Mining for Bidding Strategies on eBay*

Harshit S. Shah, Neeraj R. Joshi, Ashish Sureka, and Peter R. Wurman**

Department of Computer Science
North Carolina State University
Raleigh, NC 27695-7535 USA
{harshit@shah-family.net,neerajrj@yahoo.com,
asureka@unity.ncsu.edu,wurman@ncsu.edu}

Abstract. Millions of people participate in online auctions on websites such as eBay. The data available in these public markets offer interesting opportunities to study Internet auctions. We explore techniques for identifying common bidding patterns on eBay using data from eBay video game console auctions. The analysis reveals that there are certain bidding behaviors that appear frequently in the data, some of which have been previously identified and others which are new. We propose new attributes of bidding engagements and rules for classifying strategies. In addition, we suggest economic motivations that might lead to the identified behaviors.

1 Introduction

Once considered esoteric by the general public, auctions now engage millions of people daily. At any given time, there are millions of auction listings across thousands of categories on auction sites such as eBay, Yahoo, and uBid.

Auction sites rank high in both the number of visitors and the average time spent per visit, and there are myriad reasons to try to better understand how users interact with the service. From the system architecture point of view, it is important to know how the users' actions are distributed over the life of the auction. A market designer would like to know how the choice of auction rules affect the load on the server. Bidders may be able to use this information to improve their individual bidding strategies and eventually build intelligent software agents to support their economic activities, while sellers can use this information to improve their revenue. Economists may find the information valuable as they analyze the performance of these auction sites as social institutions. Finally, an understanding of normal and abnormal bidding behavior can help authorities track down fraud.

Although there are potentially many benefits from mining bidding data, the domain also poses some particular challenges for data mining techniques. First, the interactions between the bidder and the auction system have many attributes (time, value, number of bids, reputation, product description, etc.), some of which can be measured only imprecisely, and some of which are implicit, but useful, characterizations of the data.

* This paper is an extended version of a paper presented at WebKDD-02[1].
** Contact author.

In addition, the data has a temporal dependency that may be linked not only with the time period of the auction, but also to the greater economy. Finally, the data tends to be exceedingly sparse, as will become evident below.

In this work, we examine the actual behavior of bidders on eBay with the goal of trying to answer the following specific questions:

1. Is it feasible to classify the bidding behavior of individual bidders?
2. If so, what strategies are common on eBay?
3. Can we identify enough bidders to make it worthwhile?
4. Can we detect fraudulent behavior?

In order to address these questions, we accumulated bidding data from nearly 12,000 completed eBay auctions. In Section 2, we describe eBay's auction mechanism and it's salient features. In Section 3, we discuss how the data was collected and the approach that was adopted to analyze the data. We present the results of the analysis in Section 4. Section 5 discusses related work. The last section includes conclusion and future work.

2 Ebay - Model and Mechanism

All eBay auctions use an ascending-bid (English) format with the important distinction that there is a fixed end time set by the seller.[1] Ebay provides a standard auction and three variations:

Standard Auction: This is the most prominent type of listing. Here only one item (or group of items sold together) is being offered to the highest bidder.

Reserve Price Auction: The seller has a hidden reserve price that must be exceeded before the seller is required to sell. When a bidder's maximum bid is equal to or greater than the reserve price, the item's current price is raised to the reserve price amount.

Buy It Now Price: A variation of the standard auction in which a bidder can immediately win the item by choosing the Buy It Now option. If selected by the seller, this option is available until the first bid (or the first high bid that meets or beats the reserve price). A single item auction ends prematurely once a bidder exercises this option.

Dutch Auction: The seller offers more than one of the exact same items. The bidder enters the quantity of the items desired along with the price he is willing to pay per item. All winners pay the lowest winning bid price.

Regardless of the auction type, eBay uses a proxy mechanism for all submitted bids. The proxy mechanism allows a bidder to submit a maximum bid (i.e., maximum willingness to pay) with a guarantee that eBay will raise the bidder's active offer automatically until the bidder's maximum bid value is reached. We refer to the bid placed

[1] Other online sites may differ from this approach by providing a flexible end time for the auctions, which will greatly impact the bidders' strategies [2].

by the proxy system as the bidder's *proxy bid*. In a reserve price auction, the seller's reserve price is treated like any other bid; if the buyer's offer meets or exceeds the reserve (secret) bid set by the seller, the buyer's bid would be raised to that price immediately.

Ebay enforces a minimum bid increment that, along with the current ask price, determines a lower bound on bids the server will accept. The bid increment table specified by eBay defines a schedule in which the increments increase as the current ask price increases.[2]

3 Data Collection and Interpretation

We chose to collect data for the auctions of Sony Playstation 2 console (PS2) and Nintendo Gameboy Advanced consoles (GBA). The PS2 data was collected over a two-day period in October 2000 (PS2 launch in US) and for 3 weeks during January 2001. All of the data has been anonymized. The GBA data was collected from May 31 to July 29, 2001 (GBA launch in US - June 11, 2001). In total, details of 11,537 auctions were collected. We choose these product categories because, during the periods in question, supply greatly lagged demand. Because buyers had private values significantly greater than the retail value at the time when the data was collected, the data represents a liquid secondary market. In addition, it is reasonable to assume that a normal consumer is likely to need no more than one of either product.

