Variable Bid Fee: An Online Auction Shill Bidding Prevention Methodology

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Computer Science and Engineering

Submitted By

Dhanmeet Kaur

(Roll No. 801332007)

Under the supervision of:

Dr. Deepak Garg

Associate Professor & Head

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

July 2015

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Variable Bid Fee: An Online Auction Shill Bidding Prevention Methodology*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Deepak Garg* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Signature:

(Dhanmeet Kaur)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Deepak Garg)

Head, CSED

Countersigned by

(Dr. Deepak Garg)

Head

Computer Science and Engineering Department

Thapar University

Patiala

(Dr. S. S. Bhatia)

Dean (Academic Affairs)

Thapar University

Patiala

ACKNOWLEDGEMENT

It gives me immense pleasure to acknowledge gratefully the invaluable assistance and cooperation of all those individuals who guided me to keep me on track in completing my thesis. I express my deep sense of gratitude to my supervisor and guide Dr. Deepak Garg, Head of Computer Science and Engineering Department, Thapar University, Patiala for his constant help, patience and encouragement all along the course of the research work. His scholarship, both extensive and deep is very inspiring. My work would not have been completed without his constant supervision, critical assessment and constructive suggestion.

My special thanks to all the learned faculty members of my department who directly and indirectly assisted me in preparing this thesis report.

I am indebted to the staff of the University Library and lab assistants for their helpfulness in enabling me to collect the required data.

Last but not the least I would like to thank all my friends who spared their precious time for contributing their skills to my need.

I owe immensely to my parents for their moral support and encouragement.

Dhanmeet Kaur

ABSTRACT

A highly accelerated growth of e-market has led to a well flourished online auctions scenario. Along with the attraction of numerous users world-wide, the online auctions have also allured in multiple frauds, periodically changing in nature and strategy to accustom to the proposed fraud detection and prevention approaches. As per the Internet Crime Complaint Center report 2013, auction fraud is enlisted as the topmost fraud accounting for drastic monetary losses. Amongst the online auctions frauds, shill bidding seems to be the most prominent fraud. In this research work, we present a 'Variable Bid Fee Methodology' as a prevention technique for shill bidders. A bidder is charged for each of his bid based on the amount he bids. The winner of an auction wins back the charges he paid as bid fee; he gains an additional benefit to recover the bid fee he paid for the auctions he earlier lost in. This maintains the competitive spirit of an auction. On the contrary, the inherent nature of a shill bidder of frequently bidding in an auction and never winning one, will cause him perpetual monetary losses. We proposed this methodology based on the idea that the risk of losing money will reduce the tendency to exhibit shill behaviour. This methodology acts as a preventive measure for both single shill bidder and collusive shill bidders. The approach has been exemplified with auction scenarios involving shill bidders and experimented with bidder data set collected from one of the largest auction website – Ebay.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1 Online Auctions

Derived from Latin, the term ‘auction’ means ‘I increase’. In economic terms, auction is the process of buying and selling item/ items where numerous people vying for the item/items compete with each other to win the item.An auction is overlooked by an auctioneer. The participants of an auction continuously outbid each other till there is no - further outbid. The item on auction is finally sold to the highest bidder with the selling price being his bid i.e. the highest bid.

Auctions have said to have a history as ancient as 500 B.C. It began with the auctioning of women for marriages, expanding its domain to the auctioning of slaves and war places where the soldiers dug a spear in the ground to mark their victory during the Roman Empire.[1]Auctions found a market place while selling off the mortgaged items when the debtors failed to clear their debt but the auction market flourished in its full glory after World War II. The auction venues shifted from typical auction houses to common meeting places like coffeehouses. Special catalogs were creatively designed and distributed to publicize the items to be auctioned. This business has prospered through ages undergoing slight changes and today it has finally evolved into a huge online business.

Ebay is one of the leading online e-commerce and auction websites. Its drastic revenue growth graph from 3.27 billion dollars in 2004 to 17.9 billion dollars in 2014 is an eye catching proof of the growth of online auction market in the recent years [2].A new exciting aspect of online auctions came into light recently in April 2015 when online bidders for the first time participated in live auctions, a collaboration of Ebay and Sotheby’s (http://www.live.ebay.com/lvx/sothebys). A ‘live auction’ is slightly different from the usual online auctions. The auction duration is limited to a few minutes and has a true touch of real auctions.

1.2 Need for Online Auctions

A major flaw in the offline auctions is collusion wherein the group of people bidding for a good forms a secret alliance for their mutual benefit thereby marring the fair-play notion of auctions. Online auctions to some extent can provide a solution to this. Collusive frauds are major online auction threats yet online auctions still have an edge due to the vast platform of bidders; all the bidders cannot be colluded. Also, the auctions no longer remain limited to a particular area and a few participants. Online auctions are a boon for sellers as the sellers can reach out to a huge number of prospective buyers with varied interests, needs and paying capabilities.A huge virtual auction market has developed since the auctions have gone online. From property sales, government tenders, BSNL numbers, vehicles, appliances, décor items, clothing, children's toys, to almost every day to day utility item is being auctioned these days. Furthermore, the sellers have got rid of the overhead costs involved in organizing the auctions and paying the personnel involved. Today, there are various auction websites available to choose from and the initial investment is no more than 2-5% of the selling price of the item that needs to be paid to the website. Moreover, with everything being jelled into an e-world, online auction seems a far dream come true for the comfort and convenience of people involved in auctions.

1.3 Online Auction Scenario

Inspired by the offline auctions scenario, the online auction too involves the similar terminology. The major components in an online auction are players, goods, auctioneer, valuations, bids, price and utility.

- **Goods**: It is the set of items that have been put up for auction. The auction maybe for a single item or multiple items. Also it maybe for a fixed quantity or varied quantity of goods.
- **Players**: It is the set of all online bidders who have registered for an auction. Each player is in competition with the others to win the good/goods in an auction.
- **Seller**: The person who initiates the auction by putting up an item/multiple items for auction is the seller. The aim of the seller to sell his good at the maximum profit possible. The seller may have an additional privilege of setting the least

limit on the selling price of the good. This depends on the algorithm used for auctioning.

- **Auctioneer**: The convener who manages the whole auction process is the auctioneer. In online auctions, algorithms take over the task of the convener. The auctioning algorithm gives the final decision of the winners of the respective goods, quantities, if any, and their selling prices.
- **Valuation**: Each player when enters into an auction has an estimate of the price of the goods on auction and bids as per his valuation of the goods. It is basically an individual's perception based on knowledge and experience. In another words,
- valuation can be defined as the extent to which a player values a good. It is the maximum price a player is willing to pay for a good. Based on his valuation for a good, a player bids for it. Valuation is private to a player i.e. a player's valuation is unknown to the other players participating in the auction; the auctioneer is also unaware of the valuation of each player.
- **Bids**: It is the set of prices that the players submit when the auction process begins. An auction maybe a sealed bid auction or an open auction. In a sealed bid auction, a player is allowed to bid only once whereas in open auctions a player can submit multiple bids. Players can also change their bid multiple times in an open auction.
- **Price**: It is the amount which the player has to pay if he wins. In other words, it can be considered as the selling price of the good/goods being auctioned. The price of a good can be determined only at the end of the auction process and depends on the auction algorithm employed. For example, the price depends on valuation as is the case in Vickery auction/ The Second Price Auction; it depends on the bid in case of The First Price Auction.
- **Utility**: Also known as the payoff, it is the benefit earned by the player by participating in an auction. Each player aims at maximizing its utility. Generally a positive utility signifies profit, negative utility is loss and zero utility is a no profit and no gain situation.

  Every auction algorithm has its own utility function.

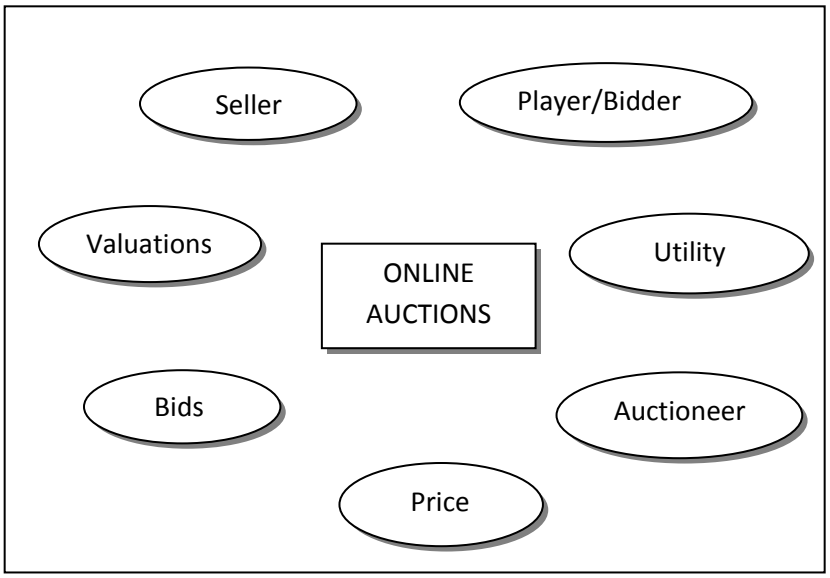


Figure 1.1: Components of an Auction Process

1.4 Types of Online Auctions

With the changing time, the goods for auction took a different form and the auctioning strategies evolved.

Following are a few different types of auctions:

- **The English Auction**: This is the most common and known type of auction in which a group of people bid against each other, raising their bids up to their maximum limit to beat out the rest of the competition to win. The bidder with the highest bid wins. An auctioneer manages the whole auctioning process. He may set a limitation on the minimum increase in the bids in every round and may also initially set a minimum selling price for the good being auctioned.
- **The Dutch auction**: In this type of auction, the auctioneer starts by setting the highest price on the goods to be auctioned. The goods may be similar but of varying quantities. The auction procedure continues by lowering the price in each round. In the end, while selling the goods, the preference is given to the bidders from higher to lower bids. The highest bidder chooses the good and the quantity, then the choice is given to the bidder with second highest bid and the process

continues. The goods remain unsold if the lowest set selling price of the goods is greater than the highest bid.

- **The Sealed Bid Auction**: In the sealed bid auctions, the bidders' bids remain unknown to the rest bidders. The bids are secret. The selling price of the good auctioned is the bid itself.
- **The Vickery Auction**: This auction forms the basis of online auctions. It has all the desirable properties of an efficient and good performance online auction. The highest bidder wins but pays the minimum bid that beats out the rest competition i.e. the selling price is the second highest bid. The bids are private as in the sealed bid auction.
- **The Combinatorial Auction**: This auction involves the sale of a group of goods and not the individual goods. The highest bidder for a good may not win as an individual but the bidders involved in the overall highest bid on a group of goods win. Deciding the winner in this auction is a complicated task.
- **The Reverse Auction**: In reverse auction, the buyer seeks to buy a good at the lowest price possible. Multiple bids are welcomed from the sellers. These bids are public. The minimum bid at any time is displayed. It ensures a fair trade.
- **The Reserved Auction**: In the reserved auction, the final decision to sell the object or not is vested in the hands of the seller. If the highest bid is not enough for the seller, he may not sell the good even to the highest bidder. This auction is safe for sellers as they can control their loss or profit margins.
- **Penny Auctions**: This is similar to the English auction but with a few modifications, the most important one being that each bid has a bid price. The auction for an item begins with zero selling price. If the first bid is 2$, the next bid will be equal to or greater than 2$ plus the predefined fixed bid fee say 0.5$. So the minimum second bid will be 2.5$. Another important element of this auction is the timer which is initially set to a fixed time. On expiration of the timer i.e. when the timer reaches 0:0, the last bidder wins the bid paying his last winning bid as the price for the item he won. This auction tends to have unexpected outcomes.

1.5 Online Auction Process

Today the sellers and buyers both have umpteen choices when it comes to auction websites in contrast to the limited auction houses prior to the e-auctions. Various auction websites can be compared based on the services provided and their reputation. Based on the type of auction the seller and the buyer want to participate in, choosing the appropriate auction website Is the first step. The seller registers himself with the website. A registration fee has to be paid to the auction website by the seller for creating a new auction on the website. This investment is negligible as compared to the investment involved in conducting auctions in the auction houses. After the completion of registration,the seller advertises his item for auction in the suitable category. The product description should be catchy enough to attract the bidders and in accordance with the actual product condition so as to avoid any dissatisfaction on the actual product delivery to the winner.The seller may fix an initial price for the auction to begin with.

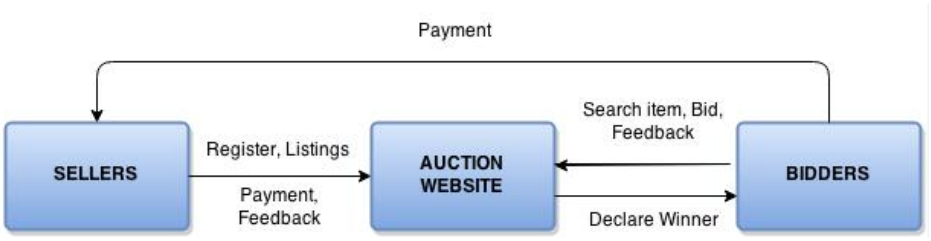


Figure 1.2: Online Auction Process

A new bidder has to sign up to create an account to participate in an auction. A bidder searches for a desirable item by comparing the prices and considering the seller credibility based on the feedback systems that most of the auction websites have incorporated these days. The auction websites provide an easy search of auction items by category and an additional option for advanced search as well. Each auction is active for a limited period of time. The participating bidders compete with each other, outbidding one another's bids to finally win the auction. Some websites like Ebay have a proxy bidding system that increments the bids on behalf of the bidders based on maximum price a bidder is willing to pay that is input when a bidder enters an auction. When the auction timer runs out, the auction website declares the winner based on the type of auction and

its rules. The winner is liable to pay the winning amount to the seller. Both the seller and the winning bidder can rate each via a feedback mechanism on the auction website. A positive feedback for a seller proves advantageous in future auctions as it builds up the reliability and credibility of the seller.

