Recognize Pretender Accounts In Social-Network-Based Online Promotions

Sk. Rehan[1], K. Sreenivas[2], Dr Y. Jahnavi[3]
[1]Dept of CSE
[2]Assistant Professor, Dept of CSE
[3]Professor, Dept of CSE
[1, 2, 3]GIST, Gangavaram, Kovur , Nellore

Abstract- *Online social networks gradually integrate financial capabilities by enabling the usage of real and virtual currency. They serve as new platforms to host a variety of business activities such as online promotion events, where users can possibly get virtual currency as rewards by participating such events. Both OSNs and business partners are significantly concerned when attackers instrument a set of accounts to collect virtual currency from these events, which make these events ineffective and result in significant financial loss. It becomes of great importance to proactively detecting these malicious accounts before the online promotion activities and subsequently decrease their priority to be rewarded. In this paper, we propose a novel system, namely ProGuard, to accomplish this objective by systematically integrating features that characterize accounts from three perspectives including their general behaviors, their recharging patterns, and the usage of their currency. We have performed extensive experiments based on data collected from Tencent QQ, a global leading OSN with built-in financial management activities. Experimental results have demonstrated that our system can accomplish a high detection rate of 96.67% at a very low false positive rate of 0.3%.*

Keywords- Online Social Networks, Virtual Currency, Malicious Accounts, Intrusion Detection, Network Security.

I. INTRODUCTION

Online social networks (OSNs) that integrate virtual currency serve as an appealing platform for various business activities, where online, interactive promotion is among the most active ones. Specifically, a user, who is commonly represented by her OSN account, can possibly get reward in the form of virtual currency by participating online promotion activities organized by business entities. She can then use such reward in various ways such as online shopping, transferring it to others, and even exchanging it for real currency [1]. Such virtual-currency-enabled online promotion model enables enormous outreach, offers direct financial stimuli to end users, and meanwhile minimizes the interactions between business entities and financial institutions. As a result, this model has shown great promise and gained huge prevalence rapidly. However, it faces a significant threat: attackers can control a large number of accounts, either by registering new accounts or compromising existing accounts, to participate in the online promotion events for virtual currency. Such malicious activities will fundamentally undermine the effectiveness of the promotion activities, immediately voiding the effectiveness of the promotion investment from business entities and meanwhile damaging ONSs' reputation. Moreover, a large volume of virtual currency, when controlled by attackers, could also become a potential challenge against virtual currency regulation. So it is essentially important to detect the accounts which are controlled by the attackers in online promotion activities we refer to such accounts as malicious accounts. The effective detection of malicious accounts enables both OSNs and business entities to take legal actions such as banning these accounts or decreasing the possibility to reward these accounts. However, designing an effective detection method is faced with a few significant challenges. First, attackers do not need to generate malicious content to launch successful attacks. Comparatively, attackers can effectively perform attacks by simply clicking links offered by business entities or sharing the benign content that is originally distributed by business partners. Second, successful attacks do not need to depend on social structures to be more specific, maintaining active social structures does not benefit to attackers, which is fundamentally different from popular attacks such as spammers in online social networks. These two challenges make the detection of such malicious OSN accounts fundamentally different from the detection of traditional attacks such as spamming and phishing.

In order to effectively detect malicious accounts in online promotion activities by overcoming the aforementioned challenges, we have designed a novel system, namely ProGuard. ProGuard employs a collection of behavioural features to profile an account that participates in an online promotion event. These features aim to characterize an account from three aspects including i) its general usage profile, ii) how an account collects virtual currency, and iii) how the virtual currency is spent. ProGuard further integrates these features using a statistical classifier so that they can be collectively used to discriminate between those accounts controlled by attackers and benign ones. To the best of our knowledge, this work represents the first effort to systematically detect malicious accounts used for online promotion activity participation. We have evaluated our system using data collected from Tencent QQ, a leading Chinese online social network that uses a widely-accepted virtual currency (i.e., Q coin), to support online financial activities for a giant body of 899 million active accounts. Our experimental results have demonstrated that ProGuard can achieve a high detection rate of 96.67% with a very low false positive rate of 0.3%.

