A Relationship Based Collusive Attack Detection Mechanism for Reputation Aggregation in Social Network

Zhang Bo
College of Information, Mechanical and Electrical Engineering
Shanghai Normal University
Shanghai 200234, PR China
e-mail: zhangbo@shnu.edu.cn

Song Feng
College of Information, Mechanical and Electrical Engineering
Shanghai Normal University
Shanghai 200234, PR China
e-mail: 1551190371@qq.com

Yuan Hailei
College of Information, Mechanical and Electrical Engineering
Shanghai Normal University
Shanghai 200234, PR China
e-mail: yhlshnu@163.com

Li Hao*
The Third Research Institute of Ministry of Public Security,
Shanghai 200234, China
Corresponding author e-mail: 13681636617@163.com

Abstract—Reputation aggregation is a significant and inevitable mechanism for ensuring the security in social network. To solve the problem of preventing collusive attack in reputation aggregation in social network, a collusive attack detection mechanism (CADM) is proposed based on users' relationships and their judgment evaluation. Firstly, the rationales of CADM include evaluations of inauthentic judgment, attack behavior similarity, similar reputation of colluders, and the close trust relationship among colluders. The construction of CADM includes four parts as social graph, trust schedule, reputation aggregation form, and collusive factor. Secondly, the four detail collusive factors, including item judgment factor, user similar factor, trust relationship factor and user malicious factor, are addressed respectively to evaluate the probability of collusion happening. And finally, a trust relationship based detection process of CADM, which is comprised by three aspects as attack happening evaluation, user detection, and relationship traversing, is present to find collusive attack through the social relationships in SNS.

Keywords- Collusive attack detection; Reputation Aggregation; Relationship; Social Network; Collusion factor

I. INTRODUCTION

Social network site (SNS) has been one of the most popular platforms for people's daily information acquiring, sharing, and social communication [1]. But in such an open network environment, dishonest individuals and their malicious behaviors are inevitable, and thus, SNS is vulnerable to attack under such environment [2]. Therefore, how to preventing SNS from malicious attacking has garnering many attentions in security researches.

Reputation system is a feasible and indispensable solution for ensuring the security of social network sites [3], which utilized a factor, named reputation, to reflect past trustworthiness and to predict the future likelihood of an individual remaining reliable. A reliable reputation system can determine whether each member of a community is trustworthy. In reputation based systems, deciding whether an individual is trustworthy can be assessed using applications that consider the context of future decisions. Most reputation systems utilize summary/average methods based on past experiences [4]. However, there are unavoidable threats which can endanger the reputation because all judgments are given equally for aggregating reputation values [5]. Mass dishonest judgments, which are inconsistent with facts, would damage the reliability of reputation system. That means reputation would show wrong trustworthiness once malicious individuals attack the reputation aggregation mechanism by commenting maliciously, including inflating or slandering. And worse still, the collusive attacks would bring more damages than single attacks because the scale of attack is larger and there are more attackers in collusive attacks [6]. Therefore, collusive attack detection is a significant challenge for reputation system security, which is also the main motivation of our work.

Many efforts have been made to evaluate, recognize, detect, and prevent collusion in reputation systems [7-9]. There are three main techniques for detecting fraud in reputation aggregation: majority rule [10], signal modeling, and trust management [11]. However, an important factor, social relationship, has not been paid sufficient attentions in traditional researches. In our consideration, collusive attack must be established through their mutual past collaborations, which means that there must be close relationships between colluders for their collusion oriented communication.

In this study, we propose a trust relationship based collusive attack detection method for reputation system in SNS. Our main works in this paper are (1) a model of collusive attack detection is proposed according to the collusion features in SNS; (2) the collusive attack detection factors, including item judgment factor, user similar factor,

trust relationship factor and user malicious factor, are addressed respectively to evaluate the probability of collusion happening, and (3) a trust relationship based detection algorithm is present to find colluders through the social relationships in SNS.

II. MODEL OF COLLUSIVE ATTACK DETECTION

Most collusive attacks are launched simultaneously by a large number of malicious nodes (users) in SNS by giving mass inauthentic judgments to the target node for attacking his/her reputation degree. Generally, there are following features of collusions in SNS as (1) inauthentic judgments are dramatically different with the existing reputation degree if the existing reputation is reliable, (2) most colluders have relative same or similar attack behaviors in past if they are in a same team; (3) most colluders keep similar reputation degrees because they execute almost similar behaviors, and (4) there must be close social relationships, which can be described as a trust link, among colluders because they need to communicate for collusions. Based on such features, we can detect collusive attacks through evaluating the three above aspects among suspicious nodes.

