Recommendation System to Identify Collusive Users in Online Auctions Using the Pollution Diffusion Method

Chen-Yang Cheng[1], Iuon-Chang Lin[2,3], Hao-Ju Wu[2]

[1] Department of Industrial Engineering and Management, National Taipei University of Technology, Taiwan
[2] Department of Management Information Systems, National Chung Hsing University, Taiwan
[3] Department of Photonics and Communication Engineering, Asia University, Taiwan
cycheng@ntut.edu.tw, iclin@nchu.edu.tw, g9729007@mail.nchu.edu.tw

Abstract

Reputation systems provide mechanisms for differentiating between honest and dishonest participants in ecommerce environments. Several reputation systems are deployed in practical electronic marketplaces. However, a considerable challenge is that malicious groups may unlawfully increase their reputation through deceitful manipulation of purchase feedback. Therefore, reputation systems must be robust and able to detect collusion, deception, and strategic manipulation. This study proposes a recommendation system to identify collusive users, inspired by observing the spread of pollution. After identifying malicious users, the suspects are likely to be identified using the proposed scheme, which efficiently reduces the relationship network size by removing relationships with positive ratings. According to the simulation results, pollution diffusion took 16 *ms* in a simulation involving 5000 users.

Keywords: Recommendation systems, Reputation systems, Purchase feedback, Collusion, Online auctions

1 Introduction

The rapid development of Internet and web technologies has resulted in the rapid growth of electronic commerce applications [1-2]. Online auctioning is one of the most successful electronic commerce applications, creating a global marketplace for participants to bid for and sell products and services over the Internet [2-4].

However, although online auctions have yielded numerous benefits, they entail potential dangers. In online auctions, the delivery of goods and payment do not occur simultaneously. Therefore, dishonest sellers may not deliver goods, and dishonest buyers may not complete payment transactions [5]. Unlike physical trading that occurs in a traditional market, sellers and buyers are virtually anonymous in the digital market [2]. Entering and leaving the market without being identified is simple. A report titled *Internet Scams: Fraud Trends January-December 2003* asserted that approximately 89% of Internet fraud complaints received by the National Consumers League (NCL) fraud center were related to online auctions until Ebay removed its website links to the NCL in the fall of 2003 [6]. The Internet Crime Complaint Center (IC3) stated that online auction fraud was the most reported offense, accounting for 35.7% of referred crime complaints, in its 2007 annual report [7]. During 2010, auction fraud was still among the top three crimes related to dollar losses [8].

To distinguish between honest and dishonest participants in digital markets, reputation systems are used to reflect user trust [3] and have become a crucial component of electronic commerce [9-10]. Studies have indicated that seller reputation exerts a slight but significant positive effect on selling price in online auctions [5, 11-13]. Therefore, sellers with high reputations can potentially earn more money. However, malicious users can exploit this feature by registering numerous alias accounts, creating "fake transactions" (i.e., creating transaction data on online auction sites, but not actually selling or buying items), and leaving positive ratings to increase reputation scores. Current reputation systems cannot solve this problem. Therefore, collusion detection requires alternative methods.

The solutions to this problem can be categorized into the following two types. First, a fee is charged for each transaction to reduce collusion. Second, recommendation systems are used to identify suspicious users. In the first method, online auction sites charge a fee for each transaction (including fake transactions), thus reducing fake transactions by increasing the cost [9]. However, although online auction sites such as Ebay charge a transaction fee, a recent study [10] indicated that fee changes have lowered the fake transaction cost. Therefore, collusive users can achieve high reputations through deceitful

*Corresponding Author: Iuon-Chang Lin; E-mail: iclin@nchu.edu.tw
DOI: 10.3966/160792642019032002004

manipulation of the purchase feedback for a small cost.

This paper proposes a novel scheme for identifying colluders. The proposed recommendation system notifies reputation centers of suspects. After identifying a malicious user (i.e., users on reputation center blacklists), suspected collusive users are identified using the proposed scheme. The suspect lists are sent to the reputation center for a detailed investigation to verify whether the suspects are malicious users.