The following is an example of an Ebay auction from a bidder's perspective. The screenshots demonstrate the various stages a bidder goes through while participating in an auction.

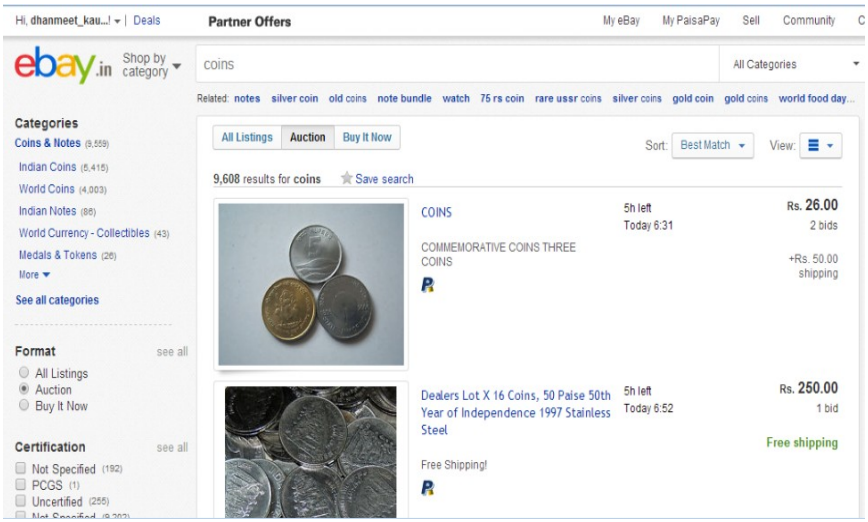


Figure 1.3: Screenshot of searching a relevant auction on Ebay

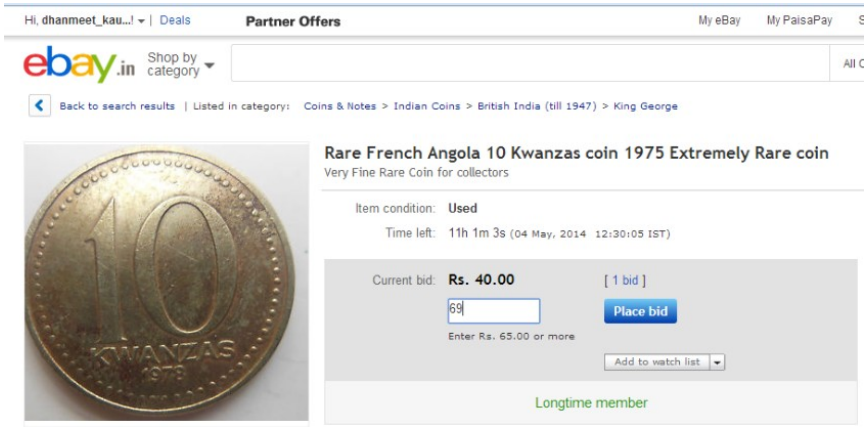


Figure 1.4: Screenshot of placing a bid

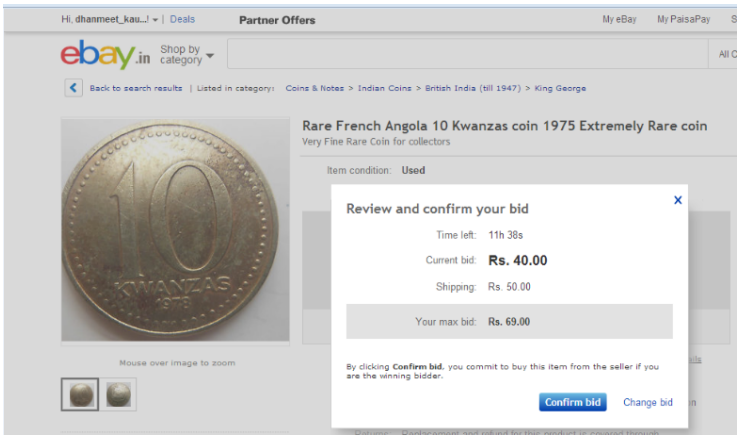


Figure 1.5: Screenshot of bid confirmation

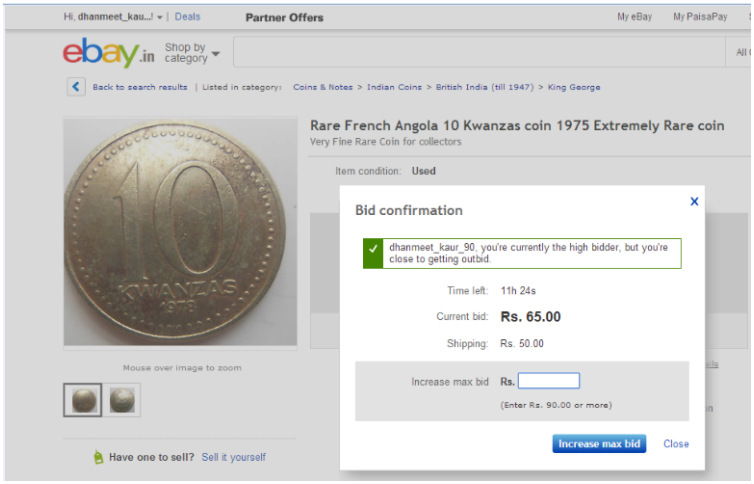


Figure 1.6: Screenshot of notification of highest bidder

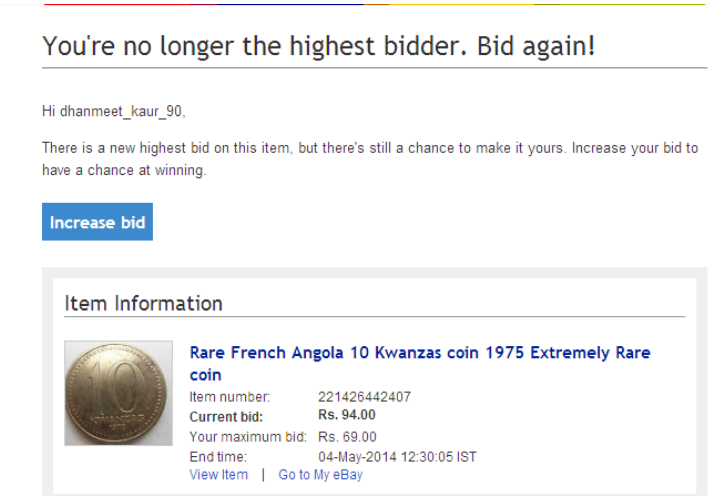


Figure 1.7: Screenshot of notification to increase the bid

1.6 Game Theory Aspect of Online Auctions

With the increasing popularity of online auctions, a new field for research is being explored for the designing of fair auctions by fusing the game theory with the auction world.

Algorithmic game theory is a branch of Computer Science that deals with the decision making in strategic environment. A strategic environment involves intelligent participants whose moves are based on calculated future predictions and perception of the behaviour of the competing participants i.e. there is a rational thinking involved in every move of the participant. There is a great extent of uncertainty involved which may lead to unexpected results. Therefore, the art of rule-making is very important and hence the application of game theory plays a vital role in understanding the online auction algorithms.

The aim is to design a system that

- Performs well with strategic participants exhibiting selfish behaviour
- Gives a desired output from designer's prospective

This research work comprises of six chapters. In Chapter 1, we discussed the basics of online auctions- the process, types and the research perspective in terms of design. A highly accelerated growth of e-market has led to a well flourished online auctions scenario. The unprecedented growth of the e-auction market has provided a huge platform where a multiple variety of fraud is nurturing in different forms. Along with the expedition of the crime, these frauds also pose a threat to the basic idea of bidding in a competitive environment to win an item. Chapter 2 focuses on the types of online auction frauds and the various suggested ways of dealing with them. The problem statement and the objectives of this research work are highlighted in Chapter 3 followed by a detailed explanation of the proposed Variable Bid Fee methodology in Chapter 4. Chapter 5 gives the experimental results and justification of the proposed methodology. Chapter 6 concludes this research work along with illuminating the future scope of the work.

2. Literature Review

Fraud is a malevolent activity wherein an intentional harm is meant to another person for a personal benefit. Fraudulent activities have crept in the e-world as well. Along with the attraction of numerous users world-wide, the online auctions have also allured in multiple frauds, periodically changing in nature and strategy to accustom to the proposed fraud detection and prevention approaches. Besides providing ease, comfort and convenient trade, this significant online auction market attracts umpteen fraudsters. As per the crime report by IC3 (Internet Crime Complaint Center), the auction fraud has drastically increased from being at ninth position in 'the top most common IC3 categories' with 5.7% complaints registered, to being the topmost complaint category as 'auto-auction fraud' in 2013 report with a total of 14,169 registered complaints accounting for a loss of $51,581,511[3]. Every one out of four complaints is related to an online auction. In a recent global study in 2014 it was highlighted that the average cost of a cyber-crime has increased by 9% and the average time required to solve a crime has increased from 36 days to 45 days in a year [4].

2.1 Types of Online Auction Frauds

E-auction market is a huge international market ranging from auctionsof a toy and story books to auction of properties, vehicles and objects of historic importance, involving a handsome amount of moneys and participation of buyers and sellers from all around the world. Such a vast extent of the online auction market is also a platform for various auction frauds – multiple in number and form. The minimal auction fee charges and the anonymity of participants lures in fraudulent groups.All types of frauds are difficult to define and hence hard to identify as well. Moreover, the types of frauds keep evolving over a period of time due to the ever changing auction market and the detection/ prevention approaches being proposed and implemented to identify and deal with the auction frauds.The auction scenario involves two major categories of participants i.e. buyers and sellers. The fraud can, therefore, be either from the seller or from the buyer side.

Some of the most common identified auction frauds are:

- **Non-Delivery of goods**: The buyer doesn't receive the good he has won in an auction and has already paid for in case of immediate payment schemes.
- **Misinterpretation**: The goods won in an auction are not as described by the seller on the auction site. The seller may have tactfully described the good or used Photoshop to enhance the product quality in the uploaded image on the site.
- **Fee-Stacking**: The seller increments the price of the good as determined at the end of the auction by applying additional charges such as shipping and handling charges which were not mentioned of during or prior to the auction process.
- **Triangulation**: It involves three actors – a buyer, a seller and a fraudster. The fraudster acts as a normal buyer to buy a commodity from a seller but uses fake credit card numbers for payment. The same fraudster acts as a seller to sell the product to a naïve buyer and immediately receives payment from him. The buyer becomes the victim of stolen product investigation.
- **Internet Fencing**: It is a seller fraud when a stolen item is put up for an auction.
- **Bid Shielding**: The buyer, through a fake account, hikes the price of an item beyond it's worth, thus driving away potential buyers. The buyer withdraws this bid at the last moment and the second highest bid wins which is the buyer's original account's bid or another buyer cooperating in the fraud.
- **Bid Shilling**: The seller creates fake accounts to act as buyers which increase the bidding price by small amounts during the auction, thus resulting the competing real buyers to further increment their bid and eventually the product gets sold at a higher price.
- **Accumulation Fraud**: The seller gains a good reputation through legal feedbacks by selling commodities of lower price. A high positive feedback wins the trust of buyers. At this stage the seller puts a high priced item for auction with an immediate payment requirement. The seller gets the payment and doesn't deliver the product.

- **Ballot stuffing**: The seller creates fake accounts that win the seller's auctions and provide him with a positive feedback. The seller reputation is increased based on the manipulated increment of the feedback score[5].
- **Buy and Switch**: The buyer receives a good quality product, replaces with an inferior one and claims a replacement or the price paid.

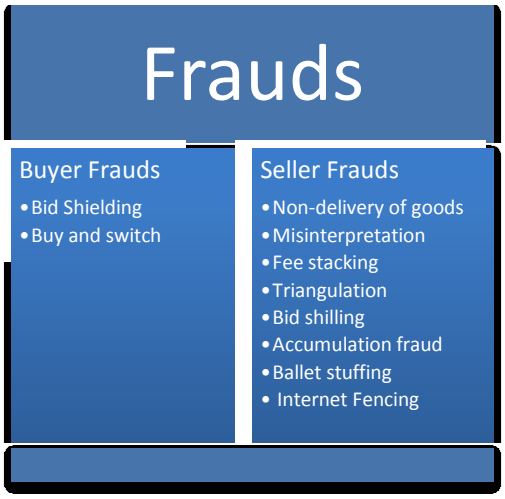


Figure 2.1: Types of Auction Frauds

By closely observing the behaviour of the buyers and the sellers individually, the various frauds can be looked into. However, fraud detection becomes complex when multiple frauds occur simultaneously, as is the current online auction fraud scenario. Nowadays, the fraudsters don't get involved in a fraud directly, an additional group of accounts called 'accomplice' are involved. Fraudsters are not directly linked to the honest buyers. Accomplice forms an intermediate layer, linking to fraudsters and protecting their identity on one end and simultaneously manipulating the honest buyers. The strategies of frauds keep changing as numerous fraud countermeasures evolve.

