

# Two-level workload characterization of online auctions

Vasudeva Akula<sup>a</sup>, Daniel A. Menascé<sup>b,\*</sup>

<sup>a</sup> *The Volgenau School of Information Technology and Engineering, George Mason University, Fairfax, VA 22033, USA*

<sup>b</sup> *Department of Computer Science, George Mason University, Fairfax, VA 22033, USA*

Received 11 February 2005; received in revised form 23 December 2005; accepted 4 July 2006

Available online 11 September 2006

## Abstract

Online auctions are rapidly becoming one of the significant forms of electronic commerce for buying and selling goods and services. A good understanding of the workload of auction sites should provide insights about their activities and help in improving the quality of the service provided to their users. This paper presents a site level and a user level workload characterization of a real online auction site using data collected by automated agents. The main contributions of this paper are as follows: (i) a detailed workload characterization of a real auction site; (ii) an analysis of the presence of heavy tailed distributions in this workload; (iii) an analysis of the bidding activity during closing minutes of auctions; and (iv) an analysis of the arrival rate process of bidders and bids within clusters based on different attributes. These results can be used to devise dynamic pricing and promotion models to improve revenue throughput of online auction sites.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Internet auctions; Workload characterization; Auction workloads; Online auction site performance

## 1. Introduction

As the Internet becomes increasingly ubiquitous, many traditional auction businesses are moving into the online auctions space joining winners in this market space, such as eBay [9] and Yahoo! Auctions [24]. Businesses with fixed price selling models are adopting online auctions to dispense excess inventory in B2C auctions. Companies that use the B2B model are using auctions to extract optimal prices from their vendors. English auctions are one of the popular categories of auctions in which bidders compete by increasing the price of an auction until either the time is over or no one else competes with a higher bid or a combination of both. These auctions can be of different types such as fixed closing time, or by varying the closing time until not enough activity on the auction is found. Some auctions allow users to buy the auctioned item immediately with a “buy it now” feature that allows users to buy the

item instead of waiting until the completion of the auction and competing with other bidders. Most online auction sites allow users to place automatic bids called proxy bids, which allow a user to specify a maximum amount and automatically place a bid if another user bids more than the current bid but below the user’s maximum bid. Alert services to indicate creation of a new auction or changes in an existing auction, and watch lists to monitor the progress of an auction are common among many online auction sites.

Online auction sites have significantly different workloads compared to other forms of e-commerce sites such as e-tailers. The activity of an auction site spikes during the closing minutes and drops immediately after the closing of each auction [2]. Building websites that can scale well is one of the challenges faced by architects of e-commerce sites [21]. Benchmarks can be used to compare competing architectures. One such available benchmark for e-commerce sites is the TPC-W benchmark [23], which mirrors the activities of e-tailers such as online bookstores. An e-commerce site developed for auctioning goods online, such as eBay, exhibits significantly different characteristics from

\* Corresponding author.

E-mail address: [menasce@gmu.edu](mailto:menasce@gmu.edu) (D.A. Menascé).

e-tailors. Thus, TPC-W is not well suited for auction-related e-commerce sites. The analysis presented in this paper will be used in designing a benchmark for online auction sites [4]. This benchmark and its workload generator will be used in our ongoing research dealing with designing and analyzing algorithms and resource allocation policies to increase the revenue throughput (i.e., dollars/s generated by a site [22]) of auction sites.

A good understanding of the workload of auction sites should provide insights about their activities and help in the process of designing business-oriented metrics and designing novel resource management policies based on these metrics, as done in [22]. This workload characterization should be taken into account in the development of realistic benchmarks and in simulation and performance studies. Our workload characterization includes detailed analysis of real data, obtained through automated agents, from online auctions to uncover patterns related to their major activities. We also analyzed how the features of the workload change within clusters determined by some specific rules. For example, it is quite likely that the distributions related to closing time will be different for popular auctions than for non-popular ones. The same might be true for expensive auctions compared to lower priced auctions.

Previous work in workload characterization of e-commerce sites in general was presented in [5,8,18–20] and was discussed in the more specific case of auction sites in [2,6]. We address in this paper different user behavior aspects that were not treated in our previous work on workload characterization [2,16] in addition to using a significantly larger dataset.

The main contributions of this paper are as follows: (i) a detailed workload characterization of a real and large auction site; (ii) analysis of the presence of heavy tailed distributions in the workload; (iii) study of sharp increases in bidding activity during closing minutes of auctions; and (iv) analysis of the arrival rates of bidders and bids within clusters based on various auction attributes. The results presented in this paper can be used in a variety of ways to improve user experience on an online auction site as well as to increase the site's revenue. Here, are some examples:

- As indicated in a benchmark study by Empirix Inc. [10], end-user response time is a significant differentiating factor in selecting auction sites. We used the workload characterization presented in this paper to develop techniques to improve end-user response time in two different ways. First, we devised an algorithm that proposes to auction creators new closing times for their auctions in a small time window around the originally proposed closing time [17]. This rescheduling algorithm was shown to smoothen the load spikes experienced by auction sites when several auctions close simultaneously. Second, we designed server-side auction caching policies based specifically on auction-related metrics [15]. These

policies showed that very high cache hit ratios can be obtained by caching a very small percentage of the auctions.

- Understanding the workload of an auction site in terms of the popularity of auctions can help sellers in devising dynamic pricing and promotion models by offering discounts to auctions that are not gaining popularity as opposed to others.
- Eighty percent of an auction site's revenues are generated by 20% of their users [7]. Our user popularity results can be used by site managers to develop user profiles for personalizing auction sites to improve usability and quality of service for the users of these sites.
- Business-oriented metrics can be developed based on our analysis to improve the revenue throughput of auction sites.

The rest of the paper is organized as follows. Section 2 describes some basic terms and notation used in the paper. Section 3 describes the motivation and approach used in our analysis. Section 4 discusses the experimental setup used in data collection and describes the characteristics of the data collected. Section 5 provides a site level workload characterization that includes cluster analysis, multi-scale analysis of auction creation activity and bid activity within auctions, closing time analysis and arrival rates of auctions and bids. Section 6 presents a user level workload characterization that includes analysis of popularity of winners, sellers and bidders in online auctions, and a detailed analysis of auctions and bids in the clusters based on unique number of bidders and closing price. Section 7 summarizes our findings and presents some concluding remarks.

## 2. Basic terms and notation

An online auction is a method of selling on the Internet in a public forum through open and competitive bidding. A bid is a prospective buyer's indication or offer of a price he or she will pay to purchase an item at an auction. Proxy bidding is the process of submitting a confidential maximum bid to an auction service. The auction will automatically increase the bid to maintain the high bid. Proxy bidding stops when the bid wins the auction or reaches the limit of the proxy bid.

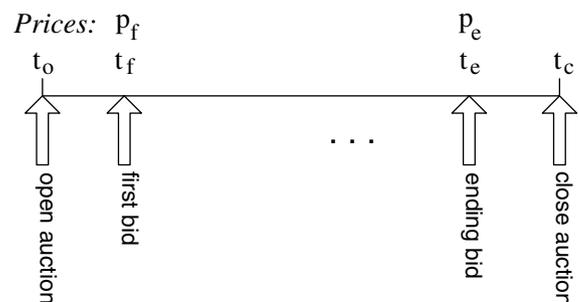


Fig. 1. Auction times and prices.

Fig. 1 introduces several terms and notation used in the paper. The life time of an auction is the difference  $t_c - t_o$  between the time,  $t_c$ , at which the auction is scheduled to close and its opening time,  $t_o$ . The first bid occurs at time  $t_f$  with a price of  $p_f$ . The last bid of the auction occurs at time  $t_e$  with a price of  $p_e$ .

The age,  $A$ , of an auction is defined as the percentage of time elapsed since the auction’s opening time relative to its life time. In other words,

$$A = (t_{\text{current}} - t_o) / (t_c - t_o) * 100. \tag{1}$$

The relative price,  $p_r$ , of an auction is defined as the ratio of the current price of the auction relative to its closing price, given as a percentage. Thus,

$$p_r = (p_{\text{current}} / p_e) * 100. \tag{2}$$

### 3. Motivation and approach

Successful deployment and operation of an online auction platform requires knowledge of its processes and workloads. Auction systems receive requests from not only individual users, but also from software programs, automating bidding and auction creation functionality. For example, companies like Andale [1] offer tools for buyers and sellers to automate some of their functions on eBay. Nearly, 40% of eBay’s listings are derived from eBay web services among which 20% are driven by sellers using third party software applications. As of June 2004, eBay received over 1 billion web service requests per month [12].

Major activities on online auction sites include creating auctions, placing bids, searching and browsing for auctions, payments, and requesting online help. Auction sites provide various miscellaneous services such as rating services, promotion tools, and notification of auction status. Auction systems depend on dynamic content generation

and typically use multi-tiered architectures. Auction systems must be designed to scale with increasing load and handle unexpected spikes. Several factors contribute to the load on auctions including time of the day, number of bidders in individual auctions, alert services, and proxy agents. The load on auctions tends to be higher for popular auctions, because they need to generate more alerts, manage a higher number of bids, and generate more proxy bids.

The proper design and sizing of auction sites requires a thorough understanding and characterization of its workload. Studying the workload of e-commerce sites at multiple levels of aggregation and from different points of view can be quite useful as illustrated by Menascé et al. [19,20]. In this paper, we present a two-level view of the workload as illustrated in Fig. 2: site level and user level. The site level workload characterization presents the workload from the site’s point of view without looking into specific auctions, bids, or users. This characterization includes cluster analysis, multi-scale analysis, closing time analysis of the activity on the site and inter-arrival times of auctions and bids on the site. The user level workload characterization takes into consideration factors such as different types of users, classes of auctions, and price ranges, and analyzes the workload specific to each class. This includes an analysis of the popularity of sellers, bidders, auctions, and categories. Analysis of bidding activity in auctions grouped by the same number of unique bidders and in different price ranges is also presented under the user level workload characterization.