Ebay keeps the data from completed auctions available on its website for the most recent thirty days. To collect the data, we wrote a spider that executes a search through the historical data for each product category. From the search results, the spider constructs the URLs to request individual auction data and the bidding history pages. The `bid history` page contains the details of all of the bids submitted to the auction. The spider caches both the `auction details` and `bid history` pages for each completed auction and parses them later. All requests are staggered to avoid putting too much load on eBay's server.

From the cached pages, the auction details were extracted and stored in a database. We refer to the complete set of data as data set D. Interpretation of the data, however, is subtle. The bid history pages show the time, bid value (e.g., maximum bid)[3], and bidder ID for each bid placed in the auction. However, the bid history page does not show us the proxy bids that were placed by eBay's system on behalf of the bidders. Thus, the ask-price (e.g., the price that eBay announced on its web page) that the bidder saw when submitting his bid is not recorded. This analysis is made even more complicated in reserve price auctions because the reserve price is *not* recorded on eBay's bid history page and cannot be recovered from the data supplied. Similar challenges apply to analyzing the Buy-it-now enabled auctions and the multi-unit Dutch auctions. Thus, for much of the analysis we restrict attention to only standard auctions.[4] We refer to this restricted data set as D_r.

---

[2] Ebay Help System - http://pages.ebay.com/help/basics/g-bid-increment.html.

[3] More precisely, we know the maximum bid of every bidder except the one who won the auction. For the winner, eBay records a bid that is one bid increment above the second highest bidder.

[4] Standard auctions account for 60% of the auctions in our data set.

To make the discussion more formal, consider an auction, k, on eBay. Let $j \in J_k \subseteq J$, where J is the set of all bidders that appear in the restricted data set D_r and J_k represents the subset of bidders who participate in auction k. Let B_j represent the set of bids placed by bidder j in the auction, and denote the i^{th} member of the set as b_{ij}. We refer to B_j as an *engagement*.

We define the time range of the auction as $t_s - t_{end}$, where t_s and t_{end} represent the start and end time of the auction, respectively. At any time t, such that $t_s \leq t \leq t_{end}$, eBay announces an ask price, denoted by π_t. The minimum bid at time t, denoted b_t is:

$$b_t = \pi_t + \theta(\pi_t)$$

where $\theta(\pi_t)$ is the bid increment from eBay's minimum bid increment table.

Clearly there are attributes of the bid that can be explicit in the data and which may be informative, such as the value of b_{ij} and the time at which it was placed. We also propose that the *context* of the bid can be used to infer strategy, especially the difference between the actual bid value and the minimum bid required by eBay. To compute this, we introduce a new parameter in the model. The *excess increment* of a bid is defined as the excess amount over the minimum bid, and represents the bidder's use of the proxy system. We denote the excess increment of bid b_{ij} as δ_{ij}, and calculate it by:

$$\delta_{ij} = b_{ij} - b_t = b_{ij} - \pi_t - \theta(\pi_t).$$

However, to compute the excess increment requires that we know the ask-price (π_t) at the time b_{ij} was placed. It is non-trivial to recover π_t from the data available on eBay. From the bid history, we can calculate the ask prices, and then the excess increment of all bids submitted to the auction, with the exception of the winning bid.

To calculate the ask price and excess increment of the bids for an auction, we replay the auction by sorting the bids according to bid-time and then processing each bid one by one in the manner eBay would. In so doing, we can calculate the excess increment of the bid, the new ask price, and the high bidder. Unfortunately, there is no official publication provided by eBay that precisely describes the bid processing algorithms. We discovered the process by reverse engineering from the data using the descriptions on eBay's Help System, and our own experience developing auction systems, with some details worked out by trial-and-error.

The procedure for admitting bids is as follows:

i. The starting price is the minimum initial bid set by the seller.
ii. The first bid does not affect the current ask price, but the bidder becomes the high bidder at the starting price.
iii. Any new bid has to at least match the minimum bid at that time.[5]
iv. When a bid is admitted, the new ask price and high bidder are determined as follows:

[5] We found one exception to this rule. If the bid is submitted within a second of the previous bid, eBay may accept the later bid before processing the prior bid. Thus, we occasionally see bids that are lower than the new minimum.

a) Compare the current high bidder's maximum bid and the new bid. The higher bid determines the new high bidder. A proxy bid is placed on behalf of the new high bidder which becomes the new ask price.
b) The high bidder's proxy bid is typically the second highest bidder's maximum bid plus one bid increment.
c) The exception to (b) occurs when the high bidder's maximum bid is less than one bid increment over the second highest bidder's maximum bid. In this case, the high bidder's proxy bid, and the ask price, is set to his maximum bid.

v. When two bids are for the same amount, the earlier one takes precedence.
vi. A bidder cannot lower his previous maximum bid.
vii. A bidder can raise his bid when he is winning, and in general, it will not change his proxy bid. The exception is when the bidder is winning and in condition (iv.c) above. In this case, if the winning bidder raises his offer, eBay would raise the current price to the minimum increment above the second highest maximum bid.

