Fraud Detection in Electronic Auction

Duen Horng Chau[1], Christos Faloutsos[2]

[1] Human-Computer Interaction Institute, School of Computer Science,
Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh PA 15213, USA
dchau@cs.cmu.edu

[2] Department of Computer Science, School of Computer Science,
Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh PA 15213, USA
christos@cs.cmu.edu

Abstract. Auction frauds plague electronic auction websites. Unfortunately, no literature has tried to formulate and solve the problem. This paper aims to tackle it by suggesting a novel method to detect auction fraudsters, which involves determining and extracting characteristic features from exposed fraudsters, through analyzing the fraudsters' transaction history which exists as a graph. We then use the features for detecting other potential fraudsters. Choosing the best features is a challenging and non-trivial task; however, with the features that we have currently selected, our method has already achieved a precision of 82% and a recall of 83% during an evaluation on some real test data from eBay. To demonstrate how our method can be used in real-world, we have developed a working Java prototype system which allows users to query the legitimacy of eBay users using our method.

1 Introduction

Electronic auction websites have created a huge virtual marketplace, where the world's population can easily buy and sell virtually any items. In an electronic auction, a person can post items for other people to bid. The person placing the highest bid will be the winner of the auction, and will contact the seller and pay for the item. The seller will then ship the item to the buyer.

Electronic auction is a thriving business – eBay, the world's largest electronic auction website, has reached a cumulative number of 147.1 million registered users at the end of Q1 2005, representing a 40% increase over that reported at the end of Q1 2004 [10].

Every day, there are millions of dollars of sales involved in electronic auction transactions. The huge sum of money has unfortunately attracted perpetrators' attention, and a large number of them commit auction frauds which are by far the most serious problems that eBay faces. Indeed, the number of reported auction frauds has been increasing over the past few years, and the trend shows that the problem is getting worse. In 2004, the Internet Crime Complaint Center (IC3) referred 103, 959 complaints, 71.2% of which were auction frauds. 87% of the victims reported monetary loss with a median of $200. [12] Most auction fraud involves sellers selling non-existent items. They do this intentionally: they typically do not

have the items, but they post them to the auction, receive payment from buyers, and yet never deliver the items to the buyers.

Auction fraud is a problem that has been getting increasingly serious. We believe that, as computer scientists, we could do something to help solve the problem. This belief has motivated our work.

2 Survey

To the best of the authors' knowledge, this is the first work that uses a systematic approach to analyze and detect electronic auction frauds.

There have been a lot of suggested common-sense approaches on websites [11] and news articles [13] that teach people how to avoid auction frauds. However, their approaches often involve asking people to invest substantial amount of time and to constantly maintain high level of vigilance. This is what average people cannot afford to do.

Little research has been done in suggesting systematic approaches in detecting or preventing auction fraud. Some researchers [1] have categorized auction fraud into different types, but they have not suggested any formalized methods in dealing with them. They have, however, suggested that an effective approach to fight auction fraud is to allow auction communities, governments and auction institutions to join forces. However, as they have pointed out, their approach can be costly both in monetary and managerial means.

There has been a lot of interest in some of the other research areas related to auction fraud detection, such as reputation systems of electronic auctions [4] [7] [8], graph mining [9], trust propagation and authority propagation [3] [6]. Also remotely related is the work on aggregating values from some other set of values in a graph [5].

3 Proposed Approach

3.1 Problem Definition

Here, we examine the problem of detecting auction fraudsters. Specifically, we define the problem as:

Given:

- The information of some electronic auction users: their profiles and their transaction history
- Some exposed fraudsters

We want to find out:

- Who else are also fraudsters

The profiles and transaction history of eBay users are readily available from the eBay website (Appendix), while the knowledge about the exposed fraudsters can be acquired from news articles, user forum on eBay, and by noting the large number of negative feedback given by other buyers or sellers saying the fraudsters have never delivered the items. Thus,

we define the auction fraud problem as given these two pieces of information, how do we identify other potential fraudsters before they carry out frauds.