2.2 Fraud Dealing Approaches

Multiple approaches have been proposed and experimented to control the immense losses incurred due to auction frauds. Over a period of time, the scope of research has shifted from fraud detection to fraud prediction. Reputation systems, social network analysis, data mining, phased model, belief propagation over MRF model and a combination of fuzzy logic and genetic algorithms are the major areas that have made significant

contribution in milestone achievements in the exploration of efficient techniques for fraud detection and prediction in online auctions.

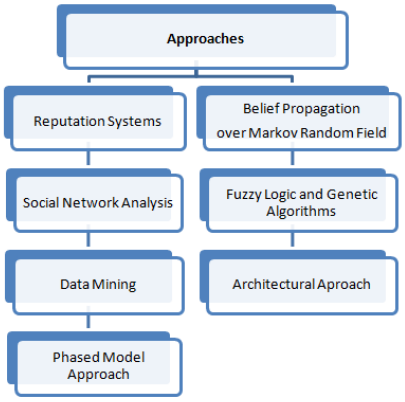


Figure 2.2: Fraud Dealing Approaches

2.2.1 Reputation Systems

The most basic and initially proposed countermeasure for online auction fraud was an integration of reputation system to the auction sites. It is based on the feedback mechanism. When an auction is over, both the seller and the buyer can leave feedback for each other in terms of rating or comments which gets accumulated in their respective profiles.Their reputation systems can be either centralized or decentralized [5]. A centralized reputation system is one wherein the participants who have direct interaction with a user leave their feedback at one place that is combined, calculated and presented by a common central authority whereas, in a decentralized reputation system, the feedback and other information is obtained from all the interacting participants with a particular user via a communication protocol. The feedback score is combined and calculated at receiver node based on the information received. Each node employs a different formula for the final feedback calculation as opposed to the centralized reputation systems where a common method of calculation for feedback score is used. The online auction websites have incorporated the centralized reputation system. Many auction sites use the reputation system because of its simplicity and comprehensibility by a layman.

This simple approach tends to have a few drawbacks. The feedback can be obtained by unfair means. A seller may create multiple accounts, participate and win its own auctions to rate itself positively. Also, an honest user may refrain itself from sending a negative feedback in fear of receiving a negative one in contempt [6]. Thus, both positive and negative ratings can be tampered with. Moreover, the comments do not add up to the final rating displayed on the user's profile. Though the content analysis on textual feedback may reveal additional information yet it is avoided on account of the amount of processing time involved. As is the case with the feedback scores, even the comments can be manipulated.

Moreover, for a relatively new user a negative response will reflect more in his score than a negative response reflection for an old user. For example, the feedback score for a user with 10 positive responses and 1 negative response will be lower than the user with 50 positive responses and 1 negative response. Therefore, the sellers tend to establish a good reputation first before initiating a fraud. To overcome this flaw, an approach that could provide incentive for the users to remain honest was suggested which involved an exponential smoothing function to calculate the reputation score [7]. A more refined approach for the calculation of reputation score has been recently suggested which other than providing incentive for honesty, also considers the selling price of the item auctioned and the reputation score of the buyer while rating the seller. It considers six various parameters for the score calculation [8].

Mikolaj Morzy proposed two algorithms for the enhancement of reputation system by the incorporation of two factors- credibility and density, derived from introspection of interaction amongst the users. These factors help us to judge the users involved in an auction in a better way than the reputation systems based on mere feedback score [9].

Another type of reputation system proposed is a reputation system based on the characteristics of a seller. Statistical models are employed to find out sellers involved in inflation of selling price of an item by unfair means. The tendency of a seller to be involved in shill bidding is also considered by one of the models. A combined result of three models gives a confidence score for each seller thus helping an honest bidder avoid participating in a potential fraudulent auction [10].

2.2.2 Social Network Analysis

Social network analysis (SNA) is the mathematical and behavioural study of interactions amongst people or groups. SNA involves the analysis of graphs where the nodes represent the people or groups under study and the interactions among them are represented by the edges between the nodes.SNA finds application in fraud detection in online auctions based on the observation that accounts involved in collusive groups tend to have a complicated relationship amongst them.

Google's Page Rank algorithm and HITS (hyperlink induced topic search) algorithm are the leading algorithms to rate the importance of a webpage on the internet. SNA along with the Page Rank algorithm has been explored to discover and rate the important accounts in a subgroup. Page rank algorithm rates the web pages based on the page rank of the pages pointing to the given page and the number of links leaving from those pages. The webpage with numerous inbound links from higher rated web pages will be an important webpage and thus, will have a higher rating as well.

To detect collusive groups, initially SNA is utilized to detect a suspicious subgroup based on the chosen characteristics of graph such as clique, k-core or k-plex. A web crawler, based on the chosen clustering parameter, helps capture potential fraudulent groups. Wang and Chu (2008) used the robbery algorithm an alternative to clustering algorithm to detect collusive groups. According to this algorithm,

- Initial score of an account is equal to the degree of the node or 1.
- If the degree of the node is greater than the degree of its neighbors, the score of neighbors becomes zero and gets added to the score of the central node.

The nodes with score greater than zero are fraudulent accounts. Experimentally, along with high precision rate robbery algorithm also provides high false positive rate.

Once the suspicious subgroup is identified, each node is assigned a score based on the modified page rank algorithm called Auction Fraud Rank algorithm which characterizes each node with a certain degree of suspiciousness parameter. The algorithm incorporates three additional parameters to the original page rank algorithm [11].

YanlinPeng et al. proposed an algorithm for collusive fraud detection in online auctions [21]. Sellers with the help of accomplices commit frauds. Based on this notion the interactions between users is observed, honest users are filtered out based on some observed behaviour and finally bipartite cores are identified from the graph. The bipartite cores represent fraudulent sellers and their accomplices.

To inspect the bipartite nature and explore the hidden relationships between sellers and accomplices, an advanced algorithm is suggested in [22] wherein the online auction scenario is modeled onto a maximum likelihood detection problem and message passing is used to solve it. The algorithm iterates for a finite number of times and provides a scalable solution for the detection of collusive frauds that involves sellers and their accomplices.

2.2.3 Data Mining Approach

Another basic approach to deal with detection of online auction fraud involves two steps, namely, detecting outliers and building a decision tree.

Figure 2.3: Data Mining Approach

Based on a set of certain attributes that play a role in analyzing and probable detection of a fraud, a dataset is created from the online auctions. The selection of attributes plays a major role in the proposed fraud detection system. The size and quality of the dataset of attributes directly affects the cost, accuracy and efficiency of the detection model being proposed. Wen-Hsi Chang and Jau-Shien Chang have categorized the attributes into three major groups – price related attributes, rating related attributes and item related attributes [12]. A major related issue is that not the whole history of users is available. Even if made available, it would lead to intensive computation and time consumption. Thus, the main focus should be on the selection of a small and relevant set of attributes that don't

degrade the system performance by exhausting the available system resources as the system has to deal with the numerous auctions occurring at any real time instance. The need of the scenario being an efficient fraud detection system, the careful selection of attributes plays a vital role in determining the success factor of the proposed system.

Once the dataset has been created, the next step is to analyze the behaviour of the users involved in the auction based upon the selected attributes and detect the outliers. The behaviour of fraudsters differs from the behaviour of normal users in some or the other aspect. Therefore, detection of outliers corresponds to the detection of potential fraudsters. Multiple methods have been proposed for the detection of outliers – an unsupervised learning technique such as a one class SVM being one of them [13]. This approach works well where shill bidding is involved.

Social network analysis on a set of user IDs is another technique to obtain a set of potential fraudsters [14]. Sometimes the normal users too might behave in a manner similar to that of the outliers or the potential fraudsters. To differentiate between the normal buyers and the cheats, data mining technique approach is incorporated. Usually C4.5, a decision tree learning method is used. Based on the pool of potential fraudsters identified in the previous step, C4.5 builds a decision tree. Each node of the tree gives a rule. Thus, these set of rules become prediction rules for the identification of fraudsters in future. Further, a close human inspection of the decision tree can help identify normal users from the fraudsters. Traversal of the tree from the root to the leaf of each branch while making a manual judgment at each node results in the bifurcation of data into fraudsters and innocent users. An additional benefit of this step is that based on the human observation, the set of attributes chosen initially can be modified to improve the overall performance of the system. It can also be considered as a countermeasure for the ever evolving techniques of the fraudsters based on the understanding of the current fraud detection mechanism employed.

The problem of separating out fraudulent users from honest users in online auctions can be mapped to an outlier detection problem. Vinicius Almendra and Bianca Roman have employed an exploratory data analysis technique viz. Andrews Curves to detect outliers[23]. The auction data is defined by a Fourier series function and the resulting

curves are analyzed to give results. Eventually, a one class classifier is developed that detects the fraudulent users from the honest users.

Clustering is a data mining technique that categorizes the input data set into different groups wherein the members of a group are more similar to each other than the members of the other groups. In [24] the transaction history of users is considered in phases and analyzed. K-means is the most popular clustering technique where the user can define k i.e., the number of clusters to be formed but this paper describes the use of X means instead of K means for clustering. X means is a variation of k means algorithm. The clusters are not defined by the number but are formed based on some characteristics which are analyzed from the phased transaction history of the users. This helps bring to light the fraudulent behaviour which is hidden and keeps changing during the initial phases of a fraud and comes forth only after the fraud is committed. Once the clustering of input data is done, C4.5 decision tree is used to formulate the rules and thus form a fraud detection model.

2.2.4 Phased Model Approach

Phased model approach is a predictive fraud detection approach. The available transaction history of the users is partitioned into smaller sections for an in-depth examination. This helps analyze the fraudulent account in detail prior and during the actual commitment of the fraud. The transaction history of a fraudulent account can be considered divided into two major portions – prior to the occurrence of the fraud known as the latency period of the account and during the occurrence of the fraud, known as the execution period. A phased model proposes analyzing the account at a few p% stages from the starting. For example, 30% phased model implies the analysis of the first 30% of the transaction history of the total lifespan of the account. Other than the timeline, even the rating can be considered as the basis for the phased model construction such as positive, negative and neutral rating phases of an account. At each phase, a learning algorithm is employed to explore the selected attributes that determines the behaviour of a user. W.H. Chang and J.S. Chang suggested a modified wrapper approach [12]. They even suggested HCM, a hybrid complement phased model aimed to improve the accuracy of the detection system considering the fact that the latter part of the transaction history

of lifespan of an account reveals more information wherein p% phase model implies the last p% of the transaction history of the account. The study of the phased model can be considered analogous to the study of the fraud development because a fraudulent account will be characterized by a noticeable change in its behaviour over its transactional lifespan from its initial imitation of an honest user's activities to the execution of a well manipulated fraud. A general behavioural pattern of a fraudster can be concluded, which forms the basis for prediction of similar frauds in future.

In another paper by Jau-Shien Chang and Wen-His Chang based on phased model approach [25], they've suggested analysis of user accounts in various phases followed by the classification using regression. A regression technique and decision tree is combined to give a fraud detection model.

2.2.5 Belief propagation over Markov Random Field Model

2LFS, 2-Level Fraud Spotting, is a technique that combines both the bidder history as well as its interactions with other users to detect a fraud prior to its occurrence. The features in combination help detect fraud patterns. The problem is designed in form of the Markov Random Field Model, followed by the application of a belief propagation algorithm to yield the results [15].

A further modification of 2LFS has been presented as an algorithm named SPAN, Score Propagation over an Auction Network, to deal with the shrewdly planned collusive frauds in online auctions [16]. This algorithm inspects various graphs based on different selected feature sets of both the sellers and the buyers, to calculate a score which is combined to obtain a final anomaly score. The calculated anomaly score is transmitted using a modified belief propagation algorithm. A significant change was the modification of the update rule to include the level of interaction amongst users.

ShashankPandit et al. proposed an online auction fraud detection system called NetProbe which is based on Markov Random Field and Belief Propagation technique [17]. The interactions amongst various users of online auction site are represented in the form of an undirected graph. The state of each node in a graph depends upon its immediate neighbours and there can be many but finite states of nodes. A propagation matrix is

formed that defines state of nodes based on some probability. By using the belief propagation method, the nodes exchange their state information via messages, defining all possible states of each node until the algorithm converges as per the pre-defined conditions. In this way, interactions of fraudulent users are observed and can be inferred in future too for similar behaviour. This paper explains in detail the design of the NetProbe and also proposed an incremental model of it for more accuracy and speed.

In [5] the users are labeled as fraudsters or honest based on their representation as Markov Random Field followed by a loopy belief propagation algorithm. The loopy belief propagation algorithm is similar to the belief propagation algorithm. The minor modification is that the cycles it forms while passing messages amongst nodes are overlooked. This algorithm has no guaranteed convergence but that is a rare case. Most of the times, it converges and provides more precise and accurate results. The paper explains the representation of users as Markov Random Field in detail.