This paper is organized as follows: Section 2 presents a brief description of related studies, Section 3 details the proposed scheme, Section 4 presents simulations, and Section 5 offers concluding remarks.

2 Related Studies

This section briefly introduces the current reputation systems and the existing collusion recommendation systems used in online auctions.

2.1 Reputation Systems

The reputation systems used in online auctions involve a centralized structure. The schemes are based on cooperative feedback mechanisms. Figure 1 shows a typical framework. Nodes A, B, C, and D represent four online users, and they could be either sellers or buyers. Once sellers and buyers complete an online transaction, they may leave feedback regarding the performance of other sellers and buyers. The reputation center collects this feedback and creates user reputation profiles, which are public. Potential sellers or buyers who wish to know whether a particular user can be trusted can access the profiles to verify their reputation and performance history [5, 9].

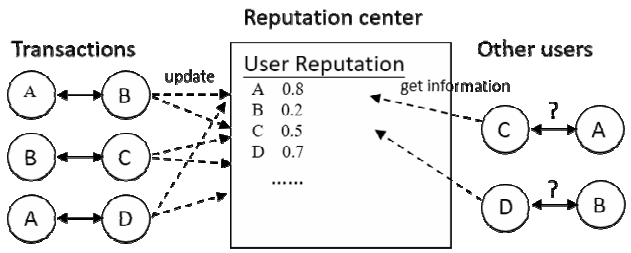


Figure 1. The framework of centralized reputation systems

On Ebay and most other online auction sites, sellers and buyers can leave feedback for each other after completing a transaction by using a numerical rating system (positive, neutral, or negative feedback) and text comments. When new feedback for a particular user is received, the reputation center updates his or her reputation score using the numerical rating. Reputation scores can be computed using the accumulative mechanism and the average mechanism.

The simplest method for computing reputation scores is to sum the ratings [9]. For example, if a user received ratings of {1, 1, 1, 1, -1}, the reputation score is computed as 1+1+1+1+ (-1) = 3 when the accumulative mechanism scheme is used. However, in the average mechanism scheme, the reputation score is the average of the ratings, computed as (1+1+1+1+ (-1)) / 5 = 0.6 (60%).

Both mechanisms are simple and do not account for the collusive user problem [14]; thus, malicious users can increase their reputation through "ballot stuffing." On the Internet, people can easily create multiple user accounts and fake transactions (although the transactions seem complete, no actual transactions occurred) to provide the particular user with numerous positive ratings. Unreliable purchase feedback is also observed in online auctions [10], in which malicious users create fake transactions to generate false positive ratings, and thus deceive other users

2.2 Collusion Recommendation Systems

Wang et al. [14] proposed a recommendation system that was tested using social network analysis, as shown in Figure 2. The system builds a network according to trading relationships and subsequently identifies suspected users by using structure analysis. The primary concept is that collusive users are frequently engaged in trading and thus form a subgroup in the network. The *k*-core indicator and core/periphery ratio are used to identify the subgroup. For example, the nodes A, B, C, and D form a three-core structure, meaning that each node in the subgroup has at least three links to members of the group. A large *k* indicates a high subgroup density.

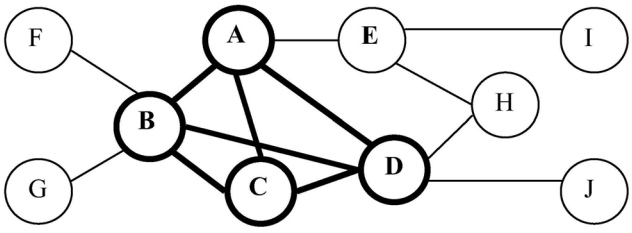


Figure 2. The k-core structure used in social network analysis

However, the complex transaction relationships created in online auctions create a large network and complex node structure. Because the social network analysis algorithm may take considerable time to run, and the goal of the scheme is to identify high-density groups, the system cannot rapidly determine whether a particular node is a malicious user. In the next section, this paper proposes a scheme that efficiently reduces the network size and calculates the "pollution value" of each node to easily discern the trust level of a particular node.