### 4. Experimental setup

We collected data for auctions created during the month of January 2003 from the Yahoo! Auctions site using auto-

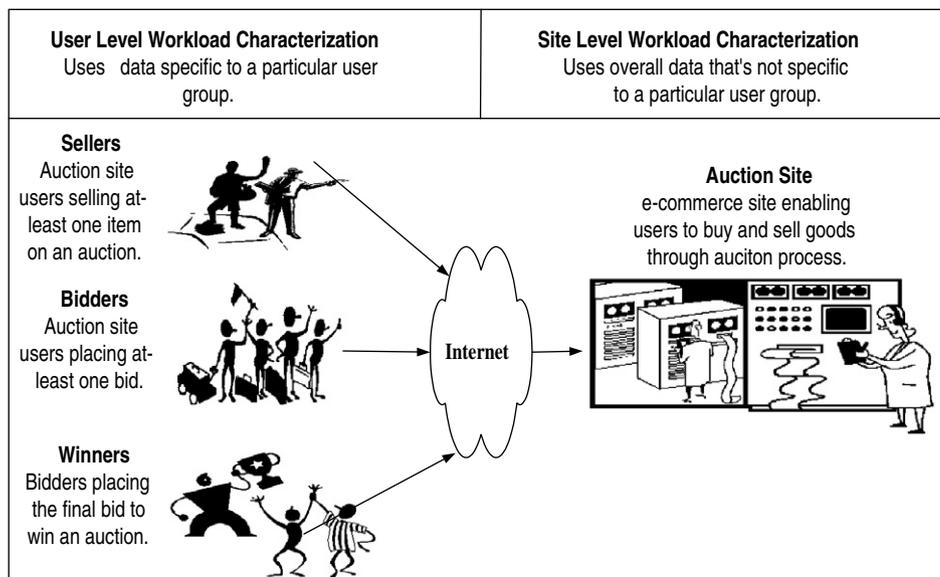


Fig. 2. Two-level workload characterization approach.

mated data collection agents. The data collection agent was designed based on the fact that for this particular online auction site, auction information was available online after the close of an auction and can be fetched using a URL that is constant except for a sequential auction ID embedded in the middle of the URL. A UNIX cron job invokes the data collection agent at programmed regular time intervals. The data collection agent is written in Java and submits HTTP requests to the auction site with a dynamically generated URL that contains the next auction ID to collect. The retrieved auction HTML pages are sent to an awk shell script that parses the HTML page to extract the auction and bid information. The output of the parsing program is used as input to the Oracle SQL Loader program, which loads the information into a relational database. After successfully extracting the auction and bid information, the auction item is marked as processed in the database so the process is not repeated for already processed auction items when the program runs next time. The information collected for each auction includes its opening and closing times, price information, the list of all bids placed during the auction, including bidder id, price and time of each bid and an indication of whether the bid was placed manually or by a proxy agent. Several SQL queries were written against the database to identify significant patterns and generate the graphs discussed in the rest of this paper.

#### 4.1. Data collection

The data collection agent gathered a total of 344,314 auction items created during the month of January 2003, belonging to over 2000 categories. A total of 1.12 million bids were placed on these auctions before their closing time, which varied from the same day of opening to 90 days from the opening date.

Table 1 shows some summary statistics for the auctions monitored during the data collection period. The table indicates that proxy agents placed 57% of all bids and bidders placed the remaining 43% bids manually. Forty-one percent of the auctions received at least one bid and 39% of the auctions had a winner. Some auctions had a reserve price, which allows the seller to cancel the auction if no bids above the reserve price are placed. Auctions that receive some bids but do not have a winner can be attributed to cancellations due to reserve prices.

Table 1  
Summary of data collected (January 2003)

Total auctions	344,314
Total bids	1,125,183
Manual bids	485,727 (43%)
Bids by proxy agents	639,456 (57%)
Auctions with at least 1 bid	140,039 (41%)
Auctions with a winner	133,121 (39%)

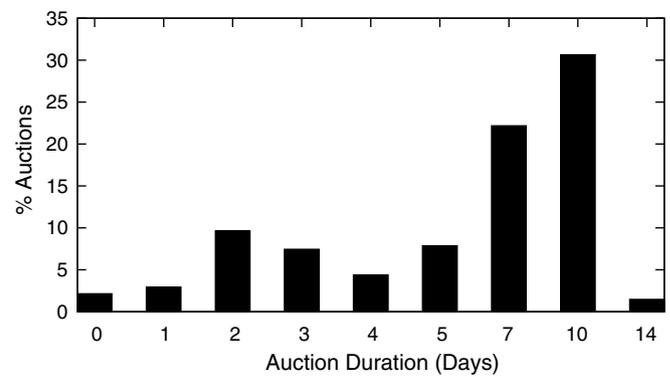


Fig. 3. Auction duration in days.

#### 4.2. Auction duration

Fig. 3 shows the percentage of auctions with a given auction duration in days. For clarity, only the durations with more than 2% of the auctions, were shown in the figure. We found that 2 days, 7 days, and 10 days are popular auction durations. As indicated by the picture, over 50% of the auctions fall into 7-day and 10-day periods.

### 5. Site level workload characterization

This section presents the site level workload characterization, i.e., the activity from the auction site's point of view. This characterization does not distinguish between different classes of auctions, bids, users, and price ranges. The site level workload characterization includes cluster analysis, multi-scale analysis, and closing time analysis of the activity on the site.

#### 5.1. Cluster analysis

We used the *k*-means clustering algorithm [11] to classify auctions based on unique number of bidders, number of bids placed, and closing prices separately. These clusters of auctions were analyzed to investigate properties of the auctions within each price cluster and compare these properties to all auctions, as discussed later in this paper.

##### 5.1.1. Clustering auctions on the number of unique bidders

A common question when using *k*-means clustering is how many clusters (i.e., which value of *k*) should be used. The best value of *k* is the one that minimizes the coefficient of variation (CV) of the intra-cluster distance (the distance between the points of a cluster and its centroid), and maximizes the coefficient of variation of the inter-cluster distance (the distance between the centroids of the various clusters) [21]. The higher the ratio between the inter-cluster CV and the intra-cluster CV, the better the clustering process is. We considered auctions for which at least two bidders participated and clustered them according to the number of unique bidders in the auctions. We varied the number of clusters from 3 to 10 and computed the coefficient of variation of the inter-cluster and intra-cluster

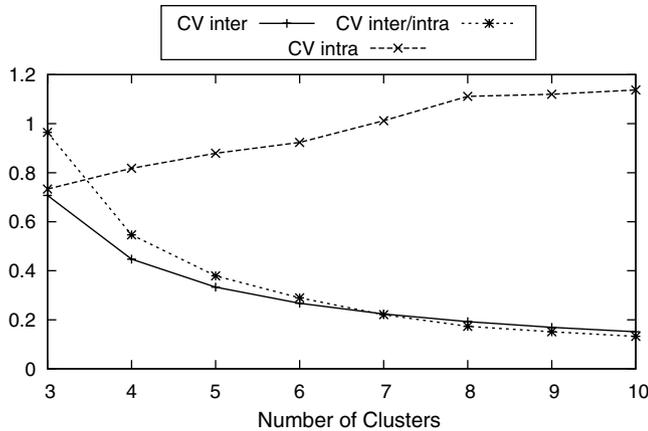


Fig. 4. Efficiency analysis of clusters based on unique number of bidders.

Table 2  
k-Means clustering of auctions on unique number of bidders

Cluster name	Minimum	Maximum	% Auctions
Low	2	3	66.62
Medium	4	6	24.81
High	7	52	8.56

distances for auctions as shown in Fig. 4. The figure clearly indicates that three clusters is the best choice.

Table 2 shows the results of the clustering process when three clusters are used. These clusters are called low, medium, and high unique number of bidder clusters and are used in the user level analysis later in this paper. It is interesting to note that as the size of the cluster decreases, the number of unique bidders increases. A very large percentage of auctions have a relatively low number of unique bidders and a very small percentage has a high number of bidders.

5.1.2. Clustering of auctions on number of bids

We performed a similar clustering analysis on auctions based on the total number of bids placed and determined that three is the optimal number of clusters when using bids as the clustering attribute. In this case, we used all auctions that received at least one bid as part of the clustering analysis. The result of the clustering process with manual bids and total bids including proxy bids is shown in Table 3. Similarly to the clustering based on unique number of bidders, as the size of the cluster decreases, the average number of bids in the cluster increases. A very large percentage

Table 3  
k-Means clustering of auctions on number of bids

Cluster name	Minimum (manual/total)	Maximum (manual/total)	% Auctions (manual/total)
Low	1/1	5/19	82.90/87.81
Medium	6/20	19/74	14.56/10.57
High	20/75	111/400	2.54/1.62

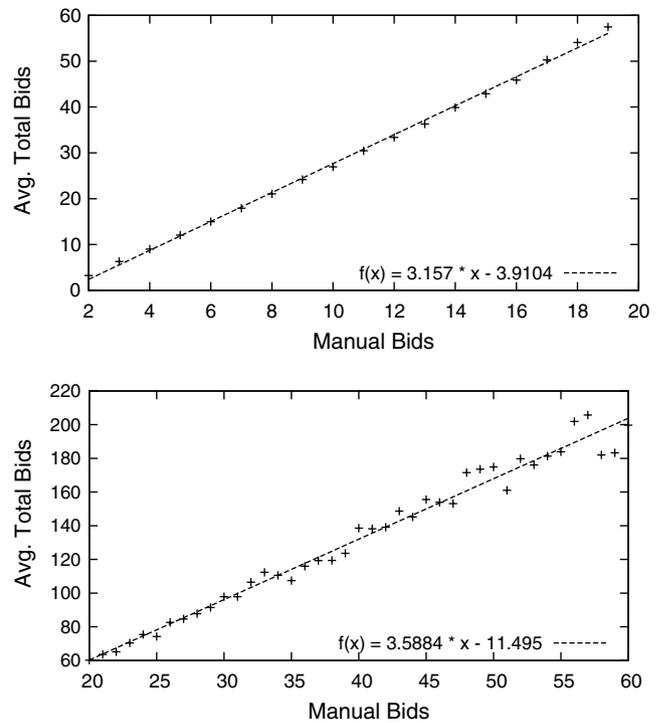


Fig. 5. Manual bids vs. avg. number of total bids. Top: Low and medium bidder clusters. Bottom: high bidder cluster.

of auctions have a relatively low number of bids and a very small percentage has a high number of bids.

There is a strong correlation between manual bids and total bids in an auction. Fig. 5 shows the relationship for low and medium bidding clusters (top) and the high bidding cluster at the bottom. Using these relationships, proxy agent bids can be predicted with good approximation, given the total number of manual bids.