2.2.6 Shill Score Algorithm

Another significant contribution to the fraud detection techniques is the proposition of SSA, Shill Score Algorithm, by Jarrod Trevanthan and Wayne Read [18]. Its main focus is on detection of shill bidding. Based on the analysis of the behaviour of an individual shill bidder or a collusive shill bidding techniques, formulae have been devised to calculate the shill score that assists in the identification of shill bidding accounts.

2.2.7 Fuzzy Logic and Genetic Algorithms

Fuzzy logic and genetic algorithms have together been experimented with, to present an altogether different approach for fraud detection in online auctions. To capture and identify the fraudulent behaviour and characteristics, social network analysis is used. For better and faster results, the inspection of fraudsters begins from the already black listed accounts. Fuzzy rules are used to depict the fraud detection rules followed by the basic genetic algorithm functions that are employed to optimize the rules, thus resulting in a fraud detection model [19].

2.2.8 Architectural Approach

Another variant of a fraud detection mechanism is the architecture approach wherein a model for the online auction scenario is suggested to provide better security against frauds [20]. A layered model is suggested which enforces high security even at the time of registration. The second layer controls all the functionality of the auction sites and the last layer is devoted to the protection and management of auction data. To ensure a fair auction, different modules in all three layers employ strict management and security constraints.

An Online Hybrid Model (OHM) is an architectural model initially proposed in [26] and further elaborated in [27]. From architecture point of view, it consists of three layers viz. users, ohm and web server. The first layer of users is responsible for fraud prevention by employing strict registration conditions and validation checks. The next layers are responsible for both fraud detection and fraud prevention. Fraud detection techniques have been explored in detail in [27] which basically involve various registration modules for both users and the web servers. The paper explains the complexity of implementing the registration modules and the OHM policies.

3. Problem Statement

3.1 Issues in Previous Work

With the widespread of e-commerce, auctions being an integral part of business market, have shifted from auction houses to auction websites. Along with the provision of ease, comfort and convenience, online auctions also provide a huge platform for multiple frauds which keep changing over a period of time. Many approaches for fraud detection and prevention as discussed in chapter 2 have been suggested and experimented with. The major issues of the previous work that we came across are

- Most of the suggested work is based on the study and analysis of types of relationship amongst auction users viz. relationship between seller-buyer, buyer-buyer and seller-seller. Social Network Analysis involves the study of patterns amongst nodes to reach conclusions. Even the belief propagation over Markov Random Field model is based on the interaction amongst users. Study of the auction user relationships requires in detail the historic data of both sellers and buyers. The data provided at the auction sites is too limited to form basis of analysis from a user's point of view. The available historic data of buyers and sellers is limited to a time period that varies from 15 days to one month. Also, all the parameters associated with a user are hard to obtain to remark out the integrity and authenticity of a user account.
- Some of the approaches as mentioned in chapter 2 employ data mining techniques to identify fraudsters from honest users. These techniques involve the availability of computational resources and time.
- The most common and widely incorporated fraud detection and prevention mechanism, i.e., reputation systems lack in various aspects. The credibility of feedback has always remained an issue of doubt. Sellers create fake accounts to give positive feedback to themselves. Users refrain from sending a negative feedback to get a negative feedback in contempt. Moreover, the user comments which tend to give a better feedback are not analyzed to contribute to the overall feedback score.

3.2 Problem Statement

Shill bidding is one of the multiple varieties of commonly encountered auction frauds. In this research work, we propose 'Variable Bid Fee: An Online Auction Shill Bidding Prevention Methodology'. As suggested by its name, this methodology caters to the collusive frauds in online auctions. It is an auction design technique to deal with the strategic participants who form a collusive group. Auction business revolves around financial gain and loss. Financial loss due to formation of a collusive group in an attempt to fraud honest buyers through shill bidding forms the basic foundation of our proposed work. The auction is proposed to be designed in a way that the fraudulent accounts involved in shill bidding undergo a financial loss. Risk of losing money reduces the tendency to exhibit shill behaviour and thus serves as a prevention methodology for shill bidding.

3.3 Objectives of the Proposed Work

The main objective of the proposed methodology is

- to provide a preventive solution for shill bidding that requires
    - No bid history and user account parameter details
    - No involvement of data mining algorithms
    - No complicated calculations of fraud parameter
    - Minimum resources and time utilization
- to provide a common preventive measure for single and collusive shill bidding

4. Proposed Work

We propose '**Variable Bid Fee: An Online Auction Shill Bidding Prevention Methodology**' to deal with the most prominent fraud in an online auction- shill bidding.

4.1 Shill Bidding

Shill bidding is an online auction fraud technique in which the seller tries to manipulate bidders into bidding more than they would have in a normal auction scenario, thus eventually increasing the selling price of the item. A shill bidder manipulates the honest bidders by bidding higher than them. An honest bidder, unaware of the fact that the increased bid was by another honest bidder or by a shill bidder, bids more and thus is trapped in a vicious cycle of bids in order to win the auction. The seller may achieve this by creating fake accounts or by colluding with other sellers to form a group. A group of shill bidders is formed in order to avoid easy detection of shill bidding. On the contrary, a single shill bidder account may catch even the naïve bidder's attention if the bid history is analyzed carefully.

4.2 Characteristics of Shill Bidders

The major characteristics of shill bidders that help identify them from honest bidders are:

1. **Minimum increment in the bid price**

A shill bidder's main aim is to instigate the honest bidders to bid higher. Thus, if it raises the current bidding price by an appreciable higher bid amount, it might never get outbid by the honest bidders. Therefore, a shill bidder bids an amount barely enough to surpass the current highest bid.

2. **Shill bidder aims to lose an auction**

A shill bidder avoids winning an auction. Winning an auction by a shill bidder would mean that the item on auction is returned to the seller himself. This is another reason that the shill bidder increments the highest bid by as little as possible. Also this makes shill bidders more active during the initial phase of an auction rather than towards the end.

Shill bidding close to the auction termination time increases the risk of shill bidder winning the auction.

3. **Minimum inter-bid interval time**

It is characteristic of a shill bidder to bid as soon as possible after a bid by an honest bidder to keep the auction process in flow. All auctions are active for a limited period of time that is determined by the seller at the time an auction is created online. Therefore, the inter-bid interval is kept minimum by the shill bidders in an attempt to buy more time for the honest bidders to bid further.

4. **Bid Frequency and Interaction with the Seller**

Two prominent parameters that can jointly be observed to indicate a potential shill bidder from the bid history are a high bid frequency and a high interaction with the seller. Usually a high bid frequency and a high interaction with the seller are features of a potential shill bidder. On observing the bid history of bidders, if there is a bidder with a high bid frequency and high interaction with the seller with no auctions won of the same seller, then that bidder tends to be a fraud accomplice rather than an honest bidder. Sellers involve accomplices to commit auction frauds. Sellers avoid getting involved directly which reflects in their bid history.

4.3 Factors Encouraging Shill Bidding

Our proposed methodology for dealing with shill bidding stems from the fact that there are two major factors responsible for the encouragement of shill bidding in online auctions viz.

- **Ease of multiple accounts creation**: Shill bidding involves fake accounts or accomplices. The ease with which the online auction websites permit the account creation is no hindrance to this seller side fraud. A user can own as many accounts as desired. To restrict this we propose that stringent account registration conditions must be applied to deal with the first factor such as linking the user ids to a phone number. Not that the individuals don't own multiple phone numbers but it can limit the multiple accounts creation on auctioning websites upto a

certain extent. One Time Passwords (OTPs) help link a user account to a phone number. Registration modules as suggested in [27] can be employed to ensure that only honest auction users have the privilege of account creation.

- **Negligible or minimal participation fee**: Shill bidding entails high frequency of bids usually during the initial phase of an auction. Usually there is no fee for bidding in an auction and this leisure of free bidding is misused by fraudulent sellers. Inorder to limit the unfair bidding and encourage bidding by honest and dedicated bidders we propose that the bidders in an auction be charged with a fee – neither large enough to dissuade bidders nor small enough to be ineffective. Therefore we propose "Variable Bid Fee Methodology".

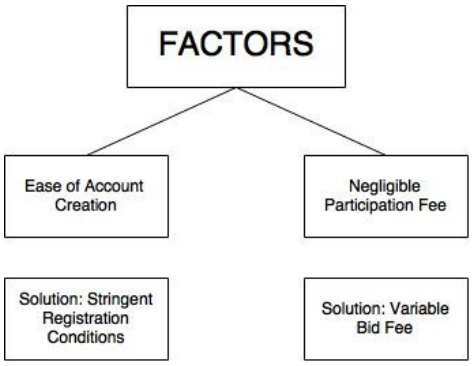


Figure 4.1: Factors Encouraging Shill Bidding

4.4 Variable Bid Fee Methodology

The key features of our proposed auction model to prevent shill bidding and encourage only honest auction buyers and sellers are:

- Every bidder will be charged for each bid in auction he participates in.
- The bid fee is neither too high to discourage bidders nor too low to be negligible.
- Bid fee is variable.
- Bid fee is directly proportional to the amount a bidder places as a bid. In this research work we have experimented with 5% of bid amount as the bid fee.
- The winner gets two benefits:
    - Bid fee of current auction is recovered.

- Prior loss in terms of bid fee in previously lost auctions can be recovered.

The main idea of our proposed methodology, in an attempt to abate the shill bidding fraud, is the introduction of a variable auction participation fee. The bidding fee is kept low so as not to discourage bidders but on the same time dissuades a single or a group of shill bidders colluding together to bid frequently in order to increase the selling price of the item. As the number of shill bids by shill accounts increase, so does the bidding fee. The risk of losing money reduces the tendency to exhibit shill behaviour. To keep up the spirit of an auction, the bidding fee scenario is so designed that the winner gets exempted from paying the auction fee. The winner's bidding charges are adjusted in the final winning bid, once he confirms to buy the item he won the auction for.

To account for the situations where an honest bidder exhibits a behaviour similar to that of a shill bidder, a provision is provided to win back the bidding charges of a previous auction. The winner of an auction gets a dual benefit. Along with concession of the bidding fee of the current auction he won, he gets to avail the opportunity to redeem the same amount of money lost in the previous auctions he didn't win. A shill bidder doesn't aim to win an auction and thus won't be able to recover the bidding charges of auctions. In an attempt to retrieve the lost amount, even if a shill bidder wins an auction that would imply a seller getting back his own item that was auctioned, eventually putting the seller at a loss whereas an honest bidder is motivated to win an auction to recover the prior losses if any due to a loss in an auction.

The following data structure and variables have been used in the further flowcharts and algorithmic representation of Variable Bid Fee methodology:

- **LA_i**: It represents the Locked Amount for each user i who bids in an auction. It stores bid fee and auction id. The bid fee paid for each bid by the bidder I is stored in this table. A bidder may participate in multiple auctions simultaneously. Therefore, auction id was required to be stored to keep and identify the bid fee charges corresponding to a particular auction.

When an auction duration ends, for that auction id

- If i = winner, the sum total of bid fee becomes benefit1 that gets subtracted from the selling price of the item.
- If i = loser, the sum total of bid fee gets transferred to the recoverable_amount$_i$.

- **Recoverable_amount$_i$**: It accounts for the benefit2 for the winner. It can be considered as a user vault that provides additional bonus equal to the benefit1 to the winner.
- **Recovered_amount**: It is the actual amount retrieved by the winner from the recoverable_amount$_i$. This amount too will be adjusted in the final selling price of the item.

Pictorially this proposed methodology can be depicted into two stages:

1. Placing a bid
2. Declaring the winner and selling the item

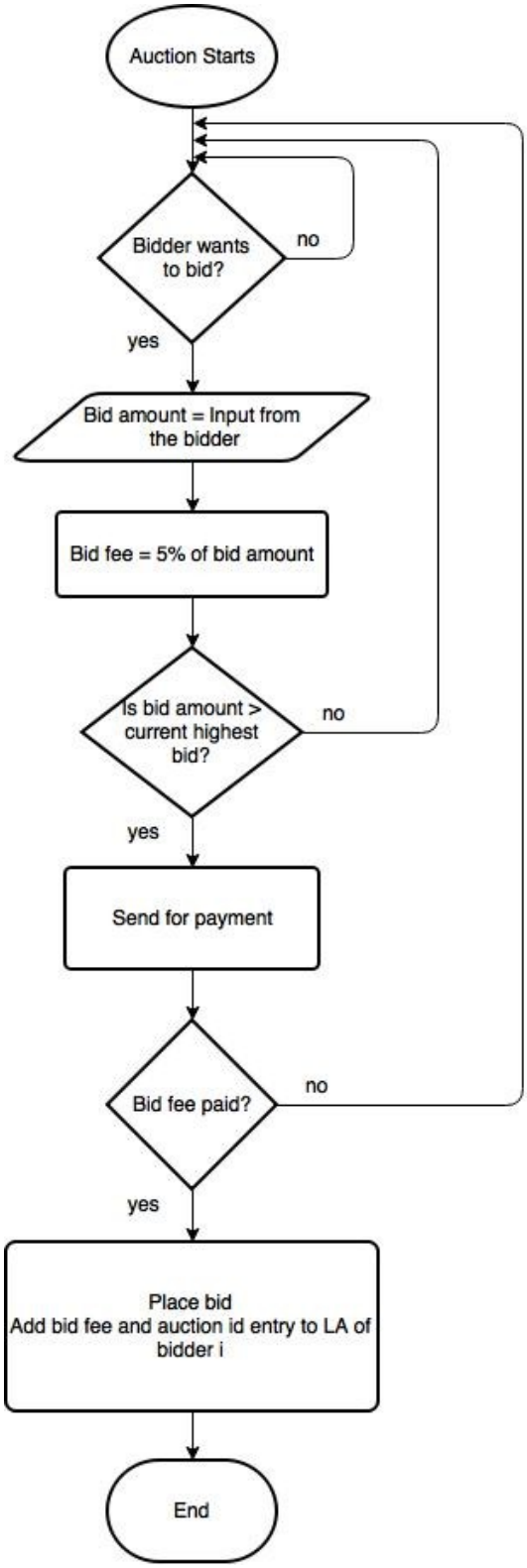
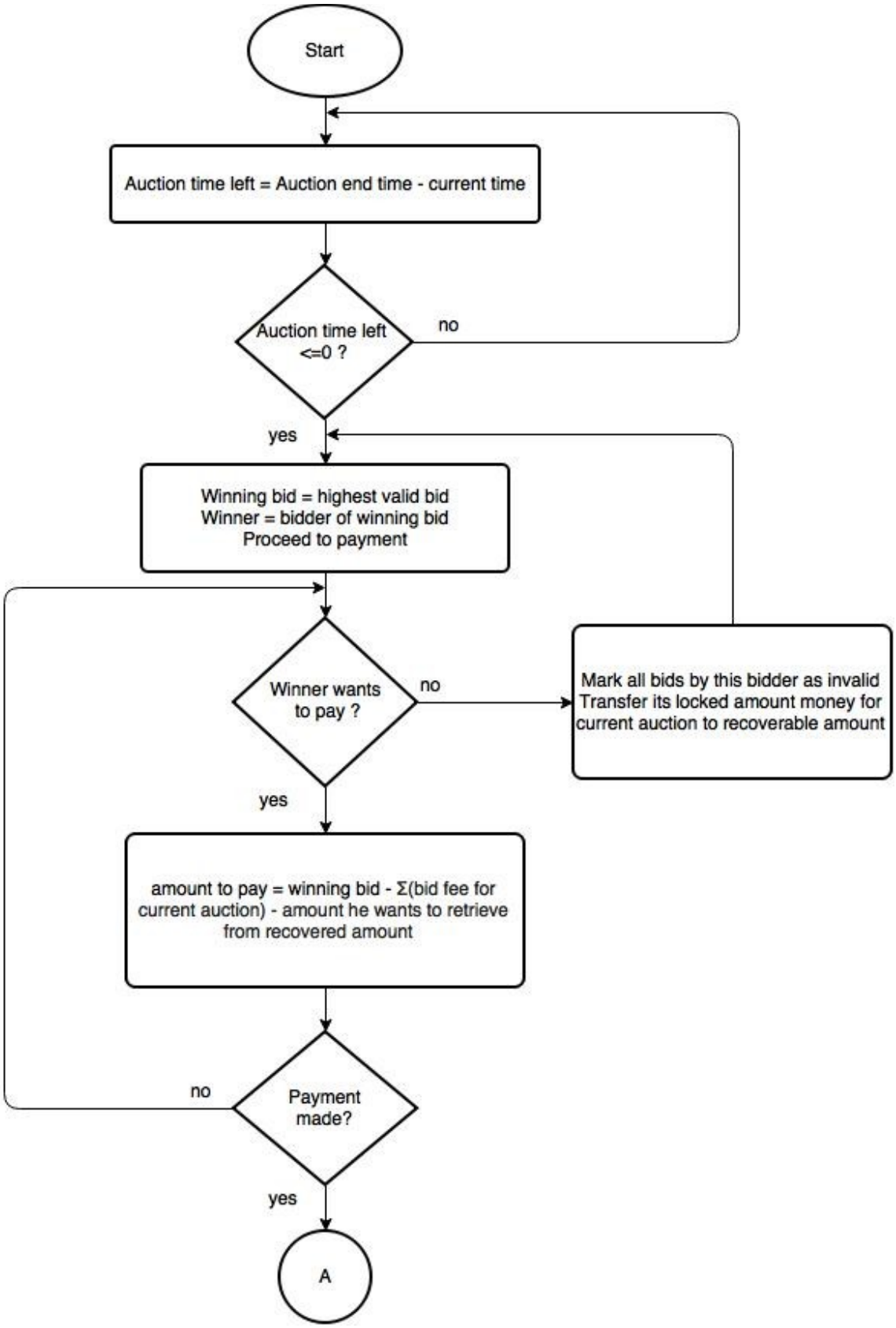


Figure 4.2: Placing a Bid


Start
Auction time left = Auction end time - current time
Auction time left <=0 ?
no
yes
Winning bid = highest valid bid
Winner = bidder of winning bid
Proceed to payment
Winner wants to pay ?
no
Mark all bids by this bidder as invalid
Transfer its locked amount money for current auction to recoverable amount
yes
amount to pay = winning bid - Σ(bid fee for current auction) - amount he wants to retrieve from recovered amount
Payment made?
no
yes
A

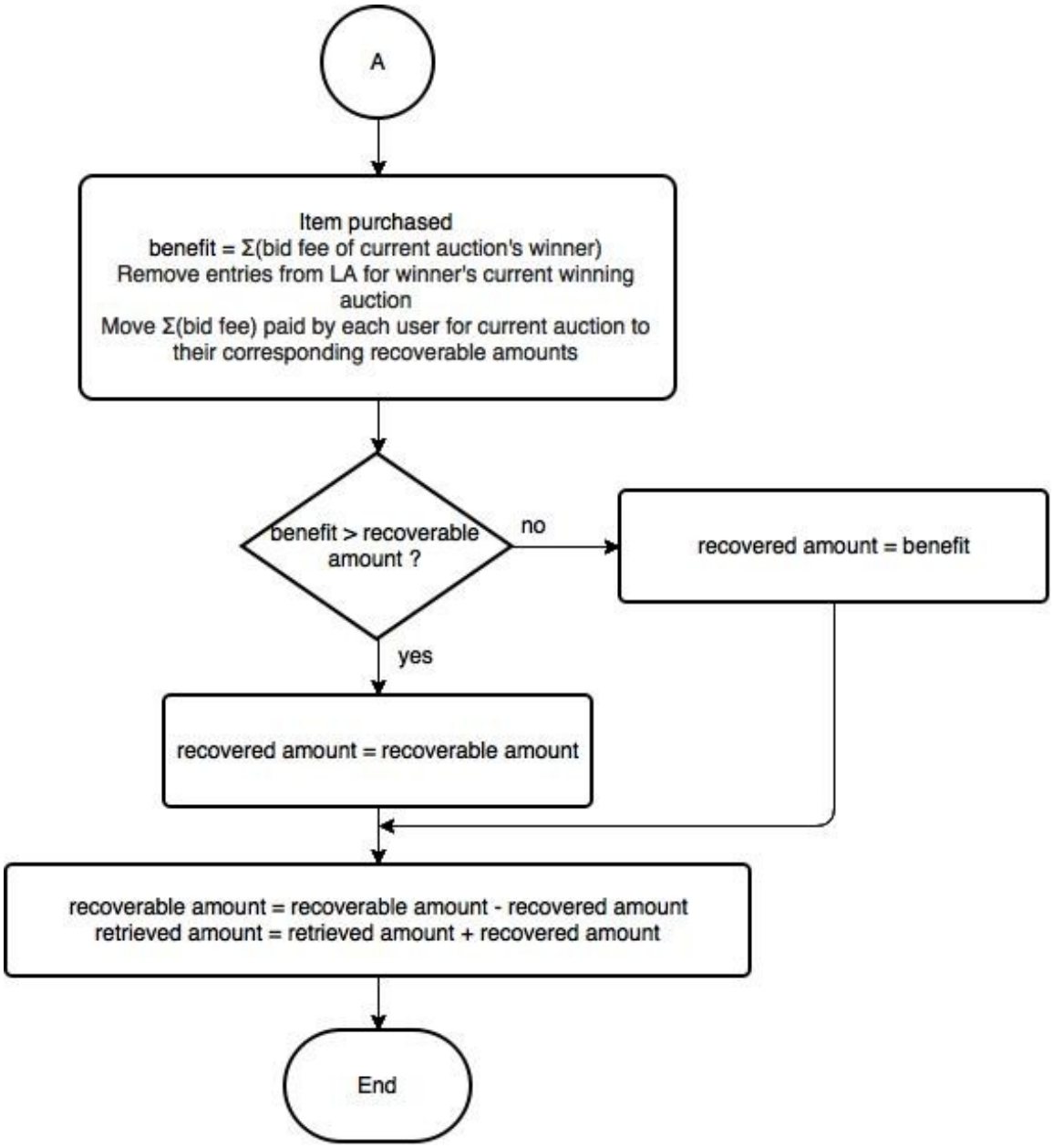


Figure 4.3: Declaring the winner and selling the item

1. Begin auction
2. auction_time_left = auction.end.time - current_time
3. while(auction_time_left)
4. if bidder i wants to bid
5. Input bid_amount from bidder
6. bid_fee = 2% of bid_amount
7. if bid_amount> current highest bid
8. Send for payment
9. if bid_fee is paid
10. bid = valid
11. LA_i: amount = bid_fee
12. LA_i: auction_id = id of the auction bidder i currently bid on
13. winning_bid = highest valid bid
14. winner = bidder of winning_bid
15. if winner wants to pay
16. amount_to_pay = winning_bid - ∑(bid_fee) for current auction_id – amount he wants to retrieve from total_recovered_amount
17. if payment made
18. item_purchased = true
19. for bidder = winner and auction_id = winning auction id
20. benefit = $\sum LA_i$: amount
21. clear LA_{winner}entries whereauction_id = winning auction id
22. for each bidder j where $j \neq$ winner
23. $Recoverable\_amount_j = \sum (LA_j: amount)$ where auction_id = id of auction bidder j lost
24. goto 15
25. else
26. for each bid of declared winner i
27. bid = invalid
28. $Recoverable\_amount_i = \sum (LA_i: amount)$ where auction_id = winning auction id
29. goto 13
30. if benefit > $recoverable\_amount_{i = winner}$
31. $Recovered\_amount_{i\ winner} = recoverable\_amount_{i = winner}$
32. else
33. $Recovered\_amount_{i = winner} = benefit$
34. $Recoverable\_amount_{i = winner} = Recoverable\_amount_{i = winner} - Recovered\_amount_{i = winner}$
35. $total\_retrieved\_amount_{i = winner} = total\_retrieved\_amount_{i = winner} + Recovered\_amount_{i = winner}$
36. End

Variable Bid Fee Algorithm

4.5 An Example

Consider the following auction scenario for a watch. Bidders b1, b2, b3 and b4 have participated in the auction. The following table shows the auction progress.

Table 4.1: An online auction scenario 1

Bidder id	**Price**	**Time**	**Bid_fee**
b1	100	0.10	5.00
b2	125	0.20	6.25
b1	127	0.21	6.35
b3	150	0.26	7.50
b2	170	0.40	8.50
b1	172	0.43	8.60
b4	175	0.55	8.75

On close observation it can be seen that bidder b1 seems to be a potential shill bidder as it exhibits the major characteristics of potential shill bidders as discussed in previous section.b1 has bid with minimum inter bid interval and minimum increase in bid amount. The following tables show the LA and Recoverable_amount of each bidder before and after the auction terminates.

LA_{b1}	
Amount	**auction_id**
5.00	watch_sale
6.35	watch_sale
8.60	watch_sale

Before the auction:

Recoverable_amount$_{b1}$
0.00

After the auction:

Recoverable_amount$_{b1}$
19.95

Let bidders b2 and b3 be honest bidders. The money they lost during the auction in the form of bid fee is transferred to the recoverable_amount which acts as a source of future bonus when a bidder wins an auction.

For bidder b2:

LA$_{b2}$	
Amount	**Auction_id**
6.25	watch_sale
8.50	watch_sale

Before the auction:

Recoverable_amount$_{b2}$
0.00

After the auction:

Recoverable_amount$_{b2}$
14.75

For bidder b3:

LA$_{b3}$	
Amount	**Auction_id**
7.50	watch_sale

Before the auction

Recoverable_amount$_{b3}$
0.00

After the auction

Recoverable_amount$_{b3}$
7.50

For the winning bidder b4:

LA$_{b4}$	
Amount	**Auction_id**
8.75	watch_sale

Before the auction:

Recoverable_amount$_{b4}$
0.00

After the auction:

Recoverable_amount$_{b4}$
0.00

As bidder b4 wins the auction, the total bid fee charged during the auction becomes benefit1 for the winner.

benefit1 = 8.75

b4 is also entitled to an additional benefit from the recoverable_amount but as that is initially also zero, there is no bonus available for b4. This implies that b4 is either a new bidder or has never lost an auction.

benefit2 = 0.00

Amount to be paid by winner b4 = 166.25

The winner get his benefits, the honest bidders get the opportunity to win back the lost amount in further auctions they win and the only bidder at loss is the shill bidder who never aims to win an auction, thus never wins back the money lost. The more a fraudulent bidders get involved in shill bidding, the more loss will he suffer.

For more clarity of the concept of recoverable_amount, let's consider another auction scenario where bidders b2, b5, b6, b7 and b8 bid to win a doll house.

