DEFENDING ONLINE STANDING ATTACK AGAINST COLLABORATIVE NETWORK AND KNOWLEDGE DISCOVERY TO PREVENT SECURITIES FRAUD

Ms. A. Santhiya[1]
[1]*Assistant professor*, Department of ECE,
Jeppiaar engineering college, Chennai, India.

Ms. A. Keerthika[2]
[2]*Assistant professor*, Department of IT,
Jeppiaar engineering college, Chennai, India.

ABSTRACT
Many small companies provide "reputation boosting" services for sellers and has taken up about three-fourths of the market share. There are also many Web sites where users can buy views for promoting their YouTube videos. Such coordinated and even profit-driven manipulations can greatly distort reputation scores, make reputation systems lose their worthiness, undermine user confidence about reputation-centric systems, and may eventually lead to system failure. Many defense solutions have been developed to protect reputation systems. Most of the defense solutions utilize signal processing techniques to differentiate normal feedback from dishonest feedback and normal users from malicious users. There is always an "arms race" or "competition" between attacks and defenses. The arms race for attacking/securing reputation systems is currently taking place and evolving rapidly.

Keywords- Whitewashing and Traitor Attacks, joint temporal analysis, user correlation analysis

INTRODUCTION
The Internet has created vast opportunities to interact with strangers. The interactions can be fun, informative, and even profitable [1]. However, there is also risk involved. Will an eBay seller ship the product in time? Is the advice from a self-proclaimed expert on Epinion.com trustworthy? Does a product from Amazon.com have high quality as described? To address these problems, one of the most ancient mechanisms in the history of human society, word of mouth, is gaining new significance in cyberspace, where it is called a standing system [2]. A standing system collects evidence about the properties of individual objects, aggregates the evidence, and disseminates the aggregated results. Here, the objects can be products (e.g., in the Amazon product rating system), businesses (e.g., hotel ratings in various travel sites), users (e.g., sellers and buyers on eBay), and digital content (e.g., video clips on YouTube). The aggregated results are called standing scores.

Most commercial systems collect user feedback (i.e., ratings/reviews) as evidence. This type of system is referred to as a feedback-based standing system. Signal processing plays an important role in standing systems, in which the standing scores are in fact the prediction of the objects' future behaviors based on the data describing their past behaviors. Various signal models are suggested for computing objects' standing scores. In Bayesian standing systems, an updated standing score (i.e., posteriori) is computed based on the previous standing score (i.e., priori) and the new feedback (i.e., observations) [3]. Belief theory, a framework based on probability theory, has been used to combine feedback (i.e., evidence from different sources) and represent standing scores [4]. Flow models, which compute standing by transitive iteration through looped or arbitrarily long chains [5], are investigated to Calculate standing scores. Standing has also been interpreted as linguistically fuzzy logic concepts [6]. Furthermore, as discussed later in this section, signal processing techniques are widely used to defend standing systems against attacks. As standing systems are having increasing influence on consumers' online purchasing decisions and online digital content distribution [7], the incentive to manipulate standing systems is growing.

STANDING SYSTEMS AND ATTACK CLASSIFICATION
STANDING SYSTEM MODEL
A standing system collects evidence about the properties of individual objects, analyzes and aggregates the evidence, and disseminates the

aggregated results as standing scores. In this subsection, we will review representative standing systems, such as those used by Amazon, YouTube, Digg, and CitySearch and build the system model as follows.

Figure1. Signal and Information Processing for Social Learning and Networking

Evidence collection: A standing system can obtain three types of evidences. The first type is direct observation, usually based on the experiences of the employees of a business (e.g., ConsumerReport.org). The second type is opinions from experts, who have verifiable expertise and provide feedback either voluntarily or for a fee. Both types of evidence are considered reliable, but they are costly to collect for a large number of objects. The third type is feedback provided by users, which have been the main source of evidence in most of today's popular standing systems, such as the product rating system on Amazon.com, restaurant ratings on Yelp.com, and customer reviews at the Apple App Store. However, user feedback is also the least reliable source of evidence because it can be easily manipulated [12].

Standing aggregation: Standing aggregation algorithms calculate the standing scores of objects based on the collected evidence. A good standing aggregation scheme should be able to compute standing scores that accurately describe the true quality of objects, even if there is fraudulent feedback. Standing dissemination: Standing systems not only make the standing scores publicly available but also release extra information to help users understand the meaning of them. For example, Amazon shows all feedback given by each reviewer. YouTube starts to provide visualization of viewing history for video clips, accompanied with some statistical features.

To manipulate a standing system, attackers can 1) obtain information about the target objects, defined as objects whose standing scores' increase/decrease is the goal of the attack, 2) insert fraudulent feedback in the evidence collection phase, and 3) aim to mislead the evidence aggregation algorithm such that it yields unfairly high/low standing scores for the target objects, misclassifies honest feedback/users as fraudulent, and misclassifies fraudulent feedback/users as honest.

