MALICIOUS ACCOUNTS DETECTING FROM ONLINE PROMOTION ACTIVITIES IN SOCIAL NETWORKS

MITTA VENKATA NARAYANA
PG Scholar, Dept of CSE, PACE Institute of Technology & Sciences, Vallur, Ongole, AP, India.

V GOPI KRISHNA
Assistant Professor, Dept of CSE, PACE Institute of Technology & Sciences, Vallur, Ongole, AP, India.

Abstract

Online social networks regularly adding to 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 probably get virtual currency as rewards by participating such events. Both Online social networks and business partners are significantly worried 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 decreases their priority to be rewarded. In this paper, I propose a novel system, to achieve 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. I performed general experiments based on data collected from Tencent QQ, a global leading Online social networks 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%.

Index Terms*—Online Social Networks, Virtual Currency, Malicious Accounts, Intrusion Detection, Network Security*

1. Introduction

Online social networks that combine virtual currency serve as an interesting platform for various business activities, where online, interactive promotion is among the most active ones. Specifically, a user, who is commonly represented by her/his Online social networks account, can possibly get reward in the form of virtual currency by participating online promotion activities organized by business entities. She/he can then use such rewards in different 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 incentive to end users, and in the meantime minimizes the interactions between business entities and financial institutions. This model has shown great promise and gained vast popularity rapidly.

However, it faces a important threat that is attackers can control a no of accounts, using by registering new accounts and compromising existing accounts, to participate in the online promotion events for virtual currency. That type of malicious activities will basically weaken the effectiveness of the promotion activities, immediately avoiding the effectiveness of the promotion investment from business entities and meanwhile damaging Online social network's. status. Furthermore, a huge volume of virtual currency, when controlled by attackers, could also become a potential challenge against virtual currency regulation [2]. It therefore becomes of essential importance to detect accounts controlled by attackers in online promotion activities. In the following discussions, I refer to such accounts as malicious accounts. The effective detection of malicious accounts enables both Online social networks and business entities to take alleviation actions such as banning these accounts or decreasing the possibility to reward these accounts. However, creating and designing an effective detection method is faced with a few important challenges. First, attackers do not need to create malicious content (e.g., phishing URLs and malicious executables) to launch successful attacks. Comparatively, attackers can effectively perform attacks by simply clicking links offered by business entities or sharing the gentle content that is originally distributed by business partners. These actions themselves do not perceivably differentiate from benign accounts. Second, successful attacks do not need to depend on social structures (e.g., "following" or "friend" relationship in popular social networks). 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. As a consequence, it is extremely hard to adopt existing methods to detect spamming and phishing accounts. In order to effectively detect malicious accounts in online promotion activities by overcoming the before mentioned challenges, I have Proposed a novel system, namely ACGuard. ACGuard employs a collection of behavioral 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. ACGuard 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 evaluate 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 ACGuard can achieve a high detection rate of 96.67% with a very low false positive rate of 0.3%. The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 briefly discusses the background of virtual-currency-enabled OSNs. Section 4 describes how data was collected and labeled. We present the system design in Section 5 and evaluation results in Section 6. Section 7 concludes.

2. 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 [3], [4], [5], [6], [7], [8], [9], [10]. 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 (e.g., phishing content or exploits), and iii) the social structure among malicious accounts and victim accounts. For example, Gao et al. [11] designed a method to reveal campaigns of malicious accounts by clustering accounts that send messages with similar content. Lee et al. [12] devised a method to first track HTTP redirection chains initiated from URLs embedded in an Online social network message, then grouped messages that led to web pages hosted in the same server, and finally used the server reputation to identify malicious accounts. Yang et al. [13] extracted a graph from the "following" relationship of twitter accounts and then propagated maliciousness score using the derived graph;

Wu et al. [9] proposed a social spammer and spam message code detection method based on the posting relations between users and messages, and utilized the relationship among user and message to improve the performance of both social spammer detection. Compared to existing methods on detecting spamming accounts in online social networks, 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 [14], [15]. For example, Olszewski et al [16] represented the user account records in 2-dimensional space of the Self-Organizing Map grid, and proposed a detection method based on threshold-type binary classification algorithm to solve problems of credit card fraud and telecommunications fraud. Lin et al. [17] 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. [18] 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. ACGuard 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.