II. RELATED WORK

Since online social networks play an increasing important role in both cyber and business world, detecting malicious users in OSNs becomes of great importance. Many detection methods have been consequently proposed. Considering the popularity of spammers in OSNs, these methods almost exclusively focus on detecting accounts that send malicious content. A spamming attack can be considered as an information flow initiated from an attacker, through a series of malicious accounts, and finally to a victim account. Despite the diversity of these methods, they generally leverage partial or all of three sources for detection including i) the content of the spam message, ii) the network infrastructure that hosts the malicious information and iii) the social structure among malicious accounts and victim accounts. For example, Gao et al designed a method to reveal campaigns of malicious accounts by clustering accounts that send messages with similar content. Lee et al devised a method to first track HTTP redirection chains initiated from URLs embedded in an OSN message, then grouped messages that led to webpages hosted in the same server, and finally used the server reputation to identify malicious accounts. Compared to existing methods on detecting spamming accounts in OSNs, it is faced with new challenges to detect malicious accounts that participate in online promotion activities.

First, different from spamming accounts, these accounts neither rely on spamming messages nor need malicious network infrastructures to launch attacks. Second, social structures are not necessary. Therefore, none of existing methods is applicable to detecting malicious accounts in online promotion activities.

To solve the new challenges, our method detects malicious accounts by investigating both regular activities of an account and its financial activities. Detecting fraudulent activities in financial transactions has also attracted significant research efforts. Lin et al ranked the importance of fraud factors used in financial statement fraud detection, and investigated the correct classification rates of three algorithms including Logistic Regression, Decision Trees, and Artificial Neural Networks. Throckmorton et al proposed a corporate financial fraud detection method based on combined features of financial numbers, linguistic behavior, and non-verbal vocal. Compared to the studied financial fraud detection problems, account behaviors of collecting and using the virtual currency in online promotion activities are almost completely different with traditional financial systems since they do not only involve financial activities but also networking and online promotion activities. To summarize, our work aims to address a new problem caused by the new trend of integrating online social networks and financial activities. ProGuard features new capability of fusing features from both networking and financial aspects for detection. Nevertheless, we believe our method and existing approaches can complement each other to improve the security of online social networks.

III. BACKGROUND

In an OSN that integrates financial activities, an OSN account is commonly associated with accounts for both online banking and virtual currency. Figure 1 presents such an example, where a QQ account, the most popular OSN account of Tencent, is associated with an online banking account for real currency and an account for virtual currency (i.e., Q coin). A user usually directly deposits real currency into her online banking account; she can recharge her virtual currency account from her banking account. By participating online promotion events, a user can also recharge her virtual currency account by collecting rewards from the promotion events. A user can expend from his accounts in two typical ways. First, she can use real or virtual currency to purchase both real and virtual goods (i.e., online shopping). Second, she can transfer both real and virtual currency to another user by sending out gifts.