Table 4. 2: An Online Auction Scenario 2

Bidder id	**Price**	**Time**	**Bid_fee**
b2	50	0.10	2.50
b5	52	0.12	2.60
b8	65	0.20	3.25
b6	68	0.22	3.40
b2	80	0.30	4.00
b7	95	0.50	4.75
b5	100	0.52	5.00
b8	120	1.10	6.00
b2	150	1.20	7.50

For bidder b5:

LA_{b5}	
Amount	**Auction_id**
2.60	doll_house
5.00	doll_house

Before the auction:

$Recoverable\_amount_{b5}$
0.00

After the auction:

Recoverable_amount$_{b5}$
7.60

For bidder b6:

LA$_{b6}$	
Amount	**Auction_id**
3.40	doll_house

Before the auction:

Recoverable_amount$_{b6}$
0.00

After the auction:

Recoverable_amount$_{b6}$
3.40

For bidder b7:

LA$_{b7}$	
Amount	**Auction_id**
4.75	doll_house

Before the auction:

Recoverable_amount$_{b7}$
0.00

After the auction:

Recoverable_amount$_{b7}$
4.75

For bidder b8:

LA$_{b8}$	
Amount	**Auction_id**
3.25	doll_house
6.00	doll_house

Before the auction:

Recoverable_amount$_{b8}$
0.00

After the auction:

Recoverable_amount$_{b8}$
9.25

For bidder b2:

LA_{b2}	
Amount	**Auction_id**
2.50	doll_house
4.00	doll_house
7.50	doll_house

Before the auction:

$Recoverable\_amount_{b2}$
14.75

After the auction:

$Recoverable\_amount_{b2}$
0.75

$Recovered\_amount_{b2}$
14.00

On observing the auction scenario 2, it can be concluded that bidders b5 and b6 seem potential bidders. The representation of LA and recoverable_amount for bidders b5, b6, b7 and b8 are the same as explained in the auction scenario 1.

Consider the bidder b2. b2 had lost an auction and already had 14.75 as the recoverable_amount. This is the amount he had lost on account of the bid fee charges in

the previous auction he had lost. Now b2 has won an auction and has earned the extra benefit from the recoverable_amount in addition to the exemption from the bid fee charges for the winning auction.

Benefit1 = 14.00

Benefit2 = 14.00

Amount to be paid by winner b2 = 122.00

Had there been lesser than 14.00 amount in the recoverable_amount$_{b2}$, that whole amount would've become benefit2, i.e., the additional bonus for the winner can be maximum up to the bid fee charges in the winning auction.

The opportunity to win back the amount lost as bid fee encourages the bidders to participate and win more auctions. The shill bidders prevent winning auctions thereby losing the amount incurred as bid charges to them. Shill bidders can never recover the loss and the loss increases as is their participation in the shill bidding fraud.

5. Experiment and Results

In this chapter, we discuss the details of testing our methodology on real auction data. Auction data is hard to obtain manually. Therefore, we created an online auction data collector for one of the leading auction websites –Ebay. In addition to the data collection, this module also calculates the major shill bidding characteristics enlisted in previous chapter and depicts the comparative behaviour of each bidder with respect to each characteristic in the form of graphs. This will help identify the potential bidders in an auction. The auction data then collected becomes the dataset for testing our proposed methodology.

5.1 Tools Used

The following tools or softwares were used for the development of the project:-

1. **Lamp:** It is a web service stack with Linux, Apache server, MySQL rdbms and PHP programming language being the major components. It is a free and open source software stack.

   Linux: It is free open source operating system belonging the UNIX family. It is written in C and Assembly. Initially released in 1991, its latest release version is 3.18

   Apache: It is the most popular cross platform HTTP server written in C/C++, XML. It accounts for 52.27% of the whole website activities currently. Initially released in 1995, the current stable release is 2.4.10. Apache2 was installed to visualize the project on localhost.

   MySQL: Developed by Oracle Corporation, it is the second most commonly used relational database management system. Written in C/C++, it is open source and works on various different platforms like Linux, Windows, and MacOS etc. With an initial release in 1995, the current stable release is 5.6.22

   PHP: Initially considered for web development as server side scripting language, nowadays it is also used as a general purpose programming language. Initially used in 1995, 5.6.3 is its latest stable release. Written in C, it is cross-platform. Php5 was installed and used to implement the project.

2. **MySQL Workbench:** MySQL workbench is a visual tool that helps design, develop and administer database. It is supported by multiple platforms. It provides a user friendly GUI for creating, managing and executing queries on database.

   6.2 CE version was installed.

   MySQL Workbench was used to create fraud schema and the tables, i.e., items, seller, interactions, bids and bidder. All the constraints were applied graphically.

3. **Sublime Text:** It is a text editor and is cross platform. It is also used to edit source code. Multiple files and multiple panels can be displayed at a time. Auto completion and syntax highlighting are its useful features.

   Sublime Text3 was used to work on the yii framework and its various files.

4. **Yii framework:** Yii stands for 'Yes It Is'. It is a PHP framework with relatively high performance and security as compared to the other frameworks. Developed by QiangXue, it was first released in 2008.Written in PHP, it is cross platform. The characteristic features of yii framework are

   - Open source
   - Object oriented
   - Component based
   - MVC

5.2 Phase 1 : Data Collection

The module runs as an extended functionality of Ebay. The home page contains a form that requires auction id and auction url to be entered by the user for the analysis of the corresponding auction. The user can open any auction on Ebay for analysis. The auction id is obtained from the url itself.

For eg: Considering an auction on Ebay.

Url to be entered:-

http://www.Ebay.com/itm/EXCLUSIVE-AUCTION-Premium-Quality-UNLOCKED-Samsung-Galaxy-Note-3-Black-32gb-/221630302354?pt=Cell_Phones&hash=item339a320892

Id:- 221630302354

The auction data required includes seller data, bidder data, items put up for auction, interaction between seller and bidder and the currently ongoing bids. The required seller data includes its id, rating, feedback and date since he is an active member of Ebay. The needed bidder data was his id, rating, feedback, total bids, total items bid on, bid retractions and bid retractions in the last 6 months. Bid data includes the bidder id, amount and time of the bid. The data related to item includes the item id, the seller of the item and the number of bids on the item at that instant. The whole required data is spread through different web pages of the website.

So phase1 collects the data from different web pages related to an auction and stores them in a relational database. It employs the following steps:

1. Accepts the url of auction under inspection
2. Verifies the url
3. Converts the source code of web pages related to the auction into objects
4. 'getAuctionInsight' uses the objects of the source web pages to extract the required information.
5. The created relational database is populated with the required auction data.

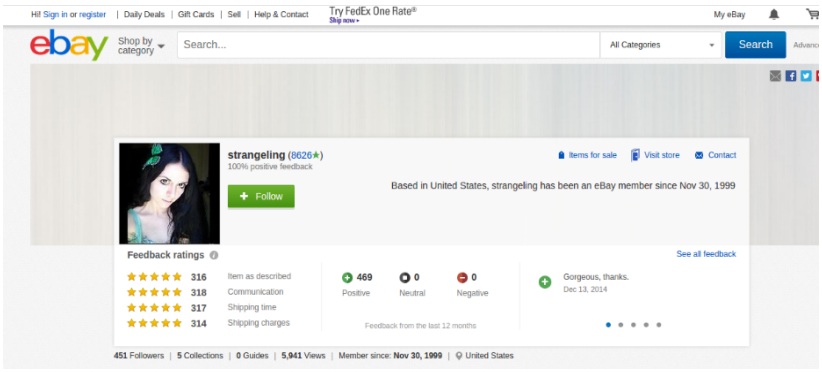


Figure 5.1: Screenshot of bidder data to be collected

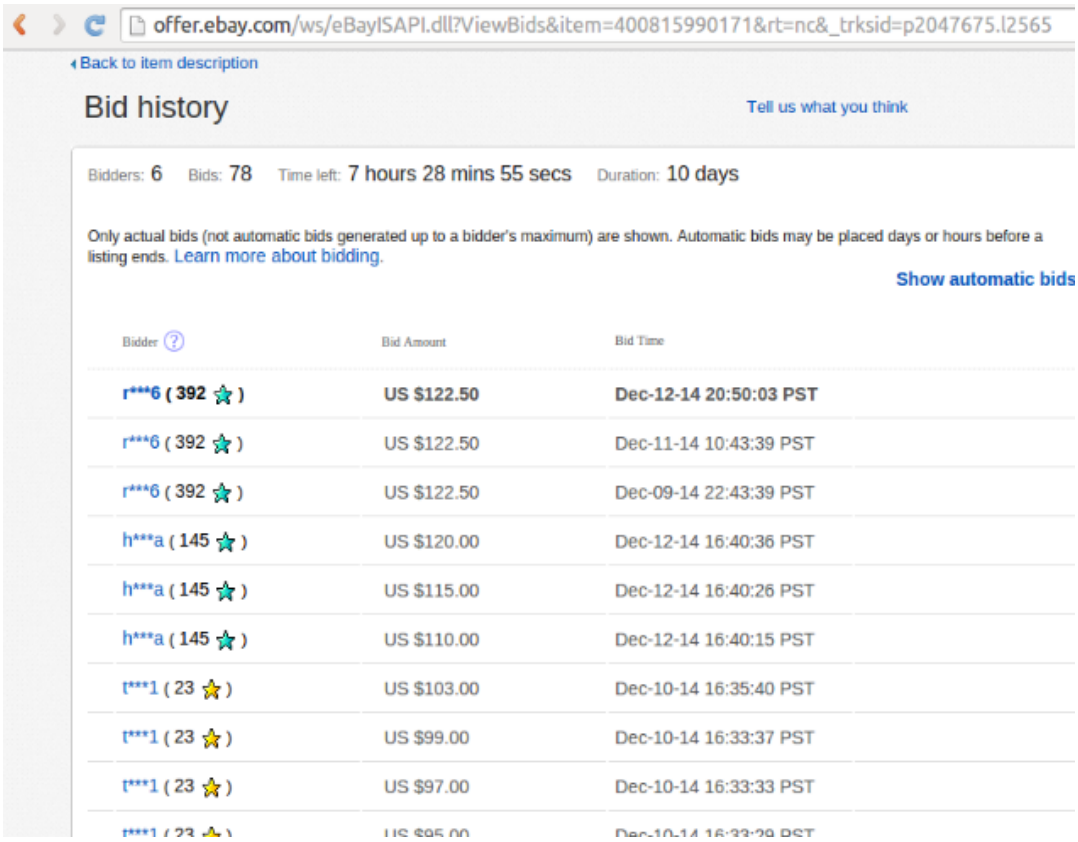


Figure 5.2: Screenshot of bidding information to be collected

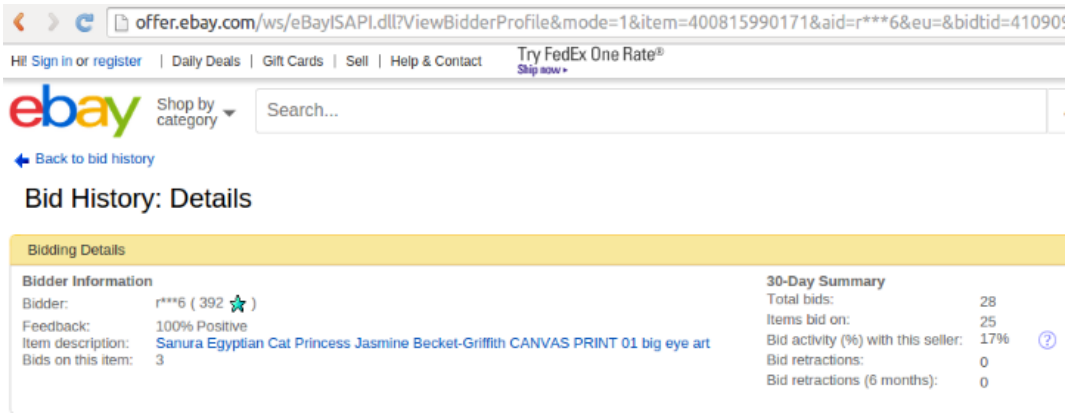


Figure 5.3: Screenshot of bidder information to be collected

Figure 5.4: Screenshot of data collection phase

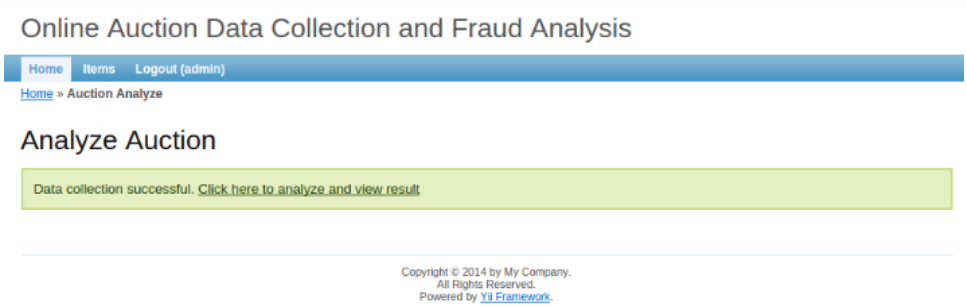


Figure 5.5: Screenshot of completion of data collection phase

5.3 Phase 2: Shill Bid Factor Analysis

The main pointers for shill bidding are as following:

1. Bid time: A shill bidder will try to bid as soon as possible after a legitimate bid so that significant time is left for the legitimate bidders to further bid on the item. Based on this, average inter-bid time interval is calculated for each bidder based on the bidding data collected for an auction. A small average inter bid time is one of the chosen factors for the shill bid fraud.
2. Bid Amount: The aim of the shill bidder is not to win the auction but to hike the eventual selling price of the item. Consistent winning of one's own auction may lead to loss. Therefore, the shill bidder increased the ongoing bid amount by a very small increment to encourage the legitimate bidders to bid further. Average inter-bid amount value is calculated for each bidder which acts as another shill bid pointer.