Therefore, our proposed model of collusive attack detection includes following parts as,

(1) Social graph. Social graph describes the users and their relationships through the model of graph in SNS. Social graph is a two-tuple as $SG =< U, E >$, where $U = \{u_1, u_2, ...\}$ and $E = \{e_1, e_2, ...\}$ denotes the sets of users and their relationships respectively.

(2) Trust schedule. Trust schedule records two kinds of reliabilities of each user, i.e., reputation and trust. Trust schedule of a user is described as $TS(u_i) =< rep(u_i), Trust(u_i, u_j) >$, where $rep(u_i)$ is the reputation degree of user u_i and $Trust(u_i, u_j)$ is the set of trust relationship values from u_i to u_j who get direct trust relationship by u_i .

(3) Reputation aggregation form. Reputation aggregation form describes user's items and their related judgments, which reflects the reputation aggregation historical data. Record in reputation aggregation form is a two-tuple as $RAF =< U, REC >$, where $U = \{u_1, u_2, ...\}$ is the set of users, while $rec(u_i) =< id, source, item, value >$ is the detail record for reputation aggregation in which id is the serial number, $source$ is the judgment source user, $item$ is the target of reputation judgment, and $value$ is the value of reputation judgment.

(4) Collusive factor. Collusive factor is a set of three factors which are used to evaluate the probabilities of collusive attack, including item judgment factor (ijf), user similar factor (usf), and trust relationship factor (trf).

In our model, all entities, including users, relationships, trust information, and behaviors, are described formally for the collusion detection. Then, we can evaluate the likelihoods of collusion happening according to the three collusive factors.

III. FACTORS FOR COLLUSIVE ATTACK DETECTION

In reputation oriented collusive attack, malicious nodes usually send a large number of dishonest judgments for inflating or slandering the reputation of attack target. Such collusion is distributed, mass, and launched in a relatively short time. Therefore, finding a collusive attack relies on evaluating user's suspicious reputation judgments, behaviors of judgment sources, and the relationships of sources. In this work, we propose three collusive factors for the detection as follows,

(1) Item judgment factor

Item judgment factor aims to evaluate the behavior judgment happening probability of collusive attack in a user's reputation aggregation. We compare the value consistency of judgment behaviors between suspicious user and common users.

Let the user be u_i , and the reputation aggregation judgment record of his/her item in RAF as $rec(u_i) =< id, source, item, value >$. Assume that the average judgment value of the item is denoted as $\overline{val}(item)$, then, the item judgment factor of user u_i can be calculated as,

$$ijf(u_i.item_k) = \sqrt{\frac{\sum_{item_k} \left[rec(u_i.item_k.value_j) - \overline{val}(item_k) \right]^2}{\sum_{item_k} rec(u_i.item_k.value_j)}} \quad (1)$$

For all items of a user, the total item judgment factor about items can be calculated as,

$$ijf(u_i) = \frac{\sum_{itme_k \in u_i} \left[ijf(u_i.item_k) \times p(item_k) \right]}{\sum_{itme_k \in u_i} ijf(u_i.item_k)}, \quad (2)$$

where $p(item_k)$ denotes the judgment probability of item $item_k$ in total reputation aggregation.

(2) User malicious factor

User malicious factor aims to evaluate the likelihood of a single user being an attacker. If a user sends judgments which are far different with reputation of targets, the user malicious factor of him/her would be large.

Let a user, u_i , provide n times of reputation judgments for other users in past. Assume that each user, u_j , which has been voted by u_i , has the reputation value $rep(u_j)$ and the judgment from u_i to u_j is $val(u_i, u_j)_k \in [0,1]$. Then, the user malicious factor, $umf(u_i)$, can be calculated as follows.

$$umf(u_i) = \sqrt{\frac{\sum_{u_j \in Vote(u_i)} \sum_{k=1}^{n} \left[\left(val(u_i, u_j)_k - rep(u_j) \right) \right]^2}{n}} \quad (3)$$

where $Vote(u_i)$ is the set of users who received judgments from u_i in past for reputation aggregation.

(3) User similar factor

User similar factor is presented to describe the similarity of users, which can be used to evaluate the colluders in collusive attacks. In this work, user similar factor is calculated according to the relevancy of users' reputation judgments. The rationale of this factor is that most colluders would have a high probability of attacking same targets if they are in a collusion team.