Table 1. A sequence of bids. The four columns on the right are computed from the data in the left columns. The actual value of the last excess increment is unknown, but a minimum can be computed.

Bid Number	Bid Time	Bidder ID	Bid Amount	Excess Increment	New High Bidder	New Ask Price	New Minimum Bid
1	June-17-01 15:06:20	63246	89.99	0	63246	89.99	90.99
2	June-17-01 15:53:20	59729	95	4.01	59729	90.99	91.99
3	June-17-01 17:51:12	59207	95	3.01	59729	95	96
4	June-17-01 17:53:06	59207	96	0	59207	96	97
5	June-17-01 17:56:57	45020	100	3	45020	97	98
6	June-17-01 17:58:32	59207	98	0	45020	99	100
7	June-17-01 17:58:45	59207	100	0	45020	100	101
8	June-17-01 18:01:08	45020	102.5	2.5*	45020	102.5	105

Table 1 shows how ask price and excess increment of bids are calculated for a sample auction. The data is from a real GBA auction, though the bidder IDs have been anonymized. The auction is a 3-day auction that started on June 14, 2001 with a scheduled end on June 17 at 18hr 3min and 14sec PST. The starting price set by the seller was $89.99.

Bidder 63246 places the first bid of $89.99. The minimum bid and ask price are equal at the start of the auction (by Step i). Hence, the excess increment for this bid is zero. After this bid is processed, the high bidder is 63246. The first bid does not change the ask price (by ii). However, the new minimum bid becomes $90.99. (in this price range, eBay requires a $1 increments; above $100 the increment rises to $2.50). Bidder 59729 places the second bid at $95 and becomes the high bidder (by iv.a). The excess increment is $4.01—the difference between the minimum bid ($90.99) and bid amount ($95)—and the new ask price is one bid increment over the second highest bid (by iv.b) i.e. $89.99 + $1.00. 59207 places the third bid and ties with 59729. Because bidder 59729 placed his bid earlier, he remains the high bidder (v), and so bidder 59207 raises his bid to become the high bidder. Later in the auction, we see that when bidder 45020 places the eighth bid, he was already the current high bidder (by the tie breaker). Generally, this would not change the ask price. However, this case it the exception to rule (vii) and eBay raises the current price to one increment over $100 i.e. $102.5. The auction ends with bidder 45020 winning the item at $102.50. Note that we do not know the actual value of 45020's winning bid. The excess increment value for the winning bid (eighth bid) is a lower bound on the actual excess increment.

4 Analysis and Results

In addition to more straightforward views of the data (i.e., bid value charted over time), we found that examining graphs of excess increment values for individual bidders to be a useful way to visualize some aspects of a bidder's strategy. Figure 1 represents a graph of excess increment values of all bids submitted by bidder 62013 across all 3-day auctions in which he participated. The horizontal axis denotes the time of the bid from the auction's end (in seconds). The vertical axis denotes the excess increment value of the bids (in $). All bids for a specific auction are connected by a dotted line and each auction is represented with a different color. An 'h' above a marker denotes that the bidder became the high bidder after this bid was processed. Analyzing bidding behavior often required going back and forth between different views to construct a complete picture of the auction context.

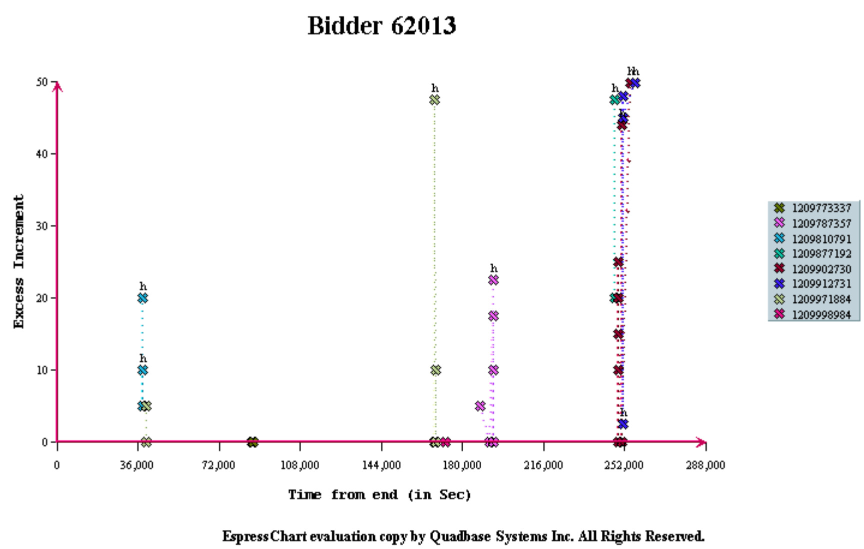

Fig. 1. Bidder 62013's bidding behavior in eight auctions.

We observe in Figure 1 that in several auctions bidder 62013 bid repeatedly in a short span of time, with increasing excess increment values, until he became the high bidder. This is one of several patterns that commonly appear in the data. To measure the actual frequency of these patterns in the restricted data, we developed tests to label individual engagements. Recall that an engagement is the set of all bids by an individual bidder in an individual auction. There are 49,523 engagements in D_r, varying in size from 1 to 24 bids. Let E denote the set of all engagements in D_r, and define $E_n \subseteq E$, to be the subset of engagements in which the bidder has placed exactly n bids in that particular auction. Figure 2 shows the distribution of engagement sizes. About 66% of engagements belong to subset E_1. The rest of the engagements exhibit multiple bids, with nearly 15% having three or more bids.