It is possible for a person to create multiple identities on eBay. However, in order to improve the clarity of our discussion in this paper, we use the terms *eBay user* and *fraudster* to refer to individual identities on eBay rather than individual persons.

3.2 Intuition

We have observed that fraudsters show typical characteristic behaviors during their lifetime on auction websites. We have also observed that before they carry out fraudulent transactions, they typically exist as seemingly-legitimate users in order to establish good-enough reputation for their up-coming fraudulent acts. Yet, their reputation building procedure is different from that of legitimate users. We believe we can identify the fraudsters' reputation build-up process by inspecting features derivable from their profiles and transaction history.

Reputation is an important measure of credibility in electronic auction websites. On eBay, a person's reputation is represented by the number of unique positive ratings given by his/her dealers, minus the number of negative ratings (Appendix). The resulting number is the user's *feedback score*.

Fraudsters usually aim to gain as much one-time profit as quickly as possible. Therefore they usually "sell" moderate value or expensive items in categories such as consumer electronics. Yet unlike legitimate users, they do not intend to deliver those items after receiving payment from the buyers.

However, before the fraudsters can carry out the fraudulent acts, they, like other people at eBay, need to establish certain level of reputation. This is because with few positive ratings, they will look too suspicious for most auction users.

We hypothesize that auction fraudsters need good reputation so as to trick people into believing that they are legitimate. We believe that fraudsters fabricate their reputation through methods that exhibit certain patterns which are different from the legitimate ones and that if we discover the fabrication process, we can essentially identify the frauds before they actually take place.

3.3 Observed Patterns

From the known and publicized frauds on eBay, we learned of several common ways that fraudsters fabricate reputation. Those patterns are typically not found among legitimate users, but are well justified from the perspective of the fraudsters:

1. Selling or buying numerous cheap items from users with good reputation.
   Justification: fraudsters want to gain many ratings at small cost
2. Selling or buying moderate value or expensive items from collaborators. The collaborators may not show fraudulent acts; they may generally act like legitimate users.
   Justification: fraudsters want to gain few but "strong" ratings virtually without incurring any cost
3. Combination of 1 and 2.
   Justification: this can create a very legitimate-looking illusion.

4. Reputation fabrication process usually occurs in short period of time.
   Justification: fraudsters are short-lived; they do not want to remain on eBay for too long to increase their risk of being discovered.

3.4 Determining Characteristic Features from Observed Patterns

From the observed patterns, we have reasoned that a number of features could effectively capture the patterns – and therefore distinguish fraudsters from legitimate users. The features are:

1. Median, sum, mean, or standard deviation of prices of items bought or sold in certain time period
2. Number of items bought or sold in certain time period
3. Ratio of the number of items bought or sold to that of all transactions in certain time period

From the set of possible features above, we have selected the following 17 features. We will evaluate the effectiveness of these features in detecting fraud in section 4.

- Median prices of items sold within the first 15, first 30, last 30, and last 15 days
- Median prices of items bought within the first 15, first 30, last 30, and last 15 days
- Standard deviation of the prices of items sold within the first 15, first 30, last 30, and last 15 days
- Standard deviation of the prices of items bought within the first 15, first 30, last 30, and last 15 days
- Ratio of the number of items bought to that of all transactions

3.5 Decision Tree

The values of the chosen features can be extracted from the profiles and transaction history of eBay users which are readily available from the eBay website.

We first download the relevant information from eBay and store it in two tables, the *profiles* and *transactions* tables. Then from the two tables, we extract the values of the features and save them to a third table, the *features* table.