3 Proposed Scheme

3.1 Primary Concept

The proposed scheme is based on observations of the spread of pollution. Whenever an area is polluted, the surrounding area is also likely to be polluted, and these new pollution sources continually spread the pollution. In online auctions, malicious users on reputation center blacklists are the pollution sources. Other users who engage in direct transactions with the pollution sources are likely to be polluted. Therefore, polluted users become new pollution sources, unknowingly spreading the pollution to users who engage in direct transactions with them.

Frequently interacting collusive users form subgroups [14]. For example, in a malicious group comprising A, B, and C, A and B, B and C, and A and C may engage in transactions with each other. Consequently, the pollution spreads among them. Thus, the degree of pollution produced by the three users is higher than that of others. The system identifies highly polluted users and notifies the reputation center, which labels them as suspected collusive users and members of the pollution source (i.e., the malicious group).

To diffuse the amount of neighborhood pollution, the system should establish a relationship network for pollution sources. At the first level, users who have engaged in direct transactions with the pollution source are included in the network. At the second level, users who have engaged in direct transactions with users at the first level are also included in the network. Therefore, as the number of layers increases, the number of nodes in the network increases exponentially. To reduce the number of network nodes, users who give the pollution source negative ratings are removed because only a user who gives a positive rating (+1) to a malicious user can be a collusive user. Therefore, all other users are removed from the network to reduce the network size.

3.2 Computing Model

Suppose that each user on an online auction site has a pollution value (*P*), which refers to the probability of being a malicious user. The *P* value is initially set to 0. For pollution sources (users on the blacklist), the *P* value is set to 1. All users who have engaged in direct transactions with the pollution sources update their *P* value according to the following formulas:

$$P_x^t = P_x^{t-1} + \Delta \boldsymbol{P}_x^t \tag{1}$$

$$\Delta P_x^t = \Delta P_y^t \times \frac{C_{xy}}{\sum_{i=1}^{n} C_{yi}} \tag{2}$$

where y is the pollution source, x is the user polluted by y, P_x^t is the pollution value of x at time t, ΔP_x^t is the change in P_x^t, C_{xy} is the number of transactions that occur between x and y and n is the number of users who have engaged in direct transactions with y.

After diffusing the neighborhood pollution, the system calculates the Z-score of the *P* value for the users, and subsequently recommends users for whom $Z(P^t) > T$. The threshold T is predefined, and $Z(P^t)$ is calculated using the following formula:

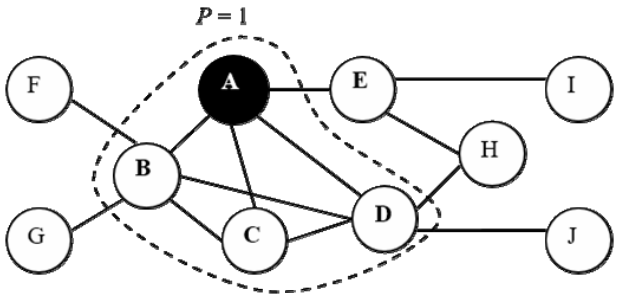
$$Z(P_i^t) = \frac{P_i^t - \overline{P^t}}{\sigma P^t} \tag{3}$$

where $\overline{P^t}$ and σP^t are the mean and standard deviation of P^t, respectively. The Z-score calculation does not include the original pollution source on the blacklist.