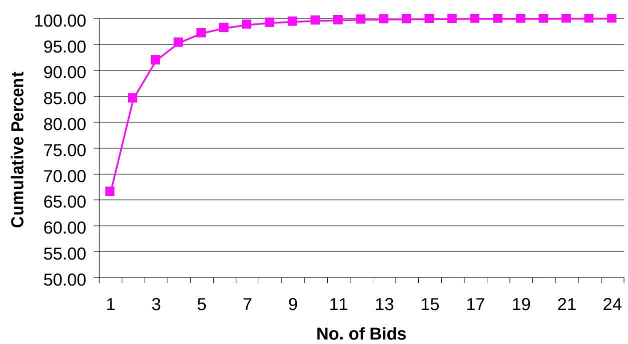

Fig. 2. Number of bids in engagements.

4.1 Single Bid Engagements (E_1)

We find that, irrespective of the auction duration, a large number of single-bid engagements have bids placed near the end of the auction, a behavior we refer to as *late bidding*. Nearly 58% of the bids were classified as late bidding. Figure 3 shows the distribution of bids during the last hour. A significant fraction of these bids are submitted in the closing seconds of the auction, a practice called *sniping*. This behavior arises despite advice from both auction theory and eBay itself that bidders should simply submit their maximum willingness to pay, once, early in the auction. Hence, it is inviting to attribute sniping behavior as being primarily due to naïve, inexperienced or irrational behavior. However, Roth and Ockenfels [3, 4] show that the sniping and late bidding need not result from either irrational behavior or common value properties of the objects being sold. They study a model of eBay's marketplace in which expert buyers have a collusive equilibrium of sniping. Although some bids are submitted too late to be accepted, the net effect is that prices are lowered enough by the decreased competition to compensate for the occasional rejected bid.

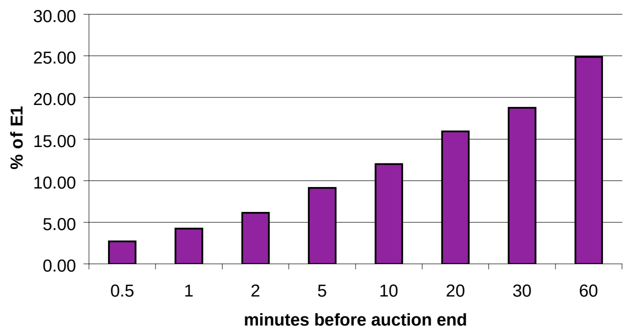

Fig. 3. Distribution of bids during the last hour.

Bidders who do seem to follow eBay's advice are also identifiable in E_1. We refer to the bidder classification by Bapna, et al. [5] in which they identify a bidder type called *Evaluators*. Evaluators are modeled as users who have a clear idea of their valuation and have the following characteristics:

– They bid once, early, and at a high value.
– Their bids are usually significantly greater than the minimum required bid at that time.

To differentiate late bidders from evaluators, we restrict our attention to only those engagements in which the single bid was placed at least one day before the end of the auction (E_1^*). However there are several reasons that make identifying evaluators non-trivial. First, it is not uncommon that the game consoles auctioned on eBay are bundled with other things like games or extra joysticks that affect buyer's willingness to pay. Second, the bidder's valuation may change over time as market price decreases.

To identify evaluator behavior, we group engagements in E_1^* according to bidder and the type of item i.e. PS2 and GBA. The standard deviation of the bids by an individual is calculated. We then consider two cases. First, if the standard deviation falls within a specified lower value, σ_{lower}, we classify the corresponding engagements as an evaluator's behavior. Figure 4 compares the bid values and the ask prices for bidder 45122's six engagements in E_1^*. The standard deviation of the bids in these six auctions is 0.52. We observe that the bids are significantly greater than the ask price at that time.

The second case involves behaviors where the standard deviation is higher than σ_{lower} but less than an upper value, σ_{upper}, and is designed to capture fluctuations due to the content of the auctioned bundles, and relies on the assumption that the final price of the auction reflects the true market value of the object. When $\sigma_{lower} < \sigma < \sigma_{upper}$, we examine the correlation of the bid amounts with the final price of the corresponding auctions (r_{final_price}) and with the ask prices of the auction at bid time (r_{ask_price}). If there is a strong correlation between bid amounts and final prices, we assume the

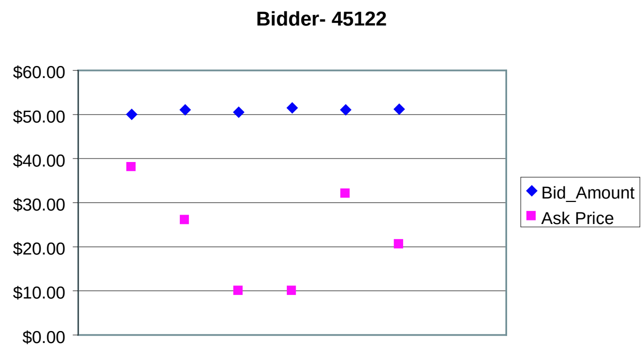

Fig. 4. Example of evaluator behavior.