3.Background Details

An online social network that integrates financial activities, an online social networks 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 online social networks 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/he can recharge her/his virtual currency account from her/his banking account. By participating online promotion events, a user can also recharge her/his virtual currency account by collecting rewards from the promotion events. A user can expend from his accounts in two typical ways. First, she/he can use real or virtual currency to purchase both real and virtual goods (i.e., online shopping). Second, she/he can transfer both real and virtual currency to another user by sending out gifts.

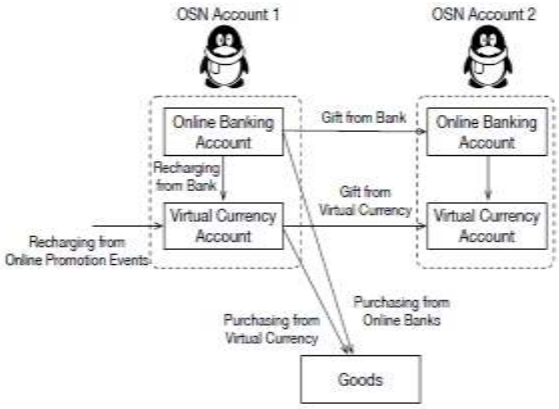

Fig. 1. The integration of OSN accounts and financial accounts

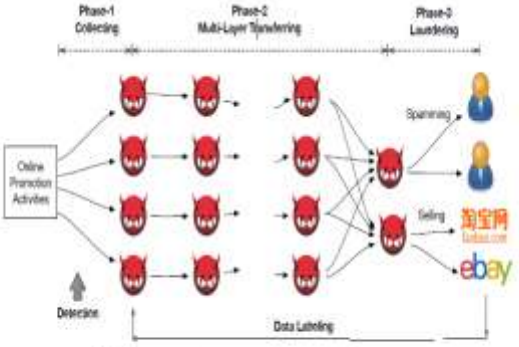

Fig. 2. Virtual Currency Flow for Malicious OSN Accounts

Figure 2 presents the typical virtual currency flow when malicious accounts participate in online promotion events. The flow is combination 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. Attackers usually employ two methods to solicit individual buyers including sending spams and advertising through major e-commerce websites such as www.taobao.com and www.tmall.com. In order to compete with regulated sources for virtual currency (i.e., purchasing virtual currency using real currency), attackers usually offer a considerable discount.

Our objective is to design a detection system capable of identifying malicious accounts that participate in online promotion events for virtual currency collection (at the collection phase) before rewards are committed. Detecting malicious accounts at this specific time point (i.e., before the commitment of rewards and at the collection phase) 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 (e.g., a few weeks). Therefore, the detected and mitigated malicious accounts cannot be immediately replaced by the newly registered accounts, thereby drastically.

Limiting attacker's 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 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 proactively mitigate the financial loss faced by business entities.

4.Data Collection

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 virtual currency- 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 at the bottom in Figure 2. It is worth noting that although both malicious and benign accounts are labelled based on their activities in Phase-2 (i.e., currency transferring) and Phase-3 (i.e., laundering), the data used for building the detection s stem are collected 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. The top of Figure 3 presents the temporal relationship among the data collection process, online promotion events, and the account label in process. Therefore, it is worth noting that an account may not have any historical financial activities (even for virtual currency collection activities) since it participates in the online promotion for the first time. 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.

5. Design of the System

ACGuard 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-labeled 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. The bottom of Figure 3 presents the architectural overview of ACGuard. As a variety of statistical classifiers have been developed and widely used, designing features capable of discriminating between malicious accounts and benign accounts becomes of central focus. 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.

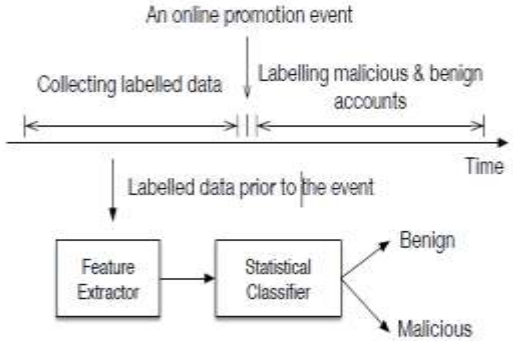
An online promotion event
Collecting labelled data
Labelling malicious & benign accounts
Time
Labelled data prior to the event
Feature Extractor
Statistical Classifier
Benign
Malicious