Unique bidders and manual bids placed on auctions also have a strong correlation as shown in Fig. 6. Using the relationship shown in Fig. 5, the total number of bids (manual + agent bids) can be predicted for a given number of unique bidders. As the number of unique bidders increases, the number of bids per auction increases. Results of analysis of one set of clusters (e.g., based on bidders) would be similar to the results of other set of clusters

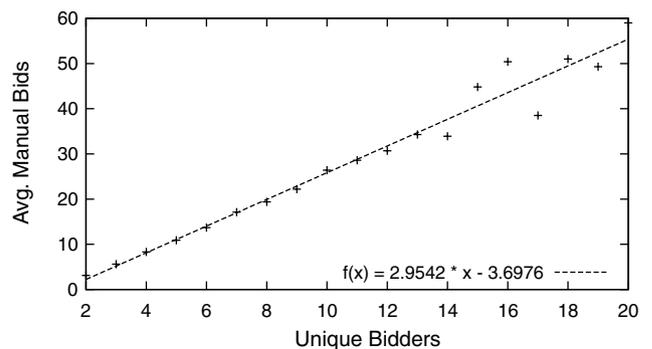


Fig. 6. Unique bidders vs. avg. number of manual bids.

(e.g., based on number of bids). In this paper, we analyzed clusters based on the number of unique bidders and skipped the analysis of clusters on number of bids because of this strong correlation.

5.1.3. Distribution of bids per auction

As shown in the clustering analysis discussed above, a large number of auctions have a small number of bids and a small number of auctions have a large number of bids. It is useful to try to establish a relationship between the number of bids (NB) and the number of auctions (NA) with that number of bids. For example, such a relationship could be used to guide caching policies for auction information. The distribution of bids per auction is presented in Fig. 7. We approximated the distribution of bids per auction using two equations, one for the low and medium bid clusters and the other for the high bid cluster. As shown in the graphs, there is an exponential relationship between the number of manual bids and the number of auctions with that number of bids. A large number of auctions have a relatively small number of bids and few auctions have a large number of bids. For example, 3.42% of the auctions received 10 or more manual bids and 0.95% of the auctions received 20 or more manual bids. Caching these small number of active auctions can improve the site’s response time [15]. The relationships depicted in the graphs of Fig. 7 can be useful for the site’s architect to determine the cache size required for the auction site.

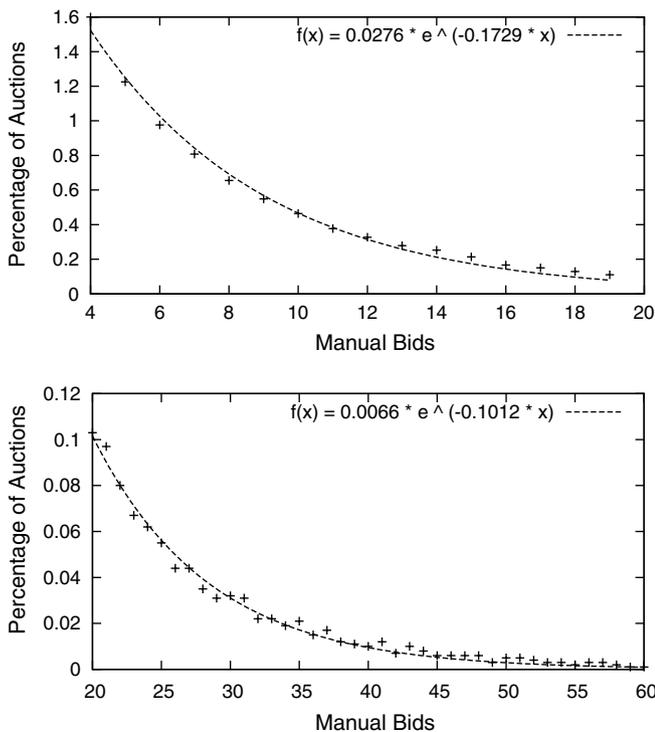


Fig. 7. Distribution of bids per auction. Top: low and medium bidder clusters. Bottom: high bidder cluster.

Table 4  
k-Means clustering of auctions on closing price

Cluster number	Minimum	Maximum	Average	% Auctions
1	0	48	15	73.74
2	48	144	82	14.73
3	144	322	207	6.08
4	325	657	443	2.52
5	660	1185	875	1.29
6	1190	1975	1499	0.78
7	1995	3200	2480	0.35
8	3450	6200	4397	0.30
9	6500	14500	8517	0.13
10	15500	32000	21867	0.07

5.1.4. Clustering of auctions using closing prices

As part of our clustering analysis, we divided auctions into 10 clusters based on their closing price. The result of the clustering process, shown in Table 4, indicates that as the size of the cluster decreases, the average closing price increases. A large percentage of auctions have relatively low closing prices and a very small, but non-negligible, number of auctions have very high closing prices. This may be an indication that the closing price distribution is heavy tailed.

These clusters of auctions were analyzed to investigate properties of the auctions within each price cluster and compare these properties to all auctions, as discussed later.

5.2. Multi-scale analysis of auctions and bids

Menascé et al. showed the importance of performing a multi-scale analysis when analyzing e-commerce site workloads. They showed that phenomena not visible at coarse time scales could be elicited when the granularity of the time scale is reduced [19]. We performed multi-scale analysis on auction and bid arrival time in the following section to identify activity at different time scales.

5.2.1. Multi-scale analysis of auctions

Fig. 8 shows the total number of auctions created each day during our data collection period in January 2003. In this figure, each point in the x-axis corresponds to a full day. Fig. 9 shows the number of auctions created during

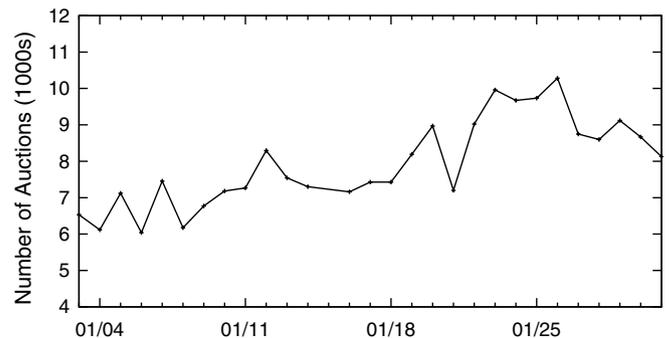


Fig. 8. Auctions per day.

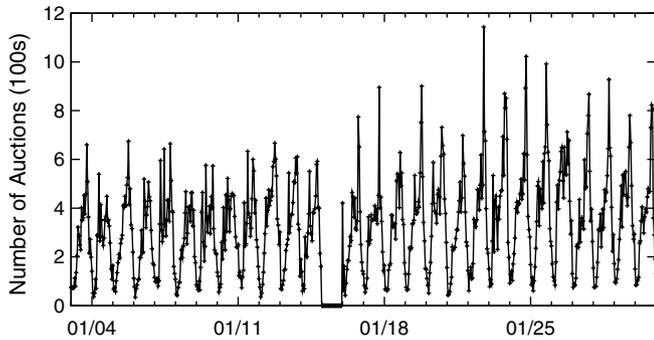


Fig. 9. Auctions per hour on each day.

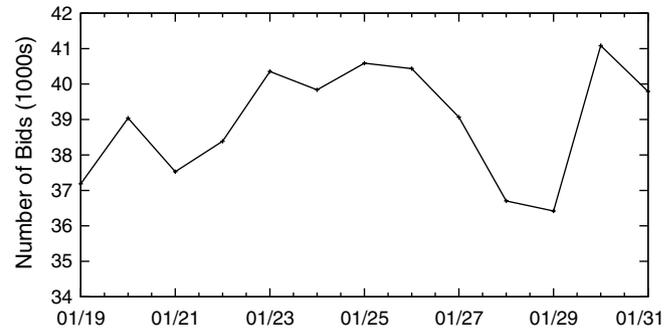


Fig. 11. Bids per day.

each hour for the same one-month period. Here, each point in the  $x$ -axis corresponds to each hour of each day. Spikes can be clearly observed in each day as the time scale is reduced from days to hours. To better understand this behavior, we show in Fig. 10 the sum, over all days, of the number of auctions created during each hour of the day. This graph clearly indicates a surge between 4 p.m. and 10 p.m. in the arrival rate of items to be auctioned. The average arrival rate in that period is almost double the average for the rest of the day. A plausible explanation is that most people create auctions at home in the evening after dinner.

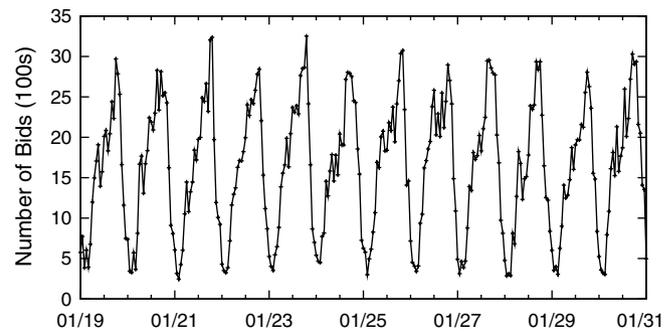


Fig. 12. Bids per hour on each day.

5.2.2. Multi-scale analysis of bids

For the multi-scale analysis on bids we used the period of 1/19/03 to 1/31/03 to make sure we had collected a steady stream of auctions in our database. Fig. 11 shows the total number of bids received each day. Fig. 12 shows the same data on the finer time-scale of hours instead of days. Fig. 13, similar to Fig. 10, shows the sum, over all days, of the number of bids submitted each hour during the 13 day period. Fig. 13, further decomposes the number of bids submitted into manual bids and proxy bids. A multi-scale analysis shows a twofold increase in the traffic of bid submissions towards the end of the day. This is true for manual and proxy bids. The reason that proxy-generated bids exhibit a behavior similar to manual bids is that proxy agents submit bids in reaction to a manual

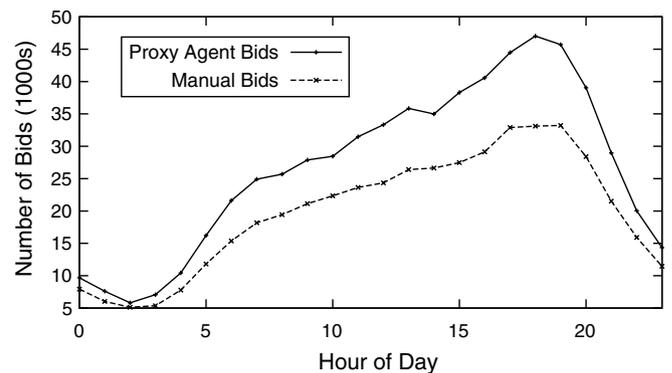


Fig. 13. Sum of bids per hour on all days.