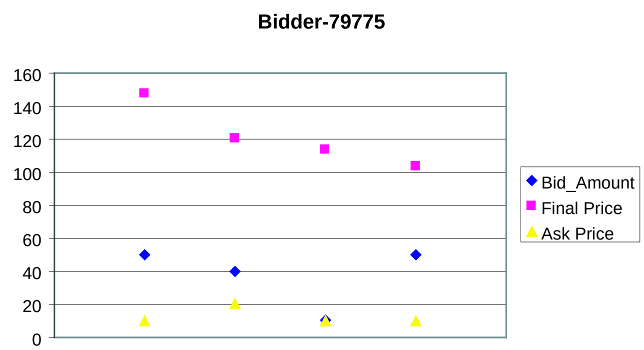

Fig. 5. Another example of evaluator behavior.

deviation in bid amounts is attributable to different options in the sellers' offerings. Further, a weak correlation between bid amounts and ask prices support classifying the behavior as evaluator behavior because it suggests that the bidder's action does not depend on the ask price. Figure 5 shows an example having such characteristics. In this case, r_{final_price} = 0.30 (Strong-Weak association) and r_{ask_price} = 0.08 (Little, if any association). Thus, we also classify this case as depicting evaluator behavior.

Keeping σ_{lower} = 10.00, σ_{upper} = 20.00, $r_{final\_price} > 0.2$ and $r_{ask\_price} < 0.2$, we have 13% of E_1 (comprising of 1074 bidders i.e. PS2 - 501 and GBA - 573), exhibiting the evaluator behavior. There is a scope for improving this analysis by using better statistical methods. It would also be interesting to explore other explanations for such behavior.

4.2 Multiple Bid Engagements

Multiple-bid engagements, $E_{>1}$, account for more than 30% of E. One behavior common in $E_{>1}$ is the *skeptic* behavior. We define the skeptic as a bidder who submits multiple bids, all of which have zero excess increment. That is, the bidder always bids the minimum increment over the current ask price. One explanation for this behavior is that the bidder might be naïve or skeptical of eBay's proxy system and hence would always bid the minimum acceptable bid. Figure 6 shows a bidder exhibiting skeptic behavior. He participated in four auctions, submitting multiple bids all of which had zero excess increment. Around 18% of E_2 (1658 out of 8936) are classified as skeptics. In the remaining group i.e. $E_{>2}$, nearly 10% (738 out of 7649) follow the skeptic pattern.

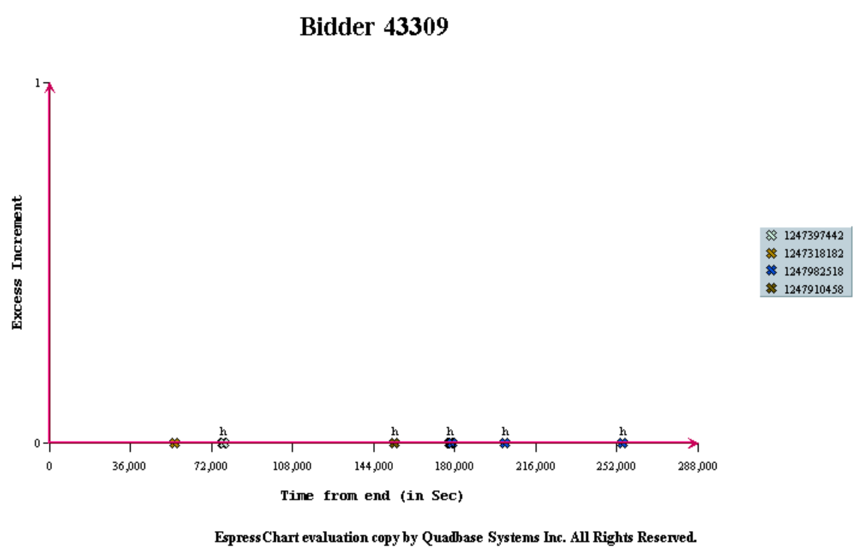

Fig. 6. Example of skeptic behavior.

Unmasking is another behavior that is common in $E_{>2}$. The pattern involves a series of closely placed bids, with variable excess increments, as shown in Figure 1. One possible rationale for this behavior is that the bidder is trying to expose the maximum bid of someone else's proxy bid. Since a bidder is informed immediately if his bid becomes a winning bid, it is relatively easy to execute this strategy quickly. There are several possible explanations for this behavior, including its role in shill bidding (described in Section 4.3). Another possible explanation is that the bidder continues the unmasking process until he becomes the high bidder, until he reaches his maximum willingness to pay, or until he has enough information to decide to move on to another auction.[6]

We identify unmasking behavior by the following characteristics:

- The bidder places multiple bids in a short span of time. (We vary the time span from 2 min to 10 min)
- There are no bids by other bidders in-between these bids.
- At least one bid in the engagement has an excess increment greater than zero.
- At least two non-winning bids are submitted as part of the sequence.

