected revenue that is incurred. To address this loss, we propose the novel pick-a-bundle strategy, in which the seller lists the items in a single bundle, but the buyers bid for the opportunity to select a predetermined number of items from this bundle, with the remaining unsold items being sold in subsequent separate single-item auctions. We showed that not only does this policy reduce the loss in revenue incurred by the complete bundle, but it usually generates greater revenue than using separate single-item auctions as well (even before the savings in terms of listing fees are considered). This occurs because the pick-a-bundle auction induces additional competition between buyers who may actually prefer distinct subsets of the items. To compute the optimal bundle swiftly and accurately we developed the novel algorithm of figure 2. This algorithm can also be used in a variety of other settings, e.g. to compute the distributions of (i) the revenue that a seller makes in a position auction (Google Adwords is a variation of a position auction) from the distributions of the bids, and (ii) the social welfare (sum of winners' valuations) in a multi-unit auction. Furthermore, we examined how the strategic behaviour of the buyers changes the expected revenue of the pick-a-bundle auction and incorporated it into our analysis.

As future work, we will extend our analysis of the bid shading to the complete case of heterogeneous items. We also plan to examine what happens if we add an extra option to the bundle auction, for the winner of the first round to be able to buy additional items, e.g. by paying a predetermined percentage of the closing price, or paying the expected profit of the seller if she were to relist these items. This has the potential of increasing the seller's revenue, especially in the static case, where the bidders shade their bids in the first round; if there exists some chance that the winner selects more items in the first round, then the probability of having a second chance to buy each item decreases, therefore bidders will shade their bids less in the first round. Furthermore, the costs (listing and shipping) would be reduced. This extension also ties as well with another extension: examining how shipping fees would affect the revenue. Furthermore, we would like to investigate the application of previous results that show how the valuation distributions of participating bidders can be learned through observation of previous auctions [8]. This will allows us to apply our autonomous auction agent in setting in which these distributions are not known.

9. ACKNOWLEDGMENTS

We would like to thank Dr. Edith Elkind for comments on this work. This research was undertaken as part of the ALADDIN project and is jointly funded by a BAE Systems and EPSRC strategic partnership (EP/C548051/1).

10. REFERENCES

[1] W. Adams and J. Yellen. Commodity Bundling and the Burden of Monopoly. *Quarterly Journal of Economics*, 90(3):475–498, 1976.

[2] Y. Bakos and E. Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management Sci.*, 45(12):1613–1630, 1999.

[3] Y. Bakos and E. Brynjolfsson. Bundling and competition on the internet. *Marketing Science*, 19(1):63–82, 2000.

[4] W. Conen and T. Sandholm. Preference elicitation in combinatorial auctions. In *ACM EC'01*, pages 256–259, New York, USA, 2001.

[5] P. Cramton, Y. Shoham, and R. Steinberg. *Combinatorial auctions*. MIT Press, 2006.

[6] A. Giovannucci, J. Rodrıguez-Aguilar, and J. Cerquides. Multi-unit combinatorial reverse auctions with transformability relationships among goods. In *WINE 2005*, pages 858–867, Hong Kong, 2005.

[7] P. Jehiel, M. Meyer-ter Vehn, and B. Moldovanu. Mixed bundling auctions. *Journal of Economic Theory*, 134(1):494–512, 2007.

[8] A. X. Jiang and K. Leyton-Brown. Bidding agents for online auctions with hidden bids. *Machine Learning J.*, 67:117–143, 2007.

[9] J. O. Kephart, C. H. Brooks, and R. Das. Pricing information bundles in a dynamic environment. In *EC-01*, pages 180–190, Oct. 2001.

[10] V. Krishna. *Auction theory*. Academic Press, 2002.

[11] A. Likhodedov and T. Sandholm. Approximating revenue-maximizing combinatorial auctions. In *AAAI-05*, pages 267–274, Austin, Texas, USA, 2005.

[12] R. P. McAfee, J. McMillan, and M. D. Whinston. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383, 1989.

[13] D. C. Parkes. iBundle: An efficient ascending price bundle auction. In *ACM EC'99*, pages 148–157, Denver, Colorado, USA, 1999.

[14] M. Said. Sequential auctions with random arrivals. *(working paper), Yale University*, 2008.

[15] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. New York, John Wiley, 1980.

[16] I. A. Vetsikas, N. R. Jennings, and B. Selman. Generating Bayes-Nash equilibria to design autonomous trading agents. In *IJCAI-07*, pages 1543–1550, Hyderabad, India, 2007.

[17] S. Wu, L. Hitt, P. Chen, and G. Anandalingam. Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach. *Management Science*, 54(3):608–622, 2008.