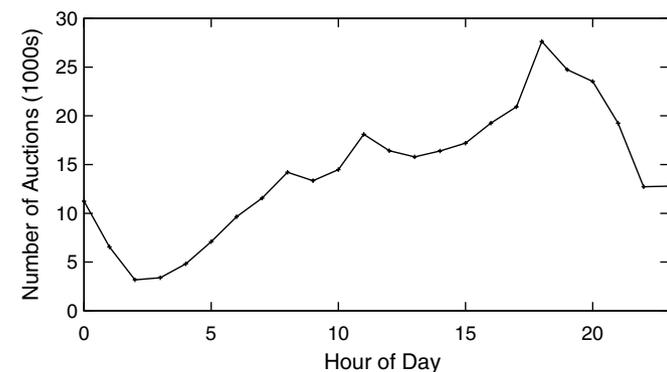


Fig. 10. Sum of auctions per hour on all days.

and other proxy-generated bids. It is also clear from the figure that the proxy agents generate more bids throughout the day than the bids placed manually by the site users.

5.3. Closing time analysis

We analyzed the number of bids received and the price variation of auctions as the closing time approached for each cluster of auctions based on unique number of bidders and auctions of popular durations. We also analyzed the bidding activity during the last few hours for auctions of popular durations to see how the bidding activity changes for auctions of different durations such as 2-day vs. 7-day auctions.

5.3.1. Closing time analysis of bidding activity

In Fig. 14, the lifetime of an auction was divided into 100 equal intervals, corresponding to 100 different values of the age of an auction. The y-axis shows the percentage of bids that were submitted in each age time slot for all auctions until they are closed. The graph shows the bidding activity for each of the three clusters based on unique number of bidders. It indicates that there is some bidding activity at the beginning stages of an auction. This activity slows down in the middle and increases significantly after 90% of the auction lifetime has elapsed. The graph shows that the activity in the low bidding activity auctions, which attracted a smaller number of unique bidders, is high in the closing minutes and low in the early stages. For auctions that attracted a higher number of unique bidders, the last minute bidding activity is lower and the early bidding is higher compared to the other classes of auctions. This indicates that auctions in which there are only a few bidders do not generate enough competition in the early stages. Bidders might wait until the last minute and submit their final bid just before the auction is closed as they do not see competition from other bidders.

We analyzed auctions with most popular durations (2, 7 and 10 days) to see how the bidding activity changes during the auction lifetime. As in the previous analysis, the bidding activity was plotted for each of the three classes, now based on auction duration. Fig. 15 shows that the percentage of bids placed throughout the lifetime of the auc-

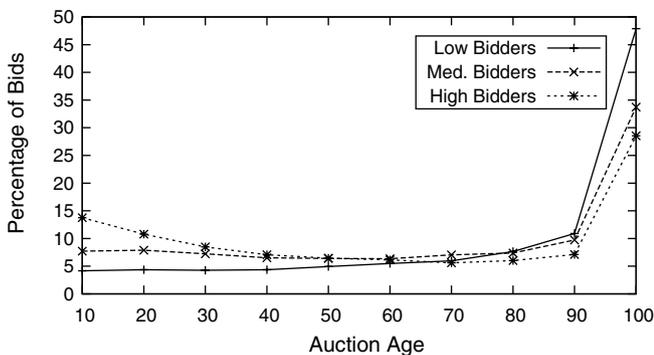


Fig. 14. Percentage of bids vs. auction age in different bidder clusters.

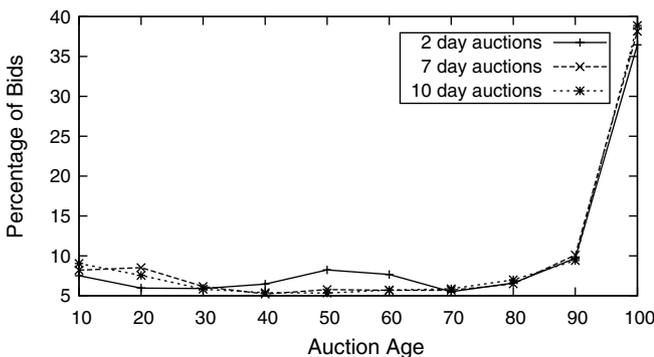


Fig. 15. Percentage of bids vs. auction age in different auction durations.

tion is virtually the same for all three classes, a clear invariant among different auction durations. Our analysis on the total number of bids vs. auction duration did not indicate that auctions with longer duration generate more bids in their lifetime. These two findings indicate that for auctions with longer duration, bidders take their own time in placing bids and shorter duration auctions attract those bids at a faster pace, in a given time interval.

5.3.2. Closing time analysis of price variation

Fig. 16 shows the variation of the price relative to the closing price as a function of the age of an auction for all auctions. The three curves in the picture are for the three clusters: low bidders, medium bidders, and heavy bidders. Prices raised faster in the first 20% of the period compared to the next 70% of the interval. However, after the age of an auction reaches 90%, the relative price increases much faster than in the two previous phases; the increase in the final phase is quadratic. The price variation data for the last 10% of the auction’s age was fitted to a second degree polynomial with a very high coefficient of determination. From Fig. 16, we can see that the auctions in the high cluster move faster towards the final price in the early stages compared to those in the medium cluster. The same is true when we compare auctions in the medium cluster with those in the low cluster. For example, auctions in the high cluster reached 46% of their closing price with an auction age of 50%, i.e., at the middle of the auction lifetime. Medium clusters reached 41% of their closing price at that age and auctions in the low clusters reached only 36% of their closing price at that point.

5.3.3. Last hour analysis

The previous analysis considered price variation as a function of auction age. In this section, we consider absolute time intervals at an hourly scale and at a minute scale to find out the bidding trend towards the end of auction. Fig. 17 shows the bidding activity at an hourly scale and Fig. 18 shows the same at a much smaller scale using a 1-min interval before the closing time of the auction. Shorter duration auctions received a higher percentage of bids compared to longer duration auctions at the last

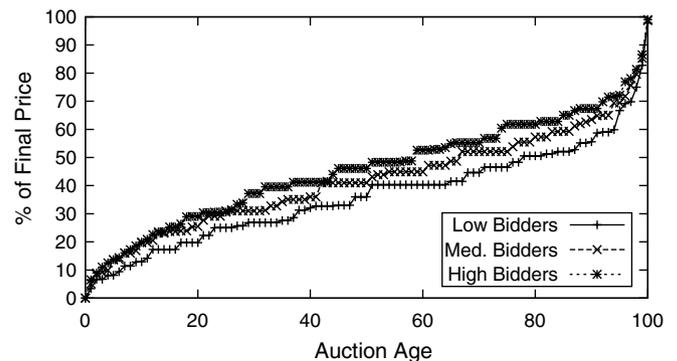


Fig. 16. Auction price variation with auction age.

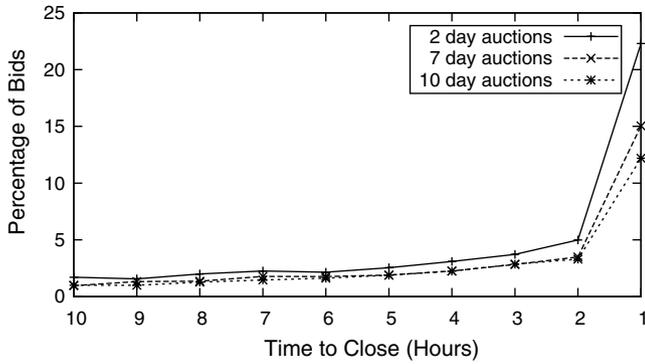


Fig. 17. Percentage of bids during closing hours.

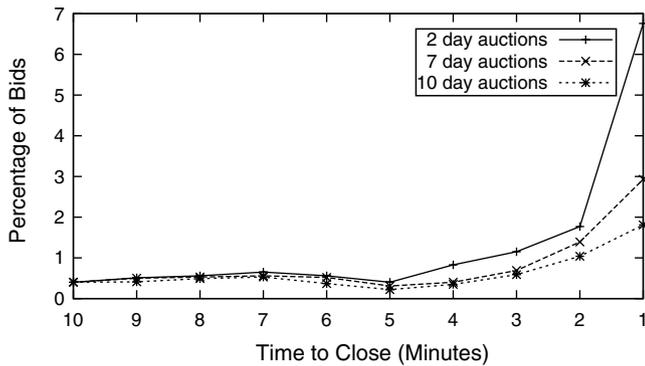


Fig. 18. Percentage of bids during closing minutes.

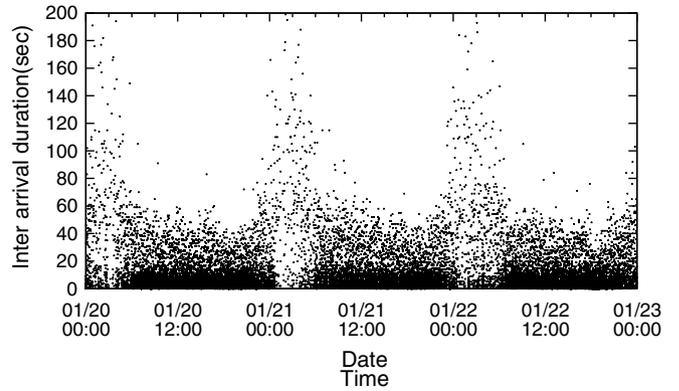


Fig. 19. Inter-arrival time of auctions.

early morning is a comparatively inactive period, with larger inter-arrival times. The figure shows that most of the inter-arrival periods are small, and there is a small number of large inter-arrival times, especially during the inactive hours. In order to better characterize the type of inter-arrival time distribution we look at periods of smaller inter-arrival times (during the day) and periods of larger inter-arrival times (during the night).

Fig. 20 shows the relative frequency of auction inter-arrival time for inter-arrival times below 15 s (top) and inter-arrival times of 15 s or more (bottom). As shown in the graphs, there is an exponential relationship between the inter-arrival time and the percentage of auctions with that inter-arrival time. A large number of auctions have

10 hours as well as at the last 10 min. This indicates that the number of bids placed on auctions with different auction durations is the same for all of them. Also, shorter duration auctions receive these bids at a faster rate compared to longer duration bids. This conclusion is similar to the one we reached in terms of percentage of bids for various auction durations as shown in Fig. 15.

5.4. Arrival process of auctions and bids

This section presents the arrival process of auctions and bids at the auction site. We first show the time between consecutive arrivals (inter-arrival times) of auctions and bids, and then model them into a standard distribution. Identifying a trend for inter-arrival time for auctions and bids is critical for accurate performance modeling and for generating realistic loads for benchmarking purposes. For example, exponential inter-arrival times allow us to use several results from queuing theory for estimating queue lengths and response times at the auction site [13].