Assume that the set of judgment target users of user u_i is $T(u_i)$. And then, for two user u_i and u_j, their user similar factor can be calculated as,

$$usf(u_i,u_j)=\frac{1}{2}\times[\frac{|T(u_i)\mathrm{I}\ T(u_j)|}{|T(u_i)\mathrm{U}T(u_j)|}+\frac{\sum_{u_k\in T(u_i)\mathrm{I}\ T(u_j)}(\overline{val}(u_i,u_k)-rep(u_k))\times(\overline{val}(u_j,u_k)-rep(u_k))}{\sqrt{\sum_{u_k\in T(u_i)\mathrm{I}\ T(u_j)}(\overline{val}(u_i,u_k)-rep(u_k))^2}\sqrt{\sum_{u_k\in T(u_i)\mathrm{I}\ T(u_j)}(\overline{val}(u_j,u_k)-rep(u_k))^2}}] \quad (4)$$

where $\overline{val}(u_i,u_k)$ denotes the average judgment value of user u_i sending to target user u_k. In above equation, user similarity is measured from two aspects: target set similarity and judgment result similarity.

(4) Trust relationship factor

Relationship reflects the tightness and trustworthiness between users according to their past behaviors, experiences or feedbacks. Most malicious users who are in a collusive attack team would keep relatively higher trust degrees with each other than honest users. We can therefore measure the likelihood of being collusive attack partners for users in the reputation aggregation.

Here, we address the trust degree calculation method between users based on their past interactions. Let there are two user u_i and u_j, and the trust value of past judgment from u_i to u_j in TS is denoted as $trust(u_i,u_j)$. Assume that the ratios of mutual judgments between the two users are $r(u_i)$ and $r(u_j)$. Then, the trust relationship factor is calculated as,

$$trf(u_i,u_j)=\left(1-\left|trust(u_i,u_j)-trust(u_j,u_i)\right|\right)^{\left(2-r(u_i)-r(u_j)\right)} \quad (5)$$

From above equation, we can see that the trust relationship factor is impacted by the distance between average judgments among users, which are traditionally calculated as trust degree in many methods.

IV. COLLUSIVE ATTACK DETECTION MECHANISM

Generally, collusive attack is launched by unfamiliar, untrustworthy and poor reputational users in SNS. From this view, our collusive attack detection mechanism contains three aspects as (1) attack happening evaluation, (2) user detection and (3) relationship traversing. That is, we first evaluate the attack happening probability of collusive attack in all reputation aggregation of user, then, evaluate whether there are colluders in the reputation aggregation, and finally traverse users through their relationships from confirming the collusive attack.

Firstly, we here address three probabilities for our mechanism as,

(1) Item attack probability

This probability aims to describe whether an item is attacked in its reputation aggregation. Let the average judgment value of a user's item $item_k$ is $ave(item_k)$, then the item attack probability can be calculated as

$$iap(item_k)=\frac{\left|ave(item_k)-rep(u_i)\right|}{\sum_{item_k\in u_i}\left|ave(item_k)-rep(u_i)\right|} \quad (6)$$

where $rep(u_i)$ is the reputation value of user u_i.

(2) User selected probability

This probability is proposed for evaluating the likelihood of being a colluder of a user who can be selected for colluder evaluation. Assume that the user u_i has trust degree $trust(u_i,u_j)$ with user u_j, and his/her reputation is $rep(u_i)$. Then, the user selected probability of u_i is as bellows,

$$usp(u_i)=1-\frac{rep(u_i)\times trust(u_j.u_i)}{\sum_{u_k\in RAF(u_j.source)}rep(u_k)\times trust(u_j.u_k)} \quad (7)$$

(3) Relationship traversing probability

Relationship traversing probability is used to select users who can be the next collusive attack evaluating user through relationships among users. This probability can be calculated according to trust relationship factor as follows,

$$rsp(u_j,u_i)=\frac{trf(u_j.u_i)}{\sum_{u_k\in Nighbor(u_j)}trf(u_j,u_k)} \quad (8)$$

Where $Neighbor(u_j)$ is the set of users who are the neighbor users of u_j in SNS. From above equation, we can see that a larger value of trust relationship factor implies a larger probability of being traversed as the next user.

Based on above points, we address our collusive attack detection mechanism as follows,

Collusive attack detection mechanism (CADM)

Step1: For a target user, denoted as tar_user, CADM collects its reputation aggregation form ($RAF(tar\_user)$), the sub-graph of social network which describes the user and relationships about tar_user, and reputation aggregation source users in $RAF(tar\_user)$, trust schedules TS of all users in tar_user reputation aggregation;

Step2: For all items of tar_user, CADM calculates the probabilities of items $iap(item_k)$. And then, the items are selected for calculating the factor of $ijf(u_i)$ according to the probability of $iap(item_k)$ repeatedly and iteratively.