Fig. 3. The Architectural Overview of the System

General Behaviors: Gentle 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 online promotion 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

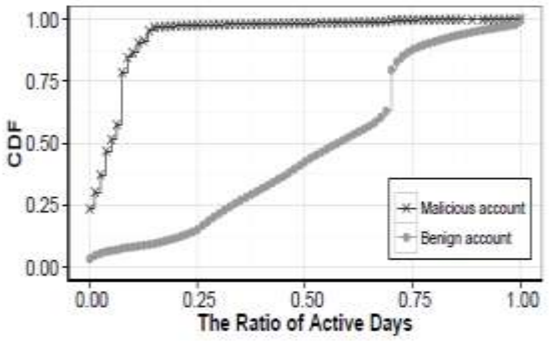

Fig. 4. Feature 1 - The Ratio of Active Days

use their virtual currency in much more diversified ways.

Behavior of 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. This feature represents the ratio of the number of active days of an account for the passed one year. Specifically, if an account is logged in at least once for a day, this day will be labeled as "active" for this account. Attackers usually login malicious accounts for participating in online promotion activities that involve virtual currency. Therefore, malicious accounts tend to be silent in the absence of online promotion activities. The availability of promotion activities is significantly influenced by timing and spatial factors. For example, promotion activities are intensive over holiday seasons, special dates, and regional events while occasionally available for other time periods. As a consequence, malicious accounts tend to be inactive generally. Comparatively, benign accounts are used by regular users and their logins are driven by the daily usage such as chatting and photo sharing. Many users configure their applications to automatically login upon the bootstrap of the underlying system (e.g., a smartphone), which further facilitates volatility of benign accounts. Figure 4 presents the distribution of feature values for both malicious accounts and benign accounts. As illustrated in the figure, the vast majority of malicious accounts (i.e., approximately 98% of malicious accounts) are active for less than 20% of total days whereas only a small percentage of benign accounts (i.e., less than 20%) experience the same active level (i.e., being active for less than 20% of one year).

Feature 2 : The Number of Friends. This feature summarizes the number of friends for each account.

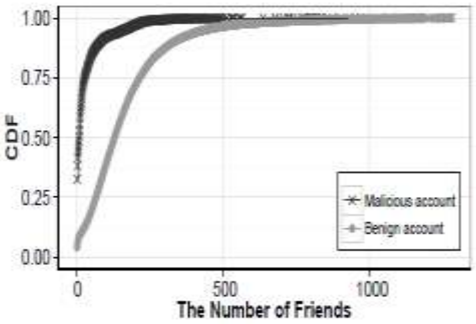

Fig. 5. Feature 2 - The Number of Friends for An Account

As a common feature for almost all online social networks, each OSN account has a list of friends. It usually implies a considerable amount of user-user interaction for one user to add another one as her friend. It is common for a benign user to maintain a relatively lengthy friend list for various social activities such as chatting and photo sharing. In contrast, an attacker usually lacks the motivation to maintain a friend list since it contributes little to promotion participation but costs significant efforts such as solving captcha challenges.

Figure 5 presents the distribution of values for this feature, where malicious accounts tend to have much less friends compared to benign accounts. Specifically, approximately 80% of malicious accounts have less than 40 friends while about 70% benign accounts have more than 200 friends.

Feature 3: The Number of Services Purchased By An Account. This feature represents the total number of types of upgraded membership that an account has paid for through all possible methods.

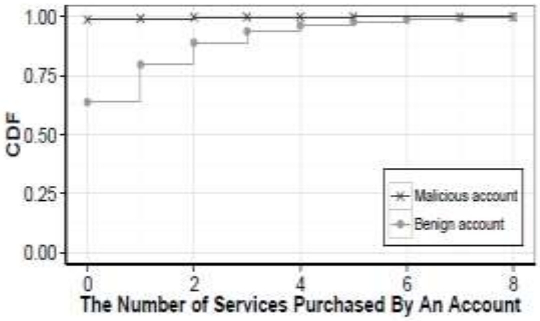

Fig. 6. **Feature 3 - The Number of Services Purchased By An Account**

It is a common feature in many online social networks that an user can upgrade his/her account by making a certain amount of payment through various ways such as credit card, wire transfer, and virtual currency. In the Tencent dataset, we consider 8 types of most popular upgraded membership including QQ VIP, Qzone, SVIP, QQ Music, Hollywood VIP, QQ Games, QQ books, Tencent Sports. An upgraded account can a wide range of paid benefits such as advanced capabilities