3. Bidding Nature: A shill bidder tends to bid more frequently at the beginning of the auction than at the end to avoid winning the auction.
4. Interaction with the seller: A high interaction with the seller solely cannot be a shill bid pointer. But if the seller has a high feedback rating in a less time duration which can be obtained from the "member since" attribute of the seller, then it can be considered a suspicious seller. A high interaction bidder account associated with such a seller can too be a shill bidder.

All the above factors are calculated for each bidder in the auction and the result is displayed to the user. The following screenshots depict the analysis phase.

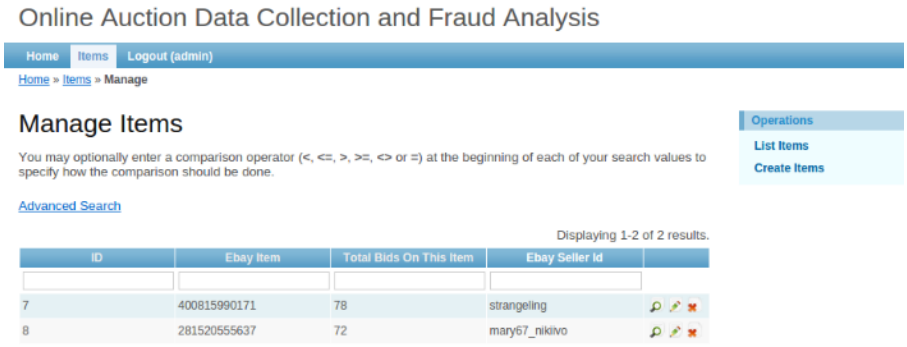


Figure 5.6: Screenshot of data analysis phase

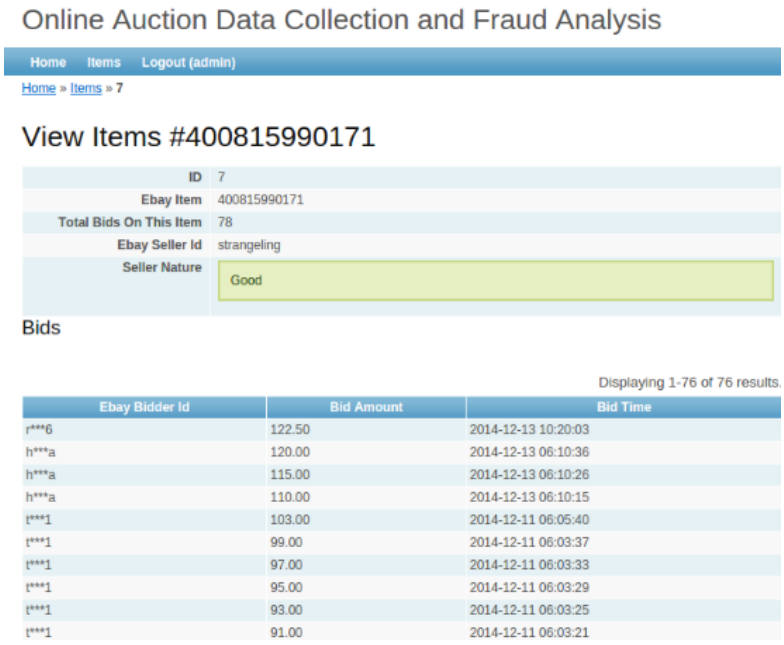


Figure 5.7: Screenshot of complete dataset of an auction

Bidder Behaviour

Displaying 1-6 of 6 results.

Ebay Bidder Id	Average Bid Amount	Average Bid Time (in Seconds)	Bid Frequency	Bidder Nature
r***6	2.50	14967.00	1.32 %	Low
h***a	5.67	57698.67	3.95 %	Low
t***1	2.09	295.74	89.47 %	High
s***h	8.32	248243.50	2.63 %	Low
t***n	9.99	89465.00	1.32 %	Low
l***r	0.00	0.00	1.32 %	Low

Figure 5.8: Screenshot of calculated shill factors for each bidder

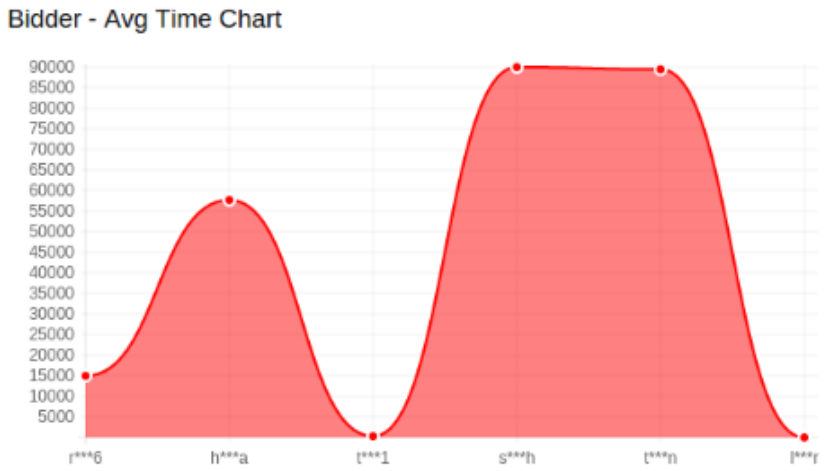


Figure 5.9: Screenshot of bidder versus average inter bid interval graph

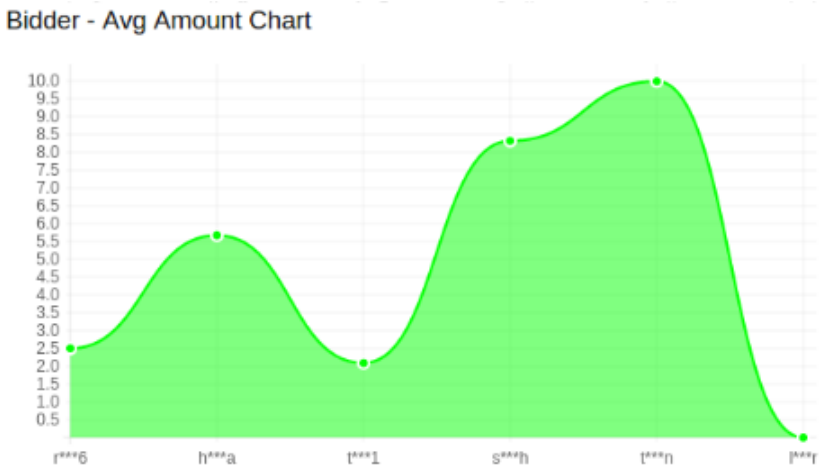


Figure 5.10: Screenshot of bidder versus average bid increment amount graph

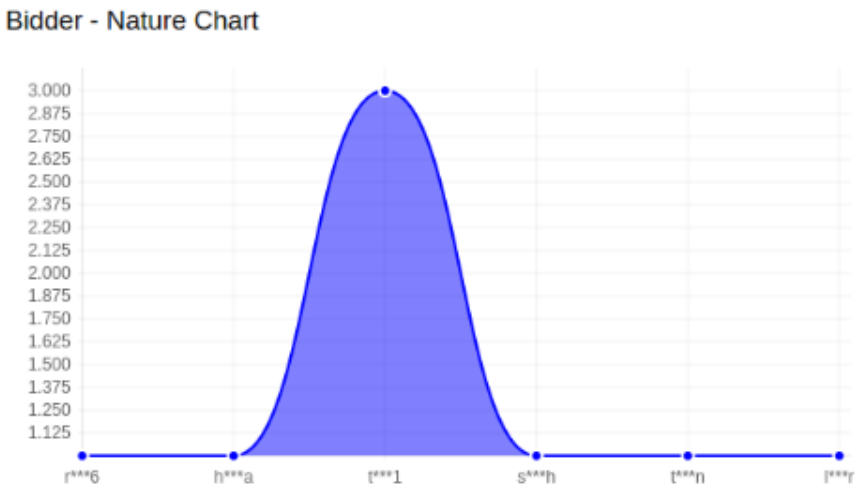


Figure 5.11: Screenshot of bidder versus nature graph

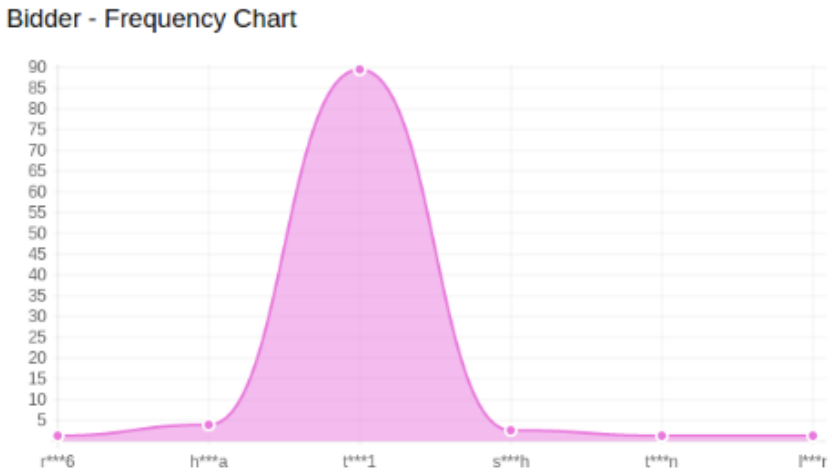


Figure 5.12: Screenshot of bidder versus bid frequency graph

On observing the calculated shill fraud characteristics, it is evident that the user with the bidder id t***1 is a potential shill bidder as it has minimum average inter-bid interval, minimum inter-bid amount increment, high bid frequency in the initial phase of the auction and high interaction with the seller.

5.4 Phase 3: Testing of Variable Bid Fee Methodology

As per our proposed methodology an aggressive shill bidder ought to go through a great financial loss. In Phase 2, t***1 seemed to be involved in shill bidding. From the data collected in Phase 1, we calculate the bid fee for each bid made by every bidder who participated in this auction.

Table 5.1: Bidding Dataset

Ebay Bidder Id	Bid Amount	Bid Fee
h***a	120	6
h***a	115	5.75
h***a	110	5.5
l***r	50	2.5
r***6	122.5	6.125
s***h	76.64	3.832
s***h	70.43	3.5215
t***1	103	5.15
t***1	99	4.95
t***1	97	4.85
t***1	95	4.75
t***1	93	4.65
t***1	91	4.55
t***1	89	4.45
t***1	87	4.35
t***1	85	4.25
t***1	83	4.15
t***1	81	4.05
t***1	79	3.95
t***1	60.01	3.0005
t***1	58.01	2.9005
t***1	56.01	2.8005
t***1	54.01	2.7005
t***1	52.01	2.6005
t***1	50.01	2.5005
t***1	48.01	2.4005
t***1	46.01	2.3005

t***1	44.01	2.2005
t***1	42.01	2.1005
t***1	40.01	2.0005
t***1	38.01	1.9005
t***1	36.01	1.8005
t***1	34.01	1.7005
t***1	32.01	1.6005
t***1	30.01	1.5005
t***1	28.01	1.4005
t***1	26.01	1.3005
t***1	24.51	1.2255
t***1	23.51	1.1755
t***1	22.51	1.1255
t***1	21.51	1.0755
t***1	20.51	1.0255
t***1	19.51	0.9755
t***1	18.51	0.9255
t***1	17.51	0.8755
t***1	16.51	0.8255
t***1	15.51	0.7755
t***1	14.51	0.7255
t***1	13.51	0.6755
t***1	12.51	0.6255
t***1	11.51	0.5755
t***1	10.51	0.5255
t***1	9.51	0.4755
t***1	8.51	0.4255
t***1	7.51	0.3755
t***1	6.51	0.3255
t***1	5.51	0.2755

t***1	4.76	0.238
t***1	4.26	0.213
t***1	3.76	0.188
t***1	3.26	0.163
t***1	2.76	0.138
t***1	2.26	0.113
t***1	1.76	0.088
t***1	1.26	0.063
t***1	0.96	0.048
t***1	0.86	0.043
t***1	0.76	0.038
t***1	0.66	0.033
t***1	0.56	0.028
t***1	0.46	0.023
t***1	0.36	0.018
t***1	0.26	0.013
t***1	0.16	0.008
t***1	0.06	0.003
t***n	60	3

Table 5.2: Bid Fee of all Bidders of an Auction

Ebay Bidder Id	**Total Bid Fee**
h***a	17.25
l***r	2.5
r***6	6.125
s***h	7.3535
t***1	109.278
t***n	3

The table 5.2 shows that the bid fee charges are the highest for a potential shill bidder. The maximum loss is undergone by a shill bidder which was the intended objective of our proposed methodology.