Step 3: For all source users in $RAF(tar\_user)$, CADM calculates the probabilities of users $usp(u_i)$. And then, CADM selects a user according to the probability of $usp(u_i)$ as the next traversing user, noted as ts_user, which is included in a set $Next\_user$ which is used to denote the user who should be traversed in future;

Step4: CADM calculates all probabilities of $rsp(ts\_user, neighbor(ts\_user)_i)$ (here, $neighbor(ts\_user)_i$ denotes the user who has relationships with ts_user and is in $RAF(tar\_user)$ as source users), and then selects user $neighbor(ts\_user)_i$ according to the $rsp(ts\_user, neighbor(ts\_user)_i)$;

Setp5: If $umf(neighbor(ts\_user)_i) \geq \alpha$, CADM calculates the user similar factor $usf(ts\_user, neighbor(ts\_user)_i)$ of ts_user and $neighbor(ts\_user)_i$. If the condition $usf(ts\_user, neighbor(ts\_user)_i) \geq \beta$ is satisfied, $neighbor(ts\_user)_i$ is listed in a set $Next\_user$;

Step6: CADM selects a user in $Next\_user$ as ts_user and repeats the Step4 and Step5 until there is no user in set $Next\_user$. And then, CADM calculates the average value of user malicious factor $umf(Next\_user)$ as,

$$umf(Next\_user) = \frac{\sum_{u_j \in Next\_user} umf(u_j)}{|Next\_user|} \quad (9)$$

Step7: CADM repeats the steps from 3 to 6 and renews $umf(Next\_user)$.

Step8: CADM returns the two values of $ijf(u_i)$ and $umf(Next\_user)$.

In above mechanism, thresholds α and β is set between 0 and 1. In CADM, there are two values for collusive attack detection, i.e., item judgment factor and user malicious factor. The higher these two values are, the higher probability the target user has suffered collusive attacks in reputation aggregation.

V. CONCLUSION

Trust management is a significant and inevitable problem in field of social network security research. Many efforts show that reputation is a feasible and effective solution for trust identification in the open and distributed network environment, such as social network. However, collusions, including inflating or slandering, brings huge damages to reputation system. In this work, we propose a mechanism for detecting collusive attack based on three facts as (1) large judgment distance between honest reputation aggregation and collusive attacks, (2) high trust relationship among collusive attackers and (3) traversing through user relationship in social network for detecting attackers. The process of CADM is addressed based on above rationales finally. In further work, we plan to examine our proposed mechanism under real data for testifying the feasibility and effectiveness of CADM.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61103069), Key Lab of Information Network Security, Ministry of Public Security (C14602), Innovation Program of Shanghai Municipal Education Commission (13YZ052), and the Program of Shanghai Normal University (DCL201302).

REFERENCES

[1] Borgatti, Stephen P., Daniel J. Brass, and Daniel S. Halgin. "Social network research: Confusions, criticisms, and controversies." Research in the Sociology of Organizations 40 (2014): 1-29.

[2] Huang, Zhen, et al. "A social network approach to trust management in VANETs." Peer-to-Peer Networking and Applications 7.3 (2014): 229-242.

[3] Audun Jøsanga, Roslan Ismailb, Colin Boydb. A survey of trust and reputation systems for online service provision, Decision Support Systems, 43(2), (2007): 618-644

[4] Kamvar SD, Schlosser MT. EigenRep: Reputation management in P2P networks. Proc. of the 12th Int'l World Wide Web Conf. (2003): 123-134

[5] Kumar, Abhishek, et al. "Social Networking Sites and Their Security Issues." International Journal of Scientific and Research Publications 3.4 (2013): 3.

[6] Wang, Yonggang, et al. "ReSpam: A Novel Reputation Based Mechanism of Defending against Tag Spam in Social Computing." Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on. IEEE, 2014.

[7] Ghosh, T., N. Pissinou, and K. Makki. Collaborative trust-based secure routing against colluding malicious nodes in multi-hop ad hoc networks. in Local Computer Networks, 2004. 29th Annual IEEE International Conference on. 2004.

[8] Marshall, J., V. Thakur, and A. Yasinsac. Identifying flaws in the secure routing protocol. in Performance, Computing, and Communications Conference, 2003. Conference Proceedings of the 2003 IEEE International. 2003.

[9] A. Whitby, A. J¿ang, and J. Indulska, "Filtering out unfair ratings in Bayesian reputation systems," in Proc. 7th Int. Workshop on Trust in Agent Societies, 2004.

[10] Staab, Eugen, and Thomas Engel. "Collusion detection for grid computing." Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid. IEEE Computer Society, 2009.

[11] Liu, Yuhong, Yafei Yang, and Yan Lindsay Sun. "Detection of collusion behaviors in online reputation systems." Signals, Systems and Computers, 2008 42nd Asilomar Conference on. IEEE, 2008.