We find the unmasking behavior in 40%-43% of the $E_{>2}$, when we vary the time of successive bids from 2 minutes to 10 minutes. The fact that the vast majority of this behavior is identifiable with a 2-minute window suggests that it is a standard tool in the bidders' bag of tricks.

Around 5% of $E_{>2}$ engagements have behavior in which the excess increment of successive bids decreases. Figure 7 is representative of engagements in this class, showing a typical behavior that appears as slanted lines. This behavior is probably attributable to a bidder who periodically attends to the auction to check whether he is outbid. The excess increment decreases as the end of the auction approaches because the price is increasing. Although this type of behavior was easy to identify by visual inspection, we have not yet identified the key attributes that form a coherent cluster for automatic detection.

4.3 Identifying Types of Bidders

The above classifications were derived largely by hand and somewhat blurred the line between classifying engagements and classifying bidders. In this section, we re-establish the distinction and investigate whether we can identify types of bidders based upon their engagements using a clustering algorithm. Using variations of the above classification rules,[7] we labeled each engagement as either an evaluator, a skeptic, a snipe,

[6] Though it is not possible to know the bidder's true valuation, it is logical to conclude that if he stopped unmasking without being a high bidder at the end, he might have reached his true valuation or switched to a different auction.

[7] Variations of the rules were used when the rules required looking at multiple engagements. In particular, we approximated the evaluator rule by assuming an engagement that consisted of a single bid that occurred at least a day before the auction's end and which was greater than 90% of the clearing price was an evaluator bid. we also extended the skeptic category to engagements with only one bid.

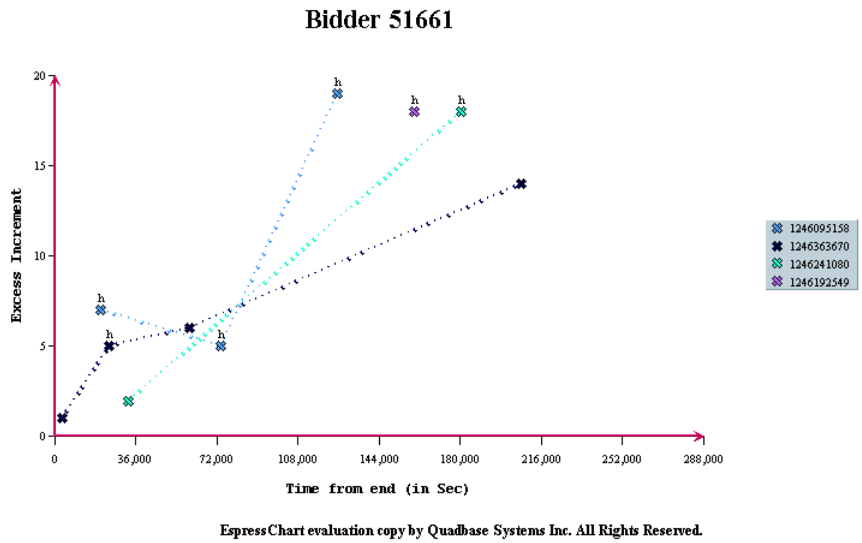

Fig. 7. Engagements of Bidder 51661.

an unmasking, or as unidentified. We then constructed a *bidder profile* for each bidder which we represented as a vector where the elements of the vector were the proportion of engagements of each type. For instance, if buyer A had two engagements that classified as skeptic, one unmasking, and one unidentified, her vector would be $\langle 0, 0.5, 0, 0.25, 0.25 \rangle$.

Using the SAS Enterprise Miner software, we ran a standard cluster analysis on the bidder profiles. Table 2 shows the results when we set the desired number of clusters to five. The first cluster consists primarily of bidders whose engagements did not fall into one of our categories. This cluster represents 88% of the bidders. However, the other 12% of the bidders do cluster nicely into the four groups with centroids firmly in one type of bidding behavior.

Table 2. Results from clustering the bidder profiles.

Cluster	Frequency	Evaluate	Skeptic	Snipe	Unmask	Unidentified
1	18939	0.0037	0.0001	0.0193	0.0150	0.9619
2	227	0.9793	0	0.0103	0.0066	0.0038
3	1377	0.0002	0	0.9711	0.0025	0.0263
4	798	0	0	0.0008	0.9772	0.0220
5	12	0	1	0	0	0

A couple of conclusions can be drawn from this analysis. First, consistent skeptic behavior was extremely uncommon, most likely because it is unlikely to be successful. Second, perhaps not surprisingly, snipers were the most common identifiable cluster. However, it is surprising that the cluster representing bidders who consistently use unmasking as a strategy was quite large—more than half as big as snipers. Finally, evaluators make up only about 1% of the sampled bidding population.

When we reduce the number of desired clusters, Enterprise Miner drops clusters from the list in decreasing order. When we increase the number of desired clusters to, for example, seven, about 1300 bidders move from cluster one into hybrid groups in which about half of the engagements are unidentified, and the other half fit in one of the other categories.