Table 1. Sample *profiles* table

User ID	Feedback score	Still registered?	Registration date	Location
theplastics21	10	FALSE	Fri Feb 11 00:00:00 EST 2005	United States
designerclothing4less404	3	TRUE	Wed Jan 05 00:00:00 EST 2005	United States
joahandbruce	12	FALSE	Thu Dec 02 00:00:00 EST 2004	United States

Table 2. Sample *transactions* table

User ID	Rating	Sell?	Dealer ID	Dealer feedback score	Feedback date	Item ID	Item $
theplastics21	1	TRUE	420stephen	21	Tue Mar 08 10:17:00 EST 2005	8171729100	7
theplastics21	-1	TRUE	rkayakr	6	Mon Mar 07 18:11:00 EST 2005	5753342945	54
designerclothing4less404	1	FALSE	bargainphone	44508	Wed Mar 02 14:57:00 EST 2005	5755903231	1.5
designerclothing4less404	1	TRUE	sunrob123	2	Sat Feb 26 16:06:00 EST 2005	3960815216	199
designerclothing4less404	1	FALSE	yeung49	467	Thu Feb 10 00:44:00 EST 2005	3953380379	19.99

Table 3. Sample *features* table

Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	User ID	Label
1.29	1.29	1.29	0	0.6	0.6	0.6	theplastics21	innocent
2.95	0.99	11.69	8.24	0.19	9.52	7.22	designerclothing4less404	fraudster
3.68	6.59	67.84	0	3.85	26.64	0	joahandbruce	innocent

The set of features are then passed into the C5.0 classification system as training dataset. C5.0 uses the C4.5 classification algorithm, to produce a decision tree based on the data. Fig. **1** shows a sample decision tree learned on some training data utilizing 16 features. With the decision tree constructed, we can use it to classify future test data.

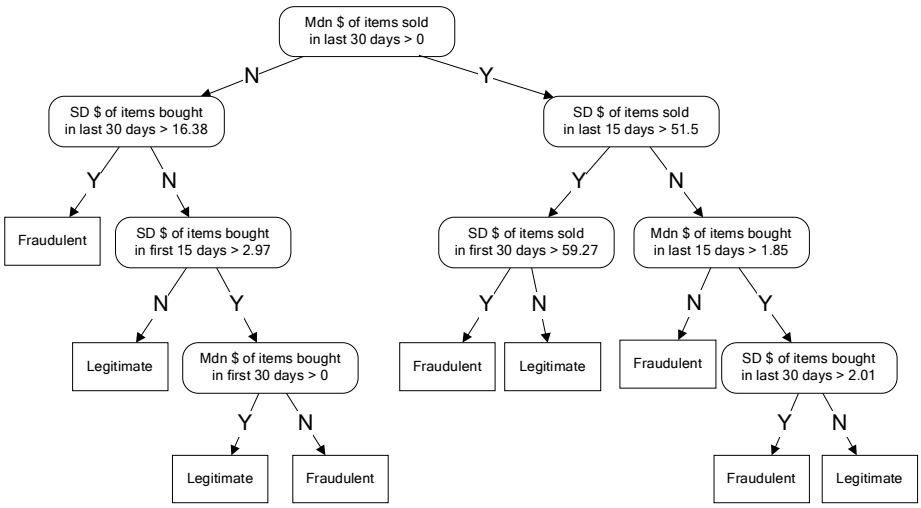


Fig. 1. Sample decision tree utilizing 16 features

4 Experiments

4.1 Tasks and Selected Features

The goal of this experiment is to evaluate the effectiveness of our proposed method in detecting fraudsters. We have drawn the profiles and transaction information of 43 fraudsters and 72 legitimate users from eBay (www.ebay.com). We created three setups for the experiment. Setup A uses 8 features; setup B uses 16 features and setup C uses 17 features.

Table 4. Features used in setups A, B and C

A (8)	B (16)	C (17)	Features
✓	✓	✓	Median prices of items sold within the first 15, first 30, last 30, and last 15 days
✓	✓	✓	Median prices of items bought within the first 15, first 30, last 30, and last 15 days
	✓	✓	Standard deviation of the prices of items sold within the first 15, first 30, last 30, and last 15 days
	✓	✓	Standard deviation of the prices of items bought within the first 15, first 30, last 30, and last 15 days
		✓	Ratio of the number of items bought to that of all transactions

We have considered many features that might differentiate fraudsters from legitimate users, and among all of them we decided that the trends (medians), and fluctuation (standard deviations) of prices of past traded items over time (first 15 days, first 30 days, etc) are some of the important ones that we should include, as they have direct relevance to fraudsters' investment costs and profits.