5.4.1. Inter-arrival process of auctions

Fig. 19 shows the time between arrivals in seconds for each new auction on the site during the days of January 20 and January 23, 2003. As explained in the multi-scale analysis section, evenings are the busiest periods on an auction site with smaller inter-arrival times, and midnight to

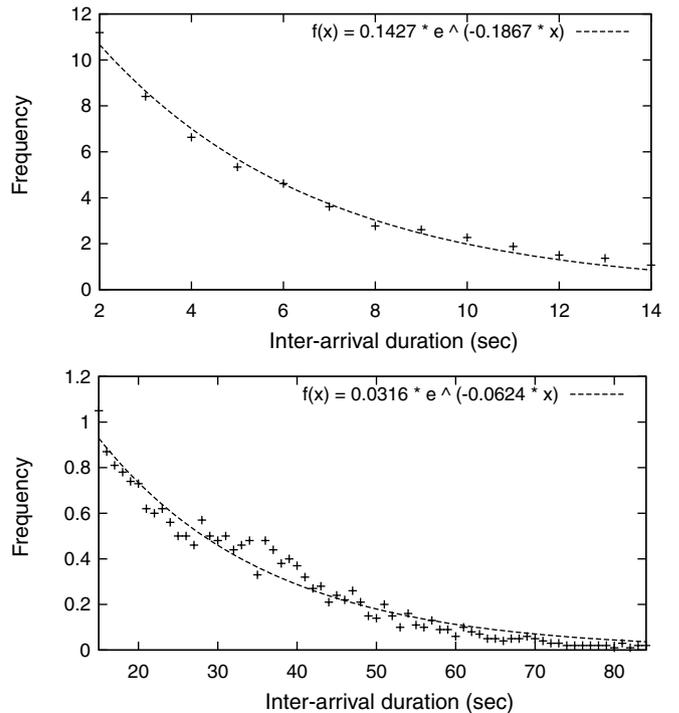


Fig. 20. Distribution of inter-arrival time of auctions. Top: inter-arrival time less than 15 s. Bottom: inter-arrival time of 15 s or more.

very small inter-arrival times and a relatively few auctions have large inter-arrival times.

#### 5.4.2. Inter-arrival process of bids

The site processes a proxy bid automatically when a manual bid is placed on an auction. Processing time for a proxy bid can be combined with the processing time of a manual bid. For this reason, we considered only manual bids. Fig. 21 shows the time between arrivals, in seconds, for each manual bid on the site during January 20 and January 23, 2003. As with the auction inter-arrival times, bid inter-arrival times are smaller during the day than in the period from midnight to early morning.

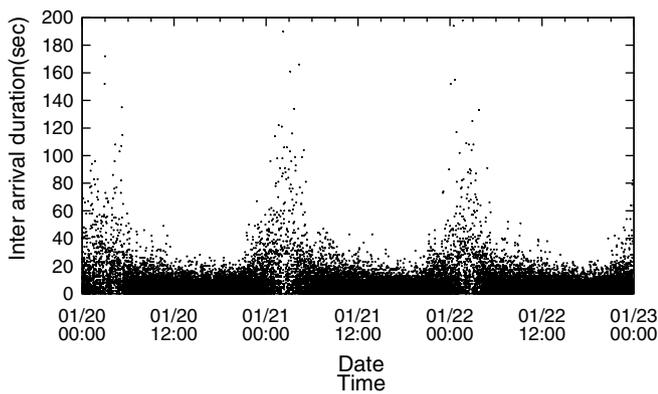


Fig. 21. Inter-arrival time of bids.

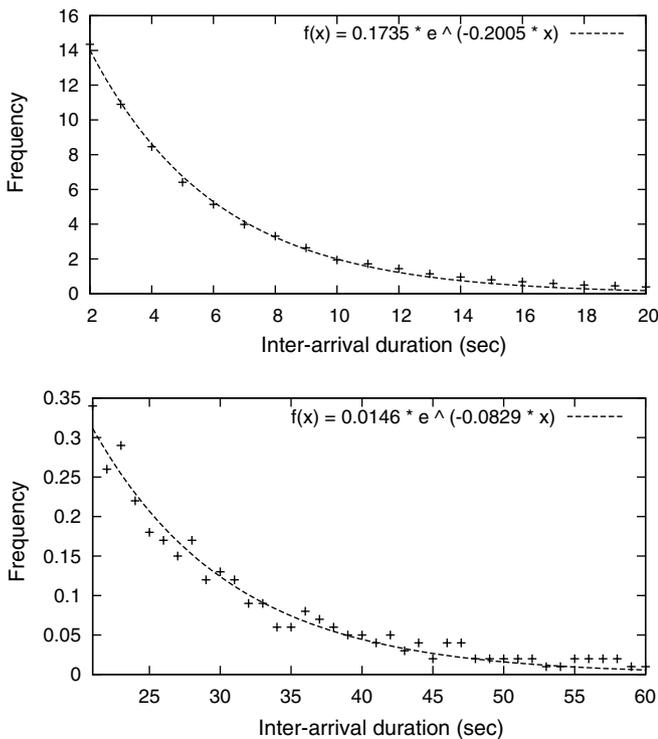


Fig. 22. Distribution of inter-arrival time of bids. Top: inter-arrival time less than or equal to 20 s. Bottom: Inter-arrival time more than 20 s.

Fig. 22 shows the relative frequency of bids with inter-arrival times less than or equal to 20 s (top) and larger than 20 s (bottom). As shown in the graphs, there is an exponential relationship between the inter-arrival times of bids and the percentage of bids received with that inter-arrival time. Again, as with the inter-arrival time of auctions, a large number of bids have very small inter-arrival times and a relatively few bids have a large inter-arrival time.

## 6. User level workload characterization

### 6.1. Popularity analysis

This section presents an analysis of the popularity (or rank) of winners, sellers, and bidders. The purpose of this analysis is to try to establish an empirical relationship between the relative frequency and rank in each case as discussed below. The basic motivation was to verify a conjecture that this relationship may follow a Zipf's Law [25] given that Zipfian distributions have been observed in many instances related to the Web [3] and e-commerce environments [20]. Zipf's Law establishes that the relative frequency,  $f$ , by which an object (or a word in a text) is accessed is inversely proportional to its rank  $r$  (also known as the popularity of the object). Thus,

$$f = K/r \tag{3}$$

where  $K$  is a normalization constant so that all frequencies add to one. According to Zipf's Law, the second most popular object (i.e., rank = 2) receives half the number of accesses as the most popular object (i.e., rank = 1) and the  $n$ th most popular receives  $1/n$  of the number of accesses of the most popular. As a result of this, a relatively few objects are responsible for the majority of the accesses. This type of property is very important for system design considerations. For example, caching a relatively small number of objects (i.e., the most popular ones) can result in significant performance improvement [14]. Taking logarithms of both sides of Eq. (3) yields  $\log(f) = \log(K) - \log(r)$ . In other words, if Zipf's Law is plotted in a log–log scale, the resulting curve is a straight line with a slope of  $-1$ . Zipf's Law is a special case of power-law distributions, i.e., distributions in which the exponent of  $r$  in Eq. (3) is not necessarily 1. Thus, in what follows we are interested in investigating whether the relationship between  $f$  and  $r$  follows a Zipf or, more generally, a power-law distribution. A straight line with negative slope in the log–log plot of  $f$  vs.  $r$  is such an indication.

#### 6.1.1. Winner's popularity

We first consider the winner's popularity analysis. For that purpose, all the winners are sorted in decreasing order of the number of auctions they won. The relative winning frequency in this case is the percentage of auctions won by the winner. The one with the highest winning frequency is assigned a rank equal to 1, the one with the second highest winning frequency receives a rank equal to 2, and so on.

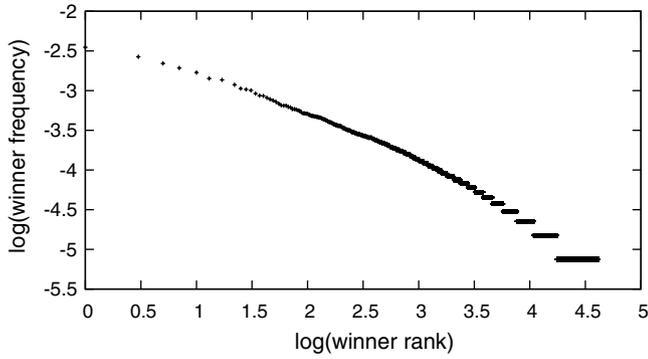


Fig. 23. Log-log plot of user's winning frequency vs. rank.

Fig. 23 shows a log-log plot of the winner frequency vs. its rank. The plot shows that the data follows a power law. In practical terms, this means that relatively few bidders are responsible for winning a large percentage of the auctions.

6.1.2. Seller's popularity

Consider now the sellers popularity analysis. In this case, all the sellers are sorted in decreasing order of the number of auctions created by a seller. The frequency  $f$  in this case is the percentage of all auctions created by the seller. As before, the one with the highest frequency receives a rank equal to one. Fig. 24 shows a log-log plot of a seller's frequency vs. its rank. As with the winner's popularity, the relationship follows a power-law distribution with a slope very close to  $-1$ , which indicates a Zipf's distribution. Thus, relatively few sellers are responsible for creating the majority of the auctions. Many of these repeat sellers are businesses which established stores to sell their goods through auction process.

6.1.3. Bidder's popularity

We now turn our attention to bidder popularity. All bidders were sorted in decreasing order of the number of bids placed among all the auctions. The frequency  $f$  in this case is the fraction of the total number of bids placed by the bidder. The one with the highest frequency receives a rank of one. Fig. 25 plots a log-log graph of bidder's frequency vs. rank. A power-law distribution is apparent in most of the

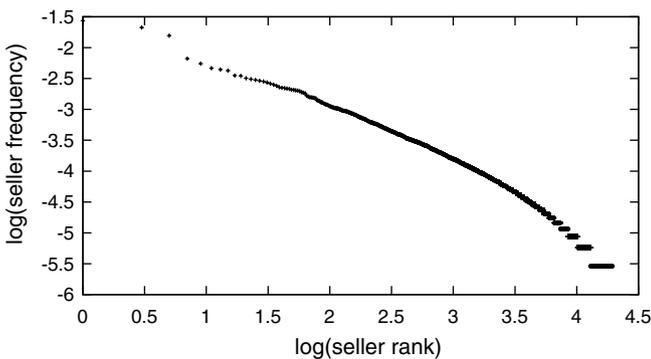


Fig. 24. Log-log plot of seller's frequency vs. rank.