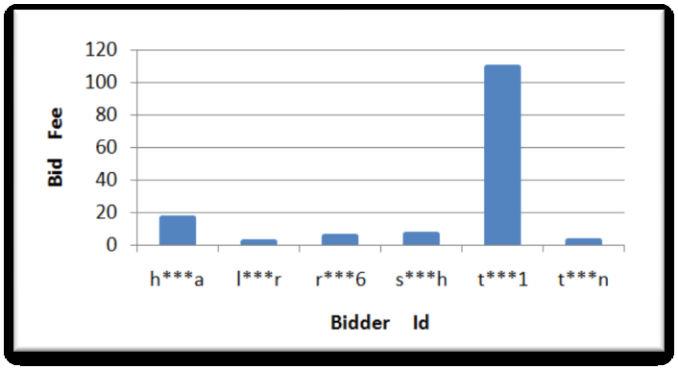


Figure 5.13: Bid Fee Versus Bidder graph

A graph between bid fee and bidders clearly depicts that the highest bid fee is charged to a potential bidder as per our 'Variable Bid Fee Methodology'.

When finances are at stake, the users refrain from getting involved in a fraud. Thus Variable Bid Fee methodology acts as a preventive measure for shill bidders in online auctions. It caters to both single and collusive shill bidders alike.

6 Conclusion and Future Scope

6.1 Conclusion

This research work proposes a variable bidding fee price methodology that targets shill bidding, the most dominant fraud strategy, in an e-auction market. It is a shill bidding prevention technique as collusive groups and individual shill bidder accounts will eventually be at a loss thereby dissuading sellers to continue with this fraud technique. We cater for the honest ferocious bidder behaviour by providing them an opportunity to recover bid fee charges of their prior auctions on winning. An additional benefit to the winner is the exemption of the bidding charges.

The major benefit of our proposed methodology over other fraud detection and prevention techniques is that our methodology works on the current active auctions of the users. In contrast to the other proposed collusive fraud detection techniques, our Variable Fee Bid methodology requires no bid history, data mining or complicated fraud parameter calculations thereby decreasing the resource requirement for large storage and complex calculations. Due to these major differences our methodology is a faster and easier preventive measure of shill bidding. A consistent shill bidder or collusive shill bidding groups will eventually be at a loss. But our proposed methodology might dissuade one time bidders from bidding because they won't get a chance to recover the bid fee charges once they have lost the only auction they have participated in.

Table 6.1: Comparison

	Requirements of a fraud detection/ prevention methodology		
	Bid History	***Data Mining***	***Fraud Calculation Parameter***
Auction Fraud Rank Algorithm	Required (Transactional dates of accounts involved)	Required (2-core clustering algorithm)	Auction Fraud Rank

	Requirements of a fraud detection/ prevention methodology		
	Bid History	***Data Mining***	***Fraud Calculation Parameter***
Shill Bidder Detection for Online Auctions	Required (Bidder attributes)	Required (Outlier detection - SVM)	Rules derived from decision tree
Detecting Collusive Shill Bidding	Required (Bid and previous auction statistics)	Not Required	Shill Score
Variable Bid Fee Methodology	Not Required (Based on current bids in an auction)	Not Required	No complex calculations required. (Risk of financial loss reduces fraud)

6.2 Future Scope

We charged 5% of the bidding price as the bidding fee. Different pricing schemes can be explored to improve the effectiveness of our shill bidding prevention mechanism.

The scope of this research work can be further extended to study the effect and efficiency of Variable Bid Fee on the automated bidding systems that are being incorporated in most of the auction sites these days.

References

[1] Wikipedia, 'Auction', 2015. [Online]. Available: http://en.wikipedia.org/wiki/Auction. [Accessed: 23- Jun- 2015]

[2] Statista, 'Ebay: annual net revenue 2004-2014 | Statistic', 2015. [Online]. Available: http://www.statista.com/statistics/266195/Ebays-annual-net-revenue/. [Accessed: 23- Jun- 2015]

[3] Internet Crime Complaint Center (2014), 2013 Internet crime report.

[4] [Www8.hp.com, 'Cyber Security, Internet and Computer Security, Ponemon Report | HP® India', 2015. [Online]. Available: http://www8.hp.com/in/en/software-solutions/ponemon-cyber-security-report/. [Accessed: 23- Jun- 2015]

[5] Zhang, Bin, Yi Zhou, and Christos Faloutsos. "Toward a comprehensive model in internet auction fraud detection." Hawaii International Conference on System Sciences, Proceedings of the 41st Annual. IEEE, 2008

[6] Jøsang, Audun, Roslan Ismail, and Colin Boyd. "A survey of trust and reputation systems for online service provision." *Decision support systems*43.2 (2007): 618-644.

[7] Fan, Ming, Yong Tan, and Andrew B. Whinston. "Evaluation and design of online cooperative feedback mechanisms for reputation management."*Knowledge and Data Engineering, IEEE Transactions on* 17.2 (2005): 244-254.

[8] Lin, Iuon-Chang, et al. "A fair reputation system for use in online auctions."*Journal of Business Research* 68.4 (2015): 878-882.

[9] Morzy, Mikołaj. "New algorithms for mining the reputation of participants of online auctions." [Algorithmica 52.1 (2008): 95-112].

[10] Christodorescu, M., et al. "An auctioning reputation system based on anomaly detection." *Proceedings of the 12th ACM conference on Computer and communications security. Alexandria, VA: The Association of Computing Machinery*. 2005.

[11] Lin, Shi-Jen, Yi-Ying Jheng, and Cheng-Hsien Yu. "Combining ranking conceptand social network analysis to detect collusive groups in online auctions."[ExpertSystems with Applications 39.10 (2012): 9079-9086].

[12] Chang, Wen-Hsi, and Jau-Shien Chang. "An effective early fraud detectionmethod for online auctions." [Electronic Commerce Research and Applications11.4 (2012): 346-360].

[13] Yoshida, Tsuyoshi, and HayatoOhwada. "Shill bidder detection for online auctions." PRICAI 2010: Trends in Artificial Intelligence. [Springer Berlin Heidelberg, 2010. 351-358].

[14] Ku, Yungchang, Yuchi Chen, and Chaochang Chiu. "A proposed data mining approach for internet auction fraud detection." Intelligence and Security Informatics. [Springer Berlin Heidelberg, 2007. 238-243].

[15] Chau, DuenHorng, ShashankPandit, and Christos Faloutsos. "Detecting fraudulent personalities in networks of online auctioneers." Knowledge Discovery in Databases: PKDD 2006. [Springer Berlin Heidelberg, 2006. 103-114]

[16] Tsang, Sidney, et al. "SPAN: Finding collaborative frauds in online auctions." [Knowledge-Based Systems 71 (2014): 389-408].

[17] Pandit, Shashank, et al. "Netprobe: a fast and scalable system for fraud detection in online auction networks." *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007.

[18] Trevathan, Jarrod, and Wayne Read. "Detecting collusive shill bidding."Information Technology, 2007. ITNG'07. Fourth International Conference on. IEEE, 2007.

[19] Yu, Cheng-Hsien, and Shi-Jen Lin. "Fuzzy rule optimization for online auction frauds detection based on genetic algorithm." [Electronic Commerce Research13.2 (2013): 169-182].

[20] Ghasemi, Hamid-Reza, and Gholam-Reza Mohammadi. "Architecture-oriented approach for detecting fraud in the online auction.″ e-Commerce in Developing Countries: With Focus on e-Trust (ECDC), 2014 8th International Conference on. IEEE, 2014.

[21] Peng, Yanlin, Linfeng Zhang, and Yong Guan. "Detecting fraud in internet auction systems." *Advances in Digital Forensics V*. Springer Berlin Heidelberg, 2009. 187-198.

[22] Zhu, Kai, Yong Guan, and Lei Ying. "Detecting hidden communities in online auction networks." *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*. IEEE, 2012.

[23] Almendra, Vinicius, and Bianca Roman. "Using exploratory data analysis for fraud elicitation through supervised learning." *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2011 13th International Symposium on*. IEEE, 2011.

[24] Chang, Wen-Hsi, and Jau-Shien Chang. "Using clustering techniques to analyze fraudulent behaviour changes in online auctions." *Networking and Information Technology (ICNIT), 2010 International Conference on*. IEEE, 2010.

[25] Chang, Jau-Shien, and Wen-Hsi Chang. "An early fraud detection mechanism for online auctions based on phased modeling." *Pervasive Computing (JCPC), 2009 Joint Conferences on*. IEEE, 2009.

[26] Mundra, Ankit, and Nitin Rakesh. "Online Hybrid model for online fraud prevention and detection." *Intelligent Computing, Networking, and Informatics*. Springer India, 2014. 805-815.

[27] Mundra, Anand, NitinRakesh, and S. P. Ghrera. "Empirical study of online hybrid model for internet frauds prevention and detection." *Human Computer Interactions (ICHCI), 2013 International Conference on*. IEEE, 2013.

List of Publications

Dhanmeet Kaur and Deepak Garg. Variable Bid Fee: An Online Auction Shill Bidding Prevention Methodology. In *2015 IEEE International Advance Computing Conference* (in proceeding). IEEE, Bangalore, June 12-13, 2015.

YouTube Link

https://www.youtube.com/watch?v=MiFMQkTwX_s&feature=youtu.be

Reflective Diary

January 2015

My journey of research work began with an attempt to ascertain my destination first. Without knowing what you want to do, where you want to go, it's as if getting stuck in a dark room with no hint of light. Thankfully I've always had my guide to help me through such situations throughout the past one year. Initially, I had no clue about the topic I wanted to work on. I talked to my guide, Dr. Deepak Garg, and he asked me to list out the topics I was interested in. I spent a lot of time surfing on internet, watching videos, going through the latest developments in various fields, trying to know what topics piqued my interest. One can never work on something that doesn't interest you. The interest and curiosity itself becomes the driving force of research. I ended up with Online Auctions, Fuzzy Logic, Genetic Algorithms and Cyber Law. I discussed it with my guide and we finally ended up short listing 'Online Auctions' as our research area.

Outcome: I had my area of research finalized viz. 'Online Auctions' but the next step was – what exactly do I want to work on where online auctions are concerned.

February 2015

This whole month I spent studying the auctions- the major terms involved in auctions such as valuation, bids, players, bidders and their strategies. I studied the various types of auctions. While studying this I came across lectures by Tim Roughgarden on YouTube. He has discussed the game theory aspect of auction design. He considers that all bidders are strategic players and the auctions should be so designed to give fair results irrespective of the strategies employed by bidders. I studied the concepts of Nash equilibrium, First Price Auction, Vickery Auction, Sealed Bid Auction and Sponsored Search Auction. As I kept studying this, the topic got more complicated and involved a lot of time and mathematical background. There was a time I started questioning myself if I had chosen the right topic. I discussed it with my guide and he advised me to study

the simpler and basic aspects of online auctions first. He asked me to download and go through the research papers on online auctions.

Outcome: By the end of this month, I had the basic knowledge of auctions – the terminology involved, the process and its scope. I still needed to streamline my research area.

March 2015

As advised by my guide, I spent this month reading various research papers of Springer, IEEE and Elsevier publications. While going through the papers, there was one aspect of online auctions that caught my attention i.e. the types of frauds and the proposed methodologies of dealing with them. When I took the print outs of the papers, I attached an extra sheet at the end of each paper where I wrote the summary of the paper or its major significant points that really helped me later on while writing 'Literature Survey' part of my paper and thesis. I also identified and jotted down the pointers for further research in each paper I read which eventually helped me in shaping up the methodology proposed in this thesis work.

Outcome: The topic of research was finalized and the major literature survey part was done.

April 2015

In this month I focused on fraud detection and prevention mechanisms. I searched videos on YouTube on fraud identification in online auctions. I observed and studied various online auctions on Ebay. To continue my work in this area, I needed a dataset of auctions – the bid history, bidder and seller characteristics. So I started working on creating a crawler for Ebay for data collection to collect all the visible data in database for further analysis. My initial version of the data collector failed after collecting 20 to 30 records. The Ebay website blocked it. I spent the next few day searching for better ways to collect

the data and I discussed the issue with my friend who is working in a company in Delhi. He helped me extract the data by accessing the data through objects. I had the required dataset. While collecting the data I realized that all the prior approaches involved a lot more data that the auction websites are not always willing to share, so I wanted to come up with a solution to deal with auction fraud based on the data available to all users.

Outcome: Created an Ebay auction data collector.

May 2015

I went through all the summaries of papers I had read and the points of potential research that I had identified earlier. While going through them again I realized what area had not been worked on. Many papers suggested that negligible bid fee is the reason of shill bidding. I focused my attention on shill bidding and bid fee. I came up with an idea of charging bid fee to prevent shill bidding. I discussed the idea with my guide and friends. It was through the discussion only that I could realize certain flaws in my initial proposal of the methodology which I further refined with the helped of my guide and thus the final version came up. I was working on my paper side by side. With this, my research paper was finalized and approved by my guide. My guide has been a great support through all the phases of my work.

Outcome: Algorithmic form of the proposed methodology was designed.

June 2015

The last month was spent writing thesis. With the help of my guide I had first decided the layout of my thesis, made appropriate chapters and then started working on them one by one. Thesis writing is a time consuming task and stretches over days. Care has to be taken not to copy any data or images. I created my flowcharts online at www.draw.io and imported them as jpeg. Most of my diagrams I created on MSWord and Paint. It takes a

whole month to write, edit and finalize the thesis report. I spent around two days on designing and creating a poster of my thesis.

Outcome: Thesis and poster were finalized.