4.2 Results and Discussion

We evaluated the accuracy of the decision tree generated in the three setups by running 2-fold cross-validations which divides the training data into 2 blocks of the same size and class distribution. For each block, decision trees are constructed from the other block and tested against the current block. For every setup, we ran the experiment 20 times, each with different folds, and averaged the error rates. Table 5, Table 6, and Table 7 show the confusion matrixes of the three setups, while Table 8 shows their precisions and recalls, and Fig. 2 shows their corresponding ROC graph. As the ROC graph is insensitive to changes in the class distribution of the training/test data [2], it provides good depiction and comparison for the relative performance of three experiment setups.

Table 5. Confusion matrix A (8)

Actual \ Predicted	Legit	Fraud
Legit	64.55	7.45
Fraud	10.6	32.4

Table 6. Confusion matrix B (16)

Actual \ Predicted	Legit	Fraud
Legit	64.55	7.45
Fraud	8.5	34.5

Table 7. Confusion matrix C (17)

Actual \ Predicted	Legit	Fraud
Legit	64.4	7.6
Fraud	7.2	35.8

Table 8. Precision and recall of the three setups

	Precision	Recall
A (8)	81%	75%
B (16)	82%	80%
C (17)	82%	83%

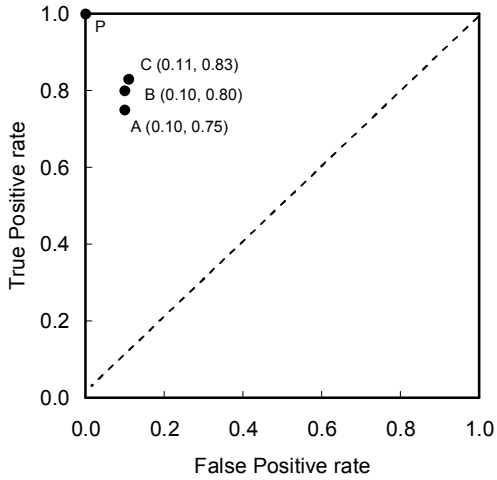


Fig. 2. ROC graph of setups A (8), B (16), and C (17)

In Fig. 2, the point P(0, 1) represents perfect classification, and any points along the dotted line represent random classification, which our method performs significantly better than. Setup C has a recall (true positive rate) of 83%, while that of setup B and A are 80% and 75 % respectively. Since the auction fraud problem has not been systematically dealt with before, no comparison can be made with other approaches. However, we believe our approach and the accuracy that it has achieved is significantly better than the heuristic approach practiced by average eBay users who want to identify suspicious dealers, which involves manual inspection of all the dealers' transaction history and profile information.

5 Working Prototype

We have built a working prototype that uses our proposed method to detect fraudsters on eBay. It is a Java application that can run on Windows, Macintosh, and Linux. Our development platform is Windows XP Professional edition.

Our system has a complex core that is responsible for retrieving and analyzing data from eBay, representing 5578 lines of code. However, we have encapsulated the complexities into a streamlined graphical user interface to promote good usability (Fig. 3).