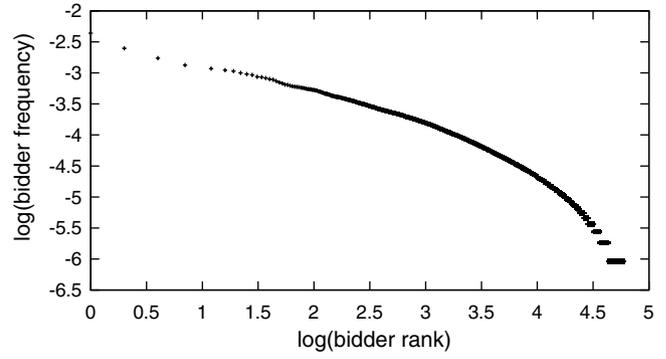


Fig. 25. Log-log plot of user's bidding frequency vs. rank.

rank range indicating that the majority of bids are placed by a relatively small number of unique bidders. The results of this section are useful for the design of resource management techniques aimed at optimizing performance and site revenue, which is the goal of our current research.

6.1.4. Popularity of auctions

We also analyzed the popularity of each auction in terms of the number of bids received. For this analysis we used all auctions that received at least one bid. The results are shown in Fig. 26. To generate this figure we sorted all auctions in decreasing order of number of bids received. The one with the most number of bids was ranked number one. This most popular auction received 400 bids. The top graph in Fig. 26 shows a log-log graph of the

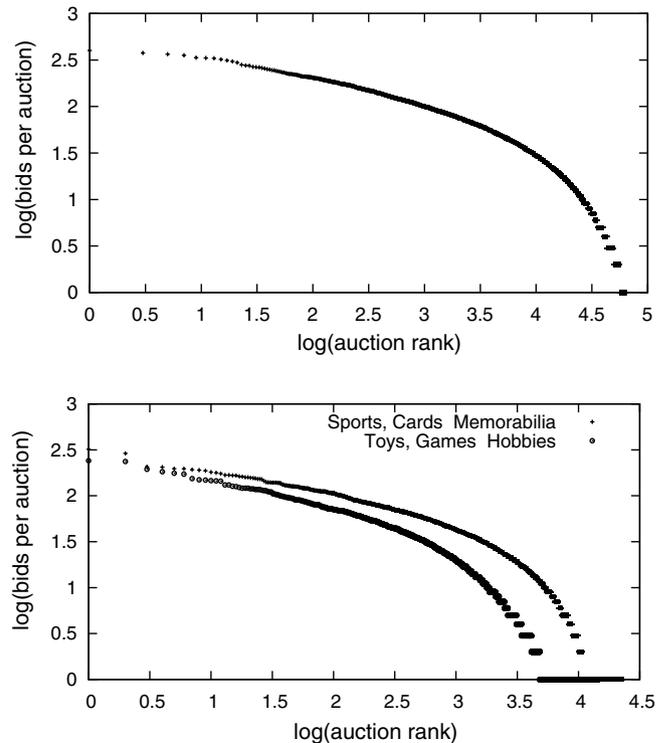


Fig. 26. Log-log plots of auction bid count vs. popularity. Top: across all categories; bottom: within categories.

number of bids per auction vs. rank for all auctions. There was a tie among a large number of auctions with a single bid. Each of these auctions was given a unique rank based on their auction number. This would show up as a horizontal line on top of the  $x$ -axis (i.e., the log of the number of auctions is equal to zero since the number of auctions is equal to one) at the end of the rank range. Once again we can see a power-law behavior for the distribution of number of bids per auction, i.e., a relatively few number of auctions attract the majority of the bids. The bottom graph in Fig. 26 shows similar results but now restricted to two of the most popular categories: sports, cards and memorabilia and toys, games and hobbies. The trend within these categories is the same as when all categories are considered.

6.1.5. Popularity of categories

The data gathered covers 2473 different categories and sub categories of auctions. Of these, 19 categories are top-level categories, i.e., categories that are not sub categories to any other category. Auctions can be created at any specific category or sub categories. However, a deeper sub category provides a more detailed classification for auctioned item. For example a Pentium 4 Dell Latitude laptop can be classified under Computers and Software → Hardware → PC → Notebooks → Dell or it can be classified at any of the categories above the final category Dell. In this example, Computers and Software is a top-level category since it has no parent category.

Fig. 27 shows the percentage of auctions, percentage of bids and bids per auction for each of the 19 top-level categories.

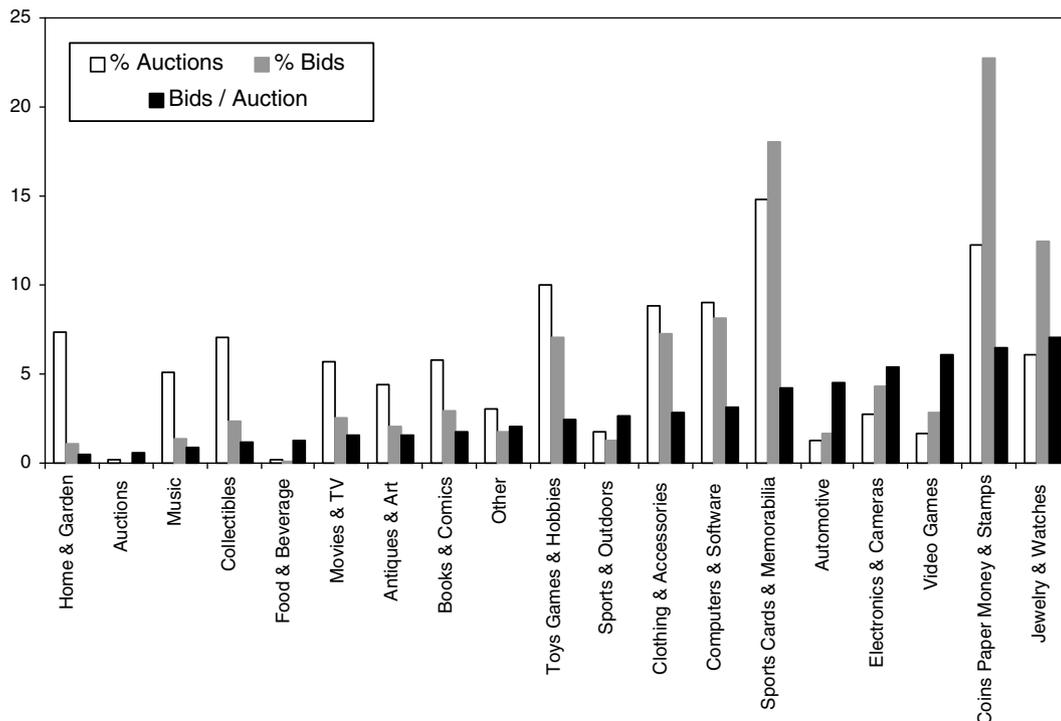


Fig. 27. Percentage of auctions, percentage of bids and bids/auctions per category.

The figure shows that the Jewelry and Watches category is the most active category with an average of 7.07 bids for each auction created in that category. Sports cards and memorabilia is the top category for number of auctions created with 14.79% auctions falling into that category. Coins, paper money and Stamps follows next with 12.21% auctions. In terms of bids placed, coins, paper money and stamps rank first with 22.75% of all the bids being placed on these auctions. Sports cards and memorabilia is the next top-bidding category with 18.03% of total bids.

Fig. 28 shows a log-log plot of auction category popularity. To generate this graph we counted the number of auctions created within each category and ranked them in sequence with the category with the highest number of auctions ranked number one. The data indicates that the number 1 ranked category (other) has 54,879 auctions and the

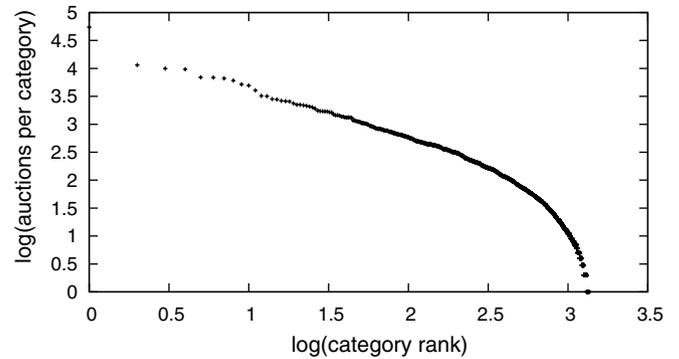


Fig. 28. Log-log plot of category auction count vs. category popularity.

next most popular category (home and garden → housewares → decorative) has 11,511 auctions. There is a tie between 36 categories for which only one auction was created. Examples of these categories include collectibles → advertising → beverage related → soda → canada dry and automotive → vehicles → vans → cargo. Each of these categories was given a unique rank alphabetically for analysis purpose. Fig. 28 indicates a power law behavior for the most part, i.e., for categories with at least 100 auctions (i.e., log base 10 of the count of auctions per category equal to 2). Similarly as before, a few categories are responsible for attracting a relatively large number of auctions.

To analyze the distribution at the top-level categories, we used the percentage of auctions in each of the 19 top-level categories. The category with the highest percent count is ranked number one. The top 11 among the 19 categories account for more than 80% of all auctions. Fig. 29 shows a log–log graph of the percent count of auctions in these 11 categories vs. their rank. The graph also shows a linear trend, which indicates a very close match to a power-law distribution.

6.2. Analysis of unique bidder clusters

6.2.1. Unique bidder arrival rate

This section presents the results of our analysis of unique bidder arrival time and bidding activity in auctions with the same number of unique bidders. We grouped auctions with two or more unique bidders per auction. Note that we used only the number of unique bidders in this analysis, not the number of bids. For example, five unique bidders could be placing a total of 50 bids in a given auction.

Fig. 30 indicates the percentage of unique bidders entering auctions during each 10% time interval of the auction age. Each unique bidder’s entry time is the time at which that user placed his first bid on the auction. Subsequent bids were not counted for this study. There are four curves in Fig. 30, one for each number of unique

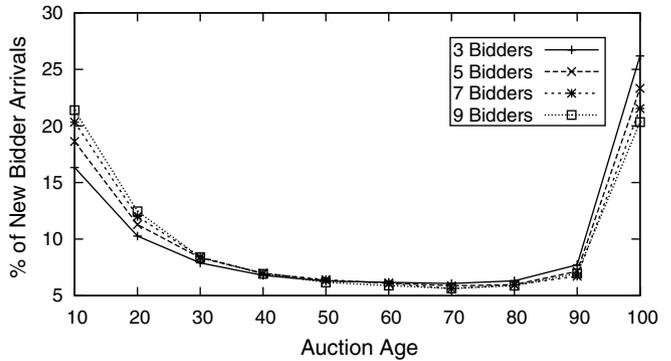


Fig. 30. Percentage of new bidder arrivals vs. auction age within auctions with different number of unique bidders.

bidders: 3, 5, 7, and 9. It is interesting to note that as the number of unique bidders increases per auction, the percentage of bidders participating in the early stages increases. Similarly, the percentage of unique bidders entering the auction during the final stages decreases. For example, for auctions with three unique bidders, about 15% of the unique bidders for all such auctions enter during the first 10% of the auction age and close to 30% enter during the last 10% of the auction age. Consider now auctions with 9 unique bidders. About 21% of the unique bidders for all such auctions enter during the first 10% of the auction age and 22% enter during the final 10% of the auction age.