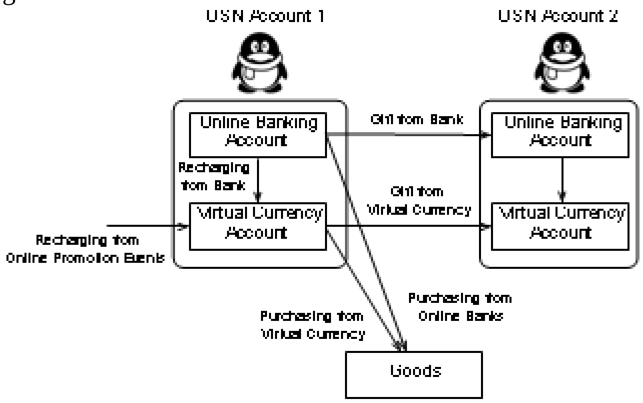

Fig. 1. The integration of OSN accounts and financial accounts

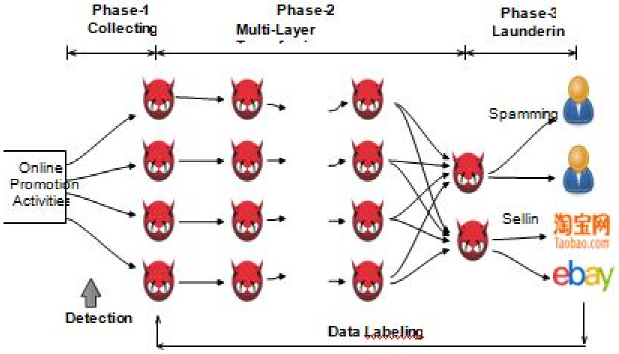

Fig. 2. Virtual Currency Flow for Malicious OSN Accounts

Figure 2 presents the typical virtual currency flow when malicious accounts participate in online promotion events. The flow is composed of three phases including i) collecting, ii) multi-layer transferring, and iii) laundering the virtual currency. In first phase, an attacker controls a set of accounts to participate in online business promotion activities and each account possibly gets a certain amount of virtual currency as return. In the second phase, the attacker will instrument these currency-collection accounts to transfer the virtual currency to other accounts. Multiple layers of transferring activities might be involved to obfuscate the identities of malicious accounts used for participating online promotion activities. At the end of the second phase, a large amount of virtual currency will be aggregated into a few laundering accounts. In the third phase, the attacker will control the laundering accounts to trade the virtual currency into real cash by selling it to individual buyers. Our objective is to design a detection system capable of identifying malicious accounts that participate in online promotion events for virtual currency collection before rewards are committed. Detecting malicious accounts at this specific time point results in unique advantages. First, as a simple heuristic to prevent freshly registered accounts that are likely to be bots, business entities usually require the participating accounts to be registered for a certain amount of time. Therefore, the detected and mitigated malicious accounts cannot be immediately replaced by the newly registered accounts, thereby limiting attackers' capabilities. In contrast, no constraint is applied for accounts used for virtual currency transferring and laundering. This implies such accounts can be easily replaced by attackers if detected, resulting negligible impact to attackers' capabilities. Second, our detection system will label whether an account is malicious when it participates in an online promotion event; this enables business entities to make actionable decisions such as de-prioritize this account from being rewarded in this event. Therefore, it can prevent the financial loss faced by business entities.

IV. DATA

We have collected labelled data from Tencent QQ, a leading Chinese online social network that offers a variety of services such as instant message, voice chat, online games, online shopping, and e-commerce. All these services support the usage of the Q coin, the virtual currency distributed and managed by Tencent

QQ. Tencent QQ has a giant body of 899 million active QQ accounts with a reportedly peak of 176.4 million simultaneous online QQ users. Tencent QQ is one of the global leading OSNs that are actively involved in virtualcurrency-based online promotion activities. Our data set is composed of 28,000 malicious accounts and 28,000 benign accounts, where all of these accounts are randomly sampled from the accounts that participated in Tencent QQ online promotion activities in August 2015. The labeling process starts from identifying laundering accounts (i.e., accounts that are associated with virtual currency spams and accounts that sell virtual currency in major e-commerce websites). Specifically, if an account transfers virtual currency to any account that engages in virtual-money laundering activities, this account will be labeled as malicious. Such "traceback" process may involve multiple layers of transferring, which is visualized. It is worth noting that although both malicious and benign accounts are labelled based on their activities in Phase-2 and Phase-3 the data used for building the detection system arecollected before the launch of the online promotion event. The reason is that the objective of our detection system is to identify malicious accounts before the rewards are committed.

Although the aforementioned "trace-back" method is effective in manually labeling malicious accounts, using it as a detection method is impractical. First, it requires a tremendous amount of manual efforts for forensic analysis such as identifying suspicious virtual-currency dealers in external e-commerce websites, correlating spamming content with user accounts, and correlating sellers' profiles with user accounts. In addition, evidence for such forensic analysis will be only available after malicious accounts participate in online promotion events. Therefore, this data labeling process, if used as detection method, cannot guide business entities to mitigate their financial loss proactively. In contrast, our method is designed to detect malicious accounts prior to the reward commitment. For each account, we collect a variety of information including 1) login activities, 2) a list of anonymized accounts that this account has sent instant messages to, 3) service purchase activities, 4) the recharging activities, and 5) the expenditure activities.

V. SYSTEM DESIGN

ProGuard is composed of two phases, namely the training phase and the detection phase. In the training phase, a statistical classifier is learnt from a set of pre-labelled malicious and benign accounts. In the detection phase, an unknown account will first be converted to a feature vector and then analyzed by the statistical classifier to assess its maliciousness. In this section, we will introduce various features and demonstrate their effectiveness on differentiating malicious accounts from benign ones. We propose three general guidelines to steer the feature design.