ATTACKS CLASSIFICATION

To secure a system, people have to first understand the attacks. Classification of various attacks is an effective way to understand the features of the attacks. In this article, we classify attacks against feedback-based standing systems from four different angles. Targeting objects or targeting system: The purposes of attackers can be divided into two categories: 1) manipulating standing scores of one or several objects and 2) undermining the performance of the entire standing system. In the first category, the attacker aims to boost or downgrade standing scores of specific target objects. These objects either gain or loss advantage due to inaccurate standing scores when competing with similar objects for users' attention or preference. In the second category, the attacker aims to mislead the standing of a sizeable proportion of objects and undermine the performance of the entire standing systems. For example, in [10], an advanced attack can overturn the standing (from positive to negative) of a large number of objects and therefore undermine users' trust in the standing system.

Direct attack or indirect attack: Many standing systems compute two types of standing: 1) object qusality describing whether or not an object has high quality and 2) feedback standing describing whether or not a user tends to provide honest feedback. Feedback standing is often used to mitigate the effect of fraudulent feedback in the evidence aggregation algorithms. For example, feedback from users with a high feedback standing can carry larger weights in the calculation of object quality standing. From this perspective, the attacks can be classified into two categories: direct attacks and indirect attacks. In direct attacks, attackers directly manipulate the object quality standing, whereas in indirect attacks, attackers boost their own feedback standing and/or downgrade honest users' feedback standing so that they can manipulate the object quality standing more effectively.

Collusion or non-collusion: In simple attacks, fraudulent feedback is provided independently. For example, when eBay users boost their standing by buying and selling feedback, the feedback is often from independent sources. These are referred to as non-collusion attacks. In advanced attacks, the attackers control multiple user IDs (referred to as malicious users) that coordinately insert fraudulent feedback. These are referred to as collusion attacks. Roughly speaking, non-collusion

attacks are easier to address because 1) fraudulent feedback that is far away from the honest opinions can be detected by various statistical methods [3], [12], [13], and 2) fraudulent feedback that resembles honest opinions usually do not cause much standing distortion. Collusion attacks, which can be strengthened by the Sybil attack [8], can use more complicated strategies and often exploit vulnerabilities in both standing systems and defense solutions. Knowledge level of attacker: Attacks can also be classified according to the amount of knowledge that the attackers need to obtain to launch an attack. We identify four knowledge levels as follows. In Level 0, attacks are launched without any prior knowledge about the standing systems. In Level 1, attackers know the general principles of the standing systems, such as more positive feedback, which usually leads to higher standing. In Level 2, attackers know the specific standing aggregation algorithm and the defense mechanism that handles fraudulent feedback. In Level 3, attackers can obtain or estimate the parameters (e.g., detection threshold) sed in the standing systems, and adjust their attack strategies accordingly.

DETECTING AND HANDLING FRAUDULENT FEEDBACK

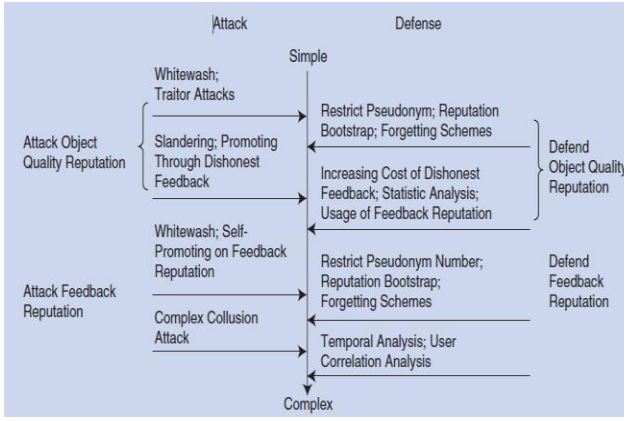

Figure2. Evolution of (a) attacks and (b)

The defense schemes against slandering/promoting attacks are designed from three perspectives

1) Increasing the cost of fraudulent feedback: Policies are in place that requires the users to have certain credentials to provide feedback. The credentials can be a record of real transactions, such as on eBay, Amazon, and App Stores [12].

2) Detection of fraudulent feedback: There are defense schemes studying statistic features of feedback. Most of them detect fraudulent feedback based on the majority rule, which considers feedback that is far away from the majority's opinions as fraudulent feedback. For example, in a Beta function-based approach [12], a user is determined as a malicious user if the estimated standing of an object rated by him/her lies outside q and (1-q) quantile of his/her underlying feedback distribution. An entropy-based approach [13] identifies feedback that brings significant changes in the uncertainty in feedback distribution as fraudulent feedback. In [3], fraudulent feedback analysis is conducted based on a Bayesian model.