6.2.2. Bidding activity

Fig. 31 indicates the percentage of bids placed during each 10% time interval of the auction age. Again this graph indicates that the bidding activity surges in the closing minutes of an auction irrespective of the number of unique bidders participating on the auction. As in Fig. 30, this graph also shows that the percentage of bids placed increases in the early stages with the number of unique bidders and decreases with the number of unique bidders in final stages. We show only four unique bidders in Fig. 30 and in Fig. 31 for clarity of the graphs.

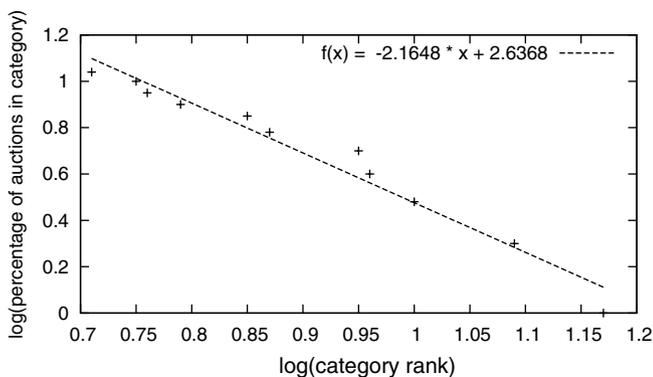


Fig. 29. Log–log plot of percent auction count per category vs. rank for the top 11 categories.

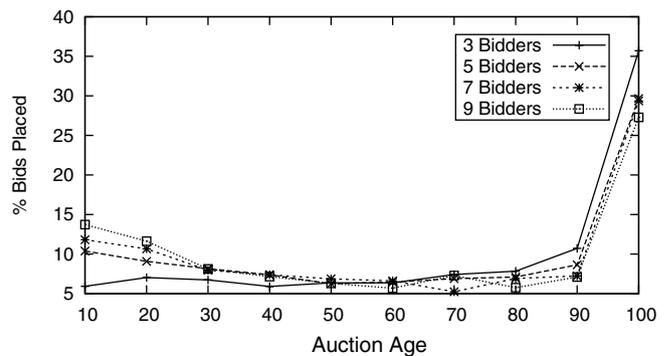


Fig. 31. Percentage of bids placed vs. auction age with different number of unique bidders.

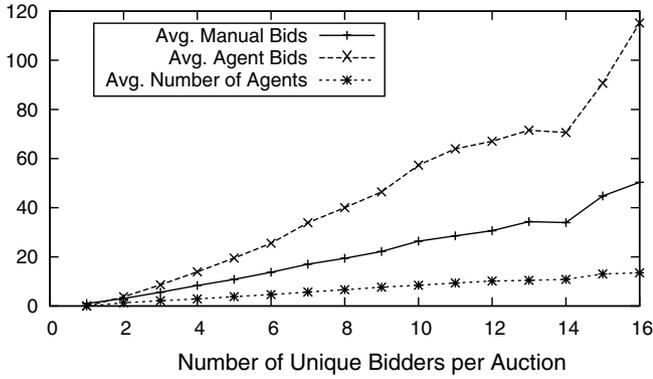


Fig. 32. Average number of bids by proxy agents, average number of manual bids and average number of proxy agents used per auction vs. number of unique bidders per auction.

6.2.3. *Bids by agents, manual bids, proxy agent usage*

This section presents an analysis of bidding activity within auctions grouped by the number of unique bidders. Fig. 32 indicates the average number of bids by proxy agents, the average number of manual bids, and the average number of users using proxy agents within auctions as a function of the number of unique bidders. As the number of unique bidders increases, the number of users using proxy agents also increases. It is clear that the difference between the percentages of bids by proxy agents to manual bids increases with an increasing number of unique bidders. This indicates that as more and more proxy agents compete, the number of overall bids they generate increases very fast resulting in higher closing prices.

6.2.4. *Closing price*

Fig. 33 shows the average closing price of auctions as a function of the number of unique bidders. The graph indicates that higher priced items attract more unique bidders. Note that this holds only up to a few thousand dollars in closing price. Very high priced items such as new and used cars attract few bids from a very small number of unique bidders. Since those auctions are very few in number, the average price of auctions with few unique bidders is still low, given that a large number of auctions have few bidders

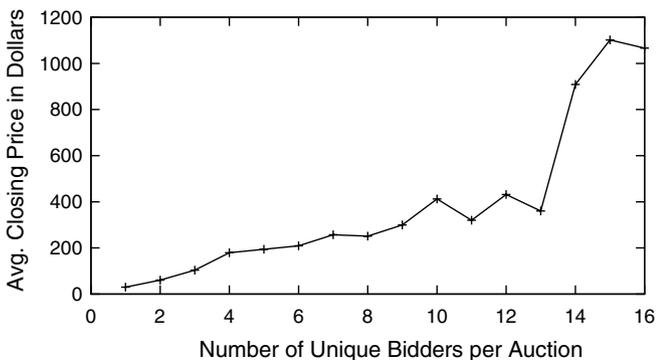


Fig. 33. Average closing price in dollars vs. number of unique bidders per auction.

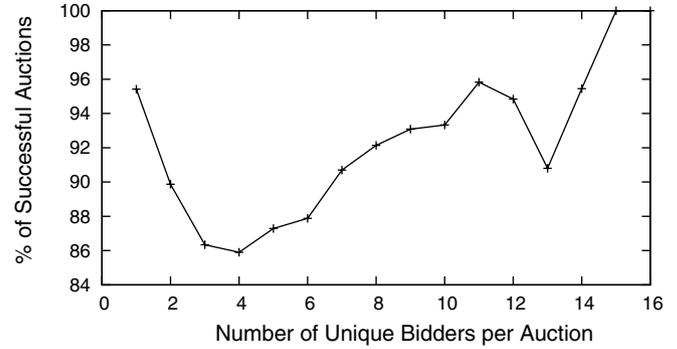


Fig. 34. Average percentage of successful auction vs. number of unique bidders per auction.

and a large number of auctions are in the lower price range, as shown earlier in this paper.

6.2.5. *Successful auctions*

Fig. 34 indicates the average percentage of successful auctions (i.e., an auction with a winner) as a function of the number of unique bidders participating in the auction. As shown in the graph, with a larger number of unique bidders participating, the success rate increases and after more than 15 unique bidders, the success rate approaches 100%. Note that the success rate in the auctions with one bidder is high since there are a large number of items (usually lower priced items) that are won with just one bid. Many of these auctions do not have reserve prices. The same explanation also applies to the large number of items with two unique bidders.

6.3. *Analysis of price clusters*

As indicated in the clustering analysis section earlier, auctions were clustered based on their closing price. This section presents the bidding activity within different price clusters in order to examine the influence of the price on the bidding activity of an auction.

6.3.1. *Bidding activity*

For Fig. 35 we divide the auction age (see definition in the background section) in 10 intervals. We selected four

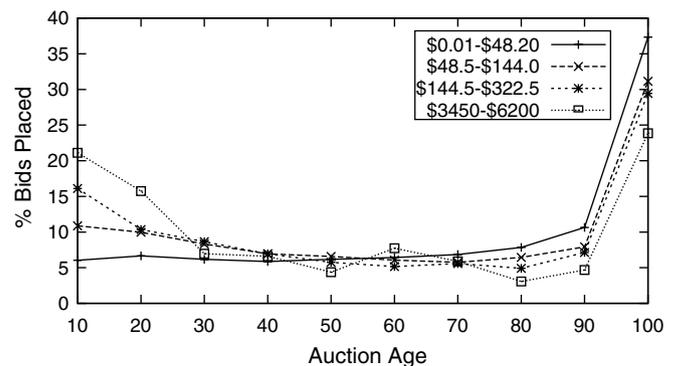


Fig. 35. Percentage of bids placed vs. auction age within price clusters.

out of the 10 clusters for better readability of the picture. The price ranges of these four clusters covers a large price range, from very cheap items to very expensive one. The figure indicates the percent of bids placed in each of 10% interval of the auction age for the four clusters. In all price ranges, the bidding activity surges in the last 10% of the auction lifetime as noted in our earlier work [2], which did not consider the effect of price. The graphs of Fig. 35 indicate that higher priced items attract relatively more bids initially and relatively few bids compared to other groups in the final stages of the auction. A possible explanation is that users tend to take more time and place bids cautiously and avoid rushing into the final closing minutes of an auction when purchasing expensive items.

6.3.2. Unique bidders and agent usage

Fig. 36 indicates the average number of unique bidders per auction, average number of unique bidders using proxy agents to place automatic bids for them, and the average auction length (in days) for auctions for all 10 price clusters. Remember that as the cluster number increases, the price range also increases. We can see that the higher priced items stay longer in an auction and the average number of unique users using proxy agents drops gradually for higher priced items. This may indicate that bidders may want to exert a closer control when buying expensive items. The number of average unique bidders participating in different price clusters raises initially, stays level for intermediately priced auctions, and drops for very high priced items. This may reflect a smaller market for very expensive items.

It should be noted that although higher priced items are being auctioned for longer periods, our analysis on the effect of auction duration on the closing price did not reveal any particular trend based on price range. This means that many lower priced items are also being auctioned for larger periods of time, which reduces the average closing price of items auctioned for longer periods.

6.3.3. Winners

Fig. 37 indicates the percentage of auctions with a winner, the percentage of auctions with a proxy agent

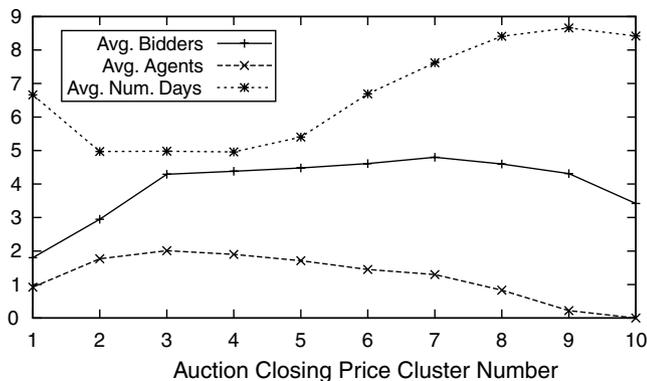


Fig. 36. Average bidders, agents, number of days on auction within each cluster.