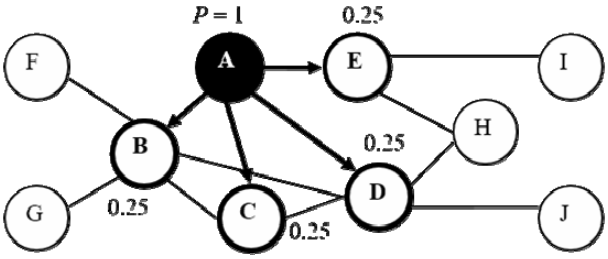
As Figure 3(a) shows, A is the malicious user included on the blacklist, and B, C, and D are the collusive users who have engaged in transactions with each other. At the first level, A diffuses the pollution to neighborhood B, C, D, and E. As Figure 3(b) shows, every node obtains a *P* value of .25. The nodes B, C, D, and E are recorded as the pollution sources at the second level. Figure 3(c) Figure 3(f) shows the pollution diffusion that occurs at the second level. Figure 3(g) shows the final *P* value. Figure 3(h) shows the Z-score of the users. Supposing that the threshold T = 0.7, all the collusive users can be accurately identified.

4 Simulation Results

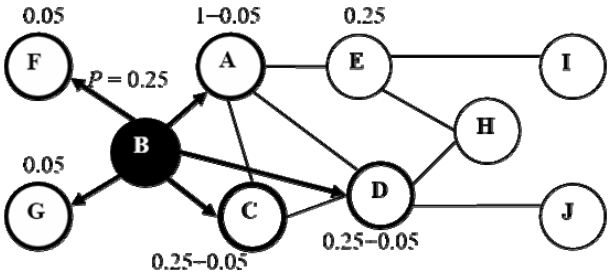
This section presents the simulations used to test the performance of the proposed scheme. In the simulations, the network comprised n users. This simulation initially established a malicious group with m users. A user was then randomly chosen to be a malicious user identified from a blacklist. The users in the malicious group were set to engage in direct transactions with each other (once or twice, selected at random). In other scenarios, including transactions between malicious users and normal users, and transactions in which both the seller and buyer are normal users, a 5% possibility of occurrence was used (only once). Starting from the malicious user on the blacklist, the pollution diffusion scheme was subsequently performed. After k layer diffusion, the threshold value T was used to determine whether the users were suspected collusive users. The simulation was implemented using the Java programming language in 10 runs.



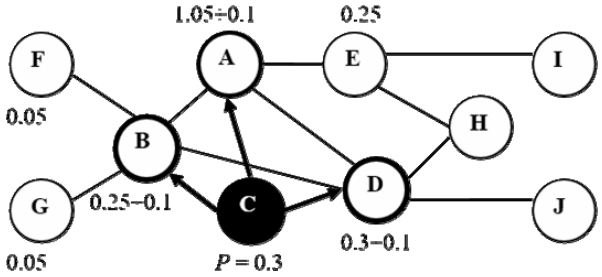
(a) The pollution source and collusive users



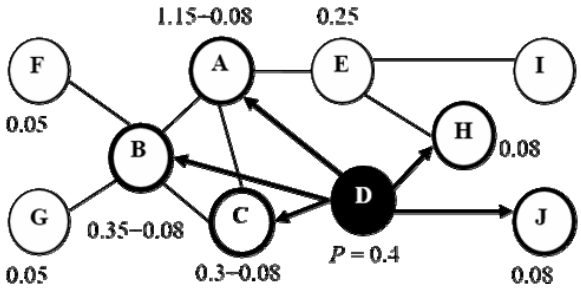
(b) The spread of pollution at the first level



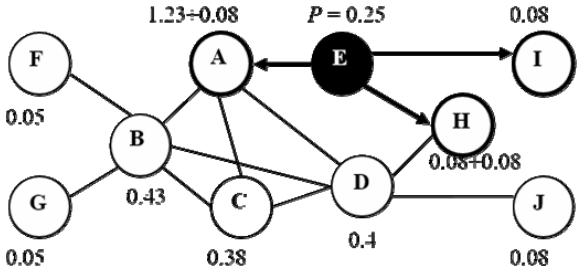
(c) The spread of pollution at the second level (for node B)