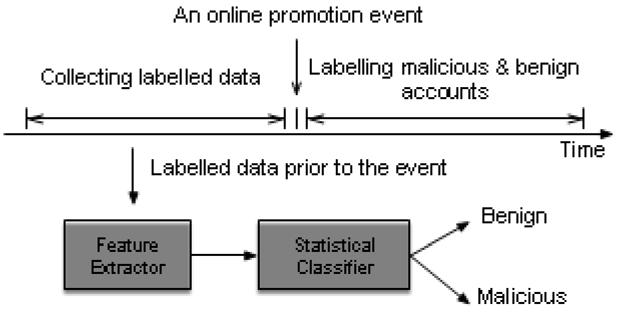
An online promotion event
Collecting labelled data
Labelling malicious & benign accounts
Time
Labelled data prior to the event
Feature Extractor
Statistical Classifier
Benign
Malicious

Fig. 3. The Architectural Overview of the System

General Behaviors: Benign accounts are usually used by regular users for variety of activities such as chatting, photo sharing, and financial activities. In contrast, malicious accounts are more likely to be driven by onlinepromotion events. Therefore, the benign accounts tend to be more socially active compared to malicious accounts.

Currency Collection: The malicious accounts under investigation focus on using online promotion activities to collect virtual currency. In contrast, benign users are likely to obtain virtual currency from multiple resources.

Currency Usage: Attackers' ultimate objective is to monetize the virtual currency. In contrast, benign users use their virtual currency in much more diversified ways.

A. General-Behavior Features

Malicious accounts tend to be less active compared to benign accounts with respect to the non-financial usage. Attackers usually control their accounts to only participate in online promotion activities. In contrast, benign accounts are more likely to engage in active interaction with other users.

Feature 1 : The Ratio of Active Days.
Feature2 : The Number of Friends. This feature summarizes the number of friends for each account.
Feature 3 : The Number of Services Purchased By An Account.

B. Currency Collection Features

In addition to collecting virtual currency by participating in online promotion activities, an OSN user can recharge her account with virtual currency through various ways such as wire transfer, selling virtual goods, and transferring from other accounts. Generally, benign users should be more active with respect to recharging their accounts. We propose two features to characterize this trend from two aspects including the amount of recharging and the important sources for recharging.

- Feature4 : The Average Recharge Amount of Virtual Currency.
- Feature5 : The Percentage of Recharge from Promotion Activities:

C. Features of Usage Activities

- Feature 6 : Total Amount of Expenditure.
- Feature 7 : The Percentage of Expenditure from Banks.
- Feature 8 : The Percentage of Expenditure as Gifts.

VI. EVALUATION

We performed extensive evaluation of ProGuard, which focuses on the overall detection accuracy, the importance of each feature, and the correlation among these features. For this evaluation, we used totally 56,000 accounts whose entire dataset is divided into 28,000 malicious accounts and 28,000 benign accounts. Such data serve as a well-balanced dataset for training a statistical classifier [19].

A. Detection Accuracy

We have used the normalized Random Forest (RF) as the statistical classifier for ProGuard and evaluated its detection accuracy. RF classifier [20] is an ensemble of unpruned classification trees, which is trained over bootstrapped samples of the original data and the prediction is made by aggregating majority vote of the ensemble. In order to avoid the bias caused by the selection of specific training set, we also performed 10-fold cross-validation. Specifically, the entire dataset is partitioned to 10 equal-size sets (i.e., 10-folds); then iteratively 9-folds are used for training and the remaining 1fold is adopted for testing. The RF classifier was trained with 3000 trees and randomly sampled 4 features for each of tree splitting [21]. The receiver operating characteristic (ROC) that characterizes the overall detection performance of ProGuard is presented in Fig. 12. The experimental results have shown that ProGuard can achieve high detection accuracy.

In practice, alternative statistical classifiers might be adopted to render new performance benefits such as scalability. Therefore, we also evaluate how ProGuard performs when alternative classifiers are used. As a means towards this end, we used Support Vector Machine (SVM) [22] and Gradient-Boosted Tree [23] to repeat our experiments. Specifically, we used 10fold cross validation for each of classifiers and calculated the area under the ROC curve (AUC) [24], a widely used measure of quality of supervised classification models, which is equal to the probability that a randomly chosen sample of malicious accounts will have a higher estimated probability of belonging to malicious accounts than a randomly chosen sample of benign accounts. Since AUC is cutoff-independent and values of AUC range from 0.5 (no predictive ability) to 1.0 (perfect predictive ability), a higher AUC of a classifier indicates the better prediction performance, irrespective of the cutoff selection. Table I lists the AUC values for all three classifiers used in the experiments. Both SVM and Gradient-Boosted Tree accomplished high detection results, comparable with the Random Forest which has the best performance on AUC. The experimental results imply that our proposed features are not sensitive to the selection of statistical classifiers.