5 Fraud Detection

Incidents of fraud in online auctions are increasing rapidly; according to the National Consumer League[8] and Federal Trade Commission,[9] online auction fraud is the number one type of Internet fraud. One of the more subtle types of auction fraud to detect is *shilling*. Shilling, also known as colluding or bid rigging, is a method by which sellers try to hike up the prices in an auction by placing buy bids under aliases or through associates. Shilling is often very difficult to detect primarily because of the sheer number of online auctions and users, and the ease with which users can create new accounts. Experienced shills can easily disguise themselves under this overload of data.

The following attributes are claimed to be indicators of shilling behavior, and are particularly applicable to our eBay data:[10]

1. There is a strong association between a seller and a buyer, or a ring of buyers. By association, we mean that the shill(s) appear very frequently in auctions hosted by the seller.
2. The shill wins the auctions infrequently (if at all). An aggressive shill may occasionally win the auction, but since he is working for the seller, the two parties need never consummate the deal.
3. The bids placed by the bidder would be significantly higher than the current ask price. This is intuitive as the sole purpose of the bid is to hike the ask price.
4. In the context of eBay, a shill is likely to apply the unmasking strategy in order to extract as much as possible from the proxy bid.
5. The shill would eschew sniping and very late bidding as this would put him at risk of winning the auction and not permit the real buyers enough time to respond to his bids.

To test the feasibility of identifying shilling, we performed association analysis between the users (sellers and buyers) participating in the auctions. We construct an abstraction of the engagement data with an entry for each buyer-auction and seller-auction pair. The user-ids of the sellers were prefixed with the letter ‘s’ while those of the buyers

[8] http://www.natlconsumersleague.org/internetscamfactsheet.html
[9] http://www.ftc.gov
[10] Descriptions of common features of shilling behavior are available on sites like AuctionWatch (http://www.auctionwatch.com).

remained the same. For example, the entry {465768718, s71497} indicates that seller 71497 participated in auction 465768718, while the entry {465768718, 59493} indicates that buyer 59493 participated in the same auction.

Table 3. Sample rules generated by association analysis

Rule	Confidence	Transactions	# won
43645 ⇒ s60993	100	26	1
s60993 ⇒ 43645	96.3	26	1
s43998s ⇒ 62466	76.92	10	0
62466 ⇒ s43998	83.33	10	0
77812 ⇒ s43998	80	8	0
49153 ⇒ s43998	100	6	0
59949 ⇒ s43998	100	5	0
56146 ⇒ s49449	100	5	0
45581 ⇒ s50336	86.21	25	2
66240 ⇒ s50336	92.86	13	1
70053 ⇒ s50336	91.67	11	1
68111 ⇒ s50336	83.33	10	0
56992 ⇒ s50336	90	9	1
68746 ⇒ s50336	100	6	0
82377 ⇒ s50336	100	5	0
42628 ⇒ s50336	100	5	2
64621 ⇒ s50336	100	5	0
67124 ⇒ s50336	100	5	1
71139 ⇒ s50336	100	5	1
47319 ⇒ s50336	83.33	5	1
55477 ⇒ s53124	100	5	3
64512 ⇒ s53666	80	12	0
50732 ⇒ s53666	100	11	0
64346 ⇒ s53666	100	10	0
77145 ⇒ s53666	85.71	6	0

We used the SAS Enterprise Miner for the association analysis. A rule like A ⇒ B suggests that the presence of user A in an auction predicts the presence of user B. The *support* is the number of times A and B appeared together in an auction divided by the total number of auctions (approximately 7000 in D_r). The *confidence* is the number of times A and B appeared together in an auction divided by the number of times A appeared in an auction. Finally, the number in the *transactions* column is the number of auctions in which A & B both participated. With a minimum confidence of 75% and a minimum of five transactions, 79 two-item rules were found representing 69 different buyers. Table 3 shows a subset of the results, clustered by seller ID. The rightmost column indicates the number of times the buyer was the high bidder among the transactions. Ten buyers appear in two rules because the complementary relationships (B ⇒ A) was also found to have high confidence. No buyer had a strong relationship with more than one seller.

It is clear from this data that there were very strong (and suspicious) relationships between some sellers and buyers. For example there is strong support and confidence for the two rules that involve seller s60993 and buyer 43645. All 26 auctions in which buyer 43645 bid were hosted by s60993, and all but one of seller s60993's auctions had buyer 43645 participating. Furthermore, 43645 won only once in 26 tries. Other sellers seem to be more sophisticated in their use of shills. Seller s50336 had strong relationships with 12 different bidders, while sellerS s43998 and s53666 each had four strongly related bidders, none of whom won any auctions.

The results of this preliminary analysis clearly suggest instances of shilling. The strong associations between some bidders and sellers are further supported by the fact that these bidders seldom won the auctions. In addition, the results show that the alleged shills are using multiple accounts to place their bids.