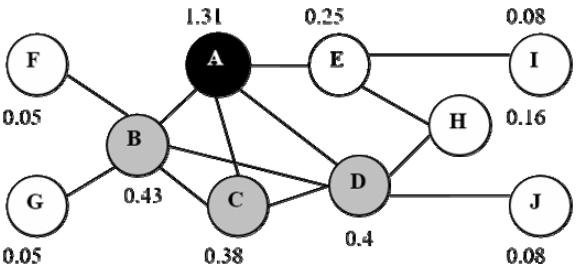
(d) The spread of pollution at the second level (for node C)



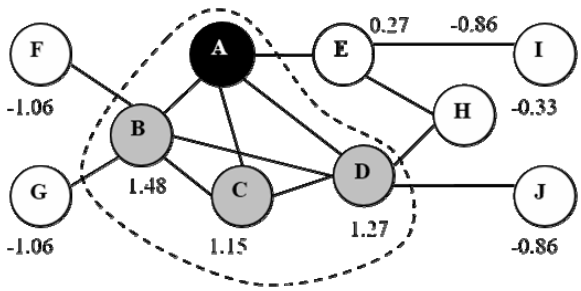
(e) The spread of pollution at the second level (for node D)



(f) The spread of pollution at the second level (for node E)



(g) The final P value distribution after two-layer diffusion



(h) The final *P* value distribution after two-layer diffusion

Figure 3. Example of the proposed scheme

We present the simulation results in the confusion matrix shown in Table 1 to demonstrate the false positive and false negative trade-off. The table with two rows and two columns presents the number of false positives, false negatives, true positives, and true negatives based on the simulation. We also performed tests with the threshold value T value as 0.4, 0.45, and 0.5. Table 2 shows that the threshold value T substantially affected the simulation results. The simulation involved 2000 users, including a malicious group containing 10 members. After two-layer pollution diffusion, the optimal T value was used to identify the collusive users between the threshold values of 0.4 and 0.5. When the T value was set to 0.4, the system correctly identified all the collusive users. However, certain normal users were falsely identified as collusive users at this value. When the T value was set to 0.5, all normal users were correctly identified as normal users; however, not all collusive users were correctly identified.

Table 1. The definitions of TP, FP, FN, and TN

	Users who are collusive users	Users who are normal users
Users identified as collusive users	TP (true positive)	FP (false positive)
Users identified as normal users	FN (false negative)	TN (true negative)

Table 2. The simulation result with $n = 2000$, $m = 10$, and $k = 2$

T	TP	FP	FN	TN
0.4	9	0	18.1	1971.9
0.45	6.4	2.6	8	1982
0.5	5	4	0	1990

The T value should be set according to the objective of the manager of the reputation center. If the manager wishes to identify all collusive users, a lower T value should be set. However, if the manager wishes to lower the cost, a higher T value should be set. For example, when $T = 0.4$, the reputation center should examine 15 suspected users on the recommendation list. Because the number is much greater than the actual number of collusive users, determining whether users on the recommendation list are collusive users may involve high costs. If the T value is set to 0.5, only five suspected users are on the recommendation list. Although all the users on the list are collusive users, identifying all four collusive users is not possible.

However, setting the T value may be a difficult decision. The simulation results demonstrate the false positive and false negative trade-off. A lower T value may enable the identification of all collusive users, but such identification may involve a high cost. A higher T value may not result in the identification of all collusive users, but guarantees that most of the users on the recommendation list are collusive users. The T value should be set according to the strategy used by the reputation center. Future research should recommend the optimal T value for users. The optimal T value in consideration of the costs of investigating suspected collusive users could be elaborated on in future study.

5 Conclusion

This paper proposes a recommendation system to identify collusive users. The primary concept of the proposed scheme was inspired by observing the spread of pollution. Whenever an area is polluted, the surrounding areas are likely to be polluted and become new pollution sources that continually spread the pollution. The algorithm calculates a pollution value for each user, which represents the possibility of being a collusive user. The spread of pollution enables the identification of collusive users. The proposed scheme efficiently reduces the relationship network size by removing the relationships with feedback ratings. The scheme is intuitive, is easy to implement, and has fast processes. Pollution diffusion takes only 16 ms in a simulation containing 5,000 users.