B. Feature Importance and Correlation

We investigated the relative importance of the proposed features in the context of Random Forest classifier, which has accomplished the best detection accuracy according to our experiments. We employed the variable importance of each feature to the Random Forest classification model using permutation test [21]. The variable importance for each feature is computed by mean decrease in accuracy, which is defined as a prediction error rate after permuting an each feature. Specifically, the ratio of active days (Feature 1), the average recharge amount of virtual currency (Feature 4), and the percentage of expenditure from banks (Feature 7) represent the most significantly for detection. It is worth noting that these top three features cover three complementary aspects including the general behaviors, currency collection, and currency usage that guide the feature design.We also performed the correlation among various features, where the correlation implies the extent to which a feature might be redundant given other features. Two widely-adopted methods have been used in our experiments. First, the upper triangular of correlation matrix is carried out for discovering if a pair of strongly correlated features appear within the features, where each column in the upper triangular matrix represents the Pearson’s r correlation coefficient [25] of a pair of two distinct features. The Pearson’s correlation coefficient r ∈ [−1,1] of two features X and Y can be defined as

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2}\sqrt{\sum(Y-\bar{Y})^2}}$$

where $^{X^-}$ and $^{Y^-}$ denote the means of the two features. Fig. 13 shows that the most of features are not strongly correlated one to each other (i.e, Pearson’s correlation coefficient |r| ≥ 0.9). For example, a pair of two features, Feature 1 (The Ratio of Active Days) and Feature 8 (The Percentage of Expenditure as Gifts) represents that the highest negative correlation score is 0.07 and the highest positive correlation between Feature 4 (The Average Recharge Amount of Virtual Currency) and Feature 6 (The Total Amount of Expenditure) is 0.82. Next, we analyzed Principal Component Analysis (PCA), which can be used to evaluate variable correlation in regard to the variance of the data [26]. Figure 14 shows the experimental result on PCA variables factor map [27]. In the variable factor map, each of features is expressed as an arrow and the angle between the two arrows of features implies the correlation among the respective features on the third and fourth principal components (PC). For example, given the angle between the two arrows of different two features goes near 90 degrees, they might not be correlated. implying a weak correlation between features. According to the correlation matrix and PCA variable factor map, which show little correlation with each other, we conclude that majority of the features complement each other given their tendency towards linearly independence.

VII. DISCUSSION

Attackers may attempt to evade our detection after they know the design of ProGuard. This represents a general challenge for all detection systems rather than a specific design flaw of the proposed system. Specifically, attackers can instrument their accounts so that their behaviors are indistinguishable from benign accounts. However, since ProGuard detection features characterize elements of malicious accounts that are critical to their success of attacks against other detection systems, the successful evasion may fundamentally constrain attackers’ capabilities. For example, attackers can significantly increase the number of active days of malicious accounts. However, it may expose malicious accounts to existing bot-account detection systems that leverage frequent login patterns of malicious accounts. Attackers can also increase the number of friends by adding malicious accounts as friends. Nevertheless, this may qualify the applicability of many detection systems that take advantage of social structures such as. Attackers can also increase the diversity for recharging sources, the amount of recharging, and the expenditure from bank accounts. However, these solutions directly increase the financial cost for launching the attacks, which could make attacks themselves meaningless. Attackers might also attempt to decrease the percentage of expenditure as gifts, which, however, fundamentally limits the bandwidth to launder the collected virtual currency.

Considering the active trend of integrating OSNs with financial capabilities, detecting malicious accounts that engage in suspicious financial activities becomes of central importance. Although the design and evaluation of ProGuard are based on real-world data collected from Tencent QQ, a leading OSN with 899 million active accounts, the features and the detection framework can be easily applied to other OSNs that integrate financial activities. Specifically, all the proposed features are based on essential financial functions such as recharging and gifting.

VIII. CONCLUSION

This paper presents a novel system, ProGuard, to automatically detect malicious OSN accounts that participate in online promotion events. ProGuard leverages three categories of features including general behavior, virtual-currency collection, and virtual-currency usage. Experimental results based on labelled data collected from Tencent QQ, a global leading OSN company, have demonstrated the detection accuracy of ProGuard, which has achieved a high detection rate of 96.67% given an extremely low false positive rate of 0.3%.