3) Mitigating the effects of fraudulent feedback: Feedback standing is proposed to measure the users' reliability in terms of providing honest feedback. The standing of an object is often calculated as the weighted average of all feedback (i.e., ratings), whereas the weight of a feedback is determined by the feedback standing of the user who provides it. As a consequence, feedback provided by users with low feedback standing will have less impact on the object quality standing. To compute the feedback standing score of a user, various methods have been developed. The iteration refinement approach proposed in [7] computes a user's judging power (i.e., weight of this user's feedback in the feedback aggregation algorithm) as the inverse of the variance in all of his or her feedback. In [8], personalized trust is introduced to measure the feedback's reliability.

WHITEWASHING AND TRAITOR ATTACKS

The self-promoting attack and the traitor attack described in the section "Whitewashing and Traitor Attacks" are closely related. They use the same approach (i.e., behaving well and badly alternatively) to achieve different goals. The former directly targets the feedback standing of users (i.e., indirect attack), whereas the latter targets the standing score of objects (i.e., direct attack). Therefore, the various forgetting schemes introduced in the section "Defense Against Whitewashing and Traitor Attacks" can be applied on user feedback standing and address the self-promoting attack.

We have seen that the attacks have evolved from a simple reentering system (i.e., whitewashing) and dynamic behavior changing (i.e., traitor), which aim to maintain standing score after conducting bad behaviors inserting fraudulent feedback (i.e., slandering/promoting), which aims to mislead the evidence aggregation algorithm manipulating the feedback standing, which is the defense mechanism used to detect fraudulent users simple collusion, in which malicious users have similar behaviors

Complicated collusion, in which malicious users carefully coordinate their tasks and the timing to conduct these tasks. On the other hand, the defense approaches have evolved from policies that remove the advantage of being a new user complicated signal processing techniques that detect abnormal feedback the usage of feedback standing Analysis in time

domain and on correlation among users. The emerging attacks have driven the development of the new defense approaches, which in turn have stimulated the advanced attacks.

DEFENSE SCHEME WITH TEMPORAL ANALYSIS

A set of statistical methods that analyze time-domain features of a rating signal [13]. These methods jointly detect the time intervals in which collusion attacks are highly likely present. Based on such detection, a trust-assisted standing aggregation algorithm is designed. When tested against attack data from real human users; such a scheme demonstrated large advantages over the defense schemes that only consider statistics of rating values but ignore temporal information.

DEFENSE SCHEME THAT COMBINES TEMPORAL AND CORRELATION ANALYSIS

Trust management, although it can capture the history of users' rating behavior, has certain shortcomings. First, there are various attacks against trust management, aiming to make honest users have lower trust and fraudulent users have higher trust [10]. Sometimes, using trust management can introduce new vulnerabilities to the system. Second, trust management only captures an individual user's past behavior but not the correlation among users' past behaviors. A defense method that identifies malicious users by combining temporal analysis and user correlation analysis [9]. Compared with the defense in the section "Defense Scheme with Temporal Analysis," this method adopts different detector for temporal analysis and uses correlation analysis to investigate users' past behavior. This scheme is called joint temporal analysis and user correlation analysis (JTAUCA).

JTAUCA contains three main components: 1) change detection, 2) user correlation calculation, and 3) malicious user group identification. In many practical standing systems, the objects have intrinsic and stable quality, which should be reflected in the distribution of normal ratings. Therefore, change detection is an intrinsically suitable tool for temporal analysis. Although the previous change detectors can catch sudden changes, they are not effective when the malicious users introduce changes gradually. Therefore, JTAUCA employs a change detector, which takes a raw rating sequence (i.e., rating values ordered according to when they are provided) as inputs, sensitively monitors the changing trend, and detects changes either occurring rapidly or accumulated over time. If the change detector is triggered by an object, this object is marked as under attack. The direction of the change, either boosting or downgrading, is called the attack direction. Once the detector is triggered, JTAUCA can estimate the starting time and the ending time of the change [9]. The time interval between the starting time and the ending time is called a suspicious interval. Figure 3i illustrates the change detection process. The x-axis is the index of ratings. Figure 3(a) shows the original rating sequence ordered according to the time when the ratings are provided. The y-axis is the rating value ranging from one to five. The honest ratings are in blue, whereas the malicious ratings are in red.