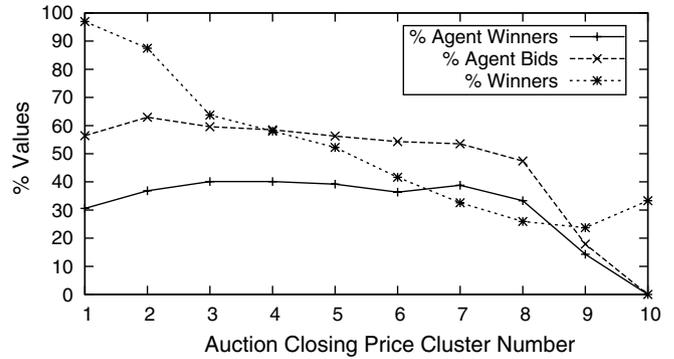


Fig. 37. Percentage of auctions with winner, percentage of winners with proxy agent bid, percentage of proxy agent bids within each cluster.

placing the winning bid, and the percentage of total bids placed by proxy agents for each of the ten price clusters. Note that if an auction does not meet its reserve price, the seller can cancel the auction, leaving the auction without a winner. It is clear from the graphs that higher priced items have lower success rate, a lower number of bids by proxy agents, and a smaller chance of proxy agents winning auctions. As more expensive items require more manual control and thought before bid are placed, these numbers tend to go down in higher priced auctions.

7. Concluding remarks

The main results of our analysis can be summarized as follows:

- *Summary statistics.* Proxy agents placed 57% of all bids and bidders manually placed the remaining 43% bids. Forty-one percent of the auctions received at least one bid and 39% had a winner.
- *Clustering analysis.* A very large percentage of auctions have a relatively low number of bids and bidders and a very small percentage of auctions have a high number of bids and bidders. A large percentage of auctions have a low closing price and a very small percentage of auctions have a large closing price.
- *Correlation between bids and bidders.* The number of unique bidders per auction and the number of manual bids placed per auction are linearly related. In addition, the total number of bids placed (manual + proxy) is linearly related to the number of manual bids. As the number of unique bidders per auction increases, the number of manual bids and agent bids placed per auction increase.
- *Multi-scale analysis.* There is a twofold increase in the auction creation and bidding activity between 3 p.m. and 9 p.m. compared to the rest of the day.
- *Closing time analysis of bidding activity.* There is some bidding activity at the beginning stages of an auction. This activity slows down in the middle and increases

considerably after 90% of the auction life time has elapsed. Auctions in the high bidder cluster attract more attention early on than those in the medium bidder cluster. The same is true when we compare auctions in the medium bidder cluster with those in the low bidder cluster. Thus, if an auction attracts a good number of bidders initially, it is likely to end up in the high bidding category.

- *Auction durations and bidding activity.* Two days, seven days and ten days are most popular auction durations. The percentage of bids placed throughout the lifetime an auction is the same for different auction durations. This indicates that shorter duration auctions attract bids and bidders quickly compared to longer auction durations.
- *Closing time analysis for price variation.* Prices rise faster in the first 20% of the auction life time than in the next 70% of its life time. In these two periods, prices increase almost linearly. However, after the age of an auction reaches 90%, the relative price increases much faster than in the two previous phases; the increase in the final phase is quadratic. This observation is true for each of the three bidder clusters.
- *Inter-arrival time of auctions and bids.* Large inter-arrival times for auctions and bids is observed during late night and early mornings. There is an exponential relationship between the frequency of auctions and their inter-arrival times. The same observation is true for inter-arrival time of bids.
- *Winner's popularity.* Winner's popularity in terms of number of auctions won by a user follows a power law. In practical terms, this means that relatively few bidders are responsible for winning a large percentage of the auctions.
- *Seller's popularity.* As with the winner's popularity, the relationship follows a power-law distribution with a slope very close to  $-1$ , which indicates a Zipf's distribution. Thus, a relatively few sellers are responsible for creating the majority of the auctions.
- *Bidder's popularity.* A power law distribution is apparent in most of the rank range indicating that the majority of bids are placed by a relatively small number of unique bidders.
- *Bidding activity within price clusters.* Higher priced items attract relatively more bids initially and relatively few bids in the final stages of the auction compared to less expensive items. Higher priced items stay longer in an auction and the average number of unique users using proxy agents drops gradually for higher priced items. This may indicate that bidders may want to exert a closer control when buying expensive items. The number of average unique bidders participating in different price clusters raises initially, stays at similar levels for intermediately priced auctions, and drops for very high priced items. This may reflect a smaller market for very expensive items.

- *Unique bidder analysis.* As the number of unique bidders increases per auction, the percentage of bidders participating and bids placed in the early stages of auction increases. Similarly, the percentage of unique bidders entering the auction and bids placed during the final stages decreases. Higher priced items attract more unique bidders. The success rate of auctions increases with an increasing number of unique bidders.
- *Proxy agent bidding activity.* The difference between the percentages of bids by proxy agents to manual bids increases with an increasing number of unique bidders. This indicates that as more and more proxy agents compete, the number of overall bids they generate increases very fast resulting in higher closing prices. Auctions with an agent placing the final bid to win the auction tend to have a larger number of total bids and unique bidders. These auctions tend to have higher closing prices as agents compete automatically to increase the final price. Therefore, use of agents within auctions results in higher revenue throughput for the auction sites.

The results of our analysis can be used to design innovative algorithms for improving the quality of the service provided to buyers and sellers through optimized resource allocation and for improving revenue throughput [22]. These results will be used to develop workload generators as well as a research testbed for auctions sites. This testbed will be used for the validation of resource management techniques for online auction sites. Auction systems need to prioritize and provide optimal performance during the last minutes of the auction as most of the bids are placed in that period. Our conclusions on popularity of winners, sellers and bidders can play a significant role in designing caching techniques and dynamic resource management techniques for online auction sites. The authors used the results presented in this paper to develop techniques to improve end-user response time through server-side caching [15] and closing time rescheduling [17].

## References

- [1] Andale. Available from: <<http://www.andale.com>>.
- [2] V. Akula, D.A. Menascé, An analysis of bidding activity in online auctions, in: Proceedings of the Fifth International Conference, EC-Web 2004, Zaragoza, Spain, August/September 2004, pp. 206–217.
- [3] V.A.F. Almeida, A. Bestavros, M. Crovella, A. Oliveira, Characterizing reference locality in the WWW, in: Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems (PDIS), IEEE Computer Society, Dec. 1996, Miami Beach, FL, pp. 92–103.
- [4] C. Amza, E. Cecchet, A. Chanda, A. Cox, S. Elnikety, R. Gil, J. Marguerite, K. Rajamani, W. Zwaenepoel, Specification and implementation of dynamic web site benchmarks, in: Proceedings of the IEEE Fifth Annual Workshop on Workload Characterization (WWC-5), Austin, TX, November 25, 2002.
- [5] M. Arlitt, D. Krishnamurthy, J. Rolia, Characterizing the scalability of a large web-based shopping system, *ACM Tr. Internet Technology (TOIT)* 1 (1) (2001) 44–69.
- [6] R. Bapna, P. Goes, A. Gupta, Online auctions: insights and analysis, *Commun. ACM* 44 (11) (2001) 42–50.

- [7] D. Bunnell, *The e-Bay Phenomenon: Business Secrets Behind the World's Hottest Internet Company*, Wiley, 2000.
- [8] M. Calzarossa, L. Massari, D. Tessera, *Workload Characterization Issues and Methodologies*, *Performance Evaluation: Origins and Directions*, 2000, pp. 459–481.
- [9] eBay. Available from: <<http://www.ebay.com>>.
- [10] Empirix Inc., *Benchmark Study of Online Auction Site Performance*, August–September 2003.
- [11] B. Everitt, *Cluster Analysis*, fourth ed., Oxford University Press, Oxford, 2001.
- [12] W. Iverson, *Web Services in Action: Integrating with the eBay Marketplace*, A White Paper from O'Reilly Media.
- [13] L. Kleinrock, second ed. *Queueing Systems Theory*, vol. 1, Wiley-Interscience, 1975.
- [14] D.A. Menascé, *Scaling web sites through caching*, *IEEE Internet Comput.* 7 (4) (2003).
- [15] D.A. Menascé, V. Akula, *Server-side caching strategies for online auction sites*, in: *Proceedings of the Sixth International Conference on Web Information Systems Engineering (WISE'05)*, New York, NY, November 20–22, 2005.
- [16] D.A. Menascé, V. Akula, *Towards workload characterization of auction sites*, in: *Proceedings of the IEEE Sixth Annual Workshop on Workload Characterization (WWC-6)*, Austin, TX, October 27, 2003.
- [17] D.A. Menascé, V. Akula, *Improving the performance of online auction sites through closing time rescheduling*, in: *Proceedings of the First International Conference on Quantitative Evaluation of Systems (QEST-2004)*, Enschede, the Netherlands, September 27–30, 2004.
- [18] D.A. Menascé, V.A.F. Almeida, R. Fonseca, M. Mendes, *A methodology for workload characterization for e-commerce servers*, in: *Proceedings of the ACM Conference in Electronic Commerce*, Denver, CO, November 3–5, 1999, pp. 119–128.
- [19] D.A. Menascé, V.A.F. Almeida, R. Riedi, F. Ribeiro, R. Fonseca, W. Meira Jr., *A Hierarchical and Multiscale Approach to Analyze E-business Workloads*, *Performance Evaluation*, North-Holland, September 2003, vol. 54, No. 1.
- [20] D.A. Menascé, V.A.F. Almeida, R. Riedi, F. Ribeiro, R. Fonseca, W. Meira Jr., *In Search of Invariants for E-business Workloads*, in: *Proceedings of the Second ACM Conference on Electronic Commerce*, Minneapolis, MN, October 17–20, 2000.
- [21] D.A. Menascé, V.A.F. Almeida, *Scaling for E-business: Technologies, Models, Performance and Capacity Planning*, Prentice Hall, Upper Saddle River, NJ, 2000.
- [22] D.A. Menascé, V.A.F. Almeida, R. Fonseca, M.A. Mendes, *Business-oriented resource management policies for e-commerce servers*, *Perform. Evaluation* 42 (Sept.) (2000) 223–239.
- [23] Transaction Processing Council, *The TPC-W Benchmark*. Available from: <<http://www.tpc.org>>.
- [24] Yahoo! Auctions. Available from: <<http://auctions.yahoo.com>>.
- [25] G. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, MA, 1949.