REFERENCES

[1] Y. Wang and S. D. Mainwaring, “Human-currency interaction: learning from virtual currency use in china,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2008, pp. 25–28.

[2] J. S. Gans and H. Halaburda, “Some economics of private digital currency,” Rotman School of Management Working Paper, no. 2297296, 2013.

[3] X. Hu, J. Tang, and H. Liu, “Online social spammer detection,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. AAAI, 2014, pp. 59–65.

[4] “Leveraging knowledge across media for spammer detection in microblogging,” in Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. ACM, 2014, pp. 547–556.

[5] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, “Detecting automation of twitter accounts: Are you a human, bot, or cyborg?” IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 811–824, 2012.

[6] Z. Chu, S. Gianvecchio, A. Koehl, H. Wang, and S. Jajodia, “Blog or block: Detecting blog bots through behavioral biometrics,” Computer Networks, vol. 57, no. 3, pp. 634–646, 2013.

[7] S. Fakhraei, J. Foulds, M. Shashanka, and L. Getoor, “Collective spammer detection in evolving multi-relational social networks,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp. 1769–1778.

[8] Y.-R. Chen and H.-H. Chen, “Opinion spammer detection in web forum,” in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015, pp. 759–762. 2169-3536 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2654272, IEEE Access 10.

[9] F. Wu, J. Shu, Y. Huang, and Z. Yuan, “Social spammer and spam message co-detection in microblogging with social context regularization,” in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015, pp. 1601–1610.

[10] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, “Twitter spammer detection using data stream clustering,” Information Sciences, vol. 260, pp. 64–73, 2014.

[11] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, “Detecting and characterizing social spam campaigns,” in Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, 2010, pp. 35–47.

[12] S. Lee and J. Kim, “Warningbird: Detecting suspicious urls in twitter stream.” in NDSS, vol. 12, 2012, pp. 1–13.

[13] C. Yang, R. C. Harkreader, and G. Gu, “Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers,” in International Workshop on Recent Advances in Intrusion Detection. Springer, 2011, pp. 318–337.

[14] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” Journal of Network and Computer Applications, vol. 68, pp. 90 – 113, 2016.

[15] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," Computers & Security, vol. 57, pp. 47 – 66, 2016.

[16] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," Knowledge-Based Systems, vol. 70, pp. 324 – 334, 2014.

[17] C.-C. Lin, A.-A. Chiu, S. Y. Huang, and D. C. Yen, "Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments," Knowledge-Based Systems, vol. 89, pp. 459 – 470, 2015.

[18] C. S. Throckmorton, W. J. Mayew, M. Venkatachalam, and L. M. Collins, "Financial fraud detection using vocal, linguistic and financial cues," Decision Support Systems, vol. 74, pp. 78 – 87, 2015.

[19] Z. Afzal, M. J. Schuemie, J. C. van Blijderveen, E. F. Sen, M. C. Sturkenboom, and J. A. Kors, "Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records," BMC medical informatics and decision making, vol. 13, no. 1, p. 1, 2013.

[20] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[21] S. RColorBrewer and M. A. Liaw, "Package randomforest," 2012.

[22] N. Cristianini and J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000.

[23] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques. Morgan kaufmann, 2006.

[24] T. Fawcett, "An introduction to roc analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861–874, 2006.

[25] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," The American Statistician, vol. 42, no. 1, pp. 59–66, 1988.

[26] I. Jolliffe, Principal component analysis. Wiley Online Library, 2005.

[27] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[28] Y. Zhao, Y. Xie, F. Yu, Q. Ke, Y. Yu, Y. Chen, and E. Gillum, "Botgraph: Large scale spamming botnet detection." in NSDI, vol. 9, 2009, pp. 321– 334.

[29] J. Song, S. Lee, and J. Kim, "Spam filtering in twitter using senderreceiver relationship," in International Workshop on Recent Advances in Intrusion Detection. Springer, 2011, pp. 301–317.

[30] T.-S. Moh and A. J. Murmann, "Can you judge a man by his friends?- enhancing spammer detection on the twitter microblogging platform using friends and followers," in International Conference on Information Systems, Technology and Management. Springer, 2010, pp. 210–220.