An Auctioning Reputation System Based on Anomaly Detection

Shai Rubin, Mihai Christodorescu
Vinod Ganapathy, Jonathon T. Giffin,
Louis Kruger, Hao Wang, Nicholas Kidd

Computer Sciences Dept.
University of Wisconsin
Madison
Online Auctioning

• Huge volume: eBay hosted 440,100,000 new listings in Q2 2005
• In this talk: trustworthiness of online auctioning
• Why do we buy in an online auction?
  A. to find a rare/collectable item
  B. to find a bargain; commodity at a “good” price

• eBay financial report (expected 2005):
  – Clothing & Accessories --- $3.3 billion (2nd)
  – Consumer Electronics --- $3.2 billion (3rd)
  – Computers --- $2.9 billion (4th)

Data suggests that most people use eBay to find bargains
Finding a Bargain is Tricky

• Inherently untrustworthy environment:
  – Pseudonymous sellers
  – Pseudonymous buyers
  – Delivery? Warranty? Quality?

• Reputation system: a tool to establish trust
Finding a Bargain is Tricky

- eBay’s reputation system provides little help
  - Based on feedback: vulnerable to “poisoning” attack
Finding a Bargain is Tricky

- eBay’s reputation system provides little help
  - Based on feedback: vulnerable to “poisoning” attack
  - Does not provide information on price

+ ANOTHER GREAT DEAL!! I WILL BE BACK!!
+ WORKS GREAT!! GREAT DEALER!! My GOD I Can't SPELL!!
+ WORKS GREAT!! DREAT DEALER!!
+ WORKS GREAT!! DREAT DEALER!!
+ WORKS GREAT!! DREAT DEALER!!
Finding a Bargain is Tricky

- eBay’s reputation system provides little help
  - Based on feedback: vulnerable to “poisoning” attack
  - Does not provide information on price
  - Does not differentiate among the majority of sellers

90% of sellers:
Positive feedback > 97.3%

50% of sellers:
Positive feedback > 99.4%

% of positive feedback

% of sellers (1545 sellers with more than 10 auctions)
Goals

• Alice—a buyer, Bob—a seller
• Develop a trustworthy mechanism that helps Alice:
  – Achieve her goal: what are the chances that Alice can find a bargain in Bob’s auctions?
  – Warn her from fraudulent activities: are the prices in Bob’s auctions artificially inflated?
  – Provide her assurance against poisoning attack: why should Alice trust the mechanism?
Contributions

• A reputation system that helps buyers avoid sellers who seem to be inflating prices
  – Formulated the “seem to be inflating prices” as an anomaly detection problem
  – Business level anomaly detection: the basic events are auctions, bidding.
  – Behavioral system: based on how human behave/act rather than on people feedback.

• Only a first step, some goals still ahead
Outline

• Motivation: find a bargain and avoid fraud
• Contributions: anomaly detection system to identify price inflating sellers:
  – The N model
  – The M model
  – The P model
• Case studies
Auctioning 101

- Pseudonymous sellers and bidders
- Auctions end after a predefined time (e.g., 7 days)
- Highest bid wins
- Seller sets minimum starting bid
- Shilling: a group of bidders that place fake bids to inflate the final price
Methodology

• Collect data from eBay
  – three weeks of data in the category: Laptop Parts & Accessories
  – 127,815 auctions, 12,331 sellers,
  – 604 high-volume sellers: posted more than 14 auctions controls 60% of the market

• Use statistical model to predict seller behavior
  – 95% of the sellers are “normal”
  – 5% are abnormal, or suspicious
Step 1: Average Number of Bids

• What is an indication that prices are high?
  – high number of bids

• Goal: identify sellers with abnormally high number of bids

• 95% of high-volume sellers have less than 7 bids per auction
  – Model is insensitive to supply: number of auctions posted by a seller
Step 1: The N Model

Correlation: many auctions implies low number of bids

Suspicious: 5% of high-volume sellers

Suspicious seller: one that posts many auctions and still attracts many bids
Outline

• Motivation: find a bargain and avoid fraud
• Contributions: anomaly detection system to identify price inflating sellers:
  – The N model: *a seller is suspicious if they post many auctions that attract many bids*
  – The M model
  – The P model
• Reputation example
Step 2: Average Minimum Starting Bid

• Legitimate explanation for high number of bids: low minimum starting bid
• Goal: identify sellers with abnormally high number of bids and high minimum bid
• Problem: how do you know that the minimum bid is high?

Relative minimum bid (RMB) = \frac{\text{winning bid} - \text{minimum bid}}{\text{winning bid}}
Step 2: The M Model

Correlation: low minimum starting bid implies many bids

M suspicious seller: starts with high minimum bid and attracts many bids

M+N suspicious seller: posts many auctions, attracts many bids, starts with high minimum bid
Step 3: Bidders’ Profile of a Seller

- Fraudulent explanation for high number of bids: shilling
- Goal: identify group of bidders that repeatedly bid and lose in a seller’s auctions
- Suspicious seller:
  - N: sellers with abnormally high number of bids and
  - M: high starting bid and
  - P: has a group of bidders that repeatedly bid and lose
Bidder Presence Curve

5% of bidders in this seller’s auctions participated in 80% of the auctions.
Bidder Presence/Win Curves

5% of bidders in this seller’s auctions participated in 80% of the auctions.

the same 5% won only 10% of the auctions.
Bidder Presence/Win Curves
(Normal case)

10% of the bidders participated in 20% of the auctions and won 20% of the times
Outline

• Motivation: find a bargain and avoid fraud
• Contributions: anomaly detection system to identify price inflating sellers:
  – The N model: a seller is suspicious if they post many auctions that attract many bids
  – The M model: a seller is suspicious if they attract many bids and start with high minimum bid
  – The P model: a seller is suspicious if they have a group of bidders that repeatedly participate and lose
• Reputation example
Reputation Example: Seller 10260

- Average bids per auction
- Average relative minimum bid for seller
- Bidders %

\[ N, M, P \]
Results Summary

- 54 sellers classified as abnormal with respect to at least one model
- 3 sellers classified as abnormal with respect to all three models
- No confirmed fraud
Summary

• Trust: do we get what we expected?
• Reputation system as anomaly detection
  – Attempt to identify price inflation
  – Work at the business level
  – Consider poisoning attack (see paper)

Thank you.
Questions?