This analysis incorporated only the simplest attributes of the relationship between buyers and sellers. Including other attributes may enable more precise classifications. For example, our analysis did not consider the time at which the auctions were held. We expect that a relationship that remains stable over a long period of time would be more suspicious. In addition, our data set was restricted to two auction categories only (GameBoy and PlayStation 2); a seller may potentially be involved with several categories and data about a seller-bidder association over these categories could be more evidence. With analysis based on a single category, it is plausible that a buyer trusts a seller, or has a geographical proximity to the seller, that makes him favor the seller. An examination of the feedback written by the buyer and seller may also be useful evidence, particularly negative feedback serving as an anti-correlant with shilling. Finally, information about whether a transaction between a seller and a buyer actually occurred is not available to us. Again, we would expect that the consummation of a transaction would be negatively correlated with shilling. In the case of these particular items, it would also be suspicious to see a buyer to continue to bid in auctions even after winning an object.

5.1 The Relationship Between Shills and Bidding Behaviors

One of the oft claimed indications of shill bidding was behavior consistent with what we have identified as unmasking. To determine if unmasking, or any other strategy, was an indicator of the presence of a shill, we examined the bidding behavior of those buyers identified as shills. The 69 bidders identified above participated in a total of 749 engagements. Table 4 shows the frequency of each type of bidding strategy among the 749 engagements.

As expected, sniping is not a common behavior among these potential shillers and it occurs much less frequently than in the larger population. However, none of the other bidding behaviors seem particularly predictive of the presence of a shill. In particular, unmasking occurs a little less frequently among the alleged shills than it does among the general population (about 4% versus 6%).

Table 4. Frequency of bidding strategies among the suspicious bidders.

Behavior	Frequency
Evaluate	12
Skeptic	0
Snipe	5
Unmask	32
Unidentified	700

6 Related Work

Internet auctions have attracted the attention of auction theorists and econometricians, and several recent papers have examined online markets. Ebay, being the most popular site, is the first choice of many such studies. David Lucking-Reiley, et al. [6, 7] analyze the effect of various eBay features on the final price of auctions. They find that seller's feedback rating have a measurable effect on his auction prices, with negative comments having a much greater effect than positive comments. Houser and Wooders [8] work finds a similar effect of the feedback ratings on the auction price. Roth and Ockenfels [3] observe late bidding in online auctions and suggest that multiple causes contribute to late bidding, with strategic issues being related to the rules about auction ending. Ünver [9] analyses the evolution of strategic multiple and last minute bidding using artificial agents. The work found support for multiple bidding in both private-value and common-value models. Bapna, et al. [5] reveal that the traditional theoretical assumptions regarding the homogeneity of bidder strategies are violated in online auctions. This conclusion is supported by our analysis as well.

7 Conclusion and Future Work

This paper serves as an exploratory analysis of the feasibility of various data mining tasks on data collected from eBay. We introduce the concept of excess increment; a useful attribute of bids that is indicative of bidding strategies. We are able to identify the following behaviors: sniping, late bidding, evaluator, skeptic and unmasking. The former three strategies have been previously suggested by other researchers, and the latter two strategies are newly identified. These behaviors account for a significant portion of engagements in our sample data and confirm that varieties of bidding strategies are common on eBay. Unmasking, in particular, is a newly identified strategy that may have an interesting economic explanation.

This paper takes an empirical approach to identify bidding strategies and shows that data mining techniques may be used to enhance the results. It is also important to note that a variety of attributes must be examined to automatically classify bidders' behavior. Some strategies, like unmasking, exist within a single engagement, while others, like evaluators, require examining a bidder's behavior across multiple engagements. However, we show that with a relatively coarse method of classifying engagements, a clustering algorithm can successfully identify groups of bidders who behave in a consistent manner.

Shill detection in online auctions is an interesting and important extension to this work. The data collected in this study, and the models of bidding strategies, can improve the realism of simulations of economic markets. These simulations might help us to develop software agents that facilitate effective participation in online markets.

Acknowledgements

This work was supported by the National Science Foundation under CAREER award No. 0092591 and the e-Commerce@NCSU program. In addition, we would like to thank Jon Doyle, for his role as committee member and data mining instructor, and the members of the Intelligent Commerce Research Group at NCSU for their encouragement and critiques.

References

1. Shah, H.S., Joshi, N.R., Wurman, P.R.: Mining for bidding strategies on eBay. Submitted for publication (2002)
2. Lucking-Reiley, D.: Auctions on the internet: What's being auctioned, and how? Journal of Industrial Economics **48** (2000) 227–252
3. Roth, A.E., Ockenfels, A.: Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. (forthcoming)
4. Ockenfels, A., Roth, A.E.: The timing of bids in internet auctions: Market design, bidder behavior, and artificial agents. AI Magazine **23** (2002) 79–87
5. Bapna, R., Goes, P., Gupta, A.: A theoretical and empirical investigation of multi-item on-line auctions. Information Technology and Management **1** (2000) 1–23
6. Lucking-Reiley, D., Bryan, D., Prasad, N., Reeves, D.: Pennies from eBay: The determinants of price in online auctions. Technical report, University of Arizona (2000)
7. Katkar, R., Lucking-Reiley, D.: Public versus secret reserve prices in ebay auctions: Results from a pokmon field experiment. Technical report, University of Arizona (2000)
8. Houser, D., Wooders, J.: Reputation in auctions: Theory. Technical report, University of Arizona (2001)
9. Ünver, M.U.: Internet auctions with artificial adaptive agents: Evolution of late and multiple bidding. Technical report, Koc University, Istanbul, Turkey (2001)