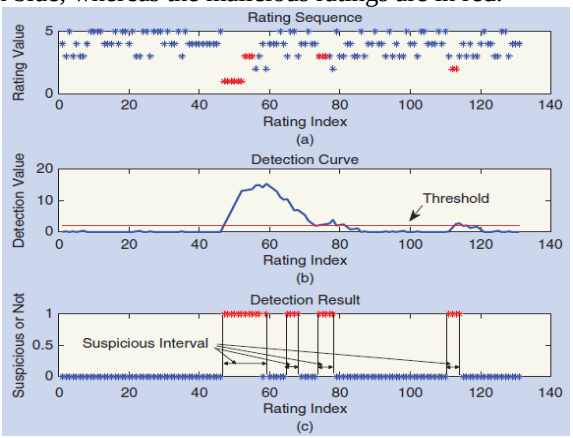

Figure3. Demonstration of change detector in JTAUCA (a) rating sequence, (b) detection curve, and (c) detection result

Figure 3(b) shows the detection curves (gk) of the change detector, as well as the detect ion threshold. Part 3(c) shows the suspicious intervals detected by the change detector. Note that once gk increases above the threshold, the detector is triggered and an alarm is set off (meaning the detection of a change). Since the detector needs some time to res pond, the time when the change starts (i.e., change starting time) is usually earlier than the alarm setting off time. Similarly, when gk drops below the threshold, the alarm is set on, which happens after the change is over. In [9], the change starting/ ending time is estimated based on the alarm set off/on time. After the change detection, JTAUCA moves from time domain to user domain. JTAUCA analyzes correlation among suspicious users, defined as users who rate in the suspicious intervals. We have observed that a larger correlation exists among colluded malicious users. After correlation analysis, suspicious users are separated into different groups/clusters. Finally, the malicious user group identification module determines which group is composed of colluded malicious users.

CONCLUSIONS

In this article, we conducted an in-depth investigation on the competition between attack and defense for feedback-based standing systems as well as introduced representative attack/defense approaches. This competition will surely continue to

evolve and lead to new research challenges. Since there is no real "conclusion" for this evolvement, we conclude this article by pointing out some useful resources and research directions. In this research, it is important to understand and obtain data describing normal users' rating behavior as well as malicious users' attacking behavior. In addition, there are still a large number of fraudulent ratings /reviews and misleading standing scores in current commercial standing systems, such as eBay and Amazon. This is partially because many advanced defense approaches have not made the way into the commercial systems. It will be important to develop tools such that users can directly benefit from the research on defense technologies, and increase their trust in online standing systems.

REFERENCES

[1] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," Commun. ACM, vol. 43, no. 12, pp. 45–48, 2000.

[2] C. Dellarocas, "The digitization of word-of-mouth: Promise and challenges of online reputation systems," Manage. Sci., vol. 49, no. 10, pp. 1407–1424, Oct. 2003.

[3] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt, "Ratings in distributed systems: A Bayesian approach," in Proc. Workshop Information Technologiesand Systems (WITS), New Orleans, LA, Dec. 2001.

[4] B. Yu and M. Singh, "An evidential model of distributed reputation management," in Proc. Joint Int. Conf. Autonomous Agents and Multiagent Systems, 2002, pp. 294–301.

[5] S. Brin and L. Page. (1998). The anatomy of a large-scale hypertextual web search engine. Proc. 7th Int. Conf. World Wide Web (WWW) [Online]. Available: http://dbpubs. stanford.edu:8090/pub/1998-8

[6] J. Sabater and C. Sierra, "Social regret: A reputation model based on social relations," SIGecom Exchanges, vol. 3, no. 1, pp. 44–56, 2002.

[7] D. Houser and J. Wooders, "Reputation in auctions: Theory, and evidence from ebay," J. Econ. Manage. Strat., vol. 15, pp. 353–369, June 2006.

[8] A. Harmon. (2004, Feb. 14). Amazon glitch unmasks war of reviewers. NY Times, [Online]. Available: http://www.nytimes.com/2004/02/14/us/amazon-glitch-unmasks- war-of-reviewers.html

[9] J. Brown and J. Morgan, "Reputation in online auctions: The market for trust," Calif. Manage. Rev., vol. 49, no. 1, pp. 61–81, 2006.

[10] A. Wei. (2009, Sept. 12). Taobao fights reputation spam in e-business boom. Beijing Today

[11] D. Cosley, S. Lam, I. Albert, J. Konstan, and J. Riedl, "Is seeing believing? How recommender systems influence users opinions," in Proc. CHI 2003 Conf. Human Factors in Computing Systems, Fort Lauderdale, FL, 2003, pp. 585–592.

[12] A. Josang and R. Ismail, "The Beta reputation system," in Proc. 15th Bled Electronic Commerce Conf., 2002, pp. 324–337.

[13] J. Weng, C. Miao, and A. Goh, "An entropy-based approach to protecting rating systems from unfair testimonies," IEICE Trans. Inform. Syst., vol. E89–D, no. 9, pp. 2502–2511, Sept. 2006.s