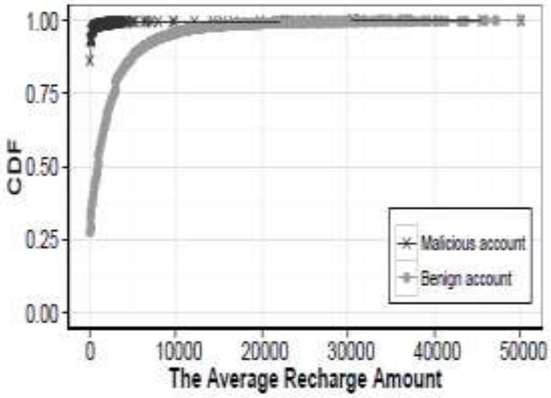

Fig. 7. **Feature 4 - The Average Recharge Amount**

for an online game avatar, enriched decoration for the account appearance, and expanded visibility of visitors. While a certain amount of benign users are inclined to be motivated to upgrade their accounts, accounts controlled by attackers are extremely unlikely to participate in such paid upgrade since the upgraded membership contributes nothing to their collection of virtual currency. Figure 6 presents the distribution for this feature: while approximately 37% of benign users purchased at least one type of upgraded membership, the vast majority of malicious accounts do not make any purchase. 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.

Feature 4 - The Average Recharge Amount of Virtual Currency. This feature represents the average amount of virtual currency for each recharge regardless of the sources for recharging. Benign users who participate in online promotion activities are usually also interested in other online financial activities. Therefore, these benign users tend to actively recharge their accounts. The recharge amount for each time by a benign user is commonly considerably large since users tend to decrease the hassle of recharging. In contrast, if a malicious account has been recharged, the amount of virtual currency for each recharge is usually bounded by a relatively small volume offered by the online promotion activity. Figure 7 presents the distribution of this

feature for benign and malicious accounts, respectively. Specifically, the average recharge amount is higher than 1100 Chinese cents1 for more than 50% of benign users, where only a small percentage (i.e., approximately 15%) of malicious users has an average amount
that is higher than 140 Chinese cents.

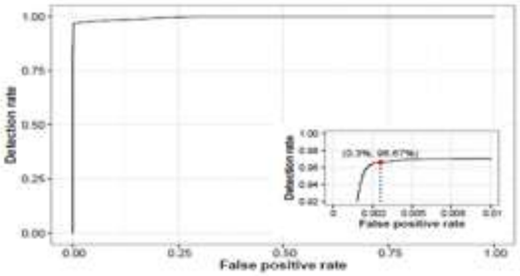

Fig. 8 ROC curve on 8 features

6.Evaluation of the System

We performed extensive evaluation of ACGuard, 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 ACGuard 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 1- fold 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 ACGuard is presented in Fig. 12. The experimental results have shown that ACGuard can achieve high detection accuracy. For example given the false positive rate of 0.3%, ACGuard can accomplish a high detection rate of 96.67%. In practice, alternative statistical classifiers might be adopted to render new performance benefits such as scalability. Therefore, we also evaluate how ACGuard 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 10- fold 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.

7.Conclusion

This paper presents a novel system, ACGuard, to automatically detect malicious Online Social Networks accounts that participate in online promotion events. ACGuard leverages 3 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 Online Social Network company, have demonstrated the detection accuracy of ACGuard, which has achieved a high detection rate of 96.67% given an extremely low false positive rate of 0.3%.

8. 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.*

[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 10thACM 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? empiricalevaluation and new design for fighting evolving twitter spammers," inInternational Workshop on Recent Advances in Intrusion Detection. Springer, 2011, pp. 318–337.*

[14]*A. Abdallah, M. A. Maarof, and A Zainal, "Fraud detection system: Asurvey," 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 Advancesin 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.

Author Profiles

Mitta Venkata Narayana, PG Scholar, Dept of CSE, PACE Institute of Technology & Sciences, Ongole, A.P

V.Gopikrishna, Assistant Professor, Dept of CSE, PACE Institute of Technology & Sciences, Ongole. He have 7 Years Teaching Experience. His Interested area is Network Security, Wireless Sensor Networks