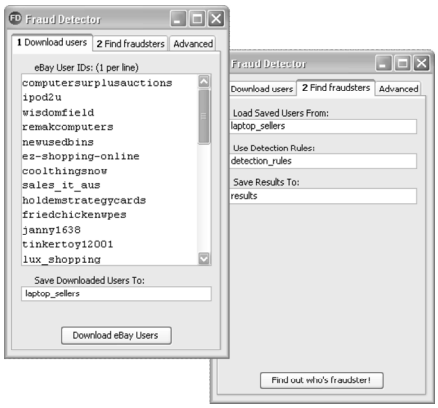


Fig. 3. Screenshots of our system

Fig. 4. Sample results of our system

Fig. 5 illustrates the system flow. Our system allows the user to input multiple eBay users' IDs to query if they are fraudulent or legitimate (Fig. 3). It then retrieves the relevant data about the identities from eBay, and extracts the pre-defined features from the data. The extracted feature sets are then run against the pre-generated decision rules which classify the identities into "fraudster" or "innocent" (Fig. 3). The results are saved to a text file (Fig. 4).

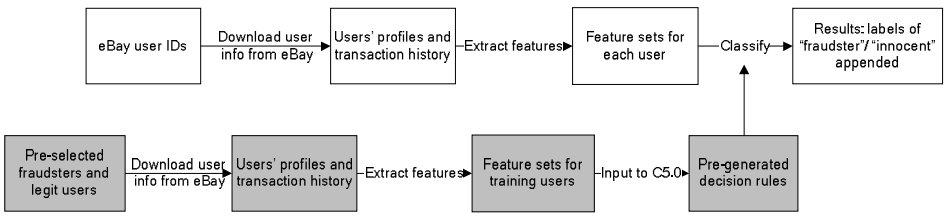


Fig. 5. How our system works

6 Discussion and Future Research Directions

The experiment results indicate that we have already achieved good accuracy with the current sets of features. However, we believe there are still more interesting features that can help boost the performance further. Currently, we focus on the dollar amount of items being bought and sold, their fluctuation, and the frequencies of transactions; some of the features that we might include in the future are the users' registration dates, times, locations, and their frequent transaction categories, etc. We believe that studying more closely and understanding better about how typical fraudsters create

and manage fraudulent activities will provide us with much insight in suggesting more effective features.

Another major improvement that we anticipant is the inclusion of graph pattern detection, which tries to detect the fraudulent patterns that might exist among the interactions between eBay users. In fact, we have already started to explore the efficacy of this method. We have constructed the visualization of some transaction graphs on eBay, which illustrate the transaction trail among users. We have observed an alarming pattern that exists among some of the fraudsters, their possible collaborators and some other legitimate users. Fig. **6** shows the pattern.

The observations from the graph are:

1. Fraudsters might exist within a few "hops" from each other.
2. Fraudsters might choose some legitimate users (usually with high feedback score) as the regular sources to gain reputation.
3. Clusters of fraudsters might be very close to each other (the top and the bottom).

The possible formation of fraud networks as shown in the figure shows that frauds might indeed be carefully planned and managed, and that a perpetrator may be controlling several identities at the same time.

We believe that if we exploit this kind of interactions during our classification process, not only could we increase the accuracy in detecting fraudsters and strengthen the proof of claims, but also discover the complete fraudster communities – which include both the fraudsters and their possible collaborators. This will significantly increase the effectiveness of the whole fraud detection process.