This study had some limitations. First, not all of the training and testing data for end price predictions were extracted from real transactions. The performance of our proposed scheme should be further researched using data from real transactions. Second, the reputation system focuses on how to force sellers to be honest.

References

[1] C.-C. Yu, A Web-based Consumer-oriented Intelligent Decision Support System for Personalized e-services, *Proceedings of the 6th International Conference on Electronic Commerce*, The Netherlands, 2004, pp. 429-437.

[2] D. G. Gregg, S. Walczak, E-commerce Auction Agents and Online-auction Dynamics, *Electronic Markets*, Vol. 13, No. 3, pp. 242-250, August, 2003.

[3] D. G. Gregg, S. Walczak, Auction Advisor: An Agent-Based Online-auction Decision Support System, *Decision Support Systems*, Vol. 41, No. 2, pp. 449-471, January, 2006.

[4] D. Ariely, I. Simonson, Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions, *Journal of Consumer Psychology*, Vol. 13, No. 1-2, pp. 113-123, January/March, 2003.

[5] D. Houser, J. Wooders, Reputation in Auctions: Theory, and Evidence from eBay, *Journal of Economics & Management Strategy*, Vol. 15, No. 2, pp. 353-369, Summer, 2006.

[6] National Consumers League, *Internet Scams Fraud Trends January-December 2003*, http://www.fraud.org/.

[7] *2007 Internet Crime Report - Internet Crime Complaint Center*, https://www.ic3.gov/media/annualreport/2007_IC3

Report.pdf.

[8] *2010 Internet Crime Report - Internet Crime Complaint Center*, https://www.ic3.gov/media/annualreport/2010_IC3Report.pdf.

[9] A. Jøsang, R. Ismail, C. Boyd, A Survey of Trust and Reputation Systems for Online Service Provision, *Decision Support Systems*, Vol. 43, No. 2, pp. 618-644, March, 2007.

[10] F. Dini, G. Spagnolo, Buying Reputation on eBay: Do Recent Changes Help?, *International Journal of Electronic Business*, Vol. 7, No. 6, pp. 581-598, January, 2009.

[11] S. Dewan, V. Hsu, Adverse Selection in Electronic Markets: Evidence from Online Stamp Auctions, *The Journal of Industrial Economics*, Vol. 52, No. 4, pp. 497-516, December, 2004.

[12] Z. Lee, I. Im, S. J. Lee, The Effect of Buyer Feedback Scores on Internet Auction Prices, *Journal of Organizational Computing and Electronic Commerce*, Vol. 16, No. 1, pp. 51-64, February, 2006.

[13] N. Bruce, E. Haruvy, R. Rao, Seller Rating, Price, and Default in Online Auctions, *Journal of Interactive Marketing*, Vol. 18, No. 4, pp. 37-50, January, 2004.

[14] J.-C. Wang, C.-C. Chiu, Recommending Trusted Online Auction Sellers Using Social Network Analysis, *Expert Systems with Applications*, Vol. 34, No. 3, pp. 1666-1679, April, 2008.

Biographies

Chen-Yang Cheng received his Ph.D. in Industrial and Manufacturing Engineering at Penn State University. He is currently an Assistant Professor in Department of Industrial Engineering and Enterprise Information at Tunghai University. Prof. Cheng's research interests include RFID in healthcare, Healthcare Systems, and Biomedical Informatics.

Iuon-Chang Lin received the Ph.D. in Computer Science and Information Engineering in March 2004 from National Chung Cheng University, Chiayi, Taiwan. He is currently a professor of the Department of Management Information Systems, National Chung Hsing University, Taichung, Taiwan. His current research interests include electronic commerce, information security, Blockchain Security, and cloud computing.

Hao-Ju Wu received the B.S. degree and M.S. degree in Department of Management Information Systems, from National Chung Hsing University, Taichung in 2008 and 2010. Her current research interests include electronic commerce, information security, and digital right management.