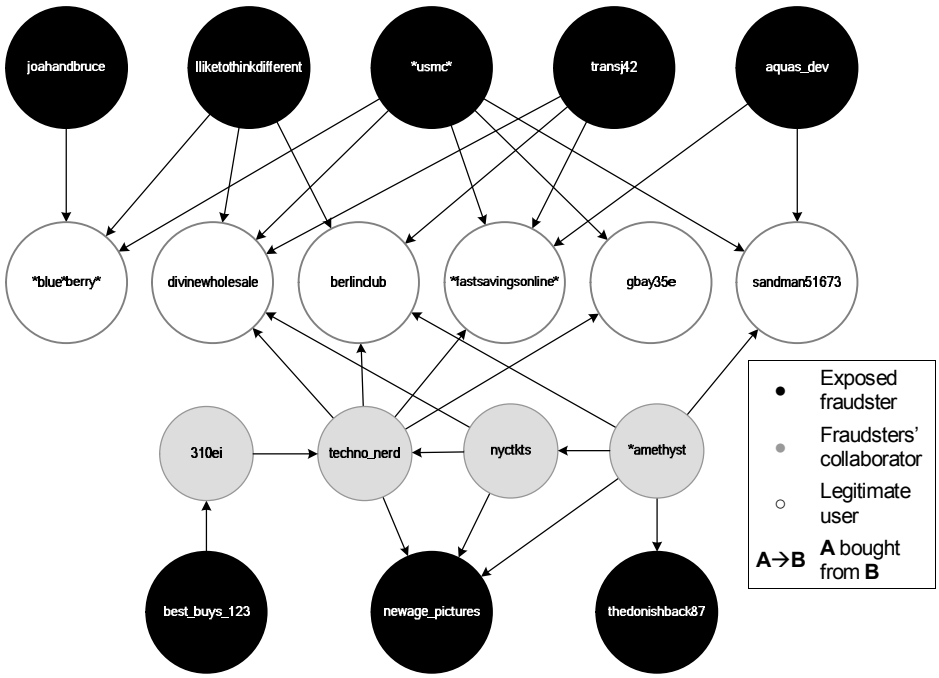


Fig. 6. Transaction graph of some eBay users

7 Conclusion

In this paper, we have presented the first method that analyzes and detects electronic auction fraudsters, as well as a working system that implements the method. We have the following three contributions:

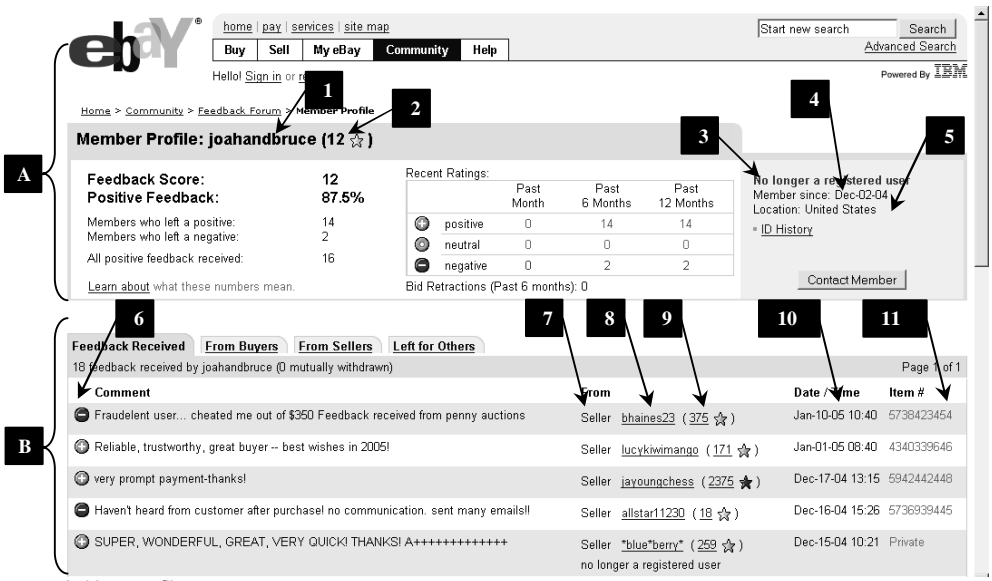
1. We formalized the auction fraud problem and proposed a method to systematically detect auction fraudsters;
2. We suggested characteristic features that distinguish fraudsters from legitimate users, where the features are extractable from the graph describing users' past transactions;
3. We evaluated our proposed method on real data from eBay and showed that it can reach a precision of 82% and a recall of 83%, representing a significant lift over random classification.

In our future research, we are going to explore the possibility of including more features and incorporating the graph pattern detection method, both of which could greatly improve the effectiveness of the whole detection process.

Reference

1. Chua, C., Wareham, J.: Fighting Internet Auction Fraud: An Assessment and Proposal, Computer, vol. 37, no. 10, October (2004) 31-37
2. Fawcett, T.: ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Technical Report HPL-2003-4, HP Labs (2003)
3. Gy¨ongyi, Z., Garcia-Molina, H., Pedersen, J.: Combating web spam with TrustRank. In Proceedings of the 30th International Conference on Very Large Databases (VLDB) (2004)
4. Melnik, M. I., Alm, J.: Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions. Journal of Industrial Economics, Vol. 50 (2002) 337-49
5. Neville, J., Jensen, D.: Collective classification with relational dependency networks. Proceedings of the 2nd Multi-Relational Data Mining Workshop, 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2003)
6. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project (1998)
7. Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K.: Reputation Systems. Communications of the ACM, 43, December (2000) 45-48
8. Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K.: The Value of Reputation on eBay: A Controlled Experiment. Working Paper (2002)
9. Zacharia, G., Moukas, A., and Maes, P.: Collaborative reputation mechanisms in electronic marketplaces. In Proceedings of the HICSS-99 Conference, Electronic Commerce Minitrack (Maui, Hawaii, Jan. 5–9). IEEE Computer Society (1999)
10. CNN: eBay posts higher profits, April (2005) http://money.cnn.com/2005/04/20/news/fortune500/ebay_analysis/
11. Federal Trade Commission: Internet Auctions: A Guide for Buyers and Sellers (2004) http://www.ftc.gov/bcp/conline/pubs/online/auctions.htm
12. Internet Fraud Complaint Center: IC3 2004 Internet Fraud - Crime Report (2005) http://www.ifccfbi.gov/strategy/statistics.asp
13. USA Today: How to avoid online auction fraud (2002) http://www.usatoday.com/tech/columnist/2002/05/07/yaukey.htm

Appendix: Annotated eBay Screenshots



A: User profile

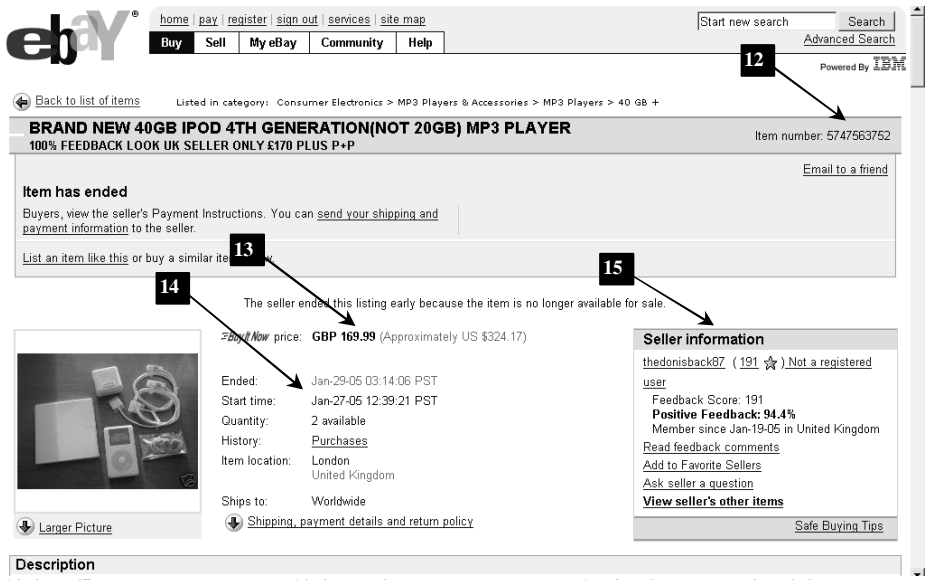
1: User ID — 2: Feedback score — 3: Registration status

4: Registration date — 5: Registration location

B: Past transactions

6: Rating — 7: Type of dealer — 8: Dealer's user ID

9: Dealer's feedback score — 10: Feedback date — 11: ID of traded item



12: Item ID — 13: Item price — 14: Auction start and end dates

15: Seller information